ON-TIME PERFORMANCE ANALYSIS

OF AIRLINE FLIGHT SCHEDULES

by

Jacqueline Fetaya

A thesis submitted to the Faculty of
Graduate Studies and Research in partial
fulfillment of the requirements for the
degree of Master of Science.

Department of Mathematics
McGill University
Montreal

November 1974

ON-TIME PERFORMANCE ANALYSIS

OF AIRLINE FLIGHT SCHEDULES

by

Jacqueline Fetaya

ABSTRACT

This thesis presents a mathematical model for the evaluation of the on-time performance of a given airline schedule. It was developped as a project within the Operational Research Division of Air Canada, and was computerized to permit interactive usage. It was tailored to suit the Company's scheduling environment and requirements.

The principal activities governing aircraft cycles are defined in terms of stochastic variables and a parametric investigation is conducted. The lognormal distribution is found to provide a good fit in most cases. The model follows an analytical rather than a simulation approach. The distributions of arrival and departure delay times for flight-legs are determined recursively using the discretization and convolution techniques. A reliability study on the model is then performed using actual flight information. Predictive models are presented for the evaluation of new flights.

The study is based on a similar investigation conducted within Lufthansa.

Department of Mathematics
McGill University
Montreal

M.Sc.
November 1974

ON-TIME PERFORMANCE ANALYSIS

OF AIRLINE FLIGHT SCHEDULES

by

Jacqueline Fetaya

RESUME

Nous présentons dans cette thèse le développement d'un modèle mathématique pour l'évaluation de la ponctualité des horaires d'une flotte aérienne. Etant membre de la Division de Recherche Opérationnelle à Air Canada, ce modèle fut programmé de façon à permettre l'usage interactif et de façon à rencontrer certaines exigences de la Compagnie pour la planification des horaires, afin de permettre son intégration parmi les systèmes existants ou en voie de développement.

Les activités principales reliées à une flotte d'avions sont définies en terme de variables aléatoires et une étude paramétrique est effectuée sur ces mêmes variables afin de représenter celles-ci par une distribution lognormale. Un modèle est alors formulé suivant une méthode analytique plutôt que de se servir du concept de simulation. Les distributions des temps de délais de départs et d'arrivées pour les arcs de vols sont déterminées par un modèle récursif, tout en utilisant les techniques de discrétisation et de convolution. Des tests de fiabilités sur des exemples réels sont décrits. Nous présentons aussi des méthodes de prédiction pour l'évaluation des nouveaux vols.

Cette thèse est basée sur une étude semblable qui a été conduite par Lufthansa.

Department of Mathematics
McGill University
Montreal

M.Sc.
November 1974

ACKNOWLEDGEMENTS

My sincerest thanks go to my thesis director, Professor T. Papaioannou for his enthusiasm in the Scheduling Punctuality Problem. His perseverance, guidance and assistance throughout the preparation of this thesis were a valuable asset.

I would also like to thank the members of the Operational Research Division of Air Canada for their support. In particular, Mr. W.J. Bartlett, Director of O.R. who made it possible to do this thesis in Air Canada, Mr. V. K. Wozniuk, Manager of O.R. and Mr. D. Girard Senior O.R. Designer who contributed to this thesis by offering their experience in the practical and technical aspects of the Airline Industry. I am also grateful to the Product Development Department of Marketing for their Scheduling expertise and to the Programming staff in the Computer and Systems Services Branch for their help.

A special thanks to Mrs. Danielle Perras for her enthusiasm and proficient typing.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The airline industry is still in a phase of rapid growth in both passenger volumes and competition between airlines. The scope of this competition is quite narrow due to the limited number of aircraft types available and to the stiff controls enforced by governments and industry bodies on air fares, air routes, flight frequencies, aircraft capacities, etc... Service thus becomes the major area open to competition and a key element of service is the on-time performance of an airline schedule. The evaluation of the on-time performance of airline schedules has therefore become increasingly necessary in recent years.

The primary concern of a study on the reliability of an airline schedule is to construct some type of a model which could identify the potential problem areas in the punctuality of a given schedule in order to increase its expected reliability. The purpose of this thesis is to provide such a model which would enable an airline, and in particular Air Canada, to evaluate the on-time performance of its schedules in both the short-term as well as the long term. The approach that was selected is an analytical type of model based on the work done by Peter Franke (1972) in Lufthansa. Due to the difference in environments and requirements of the two airlines especially regarding the type of available data and the difference in markets that each serves, the model in this study is slightly different in nature although the same in concept.

Contributions in this thesis have been made in various aspects. A mathematical representation of the model and some of its variations are given in Section 1.2.2 of this chapter. A brief review of related work is also presented showing models with similar objectives that have been developed in the past, with specific reference made to such a model previously built within Air Canada. As for predicting the means and variances of our principal variables, a method different from the one proposed by Franke is applied in this thesis. Finally, a discussion is given on the accuracy to be expected from the discretization followed by a convolution of independent lognormal random variables.

In this chapter, concepts and terminology used in scheduling are described followed by a description of the available data (Section 1.4) and a discussion on their suitability for the adoption of the proposed model. The chapter ends with a section describing the method of analysis which leads to Chapter two, the parametric analysis.

In the first section of the second chapter an intuitive discussion is given relating to the nature of the variables that are being analysed, leading to a reasoning for the use of the lognormal distribution and then describing some properties of that distribution. Some types of analyses are then discussed comparing the Kolmogorov and the Chi-Square tests for testing the sample distributions. This is followed by individual treatment and analyses of each of three variables.

The problem of predicting the means and variances of our variables is then discussed in Chapter four where an application of the method is also presented.

Chapter three deals with the model in question where some of the practical logic and a detailed formulation of the model are presented. The computational procedures used for the model are stated and this is followed by some comments on the accuracy with respect to discretization and convolutions as mentioned above. Some comments are then made on the interpretation of the results of the model. In the following section, some reliability tests for the model are performed by comparing the model results with some actual on-time performance statistics recorded for the same time period. The chapter ends with a discussion of other possible approaches that may be pursued to achieve the same objective - that is for the evaluation of the on-time performance of a flight schedule. This is then followed by the appendices and the bibliography.

1.1 Basic concepts and terminology

In planning an airline schedule some fixed times are allotted for the commencement and duration of each of an interconnecting network of activities. Basically, these activities consist of a multitude of "flight-legs" each being a single flight from one airport to another with no intermediate stop. Between these legs when the aircraft is physically inactive on the ground, other kinds of activities relating to aircraft and passenger handling occur.

One or more flight-legs of a pre-planned sequence constitute a "flight" which has an associated flight serial number and which is operational on a specified aircraft type either daily or for specific days of the week.

For instance Flight 602 may be a DC-9 scheduled to leave Ottawa daily at 6:00 a.m. arriving in Montreal at 6:30 a.m. and then departing for Halifax at 7:00 a.m. where it arrives at 9:20 a.m.:

Flight 602:	Ottawa	Montreal	Halifax
	6:00 a.m.	6:30 a.m. 7:00 a.m.	9:20 a.m.

When one aircraft is scheduled to undergo a certain sequence of flights or to follow its "routing" this is denoted as an aircraft cycle which repeats itself according to the pre-planned frequency of the flights. Thus in the former example a DC-9 was planned for Flight 602 and the same aircraft was also planned for flights 609, 227 and 296 daily as follows:

Flight 602 — Flight 609

Ottawa — Montreal — Halifax — Montreal

0600—0630—0700—0920—1035—1100

Flight 227

Montreal — Windsor — Winnipeg — Calgary — Vancouver

1100—1200—1325—1355—1510—1535—1630—1700—1715

Flight 296

Vancouver — Edmonton — Winnipeg

1715—1845—2105—2125—0005

An airline schedule, usually prepared by aircraft type, consists, in simplified form, of a number of these aircraft cycles. However, many complications occur in scheduling since one must cope with the many operational and commercial constraints while achieving the required services and also maximizing the use of the resources at a minimal cost.

The scheduling process becomes very complex when one considers the interrelated operational constraints such as curfews at stations, directional headwinds, gate and ground crew limitations, maintenance needs, time zone changes, etc... while attempting to meet the commercial requirements of frequency of service, desired departure time, and mix of non-stop, direct and connecting services.

The schedule-related elements are so complex that for analysis purposes a simplification of the entire structure is necessary. Taking the point of view of aircraft cycles as was done above enables one to follow an aircraft through its scheduled route and thus to relate to the "typical" activities within a cycle. Essentially, a cycle relates in a timewise fashion, a series of sequential activities functionally independent from one another which are each essential in forming the total service. Before departing from its home base (first departure), the aircraft must be prepared, and the amount and type of preparation depends on the type of aircraft, the station and the planned trip itself (e.g. meal preparation may or may not be necessary). From this point, its sequence of activities will take the form of flights from one airport to another as well as ground-type activities at each intermediate station.

Before proceeding to the problem definition it is essential at this point to describe some of the terminology that will be used throughout this thesis.

In observing a particular aircraft for one leg of a flight, the actual time it takes from take off to touch down is denoted as the "flight time". The addition of the taxi times upon departure as well as arrival results in the "block time".

The ground stops separating legs of the same flight are denoted as "transits" while those in between two different flights are "turns". The difference in the terminology is reflected by the difference in the nature of these activities. Turns are usually given more time on the ground since a more extensive clean up and more routine activities are planned for these. In the above example, the turns are of more than one hour duration while transits are given close to 30 minutes.

To this point the description has been centered on the functioning of a single aircraft. However passenger connections are also frequent and a "connection" will denote the event when passengers and their baggage are completely transferred from one aircraft to the other.

1.2 Problem definition

Having examined the nature of aircraft cycles and how these link to make a schedule, it is apparent that the planned time for a single activity within a cycle may not match entirely with the actual time it takes in practice. Each activity, such as block or ground activities in a cycle, has a corresponding average time span which varies as a result of some external random influences. Consequently, a schedule must allow for this variability.

The occurence of a delay in the schedule then means that the time allotted for one or for a series of activities was not sufficient to allow for the time required by these functions. Thus, in the cases when not enough time has been given for either a single activity or again for a sequence of activities, a delay is generated which is likely to be propagated over the network of activities. If the size of the delay is significant and if these delays are systematically caused by the schedule this is characterized as a potential weakness within the schedule or more specifically within the particular cycle. This thesis is primarily concerned with the study and analysis of these delays or the on-time performance of an airline schedule and the points of concentration are in the performance with respect to departures and arrivals of all scheduled passenger flight-legs, (although the model may be extended to cargo flights as well).

Normally, in attempting to overcome delays, some buffers or additional time for some activities may be built in the schedule or else some reassignment of resources may take place as in the aircraft reallocation for a flight schedule.

Large buffers would certainly increase the expected reliability of a schedule but with accompanying high costs (i.e., increasing the non-operating time of the aircraft which costs 10-30 million dollars). As for the reassignment solution, this requires frequent operational decisions often at the last minute. In practice, it is desirable to attain certain standards which are set by the airline for the reliability of a schedule and a compromise is made between both of these solutions.

One could follow a number of different approaches in analysing the on-time performance of a schedule. The most common one in the past has been in simulating the major activities constituting the schedule. This technique, of which a review of past work is described in Section 1.3, gives quite accurate results but at the cost of much data gathering and long computer processing and turnaround time. In this study, the approach taken was to divide the schedule into a number of aircraft cycles. The major variables or activities depicting each cycle were associated with some theoretical probability distributions and the validity of this association was tested using historical data.

Initially, it is important to be able to logically separate the major activities in a cycle or the major delay-causing activities. Basically a delay can be one of two types, either schedule induced, that is resulting from insufficient time allowances for the functioning of the previous legs and thus causing a delay in the leg in question, or non-schedule induced as would be the case within the station for activities other than the normal aircraft and passenger handling. In the latter case this delay-type is referred to as a station delay and examples would be an unexpected aircraft maintenance problem or a bad weather problem.

In producing any model some basic assumptions must be established concerning both the practical and theoretical aspects of the particular problem. The practical portion is pointed out in the next section for the sake of portraying some of the limitations of the model. The first ones are directed toward the assumed schedule structure required by the model, then some technical aspects are stated followed by a description of some factors that are, and others that are not considered by the model.

1.2.1 Practical assumptions. As mentioned earlier, we adopt the approach of considering an airline schedule to be made up of a set of cycles. The term "cycle" will be used throughout the text to denote both a complete cycle and a part of a cycle split where a turn of more than three hours occurs since the treatment is the same in either case.

The aircraft cycle is assumed to be composed of a number of flights (at least one) which are separated by "turns" and each such flight is made up of a number of flight legs which are separated by "transits".

In the case of the first leg in the cycle (the first leg of the flight) the aircraft is assumed to be available at scheduled time. However, any delays caused by station manoeuvring are taken into account within the station delays.

It is assumed that no aircraft leaves from an airport prior to its scheduled departure time, as this does not normally occur in practice.

Furthermore, for each cycle, the same physical aircraft is assumed to be in use for all the planned activities. No automatic re-allocations of aircraft are accepted by the model.

Delays caused by connecting passengers, where applicable, are not treated individually by this model due to the difficulty in acquiring the necessary data. Instead, they are included in the station delays. Since connection delays are not frequent and are a small percentage of all delays, the effect of including them with station delays is not of great consequence.

Our variables are assumed to be independent. This is a reasonable assumption since the activities are physically independent from one another. The block activities are mainly aircraft movement (taxi and flight), while the ground activities are for aircraft and cabin servicing.

The aspects described above have been thoroughly discussed with the schedulers in Air Canada who agree that the assumptions are both necessary and minimal in the sense that their effect on the results of the model will not be significant. The schedulers also agree that the assumptions normally hold in practice.

1.2.2 Mathematical notation and formulation. Throughout this study, the following mathematical reasoning and notation will be used. Consider one particular schedule, for instance, a schedule for the DC-9 aircraft type in the summer season (any breakdown is permitted). The schedule, according to the logic that has been followed so far, consists of m cycles. Each of these cycles is treated individually by the model and thus the model is cycle-independent. This is of great importance since the entire schedule need not be analysed completely at one time. Any number of cycles within that schedule (even one) may be studied. For this reason we need to consider only a single aircraft cycle and the method may be repeated as many times as there are cycles in the schedule.

Consider one aircraft cycle with k flights each having n_i legs for $i = 1, 2, \ldots, k$. Below is some notation used for the random variables of interest. The last in the list is the only exception. It is a set of constants defining the fixed scheduled times of departures:

B_{ij}	=	Block time variable for flight i, leg j	$i = 1, 2, \ldots, k$ $j = 1, 2, \ldots, n_i$
TS_{ij}	=	Transit time variable flight i, after leg j	$i = 1, 2, \ldots, k$ $j = 1, 2, \ldots, (n_i-1)$
TN_i	=	Turn time variable after flight i	$i = 1, 2, \ldots, (k-1)$

C_{ij}	=	Connection time variable where applicable after flight i, leg j	$i = 1, 2, \ldots, k$ $j = 1, 2, \ldots, (n_i-1)$
T_{ij}	=	Departure time variable for flight i, leg j	$i = 1, 2, \ldots, k$ $j = 1, 2, \ldots, n_i$
D_{ij}	=	Departure station delay variable for flight i, leg j	$i = 1, 2, \ldots, k$ $j = 1, 2, \ldots, n_i$
K_{ij}	=	Scheduled departure time for flight i. leg j	$i = 1, 2, \ldots, k$ $j = 1, 2, \ldots, n_i$

Using the above notation it is clear that an aircraft routing consisting of k flights has (k-1) turns, and each of the flights consisting of n_i legs ($i = 1, \ldots, k$) have (n_i-1) transits, (n_i-1) connections, n_i possible station delays and of course n_i scheduled departure times. One wishes to express the random variable T_{ij} in terms of the other random variables so that a distribution for T_{ij} may be derived, for each index i and each j. Consequently the distribution of the difference between T_{ij} and the fixed time K_{ij} may be obtained. This difference represents the departure delay random variable which need not be solely positive and which is the target of this thesis.

Suppose that a particular cycle has three flights and that the number of legs per flight are 2, 3, and 2 respectively. Then the sequence of events is the following:

Flight	Leg 1	Leg 2	Leg 3
1	$D_{11} \to B_{11} \to (TS_{11}, C_{11})$	$D_{12} \to B_{12} \to (TN_1, C_{12})$	
2	$D_{21} \to B_{21} \to (TS_{21}, C_{21})$	$D_{22} \to B_{22} \to (TS_2, C_{22})$	$D_{23} \to B_{23} \to (TN_2, C_{23})$
3	$D_{31} \to B_{31} \to (TS_{31}, C_{31})$	$D_{32} \to B_{32}$	

Consequently, the distribution of the departure time for any leg j of flight i will depend on the distributions of the component random variables over the previous legs. For the first leg, these random variables are the block time, ground time and scheduled departure time. Since some intersection between connection and ground times may occur, the maximum of the two variables is taken. The departure time for the second leg is the maximum of the corresponding variable resulting from the first leg and the station delay variable of leg 2 (when added to K_{ij}). Similarly, the departure time for other legs is also the maximum of the corresponding variable resulting from the cumulative effect of the previous legs, and the station delay variables. Following the activities sequentially, the following relation is obtained:

Let $D^*_{ij} = D_{ij} + K_{ij}$, i, j.

$$T^1_{ij} = K_{11} + \sum_{\ell=1}^{i-1} \left\{ \sum_{m=1}^{n_\ell - 1} \max\left[D^*_{\ell,m} , \max \left[(B_{\ell,m} + TS_{\ell,m}), (B_{\ell,m} + C_{\ell,m}) \right]\right] \right.$$

$$\left. + \max\left[D^*_{\ell,n_\ell}, \max\left[(B_{\ell,n_\ell} + TN_\ell), (B_{\ell,n_\ell} + C_{\ell,n_\ell}) \right]\right] \right\}$$

$$+ \sum_{m=1}^{n_{i-1}} \left\{ \max\left[D^*_{i,m}, \max\left[(B_{i,m} + TS_{i,m}), (B_{i,m} + C_{i,m}) \right]\right] \right\},$$

$$T_{ij} = \max\left[T^1_{ij}, D^*_{ij} \right]$$

$$n_0 = 0; \quad i = 1, 2, \ldots, k \quad j = 1, 2, \ldots, n_i \qquad (1.1)$$

The recursive relation is more useful computationally and is of the form given below:

$$T_{11} = K_{11} + D_{11}$$

$$T_{ij} = \begin{cases} \max\left[T_{i,j-1} + \max\left[(B_{i,j-1} + TS_{i,j-1}), (B_{i,j-1} + C_{i,j-1}) \right], D^*_{ij} \right] \\ \quad i = 1, 2, \ldots, k; \quad j = 2, 3, \ldots, n_i \\ \max\left[T_{i-1,n_{i-1}} + \max\left[(B_{i-1,n_{i-1}} + TN_{i-1}), (B_{i-1,n_{i-1}} + C_{i-1,n_{i-1}}) \right], D^*_{ij} \right] \\ \quad i = 2, 3, \ldots, k, \quad j = 1 \end{cases}$$

The first recursive relation holds for legs within the same flight while the second applies to the departure time of the first leg of a new flight.

If station delays include the connection delays, equation (1.2) reduces to the following:

$$T_{11} = K_{11} + D_{11}$$

$$T_{ij} = \begin{cases} \max\left[\left(T_{i,j-1} + B_{i,j-1} + TS_{i,j-1}\right), D^*_{ij}\right], & i = 1, 2, \ldots, k \quad j = 2, 3, \ldots, n_i \\ \max\left[\left(T_{i-1,n_{i-1}} + B_{i-1,n_{i-1}} + TN_{i-1}\right), D^*_{ij}\right], & i = 2, 3, \ldots, k \quad j = 1 \end{cases}$$

(1.3)

One of the assumptions that was made for the model was that all aircraft were not to leave prior to their scheduled departure time. This means that $T_{ij} \geq K_{ij}$ which is fixed from the schedule. Thus if $f_{ij}(t)$ is the unknown density of T_{ij} and T_{ij}^* is the departure time to be tested in the schedule we have:

$$T_{ij}^* = \begin{cases} K_{ij} & -\infty < T_{ij} \leq K_{ij} \\ T_{ij} & T_{ij} > K_{ij} \end{cases} \qquad i = 1, \ldots, k \quad j = 1, \ldots, n_i$$

(1.4)

Then, if $f_{ij}^*(t)$ is the probability density function of T_{ij}^* we have:

$$f^*_{ij}(t) = \begin{cases} 0 & -\infty < t < K_{ij} \\ \int_{-\infty}^{K_{ij}} f_{ij}(t)\,dt & t = K_{ij} \\ f_{ij}(t) & t > K_{ij} \end{cases} \qquad i = 1, \ldots, k \quad j = 1, \ldots, n_i$$

(1.5)

For simplicity of argument, let the transformation of T_{ij} to T_{ij}^* as shown above be represented by a function $G(T_{ij}, K_{ij})$. Equation (1.3) then becomes the following:

$$T^*_{11} = G((K_{11} + D_{11}), K_{11}) = K_{11} + D_{11} \qquad (\text{since } D_{11} \geq 0)$$

$$T^*_{ij} = \begin{cases} G\left\{ \max\left[T^*_{i,j-1} + B_{i,j-1} + TS_{i,j-1}, D^*_{ij}\right], K_{ij}\right\}, & i = 1, 2, \ldots, k;\ j = 2, 3, \ldots, n_i \\ G\left\{ \max\left[T_{i-1,n_{i-1}} + B_{i-1,n_{i-1}} + TN_{i-1}, D^*_{ij}\right], K_{ij}\right\}, & i = 2, 3, \ldots, k;\ j = 1 \end{cases} \tag{1.6}$$

It is now possible to obtain the distribution of the difference between the actual and the scheduled times of departure. What is of interest is to find the probability that the difference (or delay) will be at most a specified value $A \geq 0$ that is, what are the chances of having a delay of magnitude of A minutes or less? Since K_{ij} is a constant, one obtains:

$$P\left[(T_{ij}^* - K_{ij}) \leq A\right] = P\left[T_{ij}^* \leq A + K_{ij}\right] = \int_{-\infty}^{A+K_{ij}} f_{ij}^*(x)dx$$

$$i = 1, 2, \ldots, k; \quad j = 1, 2, \ldots, n_i; \quad A = \text{constant}.$$

Thus for A = 0 the probability of an on-time departure is obtained. In such applications, it is also of interest to obtain the above probability for many values of A, and in particular for increments of five minutes. In such a way it is possible to observe if each departure meets the airline standards in terms of its on-time performance, for argument sake, say 80% within five minutes.

Note that T_{ij}^* was introduced for convenience and more thorough comprehension of the logic. In actual fact, $T_{ij}^* = T_{ij}$.

1.3 Review of related work.

The important problems that come up in scheduling are not only in the on-time performance aspect of the schedule but also in determining the number and location of spare aircraft, in the required fleet size, the maintenance dispatch reliability of each aircraft type and other related problems which all have some impact on the on-time performance. Consequently in attempting to evaluate schedule reliability, it was the practice in the past to construct large multi-purpose models that encompassed many different aspects of schedule reliability so as to enable the airline in question to attain its pre-set standards.

The first of the on-time performance analyses was directed toward the aircraft assignment problem and was developed by Dequesnay in Air France (1961). This was achieved through the simulation technique as was the case for a large part of the past work on this subject. Then simulation models were reported by British European Airways (Jackson and Smith 1963), and by Air Canada. The Air Canada simulation, based on the Monte-Carlo technique, was developed by Lee and Fearnley (1963) and later revised by T. Batey and his co-workers (1967). This latter model served as a multi-purpose model that gave some answers to the following questions:

1) How would an increase or decrease in available fleet size affect the operating performance of a schedule?

2) How would changes in planned station turnaround or transit times affect the operating performances, ramp congestion and ramp overload?

3) What is the probable range of variation of departure and arrival times of flights?

4) What is the range of variation of the number of aircraft on the ramp at the same time at any station?

As may be observed from the above description, this model answers questions relating to many areas of schedule performance and operation and the results actually give quite an accurate picture of what is to be expected. The functioning of the above models require much data gathering and their nature and complexity is such that they have long computer turnaround time. Essentially these models are so complex that not much room is left for judgment and, of course, on-the-spot answers to specific questions are not possible. Most often, questions relating to schedule performance are not asked jointly but are usually asked selectively. These models take global views rather than more restricted ones.

Another problem that arose with these models was that some of the data required for their functioning wereonly available for the short term schedules and not for the longer-term ones.

As a result, airline companies have concentrated more effort in establishing more data gathering routines, setting up data bases and constructing simpler more direct models which help in answering more selective or specific areas in scheduling problems. In addition to this, instead of concentrating on complete aircraft schedules, the more recent models dealt mainly with aircraft cycles. Less data gathering requirements and quicker turnaround times resulted and much quicker answers were achieved especially with the recent interactive programming approach. A model of this sort has been developed recently by Tobin and Butfield (1970) in British European Airways. More recently a more analytical type model was developed by Peter Franke (1972) in Lufthansa which served as a core for the present study.

The advantages of these were mainly in the practical point of view. That is, in terms of simplicity (relative to the others), quick response and they were more specifically directed toward smaller problems.

1.4 Description of available data.

In Air Canada, much data relating to flight operations are kept on computer tapes for long periods of time. The data used in this study covered one year of history from January to December 1973. To begin with, the data were written in "packed" field so that they may occupy less space in storage. A program called DEPAC was written to retrieve the proper information, "de-pack" the data or translate them into a more easily readable form and restore the extracts on new tapes. In the future, such data will be easily accessible due to the recent development of a new data base within the company.

The original tapes recorded data containing information on individual flight-legs in some sequence for each day in one month of operation. One such tape represents a total of 16,000 to 18,000 records on flight-legs. From these only the scheduled passenger flights that had no irregularities were considered. The type of information contained in the original as well as in the new tapes is given in Appendix A.

As may be observed from the type of given data, enough information is given to re-construct an aircraft cycle as it occurred. In the case of an aircraft reassignment, no tracing back was possible and the cycle was broken.

The actual block minutes are explicitly given and accessing this variable is no problem. As for the transits or turns the data had to be manipulated so as to reconstruct the entire flight (for transits) or the entire cycle (for turns) and to calculate from the actual times of arrivals and departure of the legs, actual times for transits and turns. Since all times are in local times it was sometimes necessary to convert the times to GMT (Greenwich mean time). The transit and turn times together are equivalent to the single ground times considered in Franke's study.

Connection time data were however impossible to trace back from available tapes. Upon realizing this, other sources were searched, but there was difficulty to obtain these data and link them to specific departures in a cycle. Moreover, connections are not always planned in the schedules in terms of reserving time for this purpose. These problems together led to the decision of abandoning the connection times. As was pointed out before, delays caused by connections were included in the station delays.

From the definition of a station delay as given in the documentation of the original tapes (see appendix A), this variable was precisely what was needed. It is described as:

> "The delay minutes chargeable to the departure station. Arrived at by subtracting the arrival delay minutes from the departure delay minutes."

Thus these delays take into account any delay which is not caused by the previous leg. It is station-dependent rather than cycle-dependent.

1.5 Outline of the method of analysis.

Due to the difficulty in accessing connection time data as explained earlier, the model adopted for application with the Air Canada data is represented by equation (1.6) which relates the departure time variable T_{ij} of each leg to the principal time variables of a cycle excluding connection times, that is, to the block, transit and turn times as well as the station delay variables. Consequently, it is essential to have an associated probability distribution for each of these variates so as to evaluate the distribution of T_{ij} as the convolution of the other independent variables.

In the case of the three time variables (block, transit and turn times), each is analysed individually and theoretical distributions are fitted to the historical data. In observing the available data for one variable at a time, it is apparent that the data are not homogeneous. Included are measurements for all seasons, at all times of the day, for all aircraft types, all cities or city-pairs and in short for all the possible conditions that have occurred in the past.

The data need to be more finely defined and categorized so that the data in each classification may be claimed to behave in a similar way with the exception of existing random fluctuations of course. For instance, the block time variable must be broken down at least by city pair since for different city pairs the distributions of the variables are different. Moreover, the order of the pair of cities is important. For example, the mean block time for the Montreal to Paris flight-leg is known to be different from the corresponding mean block time of the Paris to Montreal leg due to the prevailing winds. One also proceeds to look for further breakdowns of the block time variable such as by season, aircraft type and so on, until a practically logical solution or one that produces the best type of fit to the theoretical distribution is obtained. This of course gives a more realistic description for the variables of the model, which in turn enables the analyst to test the reliability of a cycle with a higher degree of accuracy. This problem is extensively discussed in chapter two for each variable individually.

Once the distributions are obtained, the sample means and variances are retained since with the knowledge of these, the theoretical distribution (lognormal or normal) may be reconstructed. What is left to find is a distribution for the station delay variable. For many reasons which are discussed in the next chapter, an empirical probability distribution is used for this variate. The problem of categorizing into more "typical" distributions also comes into play.

At this point it is possible to construct, with our distributions, the set-up represented by model (1.6). In calculating convolutions however, one is faced with the problem of obtaining a distribution of the sum of three variables having theoretical distributions. The maximum of this distribution and the station delay (D^*_{ij}) must then be determined. Also, following the evaluation of the departure time distribution for each leg of the cycle, a type of truncation is required and the random variable associated with this modified distribution must be convoluted to other variables. Because of both of these aspects and of other problems which are discussed in chapter three, the distributions of the continuous variables are discretized and the numerical convolution of the variables to be summed is performed.

The analysis of the principal variables now follows in the next chapter.

CHAPTER 2

PARAMETRIC ANALYSIS

Finding the distribution of T_{ij} requires the knowledge of the distributions of B_{ij}, TS_{ij}, TN_i and D_{ij} (i=1, 2, ..., k; j = 1, 2, ..., n_i). These variables pertain to a leg within a flight of a given fixed cycle on a specified aircraft type. The cycle is scheduled for given days at specific departure (K_{ij}) and arrival (KA_{ij}) times. Due to changing schedules from period to period, historical data on such specifically defined random variables are not always available. For instance, if an aircraft (DC8 say) routed Montreal - Toronto - Vancouver in summer has a transit scheduled at Toronto at 8:35 a.m., then obtaining such specific transit time data from historical records may result in very small samples if any exist at all. Therefore all the TS_{ij}'s are classified in a way that for any given TS_{ij} pertaining to a particular flight-leg, we may say that it belongs to a pre-defined category. The same logic applies for B_{ij}, TN_{ij} and D_{ij}. In our example, we may seek the distribution of transits in Toronto on DC8's that have come from and are going to domestic stations in the summer months. This classification procedure enables one to generalize and thus adds to the flexibility of the model.

A large part of this chapter's discussion is centered on the problem of finding adequate categorization for each of our variables. Once this is established we then attempt to fit some known distribution functions to the continuous random variables B_{ij}, TS_{ij} and TN_i in their respective classes. Fitting such hypothetical distributions to empirical data is a standard type of problem in statistics for testing the goodness of fit. Initially, one must have an idea of what type of distribution may be appropriate for the data. Franke finds that his data on the B_{ij} variable are best approximated by a lognormal distribution and second best by the beta distribution. For the ground times (transits and turns in our case) the lognormal fitted best to the data but due to the small amount of skewness, he found that these could also be approximated by a normal distribution.

In Section 2.1 a discussion on fitting a lognormal distribution is given along with some properties of interest relating to that distribution. This is followed by the analyses of each variable B_{ij}, TS_{ij} and TN_i in Sections 2.3, 2.4 and 2.5 respectively. Finally Section 2.6 gives a brief discussion on the station delay variable D_{ij}.

2.1 On the lognormal distribution

Considering the nature of B_{ij}, TS_{ij}, TN_i ($i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, n_i$) it is apparent that these variables are all non-negative and that each is composed of a large number of smaller events occuring simultaneously. Another property from the practical viewpoint is that their distribution is positively skewed. For instance, the block time from Montreal to Toronto has a most likely value of one hour but the aircraft will more often be late than early because of such factors as wind or weather conditions, traffic congestion, or even an aircraft malfunction during the travel. Intuitively then, the block time distribution is skewed to the right. Similar considerations hold for transit and turn times.

Because of the above practical aspects of the distribution, we are tempted to fit a theoretical distribution having the above properties of positive skewness and non-negativity. There are several distributions that exhibit these properties. Among these, the lognormal distribution has been tested by Franke who found that it showed a good fit for all the variables, while the beta was a close second for B_{ij} and the normal for the ground variable (TS_{ij} or TN_i). As stated, our variables are composed of a large number of simultaneously occuring smaller events. According to a genesis of the lognormal distribution [Aitchison and Brown (1957)]:

> "We may suppose that at any point of time the existing distribution of the variate arises from a large number of causes which operate simultaneously."

This gives further intuitive support in fitting the lognormal distribution to our data.

The choice of the family of distributions is an open problem for the data analyst. One may proceed on the basis of intuitive considerations, histograms, moment properties, or other characterizations in order to discover the most suitable family of distributions. In our case the successful use of the lognormal distribution by Franke and the positive results obtained in the goodness of fit tests provide enough support for the utilization of this distribution.

A lognormal variable is basically one whose logarithm is normally distributed. Let X be a random variable such that

1) $0 < X < \infty$

2) $Y = \log X \frown N(\mu,\sigma^2)$

then, $X \frown \Lambda\ (\mu,\sigma^2)$

That is, X is a lognormal variate. The density function for X is of the form:

$$f(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\log x - \mu)^2\right] & x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

The mode, median and moments about the origin of X are as described below:

single mode at $x = e^{\mu-\sigma^2}$ (2.1)

median at $x = e^{\mu}$ (2.2)

j^{th} moment about the origin is

$$\lambda_j' = \int_0^\infty x^j \, d\Lambda(x) = \int_{-\infty}^\infty e^{jy} \, dN(y) = e^{j\mu + \frac{1}{2}j^2\sigma^2} \quad (2.3)$$

The mean α and variance β^2 are then obtained from (2.3) as:

$$\alpha = e^{\mu+\frac{1}{2}\sigma^2} \quad (2.4)$$

$$\beta^2 = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) \quad (2.5)$$

Now, let $\eta = \frac{\beta}{\alpha}$ be the coefficient of variation of X. Then by (2.4) and (2.5) we have:

$$\eta^2 = e^{\sigma^2} - 1 \quad (2.6)$$

In other words the coefficient of variation of X depends only on the variance σ^2.

Also, let γ_1 and γ_2 be the coefficients of skewness and kurtosis respectively. Then by (2.3), (2.4) and (2.5) we have:

$$\gamma_1 = \frac{\lambda_3}{\beta^3} = \eta^3 + 3\eta \quad (2.7)$$

$$\gamma_2 = \frac{\lambda_4}{\beta^4} - 3 = \eta^8 + 6\eta^6 + 15\eta^4 + 16\eta^2 \quad (2.8)$$

Because η depends solely on σ^2[by (2.6)] it follows that γ_1 and γ_2 are always positive and the degree of skewness of the distribution is an increasing function of the variance σ^2.

As we know in the normal case the sum of two independent normal random variables is also a normal random variable with a mean equal to the sum of means and variance equal to the sum of variances. Let X_1 and X_2 be two independent lognormal variables with parameters (μ_1,σ_1^2) and (μ_2,σ_2^2) respectively. In general the distribution of $X_1 + X_2$ cannot be obtained in closed form. However, the distribution of the product X_1X_2 is lognormal with parameters $(\mu_1+\mu_2,\ \sigma_1^2+\sigma_2^2)$. This is true because

$$Y = \log X_1X_2 = \log X_1 + \log X_2$$

is a sum of two independent normal variables with means μ_1, μ_2 and variances σ_1^2 and σ_2^2 respectively.

2.2 Goodness of fit test.

In testing the goodness of fit of certain distribution functions to our data, a modification of the Kolmogorov-Smirnov test is used. Essentially, the statistic D_n used for this test is the maximum absolute deviation between the specified continuous hypothetical cumulative distribution $F_0(x)$ and the sample (empirical) distribution $S_n(x)$:

$$D_n = \sup_x \left| S_n(x) - F_0(x) \right|$$

When comparing this test to the Chi-Square test (based on the comparison of observed frequencies to the expected frequencies on specified classes of data) two observations may be made [M. G. Kendall and A. Stuart, (1973)]:

1) "D_n is a very much more sensitive test for the fit of a continuous distribution."

2) The Kolmogorov-Smirnov test requires smaller sample sizes than for the Chi-Square test to achieve the same efficiency:

"D_n asymptotically requires sample sizes to be of order $n^{4/5}$ compared to n for the X^2 test and is asymptotically very much more efficient. In fact the relative efficiency of X^2 will tend to 0 as n increases."

Because of the above reasons of continuity of the distribution and the efficiency using smaller sample sizes, it was thought best to use the Kolmogorov-Smirnov test.

As may be observed from the properties of the lognormal distribution, the goodness of fit test in this case is equivalent to testing the logarithms of the data in the sample for normality. We are thus testing to find out whether the parent distribution is normal. Since the two parameters μ and σ^2 are estimated from the sample by $\bar{y}, s_y^2$, respectively, the critical values of D_n used are those given by Lilliefors (1967). These values are tabulated for α = .01, .05, .10, .15 and .20 and for sample sizes of n = 1, 2, ..., 20, 25, 30, whereas for larger values of n an approximation is given.

To test for lognormality, the natural logarithms of the sample data are calculated. From these, some outliers are deleted, that is, values lying on either side of three standard deviations away from the mean. This is done so as to attempt to eliminate misleading values arising from such irregular events as strikes. Since these data are obtained from historical records we do not have an experimentally controlled situation. Abnormalities are thus expected, and evident outliers are rejected.

The sample parameters are then adjusted so as not to include the outlier values. The resulting sample values are tested for normality through the Kolmogorov-Smirnov test with Lilliefors' critical values as described above.

The D_n statistic is first compared to the critical value D_n* for $\alpha = .20$. If $D_n < D_n*$ the hypothesis is not rejected and one does not need to go further. Otherwise the comparison is made for other levels of $\alpha = .15, .10, .05$ and $.01$ in that order and if $D_n > D_n*$ for $\alpha = .01$ the hypothesis is "rejected" for our purpose. A sample output showing the results of this procedure is given in the discussion of the block time analysis.

To do these tests, subroutines were written in Fortran IV for an IBM 360 model 50 computer. They were written in a generalized form to serve for the analyses of all three variables. These subroutines are called by a major one that computes the means and standard deviations of the raw data as well as the logged data. The other subroutines delete the outliers, and perform the Kolmogorov-Smirnov test. The Kolmogorov-Smirnov routine accepts as input the values corresponding to $S_n(x)$ as well as the matrix (Z) of critical values. The output consists of the D_n value and the D_n* value associated with the highest α value for which $D_n \leq D_n*$. If such a D_n* does not exist in the table, the word "REJECT" appears.

2. Block time.

The block time data were accessed directly from existing tapes containing flight-type information on the Air Canada flight legs. These legs correspond to a large variety of city pairs on various aircraft types at all times of the year while under many different conditions such as wind factors, differences in traffic conditions and so on. In order to sample the block time data, the legs had to be sorted initially into categories which would contain data with similar characteristics or data that belong to typical populations.

As it was mentioned earlier (Section 1.5), the most logical first breakdown is by city-pair, that is, by departure and arrival stations. The Montreal to Toronto block times for instance, are certainly different in nature and in magnitude than the Montreal to Paris ones. Due to prevailing winds and the routes taken, the direction is also an important factor. Thus the order of city pair is taken into account.

The difference in performance of various aircraft types is quite significant as far as block times are concerned especially for long-haul legs. Toronto to London, England has an average block time of 401 minutes on a 747 and 428 minutes on a DC8L. Both these averages were calculated from data obtained from the same summer months of July and August and both aircraft were scheduled to leave at identical departure times (8:00 p.m.). The aircraft type distinction is therefore necessary.

Another important factor was the seasonal behavior of the data and the winter and summer seasons were treated separately. These "seasons" were defined to be of six months each, the summer ranging from May to October inclusive and November to April is defined as the winter season.

One extra refinement was made. It was observed, especially when either the departure or arrival stations were major airports, such as Toronto, Montreal, New York or Chicago, the block times were different for departure times in different sections of the day. This may be explained by the possibility of having heavy traffic congestion at specific times of the day when the incoming and outgoing rush occurs. Such situations would affect the taxi time since the aircraft would be waiting for a runway for departure. Upon arrival, if it happens to be a peak period at the station, the aircraft may be forced to circle above the airport until permission is given to land. Thus, the holding time which is part of the block time is affected.

The sections of the day were selected according to the suggestions made by the Schedulers in Air Canada. These sections typify the peaks and valleys in the incoming and outgoing traffic of airports in general. The ranges of hours were 0:00 - 5:59 hrs, 6:00 - 8:59 hrs, 9:00 - 11:59 hrs, 12:00 - 14:59 hrs, 15:00 - 19:59 hrs, and 20:00 - 23:59 hrs which were labelled by a section of day indicator ranging from 1 to 6 in the order shown above.

It is worth mentioning that most of Air Canada's statistical summaries on block times are classified in the way so far described. That is, by departure and arrival city pairs in their respective order, by aircraft type, season and section of day for the departure. This consistency offered further support for adopting this categorization.

The data (on tape) are then sorted according to the above groups of data so that collecting the data for each category became a simple matter.

The program first collects all observed block times within a specific group then tests the sample for lognormality as well as for normality in some cases. The results are the following:

1. Grouped by city-pair, aircraft type and season:

   72% of samples did not reject lognormality.

2. Grouped by city pair, aircraft type, season and time of day:

   a) Summer: 85% of samples did not reject lognormality

   82% of samples did not reject normality

   b) Winter: 90% of samples did not reject lognormality.

The summer data alone consisted of a total of 743 samples of which 631 did not reject lognormality. The number of samples tested in the overall being approximately 1,500 resulted in very large printouts. Consequently these are not included in this thesis but they may be obtained from the author.

The program that collected, grouped and tested the block data made use of an existing sorting system as well as of a Fortran IV main program also on an IBM 360 Model 50 written for the analysis. This program collects the block data for one category at a time then calls the series of subroutines which test the sample (as discussed in section 2.2). A sample output is shown on the next page where each sample of data is titled by the city pair, aircraft type, flight number and scheduled time of departure. (The flight number is given but corresponds to a specific departure time.) The maximum deviation as described in the Kolmogorov-Smirnov procedure is shown as D-MAX along with the associated critical value. On the extreme right shows at what level of confidence α, the test

```
      YEG-YYZ     AC-TYPE IS B    FLIGHT NO. 164.   STD=    110.
SAMPLE SIZE=  60  MEAN=      205.5666656  SAMPLE S.D.=      8.7688999

LOGGED DATA:      MEAN=        5.3248568  SAMPLE S.D.=      0.0422014

NUMBER OF OUTLIERS=   0 ADJUSTED MEAN=       5.3248568 ADJUSTED S.D.=       0.0422014

D-MAX.=       0.1048150  CRITICAL VALUE=      0.1143821                                   0. 5. 0. 0. 0.


      YFC-YHZ     AC-TYPE IS D    FLIGHT NO. 140.   STD=   1265.
SAMPLE SIZE=  58  MEAN=       37.5689545  SAMPLE S.D.=     10.8273554

LOGGED DATA:      MEAN=        3.6024618  SAMPLE S.D.=      0.1936192

NUMBER OF OUTLIERS=   2 ADJUSTED MEAN=       3.5702817 ADJUSTED S.D.=       0.0879856

D-MAX.=       0.1302555  CRITICAL VALUE=      0.1377730                                   1. 0. 0. 0. 0.


      YFC-YHZ     AC-TYPE IS D    FLIGHT NO. 644.   STD=    980.
SAMPLE SIZE=  58  MEAN=       36.6379242  SAMPLE S.D.=      4.0161762

LOGGED DATA:      MEAN=        3.5953741  SAMPLE S.D.=      0.1069234

NUMBER OF OUTLIERS=   0 ADJUSTED MEAN=       3.5953741 ADJUSTED S.D.=       0.1069234

D-MAX.=       0.1286829  CRITICAL VALUE=      0.1353768                                   1. 0. 0. 0. 0.


      YFC-YQB     AC-TYPE IS F    FLIGHT NO. 655.   STD=   1195.
SAMPLE SIZE=  58  MEAN=       70.7930908  SAMPLE S.D.=      6.7427835

LOGGED DATA:      MEAN=        4.2553549  SAMPLE S.D.=      0.0944309

NUMBER OF OUTLIERS=   0 ADJUSTED MEAN=       4.2553549 ADJUSTED S.D.=       0.0944309

D-MAX.=       0.0806792  CRITICAL VALUE=      0.0966415                                   0. 0. 0. 0.20.


      YFC-YSJ     AC-TYPE IS F    FLIGHT NO. 632.   STD=    715.
SAMPLE SIZE=  60  MEAN=       26.3999939  SAMPLE S.D.=      4.9306965

LOGGED DATA:      MEAN=        3.2542439  SAMPLE S.D.=      0.2051723

NUMBER OF OUTLIERS=   1 ADJUSTED MEAN=       3.2687578 ADJUSTED S.D.=       0.1730952

D-MAX.=       0.1020397  CRITICAL VALUE=      0.1048020                                   0. 0.10. 0. 0.
```

TABLE 2.1: SAMPLE OUTPUT FOR BLOCK TIME DISTRIBUTIONS

```
 AC-TYPE IS B    FLIGHT NO. 164.    STD=    110.
 MEAN=     205.5666656  SAMPLE S.D.=     8.7688999

 MEAN=       5.3248568  SAMPLE S.D.=     0.0422014

S=   0 ADJUSTED MEAN=       5.3248568 ADJUSTED S.D.=      0.0422014

048150  CRITICAL VALUE=      0.1143921                          0. 5. 0. 0. 0.


 AC-TYPE IS D    FLIGHT NO. 143.    STD=   1265.
 MEAN=      37.5689545  SAMPLE S.D.=    10.8273554

 MEAN=       3.6024618  SAMPLE S.D.=     0.1936192

S=   2 ADJUSTED MEAN=       3.5702610 ADJUSTED S.D.=      0.0879856

C2555  CRITICAL VALUE=       0.1377730                          1. 0. 0. 0. 0.


 AC-TYPE IS D    FLIGHT NO. 644.    STD=    980.
 MEAN=      36.6379242  SAMPLE S.D.=     4.0161762

 MEAN=       3.5953741  SAMPLE S.D.=     0.1069234

S=   0 ADJUSTED MEAN=       3.5953741 ADJUSTED S.D.=      0.1069234

286829  CRITICAL VALUE=      0.1353768                          1. 0. 0. 0. 0.


 AC-TYPE IS F    FLIGHT NO. 655.    STD=   1195.
 MEAN=      70.7930908  SAMPLE S.D.=     6.7427835

 MEAN=       4.2553549  SAMPLE S.D.=     0.0944309

S=   0 ADJUSTED MEAN=       4.2553549 ADJUSTED S.D.=      0.0944309

06792  CRITICAL VALUE=       0.0966415                          0. 0. 0. 0.20.


 AC-TYPE IS F    FLIGHT NO. 632.    STD=    715.
 MEAN=      26.3999939  SAMPLE S.D.=     4.9306965

 MEAN=       3.2542439  SAMPLE S.D.=     0.2051723

S=   1 ADJUSTED MEAN=       3.2687578 ADJUSTED S.D.=      0.1730952

20397  CRITICAL VALUE=       0.1048020                          0. 0.10. 0. 0.
```

was not rejected (1, 5, 10, 15, or 20%). If the D-MAX value was greater than the critical value at the 1% level, then the word "REJECT" appeared instead of the α-level. The total number of non-rejects over the total number of samples tested gave us the ratio of the samples that did not reject.

2.4 Transit Times.

The data collection for transit times was a more difficult task than for block times. Ground times are not explicitly recorded on tapes. Enough information was collected for each leg so as to reconstruct the flights and extract the transits times calculated as the difference between the actual time of departure of one leg and the actual time of arrival of the previous leg.

The transit time, for our purpose, is basically the time required for the performance of a minimal number of activities to enable the aircraft to be prepared for the next leg. This will be referred to as the minimal transit time.

In breaking down the transit time variable into classificatory categories we may examine the types of servicing required in various situations. The work categories for servicing in a ground stop (transit or turn) may be divided into two types. The external servicing includes aircraft handling and preparation as well as loading and unloading of baggage, freight and other objects. These functions are quite essential and are done no matter the amount of available ground time. The second type of functions that vary largely with available time and resources is internal or cabin servicing. These include such cabin functions as:

1) Removing litter and used literature (Cabins, galleys, washrooms, flight deck)
2) Removing and replacing used equipment (Cabins, galleys, & washrooms)
3) Cleaning (Cabins, galleys, washrooms and flight deck).

The time required for such servicing will of course depend greatly on the aircraft size and therefore on the aircraft type. Also the equipment and resources available will depend on the volume of traffic and on the size of the particular station. The most necessary initial breakdown is thus by aircraft type and by station of transit.

A transit, being a stop between two consecutive legs of one flight, may vary in time according to the station the aircraft comes from and the station it is going to. Consider the following transits:

Flight No.	Aircraft Type	From Station	Transit (.)Sched.time	To Station
790	DC8	Los Angeles	Toronto (60 mins.)	Montreal
792	L-1011	Los Angeles	Toronto (70 mins.)	Montreal
148	L-1011	Vancouver	Toronto (35 mins.)	Montreal

In Flights 790 and 792, the aircraft are scheduled to undergo precisely the same routing but a difference of 10 minutes appears in the scheduled transit times in Toronto. This is due to the difference in aircraft types. Flights

148 and 792, however, are both on an L-1011 aircraft. The two flights are similar in that Los Angeles - Toronto and Vancouver - Toronto are approximately of the same distance and the scheduled destination following Toronto is Montreal in both cases. A difference of 35 minutes appears in the transit times scheduled for Toronto which is purposely planned to allow passengers to pass through customs when arriving from Los Angeles, being a Transborder Station. Schedulers take into consideration the "service" type prior to and after the stop in alloting a transit time. The services considered are:

1 = Domestic

2 = Atlantic (European)

3 = South (South bound destinations as defined in Air Canada)

4 = Transborder (U.S.A.)

Each such pair [(1, 1) and(1, 2) are each considered as a pair] is treated separately since the passenger and baggage handling procedures are somewhat different.

Domestic legs, being city-pairs within Canada, range from short Montreal - Ottawa (94 miles) trips to long-haul legs such as Montreal - Vancouver (2,287 miles). Thus on domestic routes the distances covered prior to and after the transit may have an impact on the transit time since more aircraft preparation or more cabin activities may have to be performed.

The distribution of TS_{ij} is thus determined with the same classification as B_{ij} except for the time of day. Moreover, B_{ij} is associated with a city pair which corresponds to the city triplet associated with TS_{ij}. Here, rather than considering the precise triplet, we consider the city of transit and we classify the two stations before and after the stop in classes of of "similar" legs - that is, differentiated according to the service type and distance category (0 - 500, 501 - 1000 and > 1000 miles)for Domestic legs.

The distribution sought for our TS_{ij} variable is one that would typify the minimum transit time variable, that is, the minimal time required on the ground when the available time is small. Being under time pressure, the ground activities done are the most necessary ones.

When the available time (scheduled time of departure - actual time of arrival of previous leg) is large the actual transit time is approximately equal to the time at hand. Even if all activities are completed, the aircraft cannot leave prior to its scheduled departure time (as stated in the practical assumptions in section 1.2.1). Thus the measured transit time (actual time of departure - actual time of arrival of previous leg) may be greater than the actual transit, and is thus misleading. Moreover, when much time is available the variable internal servicing may be done completely. Consequently in considering all transits we risk obtaining a misleading distribution characterized by two different events or distributions.

A. The distribution of transit times when the available time is large which is approximately the same as the distribution of available times. This is of no interest in the model.

B. The minimum transit time distribution.

When transits with any size of available time (A. and B. above) are tested for lognormality, about 78% of the samples do not reject the test. The same type of results hold true when testing for normality.

If an aircraft arrives late (after the scheduled time of arrival of the preceding leg), it does not have much ground time at hand. This increases the likelihood of incurring a delay. In such cases the ground activities performed in reality are kept to a minimum. Therefore, considering the minimal transit time variable makes the model more realistic and sensitive in the instances of most interest, when a significant delay is likely to occur.

Much experimentation was done for defining and thus accessing these minimum transit times from historical data. Initially, some minimum standards for transit times, obtained from Air Canada documentation were used to define the upper allowable limit of the available times. For available times greater than that limit, it could be said that the aircraft had much time at hand in which case a distribution of type A above would be obtained for the TS_{ij}. These standards were classified mainly by aircraft type:

20 minutes: Dc9, D9S, Viscount

30 minutes: DC8, D8S

40 minutes: L-1011

45 minutes: 747

Using the above standards, all actual transit times within one category for which the available time was less than or equal to the minimum standard were collected and tested for lognormality and normality. The results were not very encouraging. For some transit classifications, no such data were found in the sample while for other groups the samples were too large. It was thought that some of these large samples did not really typify the minimum transit time variable. This, together with the fact that some classifications were not sampled from at all led to the conclusion that this straightforward way of defining the available time limit was not adequate.

It was apparent that these limits must be calculated according to the individual sample, rather than treating all samples in a likewise manner.

This concept is discussed by Franke who proposes the implementation of one of three methods depending on the particular sample. The method employed above is a modification of one of these three. For our analysis it was finally decided to use another of the three, in a modified form. For each transit category we compute the limit for available time of the sample of n observations as follows:

1. AVL_i = available time of observation i, i = 1, 2, ..., n
   = (scheduled time of departure - actual time of arrival of previous leg)

   ts_i = i^{th} observed transit time in sample, i = 1, 2, ..., n.

2. $$\overline{AVL} = \sum_{i=1}^{n} \frac{AVL_i}{n} \quad ; \quad s^2_{AVL} = \sum_{i=1}^{n} \frac{(AVL_i - \overline{AVL})^2}{n-1}$$

   $$\text{Limit}_{AVL} = (\overline{AVL} - s_{AVL}). \qquad (2.9)$$

3. If ($AVL_i \le \text{Limit}_{AVL}$) and ($ts_i \ge AVL_i$) then select observation ts_i.

In step 3 all transits for which the actual time ts_i is at least as large as the available time AVL_i and this one is smaller than or equal to the limit, are taken in the new sample of minimal transit times. This situation depicts a transit when it is truly under time pressure.

In performing the above steps, it was noticed that the limits calculated by (2.9) were sometimes negative and thus very small samples were obtained. The following check was thus made:

4. If $Limit_{AVL}$ < 5 mins., then $Limit_{AVL}$ = max ($\overline{AVL}$, 5 mins.)

In this fashion, the available time limit was never allowed to be less than 5 minutes. This ensures that all classes of data are represented in the samples tested and that these are not too small.

The above precedure led to the following results:

Hypothetical Distribution	Total No. of Samples	No. of Samples Rejected	% of Samples Not Rejected
Lognormal	146	22	84.93%
Normal	146	31	78.77%

The level of significance for these tests was the same as in the case of block times, that is, $\alpha = .01$.

2.5 Turn Times

In the case of the turn times the same problem occured for the data collection as in the transit times. The aircraft cycles had to be reconstructed from the leg-data and to achieve this all time measurements were converted to a standard (GMT) so as to overcome differences due to the many time zones. Following the aircraft through the cycle made it possible to calculate turn times as the difference between the actual departure time of the first leg of one flight and the actual arrival time of the last leg of the preceding flight.

In selecting the appropriate classification for grouping the turn times in classes of similar types, the same type of reasoning used in the transit discussion was applied. Although the type of servicing in a turn is more of a "major" or "intermediate" one compared to the "minor" servicing in a transit, the type of classification remains the same although more pronounced for turns. These are grouped by aircraft type, station of turn, and services prior to and after the turn.

Again, the distributions of the minimal turn times were sought. Since the turn and transit analyses were performed simultaneously, the same sequence of experiments were performed with similar (but more optimistic) results. The standards for minimum turn times were classified differently than the transit standards. They were classified by aircraft type as well as by type of station which explains the more optimistic results.

Minimum Standards for Turns (Minutes)			
Aircraft Type	Domestic Station	U.S. Station	International Station
747	90	90	90
L-1011	90	90	90
D8S	60	75	90
DC8	45	60	90
D9S	45	45	75
DC9	45	45	75

The procedure used was identical to the transit procedure and gave very favorable results. However, .5 rather than one standard deviation was subtracted from the average available time in order to obtain the limit, so as to increase the sample sizes which are considerably smaller than transits. Moreover, when negative or small limits occured, the minimum allowable limit of 5 minutes was increased to 15 minutes since turns are much longer than transits. Also a turn of less than 15 minutes rarely occurs. Out of a total of 114 samples tested only 10 rejected lognormality. Thus 91.2% of samples did not reject the lognormal distribution (α = .01).

2.6 Station Delays.

Station delays play an important role in determining the on-time performance of a cycle since these are mainly station-dependent. Because of the differences of equipment, crew, mean volume of traffic in each station, this type of delay may be more likely to occur in some stations than in others. Examples of such delays are the mechanical, sales or passenger delays occuring at a given station.

Station delays (defined in section 1.4), D_{ij}, are measured in minutes rounded to the nearest minute. If no such delay occurs, then $D_{ij} = 0$ for that particular leg. Empirical observations have shown that these delays vary anywhere between 0 and up to about 180 minutes (3 hours). Franke attempts to fit hypothetical distribution functions, but due to the difficulties encountered, an empirical distribution was used. Adopting the empirical sample distribution does not allow the generalization obtained in having a parametric distribution which could be updated with the use of new data. However when no good approximation is found through a theoretical distribution, it is best to use the one at hand - the empirical one. Table 2.2 gives frequency counts of station delay minutes for selected stations. The most logical breakdown which was agreed upon is by aircraft type, departure station and time of day within a given season. In our example the observable frequencies for three stations YXE, YWG, and YVR (Saskatoon, Winnipeg

TABLE 2.2 SELECTED DISTRIBUTIONS OF STATION DELAYS

YXE: D, 3, 92, 31*

MINS.	FREQ.
1	6.
2	7.
3	13.
4	10.
5	8.
6	2.
7	3.
9	3.
10	1.
14	2.
15	1.
18	1.
19	1.
31	1.
35	1.
92	1.

YWG: D, 1, 31, 3

MINS.	FREQ.
1	5.
2	4.
3	3.
4	1.
5	2.
6	2.
8	4.
10	1.
11	1.
12	1.
14	1.
22	1.
23	1.
47	1.

YVR: B, 6, 36, 0

MINS.	FREQ.
3	2.
5	1.
7	1.
8	2.
10	3.
11	1.
12	3.
13	2.
15	1.
17	2.
19	1.
20	3.
22	1.
25	1.
28	1.
29	1.
30	1.
32	1.
33	2.
35	1.
50	1.
59	1.
80	1.
140	2.

YXE: D, 5, 93, 23

MINS.	FREQ.
1	17.
2	16.
3	9.
4	8.
5	5.
6	4.
7	2.
8	2.
9	2.
10	1.
11	1.
14	1.
15	1.
29	1.

YWG: D, 2, 124, 65

MINS.	FREQ.
1	12.
2	18.
3	9.
4	3.
5	2.
6	3.
7	4.
8	1.
9	2.
10	2.
11	1.
30	1.
76	1.

YVR: D, 2, 31, 11

MINS.	FREQ.
3	2.
4	3.
5	2.
8	2.
11	2.
13	1.
14	1.
15	2.
17	1.
19	1.
96	1.
120	1.
150	1.

* YXE = Station code, D = aircraft type code,

3 = Section of day index, 92 = Total number of delays,

31 = Total number of zero delays.

and Vancouver) are given on specified aircraft types and sections of day for a summer month. The example titled "YVR B 6 36 0" represents departures from the Vancouver Airport on an aircraft of type B (D8S) during the 6th period of the day (20:00 - 24:00 hrs). The sample size is n = 36 from which none have delays = 0 minutes. Beneath the title are two columns, the left denoting the number of delay minutes and in the right are the corresponding frequencies. The sections of day used are consistent with the breakdown given in Section 2.3.

The empirical distributions used were calculated by computing the actual probabilities of occurence for each size of delay minutes in increments of one minute from 0 to 149 and any delay of 150 minutes or more are treated in the 150th interval. This limit is more than sufficient since greater delays are very rare.

Empirical distributions of this type were constructed for each category of data. The classification is consistent with the ones for block time and transit times, that is by aircraft type, season, station of departure and time of day.

CHAPTER 3

DISCUSSION OF MODEL

3.1 Formulation and practical logic

As previously explained, the mathematical models of Section 1.2.2 are applicable to individual cycles. The entire airline schedule is considered to consist of m cycles. Therefore, the procedure and analysis to be adopted for a single cycle will have to be repeated m times if one is interested in the performance of the entire schedule.

Consider one particular aircraft cycle having k flights and let n_i be the number of legs in each flight. The scheduled information relating to that cycle pertains to a particular season and a specified aircraft type. The information consists of the following:

F_i = ith flight serial number,

CTY_{ij} = city of departure pertaining to the jth leg of flight i,

CTY_{k,n_k+1} = city of arrival of the last leg of the cycle,

K_{ij} = scheduled departure time of jth leg of ith flight,

KA_{ij} = scheduled arrival time of jth leg of ith flight.

From the above information some simple calculations of the scheduled block times and the scheduled ground (transit or turn times) may be performed:

Scheduled block time (i, j) = $KA_{ij} - K_{ij}$,

Scheduled transit time (i, j) = $K_{i,j+1} - KA_{ij}$,

Scheduled turn time (i) = $K_{i+1,1} - KA_{i,n_i}$.

For a fixed K_{11}, K_{ij} may be also calculated from the scheduled block, transit and turn times as follows:

$$K_{ij} = \begin{cases} \text{Scheduled block } (i, j-1) + \text{scheduled transit } (i, j-1) + K_{i,j-1}, \\ i = 1, 2, \ldots, k; \quad j = 2, 3, \ldots, n_i \\ \\ \text{Scheduled block } (i-1, n_{i-1}) + \text{Scheduled turn } (i-1) + K_{i-1,n_{i-1}}, \\ i = 2, 3, \ldots, k; \quad j = 1 \end{cases}$$

As it was mentioned in Section 1.2.2 we seek the distribution of T_{ij}^* as given by equation (1.6). In view of this equation, the distribution of T_{ij}^* depends on the distributions of D_{ij} and either $T_{i,j-1}^*$, $B_{i,j-1}$ and $TS_{i,j-1}$ or $T_{i-1,n_{i-1}}^*$, $B_{i-1,n_{i-1}}$ and TN_{i-1} as is explained in Section 1.2.2. In chapter two, it was found that the variables B_{ij}, TS_{ij} and TN_i fit the lognormal distribution for every i, j. Moreover, instead of testing the goodness of fit of each B_{ij}, TS_{ij} and TN_i separately, a classification procedure was adopted. Thus, all B_{ij} that belong to the same class are said to belong to the same population. The parameters estimated (according to the procedures in Chapter 2) for each class of the block, transit and turn variables have been stored in a computerized file system. In this manner, if we seek the distribution of a particular TN_i, say, we must first determine the class to which that TN_i belongs. This TN_i is associated with a given station, a specific aircraft type and is scheduled for a given season. Also the legs prior to and following the turn, belong to defined service types. This is identified by the pair (i, j), i, j = 1, ..., 4 as it was explained in Section 2.4. This information completely specifies the class to which TN_i belongs and, in turn, permits the retrieval of the estimated parameters pertaining to that class from the file system (see the schematic representation in Table 3.1). The same procedure applies to the transit and block variables:

TN_i = f (season, aircraft type, station, service type of previous leg, service type of following leg),

TS_{ij} = f (season, aircraft type, station, service type and distance of previous leg, service type and distance of following leg),

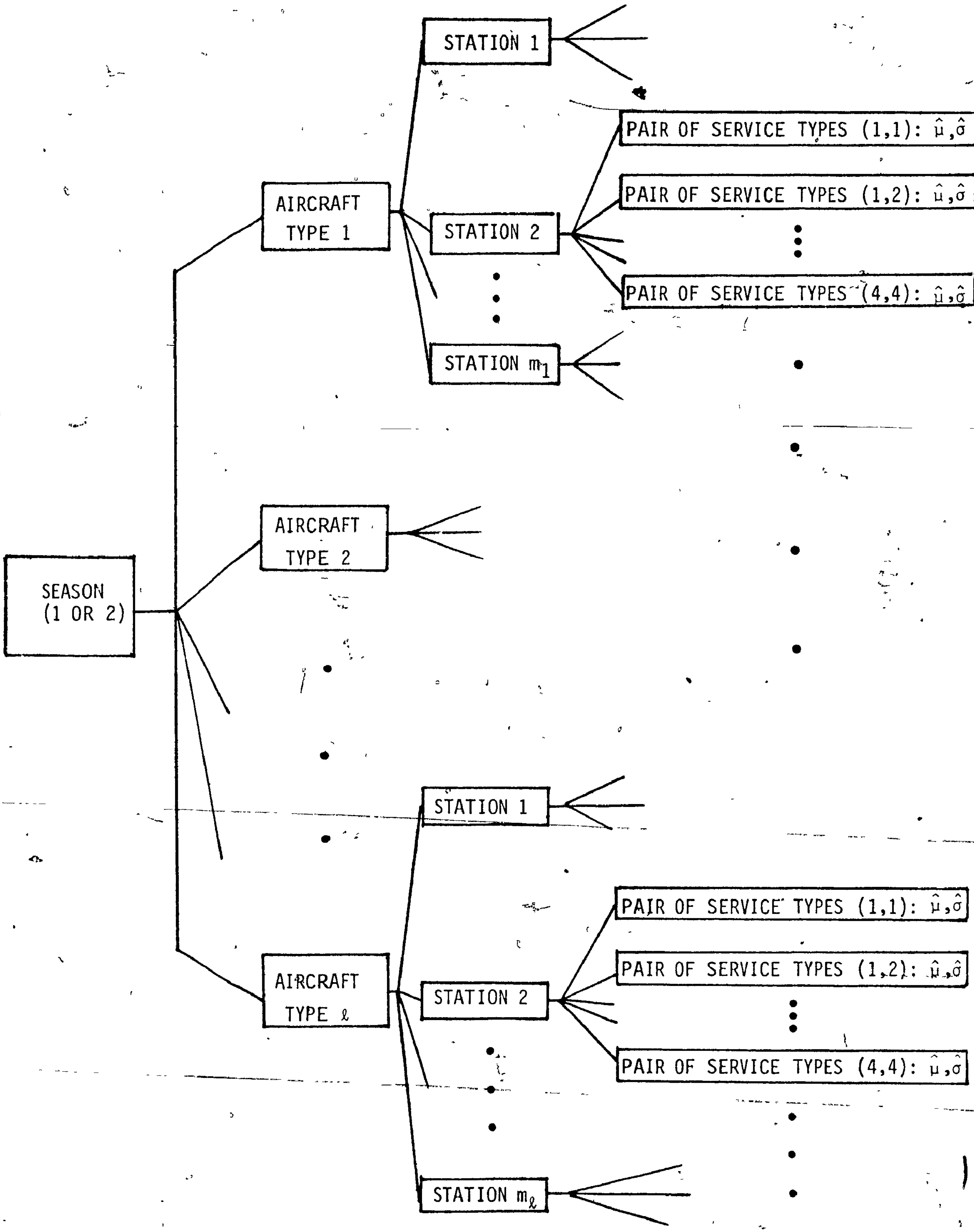
B_{ij} = f (season, aircraft type, station of departure, station of arrival, section of day for departure),

where f indicates that the corresponding variable has been classified according to the attributes listed as arguments of f. On the basis of this classification system, TN_i, TS_{ij} and B_{ij}, $i = 1, \ldots, k$; $j = 1, \ldots, n_i$ can be classified into one and only one class.

The file system containing the estimated parameters for the classes of the variables is constructed in such a way as to allow quick retrieval of the required parameters.

Basically, the system consists of separate files for block transit and turn parameters. Each file is first divided into two season sections and then subdivided by aircraft type. Since schedules are presently prepared according to these criteria, this allows quick reference to the estimates of interest. These files will be revised yearly so that the lag between the time when the estimates are obtained from historical data and the time when the model is implemented does not exceed one year.

Table 3.1



SEASON (1 OR 2)
AIRCRAFT TYPE 1
AIRCRAFT TYPE 2
AIRCRAFT TYPE ℓ
STATION 1
STATION 2
STATION m_1
STATION 1
STATION 2
STATION m_ℓ
PAIR OF SERVICE TYPES (1,1): $\hat{\mu},\hat{\sigma}$
PAIR OF SERVICE TYPES (1,2): $\hat{\mu},\hat{\sigma}$
PAIR OF SERVICE TYPES (4,4): $\hat{\mu},\hat{\sigma}$
PAIR OF SERVICE TYPES (1,1): $\hat{\mu},\hat{\sigma}$
PAIR OF SERVICE TYPES (1,2): $\hat{\mu},\hat{\sigma}$
PAIR OF SERVICE TYPES (4,4): $\hat{\mu},\hat{\sigma}$

The estimates may thus follow the trend of changing conditions, such as the implementation of new routes, new or improved aircraft types, and changes in service requirements. The distribution tests should be repeated to ensure that the distributions of the possibly changed variables still fit the lognormal distribution. These revisions are quite simple with the availability of the programs discussed in Chapter 2 which perform all the necessary steps at one time.

For each variable B_{ij}, TS_{ij} and TN_i, we now have its scheduled time as well as the distribution of each actual time as obtained from historical data:

$$\text{scheduled block } (i, j) \longleftrightarrow B_{ij} \frown \Lambda(\hat{\mu}_{B_{ij}}, \hat{\sigma}_{B_{ij}}),$$
$$\text{scheduled transit } (i, j) \longleftrightarrow TS_{ij} \frown \Lambda(\hat{\mu}_{TS_{ij}}, \hat{\sigma}_{TS_{ij}}),$$
$$\text{scheduled turn } (i) \longleftrightarrow TN_i \frown \Lambda(\hat{\mu}_{TN_i}, \hat{\sigma}_{TN_i}).$$

The only additional variable which is not scheduled is the station delay variable, for which the use of the empirical distribution function is recommended (c.f. Section 2.6). Again, the distributions of the classified station delays are stored in a computerized file according to the following subdivisions:

$$D_{ij} = f(\text{season, aircraft type, station, section of day})$$

Again there is a unique classification for each D_{ij}.

Now that the distributions of the variables in equation (1.6) have been identified, we may proceed with the logic in the model. Since equation (1.6) is a recursive relation, the reliability of the cycle has to be evaluated sequentially starting from the first leg. The steps below follow an aircraft through its planned cycle. The rationale is similar to that of a flow diagram:

A. First departure: set $i = 1$, $j = 1$,

$T_{ij} = K_{ij} + D_{ij} = T_{ij}^*$.

B. Probability that departure is late A minutes or less, (due to single leg i, j):

$P\left[D_{ij} \leq A\right]$, $A \geq 0$.

C. Time of arrival

i) Assuming punctual departure at K_{ij} (due to leg) = $B_{ij} + K_{ij}$.

ii) Actual departure (due to cycle) = $T_{ij}^* + B_{ij}$.

The empirical distribution of D_{ij} is in fact the distribution of a departure delay for the jth leg of flight i, independent of previous cycle activities. Thus the on-time performance of any leg as caused by the leg in question may be calculated by step B which in fact gives the cumulative probability distribution of D_{ij}.

Step C above examines the time of arrival. Although the study was centered around the problem of evaluating the punctuality of departures, the same may be done for arrivals. If the scheduled time of arrival is KA_{ij}, then the arrival reliability may be obtained in determining the distribution of $(T_{ij}^* + B_{ij} - KA_{ij})$ in the same way as for departures. The only difference arises from the fact that the arrival distribution is not "cut-off" at the scheduled time as in the case of departures which are assumed not to occur prior to K_{ij}, the planned time of departure. Thus continuing with the above logic, we have:

D. Probability that the arrival is late A minutes or less:

i) due to leg = $P\left[(K_{ij} + B_{ij} - KA_{ij}) \leq A\right]$, $A \geq 0$,

ii) due to cycle = $P\left[(T_{ij}^{*} + B_{ij} - KA_{ij}) \leq A\right]$, $A \geq 0$.

If the next leg is the start of a new flight, the next event is a turn and we proceed to step F. Otherwise the next event is a transit as in E:

E. Complete leg j: $E = T_{ij}^{*} + B_{ij} + TS_{ij}$,

$E_1 = B_{ij} + TS_{ij} + K_{ij}$,

start next leg: $j \rightarrow j + 1$, proceed to G.

F. Complege leg j: $E = T_{ij}^{*} + B_{i} + TN_{i}$,

$E_1 = B_{ij} + TN_{i} + K_{ij}$,

start next leg: $i \rightarrow i + 1$, $j \rightarrow 1$.

G. New departure: $T_{ij} = \max\left[E, D_{ij}^{*}\right]$,

$T_{ij}^{*} = G\left[T_{ij}, K_{ij}\right]$.

H. Probability that the departure is late A minutes or less:

(i) due to leg = $P\left[D_{ij} \leq A\right]$, $A \geq 0$,

(ii) due to cycle = $P\left[(T_{ij}^{*} - K_{ij}) \leq A\right]$, $A \geq 0$,

(iii) due to previous leg = $P\left[(E - K_{ij}) \leq A\right]$, $A \geq 0$.

While steps E and F complete the events of the particular leg, step G gives the departure time variable of the following leg. In H, we calculate the probabilities of incurring up to an A minute delay at departure as caused by the leg in question, the entire cycle, or by the previous leg alone. At this point we may return to step C and continue the sequence until the last leg in the cycle is reached, where the sequence is terminated at step D, the arrival of the last leg.

As may be observed from the above discussion, we may evaluate the on-time performance of any leg in a cycle according to more than one criterion. This further refinement is quite important since, in addition to evaluating the punctuality of the leg, it also enables one to determine where the problem was initiated. In fact, the delay may be attributable mainly to the leg itself, the previous leg or to the cycle. The model adopted in this thesis gives the following probabilities for each leg within the cycle [except the first leg which only has (3.1) and (3.4)]:

1. On-time departure probabilities as caused by:

   (i) leg itself, (assuming a punctual departure at K_{ij}):

$$P(A) = P(D_{ij} \leqslant A), \quad A \geqslant 0. \qquad (3.1)$$

   (ii) previous leg, (assuming a punctual departure for previous leg):

$$P(A) = \begin{cases} P\left[(K_{i,j-1} + B_{i,j-1} + TS_{i,j-1} - K_{ij}) \leqslant A\right], & i = 1, 2, \ldots, k, \quad j = 2, 3, \ldots, n_i \\ P\left[(K_{i-1,n_{i-1}} + B_{i-1,n_{i-1}} + TN_{i-1} - K_{ij}) \leqslant A\right], & i = 2, 3, \ldots, k, \quad j = 1, \end{cases}$$

$$A \geqslant 0 \qquad (3.2)$$

   (iii) cycle:

$$P(A) = P\,(T_{ij}^{*} - k_{ij}) \leqslant A\,, \quad A \geqslant 0. \qquad (3.3)$$

2. On-time arrival probabilities as caused by

   leg itself:

$$P(A) = P\left[(K_{ij} + B_{ij} - KA_{ij}) \leqslant A\right], \quad A \geqslant 0. \quad (3.4)$$

As requested by the users of the model, the above probabilities are calculated for A = 0, 5, 10, 15, 30 and 60 minutes.

Most of the above probability evaluations are based on the distribution of sums of random variables which have been assumed to be statistically independent. Since we have the distributions of the component variables, we may perform a convolution to obtain the distribution of the sum. This problem is discussed in the next section in terms of the methodology used as well as the accuracy to be expected.

3.2 Determination of distributions

3.2.1 Discretizations and Convolutions. There are various distributions that need to be determined when we are sequentially following the recursive relation (1.6). In the previous section, a number of steps (A - H) were described which are analagous to the steps taken in (1.6). Here, the first step A evaluates the sum of a constant and a random variable D_{ij}, which has an empirical distribution. Step B and similarly steps D and H simply calculate the cumulative probabilities of distributions as obtained from previous steps. In step E and F however, we need to calculate the distribution of the sum of three variables, one with a pre-calculated distribution (T_{ij}^*) and the others having a lognormal distribution with given estimated parameters.

One is thus faced with the problem of determining the distribution of a sum of independent random variables whose individual probability distributions are known. This is a problem which frequently arises in statistics. One is first inclined to seek a theoretical solution through the theory of characteristic or moment generating functions. The characteristic function of a sum of independent random variables is known to be the product of the individual characteristic functions. In the present case, one could theoretically proceed in this manner and then use the inversion theorem to obtain the probability distribution. [See for example, Rao (1965)].

The characteristic function for the lognormal distribution is not available in closed form. Thus, the above technique is not applicable unless one wishes to consider approximation techniques. Here, the use of the inversion formula is required which is quite complicated even when performed numerically. Another problem with the above procedure is that we sometimes have variables in a sum which have an empirical distribution or a modified distribution (T^*_{ij}) for which the characteristic functions in closed form are not obtainable. The method proposed below is intuitively simple and is convenient since it may be applied to any situation.

A way to find the distribution of the sum of several independent random variables was discussed in section (1.5) where the idea of using the numerical convolution technique was introduced. Each of the distributions of the component variables is discretized and then some convolutions of the variables are performed. Let X_i, i = 1, 2 be two random variables and let $V = X_1 + X_2$. If X_i, i = 1, 2 are discrete with probability distributions p_{X_i}, the distribution of V is given by [B. Harris (1966)]:

$$p_V(v) = \sum_z p_{X_1}(z) \cdot p_{X_2}(v-z), \qquad (3.5)$$

where the range of V depends on the ranges of X_1 and X_2.

If X_1 and X_2 are continuous with probability distributions P_{X_1}, P_{X_2} respectively, one way to discretize them is to chose n_1 and n_2 equidistant points with intervals between them of size 2c respectively, as follows:

$$a_i < a_i + 2c < \ldots < a_i + 2c(n_i-1), i = 1, 2.$$

Then X_i is equivalent to a discrete random variable over the above points with the probabilities given by:

$$p_{X_i}(x) = \begin{cases} P_{X_i}(X_i \leq x + c), & x = a_i \\ P_{X_i}(x - c < X_i \leq x + c), & x = a_i + 2c\cdot k,\ k = 1, 2, \ldots, n_i - 2 \\ P_{X_i}(X_i > x - c), & x = a_i + 2c(n_i - 1) = b_i. \end{cases} \quad (3.6)$$

In this case the distribution of V may be approximated by formula (3.5). For specific cases, one of course has to determine a_i, n_i and 2c. This depends on the particular distributions of X_1 and X_2.

If X_1 and X_2 are lognormal as in some instances in our study, one criterion to select a_i, b_i is by equating them to the exponentials of $\hat{\mu}_i - 3\hat{\sigma}_i$ and $\hat{\mu}_i + 3\hat{\sigma}_i$ respectively, which, by normal theory accounts for about 99% of the area under the normal density curve. The same amount is accounted for under the lognormal density curve as well. The reason why we do not define the limits of our range distribution by $\bar{X} \pm 3S_X$ is because of the skewed nature of the lognormal distribution, which would thus lead to a bad approximation. Other choices of a_i and b_i may also be made. For instance, since the lognormal distribution has a natural lower bound at 0, we may select $a_i = 0$ and choose an appropriate upper bound $b_i = \exp(\hat{\mu}_i + 3\hat{\sigma}_i)$, say. This choice would not improve the accuracy to a great extent since we have already covered over 99% of the distribution. In addition since numerical convolutions of variables have a cumulative effect on the resulting number of intervals, the latter choice of a_i may cause much uncontrollable expansions. Hence a_i and b_i are given by:

$$a_i = \exp(\hat{\mu}_i - 3\hat{\sigma}_i)$$
$$b_i = \exp(\hat{\mu}_i + 3\hat{\sigma}_i). \quad (3.7)$$

These values of a_i and b_i are rounded to whole minutes (a_i and b_i are rounded down and up respectively to ensure a range of at least $\hat{\mu}_i \pm 3\hat{\sigma}_i$).

Our step size is chosen to be $2c = 1$ minute which thus generates integral values for the discretized versions of the X_i's. Then, equation (3.6) becomes:

$$p_{X_i}(x) = \begin{cases} P\left[\dfrac{Y_i - \hat{\mu}_i}{\hat{\sigma}_i} \leq \dfrac{\ln(x + c) - \hat{\mu}_i}{\hat{\sigma}_i}\right], & x = a_i \\ P\left[\dfrac{\ln(x - c) - \hat{\mu}_i}{\hat{\sigma}_i} < \dfrac{Y_i - \hat{\mu}_i}{\hat{\sigma}_i} \leq \dfrac{\ln(x + c) - \hat{\mu}_i}{\hat{\sigma}_i}\right], & x = a_i + k \\ & k = 1, 2, \ldots, n_i - 2 \\ P\left[\dfrac{Y_i - \hat{\mu}_i}{\hat{\sigma}_i} > \dfrac{\ln(x - c) - \hat{\mu}_i}{\hat{\sigma}_i}\right], & x = b_i \\ 0 & \text{elsewhere,} \end{cases}$$

where $Y_i = \ln X_i$. (3.8)

Letting $Z_i = \dfrac{Y_i - \hat{\mu}_i}{\hat{\sigma}_i}$, we have that Z_i is standard normal (0, 1). Thus the discrete probabilities may be obtained from standard normal tables or with the aid of standard computer routines (usually functions).

Now, if we have two independent random variables X_1 and X_2, whether discrete or continuous, we may proceed to evaluate the distribution of their sum V by (3.5). The fact that $2c = 1$ greatly simplifies the technical procedures required for (3.5), in Fortran programming. For programming simplicity, equation (3.5) is rewritten in the following way:

Let $p_{X_i^*}(j) = p_{X_i}(x)$, $j = x - a_i + 1$, $i = 1, 2$, and

$p_{Y*}(y) = p_Y(v)$, $y = v - a + 1$, $a = a_1 + a_2$. Then (3.5)

becomes:

$$p_{Y*}(y) = \begin{cases} \sum_{K=K1}^{K2} p_{X_1}(K)\, p_{X_2}(y - K + 1), & y = 1, 2, \ldots, n \\ 0 & , \text{ elsewhere} \end{cases} \qquad (3.10)$$

where $K1 = \max(1, y - n_2 + 1)$, $K2 = \min(y, n_1)$, $n = n_1 + n_2$. These limits are determined from the following inequalities: $1 < K < n$ and $1 < y - K + 1 < n_2$, since p_{X_1} and p_{X_2} are zero outside these ranges, respectively.

3.2.2 Accuracy in discretization and convolution. When the range and the distribution of a continuous random variable are broken up into a number of disjoint intervals and discrete probabilities respectively, much information relating to the continuity of the variable is lost. It is obvious that the finer the partition is, i.e. the larger is the number of disjoint intervals, the more accurate is the approximation of the continuous distribution by its discrete counterpart. The accuracy also depends on the type of partition (i.e., intervals of equal size or intervals of variable size, etc....) and on the range over which most of the probability mass of the continuous random variable is assumed to be distributed. This in turn is connected with the skewness of the distribution. In the present case where the distribution of our random variables is lognormal, we selected to use intervals of equal size $2c = 1$ minute and a range given by (3.7).

The above aspects become more critical when one seeks to convolute two or more independent random variables and follows the discretization approach. The major problem here is the propagation of errors when the individual discretized distributions are numerically convoluted. To obtain a measure of the discretization-convolution errors we study first the convolution of two independent normal variables and then apply the results to lognormal distributions.

Let $Y \frown N(\mu,\sigma^2)$ and suppose that the discretization of Y is performed as indicated in (3.6) and is over the interval $(\mu - 3\sigma, \mu + 3\sigma)$. Then, since the discretization step is 2c, the number of intervals is $n = \frac{3\sigma+1}{c}$ (rounded-up). Thus the accuracy of discretization will improve with a large σ and/or a small c. Let $Z = \frac{Y-\mu}{\sigma}$. Then Z is a standard normal variable and the standardized intervals of Y are $\left(\frac{Y-c-\mu}{\sigma}, \frac{Y+c-\mu}{\sigma}\right)$, except for points a and b for which the intervals are $\left(-\infty, \frac{a+c-\mu}{\sigma}\right)$ and $\left(\frac{b-c-\mu}{\sigma}, +\infty\right)$ respectively. Considering only the points x for which $a < x < b$, we have intervals between the consecutive points of equal length $I_Z = \frac{2c}{\sigma}$. Thus, the smaller the interval size I_Z, the larger the number of intervals, and the better the accuracy we may expect. However, if we wish to vary I_Z by fixing $c = .5$ and varying σ depending on the population, the accuracy of discretization will then be dependent on the size of the variance and it will be better with a large σ.

To obtain a measure of the possible sizes of the discretization and convolution errors, two independent variables following normal distributions from populations with different means and variances were first individually discretized and then the distribution of their sum was obtained through a numerical convolution. The results were then compared to the discrete probabilities obtained from the normal random variable having a mean equal to the sum of the two means and a variance composed of the sum of the two variances.

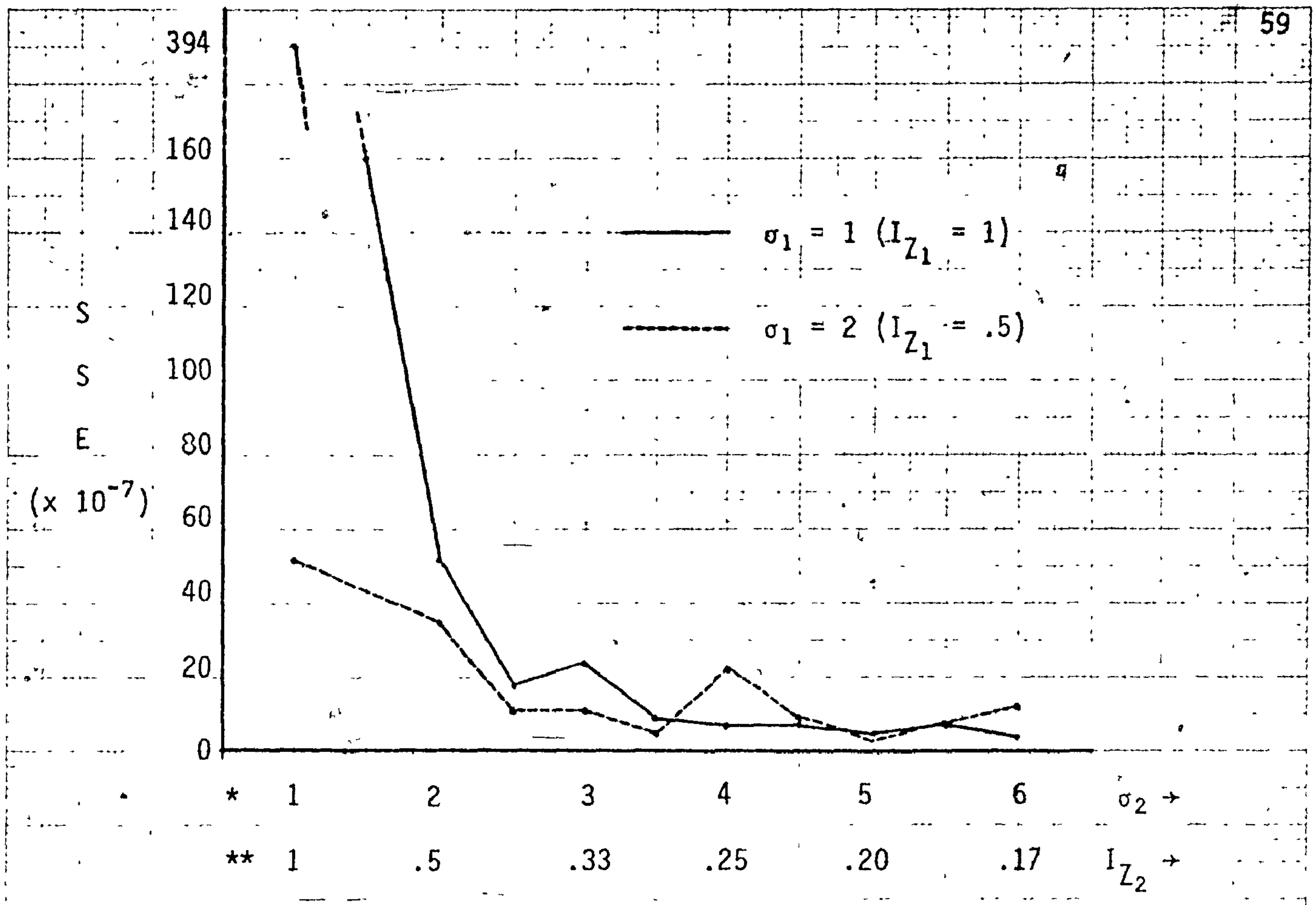
Let $Y_i \frown N(\mu_i, \sigma_i^2)$, i = 1, 2 and let $f_1(y)$ and $f_2(y)$ be the probability density functions of Y_1 and Y_2 respectively. Let also $Y_3 = Y_1 + Y_2$. Then $Y_3 \frown N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Following the discretization approach outlined above we obtain the discrete probability functions $p_i(t)$ as described in (3.6) for a range defined by $a_i = \mu_i - 3\sigma_i$ and $b_i = \mu_i + 3\sigma_i$, i = 1, 2, 3. Moreover, let $p_4(t)$, $t = a_3(2c)b_3$, be the probabilities of Y_3 obtained from the numerical convolution expression given by (3.10).

Ideally $p_4(t)$ and $p_3(t)$ should coincide for all values of t. The closeness of these two probability distributions determines the accuracy of the convolution. The following sums of squares may serve as a measure of this closeness:

$$SSE(\sigma_1, \sigma_2 | c) = SSE(I_{Z_1}, I_{Z_2}) = \sum_{t=a_3}^{b_3} \left[p_4(t) - p_3(t)\right]^2$$

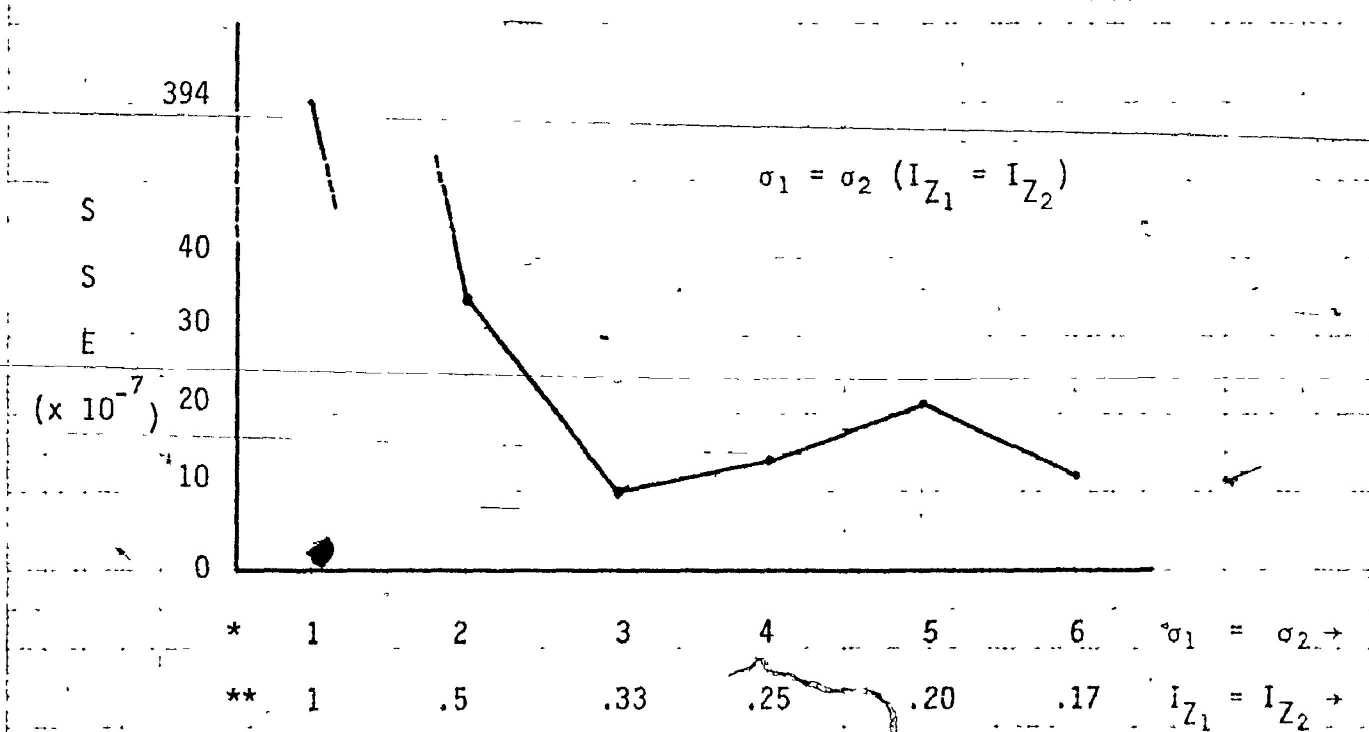
The symmetry of the normal curve and a limited numerical investigation indicated that the above SSE is functionally independent of μ_1 and μ_2.

Normal variables with varying standard deviations σ_1, σ_2 were used and in each case the sums of squares of errors as defined by $SSE(\sigma_1, \sigma_2 | c)$ above was recorded. On the following page two graphs are shown, the first demonstrates the decrease in SSE with the increase of σ_2 (I_{Z_2}) while having σ_1 (I_{Z_1}) fixed (σ_1 = 1 and 2) and the second relates SSE with simultaneous



* σ_2 : standard deviation of Y_2

**I_{Z_2} : standardized intervals for Y_2



* $\sigma_1 = \sigma_2$: standard deviations of Y_1, Y_2

**$I_{Z_1} = I_{Z_2}$: standardized intervals for Y_1, Y_2

increases of σ_1 and σ_2 (I_{Z_1} and I_{Z_2}). From these graphs we may claim that for values of σ_1 and σ_2 below 2, the SSE expands very significantly, while for values above two, the error (SSE) remains in a low range. Thus the interval size should be $\frac{2c}{\sigma} = \frac{2(.5)}{2} = .5$ at most, to ensure good accuracy. That is, we want:

$$I_Z \leq .5 \quad , \forall Z \qquad (3.11)$$

In the lognormal case, the discrete steps of one minute are taken on the variable X, and not on Y = ln X which is normal. Thus we have the following situation:

$$X \sim \Lambda(\mu, \sigma^2),$$

$$a = \exp(\mu - 3\sigma),$$

$$b = \exp(\mu + 3\sigma),$$

$$\text{range} = e^{\mu+3\sigma} - e^{\mu-3\sigma},$$

$$n = \text{No. of intervals} = \left[\frac{e^{\mu+3\sigma} - e^{\mu-3\sigma}}{2c}\right] + 1 \text{ (rounded up).}$$

As may be observed in this case, the accuracy of the discretization will not only be dependent on c and σ, but on μ as well. The discrete probabilities are expressed by (3.8) and the interval size about Z is:

$$I_Z = \frac{1}{\sigma}\left[\ln(x + c) - \ln(x - c)\right] = \frac{1}{\sigma} \ln\left[\frac{x + c}{x - c}\right]. \qquad (3.12)$$

Although the intervals are of equal size on the variable X, they are of variable length on Z which is N(0, 1). Let $I_Z^* = \max_Z I_Z$. Suppose we have a normal random variable Y which is discretized with intervals of equal size, namely of size I_Z^* (on the standardized Y). Then, by (3.11) we require $I_Z^* \leq .5$ to obtain a good accuracy for convolutions.

This implies that the accuracy of discretization and convolution of two lognormal variables such as X will be at least as good as the accuracy to be expected for Y, since the discrete steps associated with X are $I_Z \leqslant I_Z^* \leqslant .5$. We may thus use the same criterion as in the normal case, provided that we show that $I_Z^* \leqslant .5$. Now, since c is a constant we have:

$$\lim_{X \to \infty} \ln \left(\frac{X + c}{X - c} \right) = 0$$

This implies that a large value of X and/or a large value of σ will cause a small value of I_Z. In a given distribution the maximum over Z of I_Z will occur at $x = a_1$, the smallest possible value of X. The criterion now becomes the following:

$$I_Z^* = \frac{1}{\sigma} \ln \left(\frac{a_1 + c}{a_1 - c} \right) \leqslant .5$$

Replacing $a_1 = e^{\mu - 3\sigma}$, the above implies that:

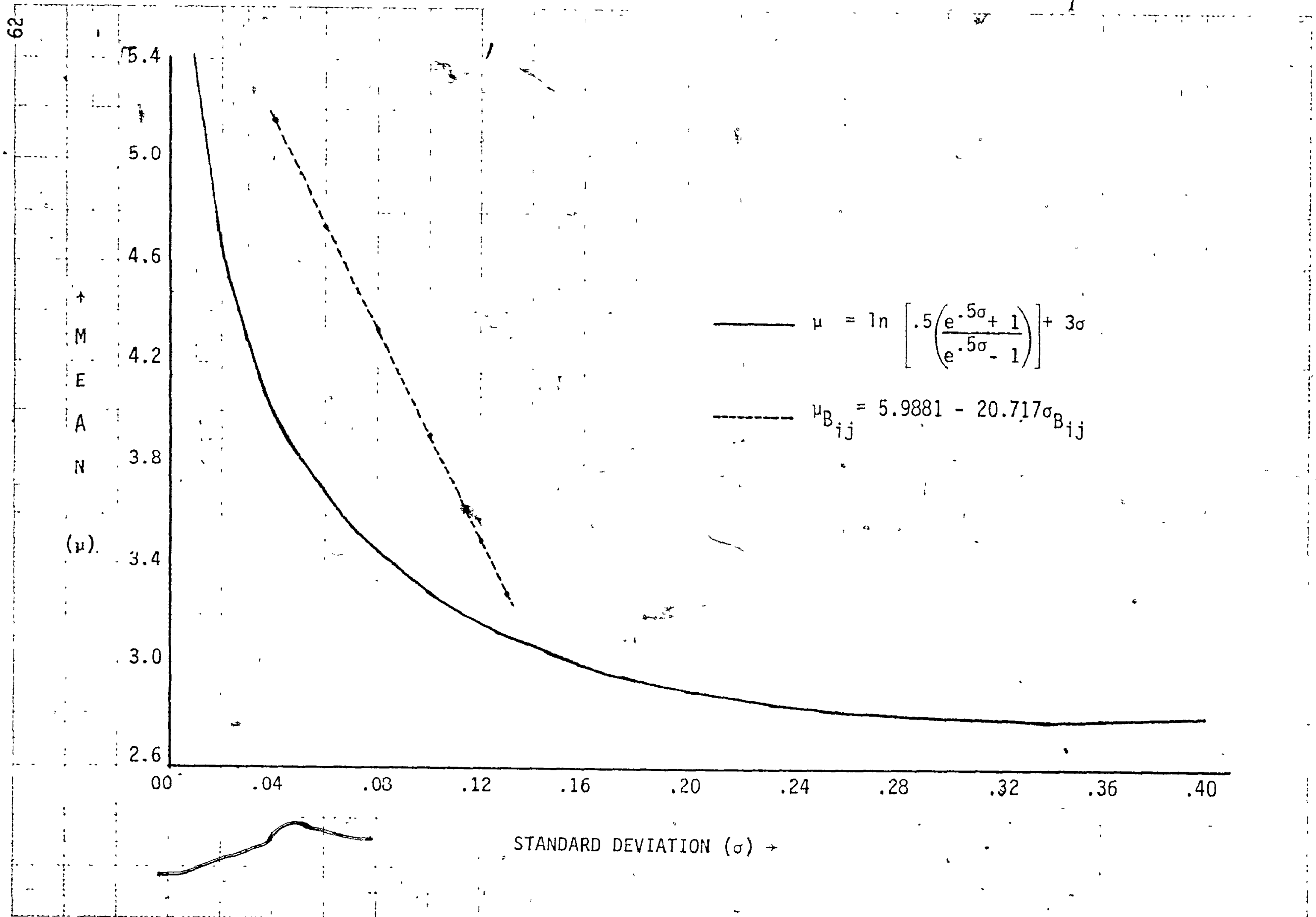
$$e^{\mu - 3\sigma} \geqslant .5 \left(\frac{e^{.5\sigma} + 1}{e^{.5\sigma} - 1} \right)$$

or equivalently,

$$\mu \geqslant \ln \left[.5 \left(\frac{e^{.5\sigma} + 1}{e^{.5\sigma} - 1} \right) \right] + 3\sigma \, . \qquad (3.13)$$

If (3.13) holds true for our distributions we may be reassured on the resulting accuracy of our convolutions. A graph was thus plotted showing the equality relation of (3.13). If, in our distributions, $\hat{\mu}$ is above or on the plotted line for its corresponding $\hat{\sigma}$, then (3.13) is satisfied.

Since block, transit and turn times were analysed individually, we may examine each case separately. Also, since the worst cases are those when μ and/or σ are small, we observe the instances when the smallest of each occurs.


↑ MEAN (μ)
5.4
5.0
4.6
4.2
3.8
3.4
3.0
2.6
00
.04
.08
.12
.16
.20
.24
.28
.32
.36
.40
STANDARD DEVIATION (σ) →
$\mu = \ln\left[.5\left(\frac{e^{.5\sigma}+1}{e^{.5\sigma}-1}\right)\right] + 3\sigma$
$\mu_{B_{ij}} = 5.9881 - 20.717\sigma_{B_{ij}}$

A. Block times

Fortunately in the prediction of the block parameters a negative correlation was found between the means and standard deviations of our samples of logged data. Thus a small standard deviation σ which may cause error propagation in discretization and convolutions has a corresponding large μ which has an opposite effect of increasing the accuracy. Some 68 samples were selected from the analysis made in chapter two, and using the sample estimates of the logged data, a regression of the means on the standard deviations was performed.

Normally, one could perform such a regression using not just the 68 samples, but in fact, using one sample from each classification of block times. This procedure is too lengthy for our purpose since we have 2 seasons x 6 sections of the day x over 7 aircraft types and a large number of city pairs. The 68 samples were however carefully selected to ensure a good representation of the bulk of classifications.

Let $\overline{Y}_i$ be the mean of the ith sample of logged block times, $s^2_{Y_i}$ be the variance of the ith sample, and n_i be the size of the sample i. Then $\overline{Y}_i \sim N\left(\mu_i, \frac{\sigma^2_{Y_i}}{n_i}\right)$. A regression is assumed between $\overline{Y}$ and s_{Y_i} as follows:

$$E(\overline{Y}_i \mid s_{Y_i}) = \gamma_1 + \gamma_2 \bullet s_i$$

On performing such a regression, the BMD02R program of the BMD Statistical Package was used, and the results were the following:

$$E(\overline{Y}_i) = 5.9881 - 20.71715\ s_{Y_i}, \qquad (3.14)$$

with the correlation coefficient r = -.8117 and the F-ratio with 1 and 66 degrees of freedom, $F_{1,66} = 127.509$ (see Appendix B for details of the regression).

As may be seen, a high negative correlation exists and since the F-ratio is significant at $\alpha = .01$, we may rely on the relation (3.14). This regression and its implications will be discussed more thoroughly in the next chapter where the question of predicting the parameters is discussed.

Among the 68 samples that were used in the above analysis, the sample with the smallest s_{Y_i} as well as the sample with the smallest $E(\overline{Y}_i)$ were selected to see if their respective s_Y and $\overline{Y}$ would satisfy (3.13):

a) min $s = .0237$, $\hat{\mu} = 5.6127$

graph requires, $\mu \geqslant 4.67$, which holds.

b) minimum $E(\overline{Y}_i) = 3.27$, $s = .13$

graph requires $\mu \geqslant 3.11$, which holds.

Thus, in both of the worst cases, we would still expect good accuracy.

B. Transit times

a) min $s = .131$, $\hat{\mu} = 3.504$

graph requires $\mu \geqslant 3.1$ which holds.

b) small $E(\overline{Y}_i) = 2.86$, $S = .265$

graph requires $\mu \geqslant 2.81$.

C. Turn time

a) For a sample with a small standard deviation

$s = .06$, $\hat{\mu} = 4.512$

required $\mu \geqslant 3.67$ is satisfied.

b) For a sample with a small $\hat{\mu}$:

$s = .19$, $\hat{\mu} = 2.98$

requires $\mu \geqslant 2.92$, which is satisfied.

3.2.3 Determination of the distribution of extreme values. We now describe the procedures used in evaluating the distribution of the maximum of the two variables E and D_{ij}^* which is required for step G in Section 3.1. In this case both E and D_{ij}^* are discrete. In general, if we seek the distribution of $W = \max(X_1, X_2)$ where X_1 and X_2 are independent discrete random variables, we obtain the following:

$$P(W \leq w) = P(X_1 \leq w)\, P(X_2 \leq w)$$

$$= \sum_{t_1=a_1}^{w} p_{X_1}(t_1) \sum_{t_2=a_2}^{w} p_{X_2}(t_2)$$

$$\text{and } p_W(w) = \begin{cases} 0 & w < \max(a_1, a_2) \\ p_{X1}(w) \sum\limits_{t_2=a_2}^{w} p_{X2}(w) + p_{X2}(w) \sum\limits_{t_1=a_1}^{w-2c} p_{X1}(t_1), & \\ \qquad \max(a_1, a_2) \leq w \leq \min(b_1, b_2) & \\ p_{X_1}(w) & b_2 < w \leq b_1 \\ p_{X_2}(w) & b_1 < w \leq b_2 . \end{cases} \qquad (3.15)$$

The application of standard theory is straight forward and no special problems arose in the actual implementation of the model.

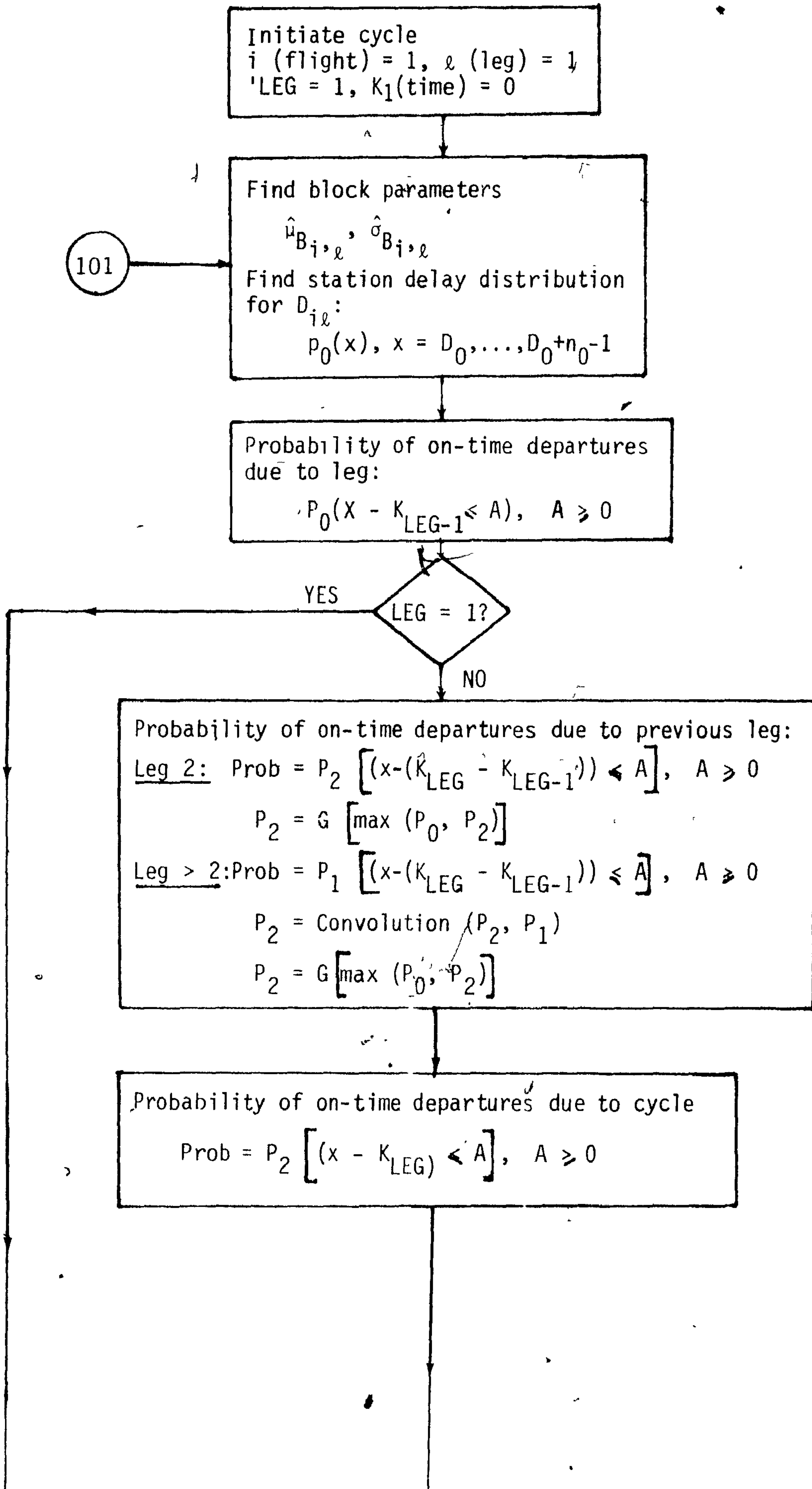
3.3 Implementation of model

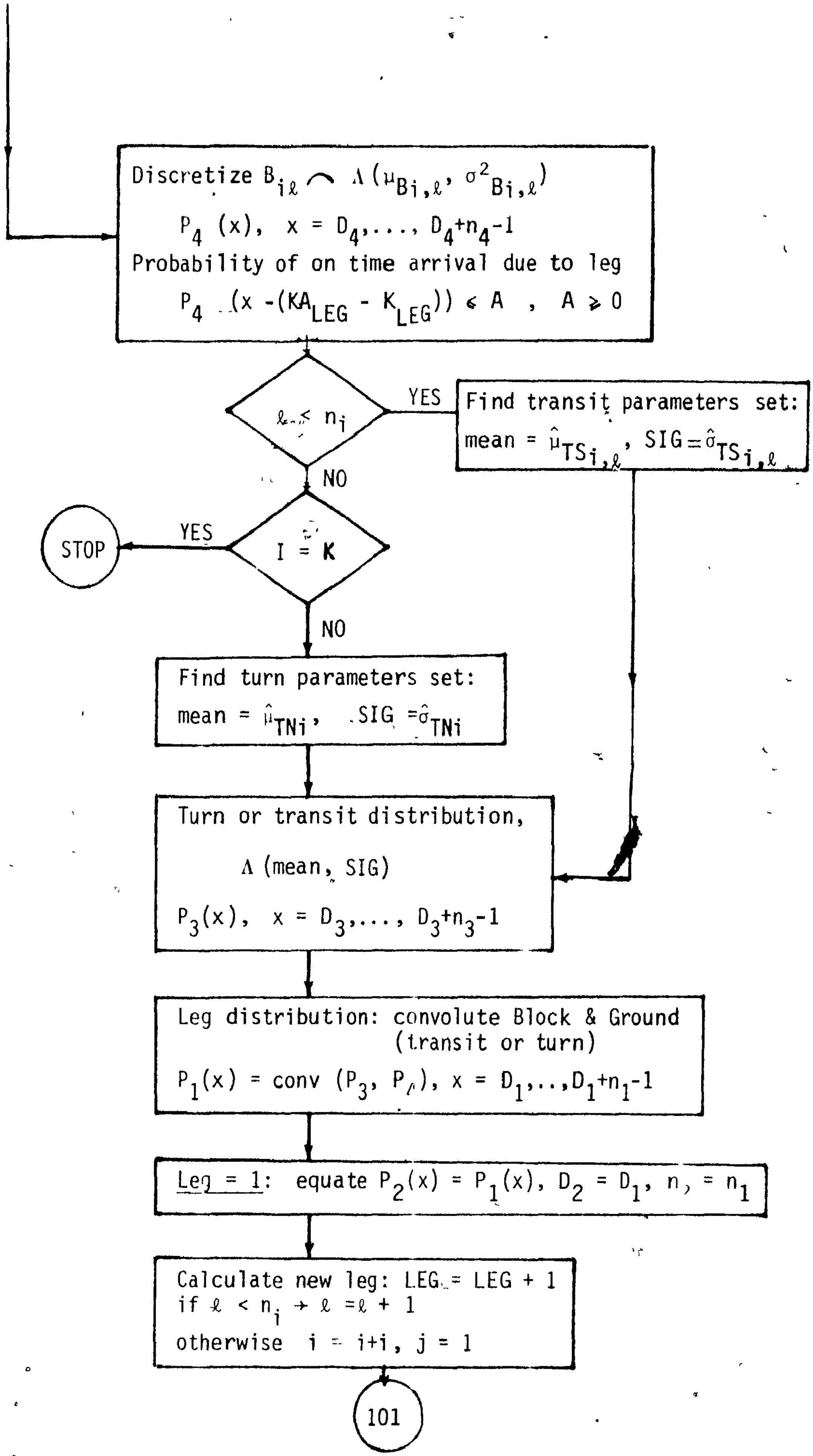
The computer program developed for the model to evaluate the on-time performance of cycles was written in Fortran IV and implemented on the Honeywell 6000 computer. Presently it runs on the time-sharing system in an interactive fashion, where the user may manually input the schedule related data for one or more cycles. This allows quick manipulation and enables the user to get some answers to "what if ...?" type of questions.

Basically, there are two types of input required by the program. First, there is the information describing the necessary distributions of block, transit, turn and station delay variables which are automatically (internally) retrieved from the computerized file system, described in Section 3.1. The second type of input is of course the scheduled information which includes the identification of the season and aircraft type, as well as data relating to each cycle to be tested. The cycle information corresponds to the description given in Section 3.1.

Once the input is obtained, the procedure is initiated within the program. A flow chart is shown on the next page where the principal computations made in the program for one cycle are followed sequentially. The flow chart assumes that all the required data have been inputted. Once the cycle is initiated, the sequence of steps is followed for each leg. For the following leg, the sequence continues at 101 until the last leg of the cycle where the flow is broken at 999.

The main program calls a series of subroutines designed to evaluate convolutions (3.10), discretizations (3.8), the distribution of the maximum of two variables (3.15), the transformation of the departure distribution $G(T_{ij}, K_{ij})$ as in (1.4) and (1.5), and the cumulative probabilities for any of the calculated distributions (steps B, D, H in Section 3.1).


Initiate cycle
i (flight) = 1, ℓ (leg) = 1
LEG = 1, K_1(time) = 0
Find block parameters
$\hat{\mu}_{B_{i,\ell}}$, $\hat{\sigma}_{B_{i,\ell}}$
Find station delay distribution
for $D_{i\ell}$:
$p_0(x)$, $x = D_0, \ldots, D_0+n_0-1$
101
Probability of on-time departures
due to leg:
$P_0(X - K_{LEG-1} \leq A)$, $A \geq 0$
LEG = 1?
YES
NO
Probability of on-time departures due to previous leg:
Leg 2: Prob = $P_2\left[(x-(K_{LEG} - K_{LEG-1})) \leq A\right]$, $A \geq 0$
$P_2 = G\left[\max(P_0, P_2)\right]$
Leg > 2: Prob = $P_1\left[(x-(K_{LEG} - K_{LEG-1})) \leq A\right]$, $A \geq 0$
P_2 = Convolution (P_2, P_1)
$P_2 = G\left[\max(P_0, P_2)\right]$
Probability of on-time departures due to cycle
Prob = $P_2\left[(x - K_{LEG}) \leq A\right]$, $A \geq 0$


Discretize $B_{i\ell} \frown \Lambda(\mu_{B_{i,\ell}}, \sigma^2_{B_{i,\ell}})$
$P_4(x)$, $x = D_4, \ldots, D_4+n_4-1$
Probability of on time arrival due to leg
$P_4(x - (KA_{LEG} - K_{LEG})) \leq A$, $A \geq 0$
$\ell < n_i$
YES
Find transit parameters set:
mean = $\hat{\mu}_{TS_{i,\ell}}$, SIG = $\hat{\sigma}_{TS_{i,\ell}}$
NO
$I = K$
YES
STOP
NO
Find turn parameters set:
mean = $\hat{\mu}_{TNi}$, SIG = $\hat{\sigma}_{TNi}$
Turn or transit distribution,
Λ (mean, SIG)
$P_3(x)$, $x = D_3, \ldots, D_3+n_3-1$
Leg distribution: convolute Block & Ground (transit or turn)
$P_1(x) = \text{conv}(P_3, P_4)$, $x = D_1, \ldots, D_1+n_1-1$
Leg = 1: equate $P_2(x) = P_1(x)$, $D_2 = D_1$, $n_2 = n_1$
Calculate new leg: LEG = LEG + 1
if $\ell < n_i \rightarrow \ell = \ell + 1$
otherwise $i = i+i$, $j = 1$
101

In the convolution subroutine, since the number of intervals of the resulting discrete variable, is the sum of the number of intervals of the individual distributions, some care must be exercised to control the expanding sizes. In the first place, the distribution is truncated at .001 and .999 probability points. After this is done, if the number of intervals exceeds 150 units which corresponds to 2:30 hrs., then the distribution is again truncated on either side at points of equal probability so as to include a maximum of 150 intervals (minutes).

The subroutine which transforms the continuous lognormal distribution of a variable to discrete probabilities, DSCRLN (discrete - lognormal) proceeds in a way analogous to (3.8). The range of the discretization is for ($\mu \pm 3\hat{\sigma}$) and therefore covers about 99.84% of the area under the lognormal curve.

All the cumulative probabilities required for steps B, D, and H in Section (3.1) are evaluated using another subroutine which also prints out the output. On Table 3.2, an example of the output obtained from the model is shown. This is described below:

In observing the table, we note that the aircraft type (D9S) and the season (summer) are explicitly shown underneath the title. Following this heading is a legend describing the abbreviations used. At this point, the remainder of the printout takes the form of a cycle heading defining all the sequence of stations within the cycle, and then the on-time performance probabilities for each leg are shown. The maximum delay minutes, of course, refer to the constant A used in steps B, D, and H

Table 3.2

ON TIME PERFORMANCE
PROBABILITIES

****
*DOS SUMMER 1973 *
****

D A = DEPARTURE ARRIVAL DELAY.
L = DUE TO LEG.
PL = DUE TO PREVIOUS LEG.
C = DUE TO CYCLE.

CYCLE YAM-YYZ-YOW-YUL-YOG-YWG-YYC-YVR-YEG-YWG-

MAXIMUM DELAY MINUTES

FLIGHT NO.	LEG	D A	DUE TO	0	5	10	15	30	60
346	YAM-YYZ	D	L	0.600	0.700	0.767	0.867	0.933	1.000
		A	L	0.381	0.887	0.995	1.000	1.000	1.000
346	YYZ-YOW	D	L	0.364	0.709	0.836	0.891	0.964	1.000
		D	PL	0.402	0.768	0.951	0.995	1.000	1.000
		D	C	0.146	0.545	0.795	0.887	0.964	1.000
		A	L	0.942	0.998	1.000	1.000	1.000	1.000
346	YOW-YUL	D	L	0.305	0.699	0.867	0.930	0.972	1.000
		D	PL	0.687	0.845	0.929	0.969	0.999	1.000
		D	C	0.154	0.415	0.643	0.779	0.927	1.000
		A	L	0.181	0.741	0.977	1.000	1.000	1.000
227	YUL-YOG	D	L	0.376	0.652	0.759	0.837	0.922	0.979
		D	PL	0.993	0.997	0.999	1.000	1.000	1.000
		D	C	0.357	0.640	0.759	0.837	0.922	0.979
		A	L	0.912	0.995	1.000	1.000	1.000	1.000
227	YOG-YWG	D	L	0.344	0.813	0.906	0.938	1.000	1.000
		D	PL	0.651	0.811	0.908	0.959	0.999	1.000
		D	C	0.147	0.475	0.639	0.742	0.926	1.000
		A	L	0.696	0.920	0.988	1.000	1.000	1.000
227	YWG-YYC	D	L	0.419	0.790	0.855	0.919	0.984	1.000
		D	PL	0.969	0.999	0.997	1.000	1.000	1.000
		D	C	0.305	0.649	0.773	0.905	0.984	1.000
		A	L	0.593	0.949	0.964	0.994	1.000	1.000
227	YYC-YVR	D	L	0.344	0.817	0.935	0.969	1.000	1.000
		D	PL	0.367	0.621	0.822	0.935	1.000	1.000
		D	C	0.065	0.317	0.546	0.720	1.000	1.000
		A	L	0.725	0.960	0.999	1.000	1.000	1.000

(Section 3.1), for A = 0, 5, 10, 15, 30 and 60. The flight number column gives the serial number associated with the flight. If there are more than one leg associated with a particular flight, then the serial number will also appear more than once as in the case of flight No. 346. The column falling under the title "LEG" gives the alphabetic codes of the stations of departure and arrival of the leg.

If we now examine the chart horizontally, we note that the abbreviations defined in the legend are used to identify what type of probabilities are given. The probabilities following the pair of abbreviations D-L are analogous to the set-up (3.1). Also D-PL, D-C and A-L correspond to (3.2), (3.3) and (3.4) respectively.

Since the first leg in a cycle does not have any previous history, the probabilities evaluated are only the D-L, and A-L, i.e. the leg-dependent probabilities. All other legs in the cycle have the entire foursome D-L, D-PL, D-C and A-L.

As may be observed from the example, the first leg of Flight # 346 has a 70% chance of departing within 5 minutes of the scheduled time, and an 89% chance of arriving within 5 minutes of schedule. It may also be seen that Flight 346's YOW-YUL leg has a 64% probability of departing within 10 minutes of schedule. If we wish to improve on this performance, the probable initiation of the delay must first be determined. Looking down the 10 minute column, the latter leg has a D-L probability of 87% and a D-PL probability of 93%. This implies that the delay is cycle dependent. Consequently, a buffer may be added to the YOW scheduled transit time so as to improve on the departure performance. When a

5 minute buffer is added, the probability of departing within 10 minutes of schedule in fact increases to 73%.

The model output thus serves as a tool to enable the user to identify where a problem is likely to occur and to consequently modify the schedule so as to improve its on-time performance.

3.4 Reliability of model.

One must consider the actual intention of the model before determining the type of reliability test to be used. What questions do we expect the model to answer and how precise must these answers be? The model is intended to be used as a tool to assist in the planning processes required in the construction of multi-cycle schedules. The model results, together with other criteria such as judgement based on personal experience, could help in determining whether some cycles are too tight and thus are expected to perform very badly. In such a case something may be done at this early stage to modify the schedule while acting within the limiting constraints.

It is of interest to compare the predictions of the on-time performance of aircraft cycles as given by the model to what actually happens in reality. Thus we would like to know how well the model duplicated the actual process. The model will be reliable if its results are consistent with the actual performance and if it helps to determine whether the cycle is a potential problem cycle or not. To perform such tests we may select some aircraft cycles and compare the results to some actual statistical summaries reported for that particular period.

Two D9S cycles for July and August 1973 were selected. These examples are also used to illustrate the overall process required for the functioning of the model. In the Appendix C the cycles are shown as cycles A and B where the scheduled times of departure and arrival are pre-converted to a standard time (Toronto and Winnipeg times respectively).

The estimated parameters and the empirical distributions of the variables required for the model are gathered from the distribution files (Section 3.1). The classification associated with each variable is determined using the set-up described in Section 3.1, which permit the retrieval of the relevant data from the files. The retrieved estimated parameters for the necessary B_{ij}, TS_{ij} and TN_i are summarized in Appendix C (Tables 2, 3 and 4) along with Table 5 showing the station delay empirical distributions.

Using the scheduled data together with the retrieved distribution data as in section 3.1, model (1.6) was applied. The results obtained are shown in Appendix C (Tables 6 and 7).

Below are the actual on-time performance statistics as reported for that period of time as compared to the model results:

Cumulative probabilities of
incurring up to an A minute
delay, A=5, 15, 30, 60.
Model vs. Actual results.

M = Model results

A = Actual statistics

Cycle A.

M/A	Flt No.	City Pair	Minutes (A) 5	15	30	60
M	346	YAM - YYZ	.70	.87	.93	1.00
A			.68	.84	.90	.97
M	346	YYZ - YOW	.55	.89	.96	1.00
A			.45	.78	.87	.94
M	346	YOW - YUL	.42	.78	.93	1.00
A			.41	.67	.85	.89
M	227	YUL - YQG	.64	.84	.92	.98
A			.56	.81	.90	.97
M	227	YQG - YWG	.48	.74	.93	1.00
A			.55	.78	.94	1.00
M	227	YWG - YYC	.65	.91	.98	1.00
A			.71	.78	.90	.97
M	227	YYC - YVR	.32	.72	1.00	1.00
A			.52	.68	.87	1.00

Cycle B.

M/A	Flt No.	City Pair	Minutes (A) 5	15	30	60
M	271	YWG - YQR	.89	.98	1.00	1.00
A			.74	1.00	1.00	1.00
M	271	YQR - YYC	.76	1.00	1.00	1.00
A			.71	1.00	1.00	1.00
M	271	YYC - YVR	.43	.95	.99	1.00
A			.65	.90	1.00	1.00
M	280	YVR - YQR	.57	.81	.91	1.00
A			.78	.87	.94	1.00
M	280	YQR - YWG	.38	.74	.90	1.00
A			.39	.74	.84	.94
M	289	YWG - YQR	.65	.84	.98	1.00
A			.84	.87	.90	.94
M	228	YQR - YWG	.71	[illegible]	.92	1.00
A			.81	.94	.94	.97
M	228	YWG - YOW	.51	.80	.91	.95
A			.61	.87	.90	.94

As may be observed from the above tables the model results are consistent with the actual statistics in that deteriorations or propagation of delays or on the other hand improvements of the on-time performance are depicted by the model when they occur in reality.

3.5 General comments on the approach.

Initially when one is attempting to solve a problem, one wishes to satisfy a set of objectives. Once the objectives are well defined, one is faced with the problem of selecting one of the possibly many approaches which are appropriate for the situation. One may proceed in this selection by examining the literature on historical case studies or on other relevant material which may guide one to an appropriate approach. The success and failures of others are helpful in determining the right direction.

The objective in our case was to provide a tool which may be used in scheduling to evaluate the on-time performance of flight schedules. Moreover, upon examining the many possible approaches, we considered our own requirements and,benefiting from the successful use of one model by another airline, we made a choice of the approach and the model. A particular model may be selected in such a way, not necessarily because it is the "best" one. Alternatively, the model may be appealing because of its simplicity, low cost in running, or on the other hand because it achieves high accuracy and good reliability. The selection of the appropriate model greatly dependson the purpose intended for its use.

Below we shall examine other possible approaches that could have been followed to achieve our objectives. We shall also see how other variations the model used in this thesis could have been designed.

As stated in Section 1.3, the approach that was most often adopted in the past is the simulation technique. Here, one must regenerate or simulate the sequence of events that actually happen according to some assumed probability distribution. Besides the block and ground time variables, connections must be simulated. The delay-causes are treated individually as well, where each delay type has a certain frequency distribution of occurance. Thus the schedule may be simulated within the computerized procedure many times so as to obtain some departure delay distributions. When a complete schedule is simulated other variables may also enter the picture. Starting with an assumed number of available aircraft per type at each station (≥ 0), aircraft reassignment may be performed. That is, if an aircraft arrives so late that it may create a departure delay of more than a specified limit, and if an aircraft of the type required is available at that station then the available aircraft is used. This actually occurs in practice. Thus simulation, when it uses proper assumptions, gives a very realistic picture of the problem. However, to run such a simulation, this requires very detailed and good data, and as mentioned in Section 1.3, this method requires long computer turnaround time.

Another possibility which is very similar to the above is to consider cycle simulations instead of schedule simulations. The former approach simplifies the problem. Since the complicated schedule network is reduced to simple aircraft cycles, many variables are dropped, such as the aircraft availability variable and hence the problem becomes less complex.

We may also choose to pursue a univariate model where the departure delay variable may be represented as a function of the other factors such as: airport, aircraft type, number of legs prior to given leg, types of services prior to and following the leg, distances covered prior to and following the leg, season, time of day and so on. Such a model would purely be based on historical observations. In such a case, much care must be taken in selecting a representative sample of observations.

Alternatively, an approach similar to ours may be adopted but based on a different structure. Other variables may be added, such as the connection time. Some spare aircraft distribution may also be appropriate for certain stations. Also, the block time variable may be split into the flight and taxi times.

We have adopted one classification procedure for each of our variables out of a possible few. Also we have made some assumptions regarding the distributions of our variables - lognormality for B_{ij}, TS_{ij} and TN_i; some other distribution may have been appropriate as well. Moreover, we could have approximated the station delay variables through some parametric distribution rather than using the empirical one.

CHAPTER 4

PREDICTION OF THE PARAMETERS

In the case where a new T_{ij} is introduced, that is when a departure with a set of conditions occurs that has not happened in the past, the same model is still applicable in testing the on-time performance. An example of a new T_{ij} is when a new route or a new destination is introduced. If some variables (B_{ij}, TS_{ij}, TN_i, D_{ij}) that are used in determining the distribution of T_{ij} are historically available, one could use their pre-determined parameters (Chapter 2) and thus find their distributions. However if for one or more variables in the model we lack historical data or if the available data are few or bad for some reasons, then one has to resort to methods for predicting the parameter(s) of the distribution.

The discussion below is restricted to block time parameters and similar methodology may be applied in predicting parameters for the transit or turn time variables.

When a new B_{ij} is under consideration and no data are available, one must make certain assumptions regarding the distribution of B_{ij} and its parameters, in order to predict them as required by the model. Because of the evidence of lognormality acquired in Chapter 2, it seems reasonable to assume that any new B_{ij} has a lognormal distribution as well. Two prediction relationships are now needed for μ and σ^2 (or equivalently for α and β^2). Several possibilities are open to consideration. Naturally, it is much simpler to predict the block time mean α than the variance β^2, since one could relate α to fixed factors such as distance, etc... One could also look for a relationship between means and variances μ and σ^2 (or α and β^2). This problem has been discussed by P. Franke (1972) and reviewed by M. M. Etschmaier and M. Roshteim (1974), where, indeed, a linear relationship between μ and σ was found.

In Section 3.2.2, a regression of the means on the standard deviations calculated from 68 samples of logged block times was performed. Since each set of parameters "belong" to a different population, we may expect a non-zero correlation coefficient between the sample means and standard deviations. As may be observed from the discussion in Section 3.2.2, the correlation is -.8117 and the slope of the line (regression coefficient) is naturally negative. The F-ratio for testing a zero slope is highly significant which gives further evidence of the dependence of the means on their standard deviations.

The high negative correlation could be explained in observing that much of the block variance is due to the taxi time since this is dependent on the traffic congestion upon departure or arrival. Moreover, the taxi time accounts for a large percent of the block time for short trips, while the inverse relation exists in the case of long block times. In addition, there is the possibility of making up for lost time in long trips. Thus a larger variance is expected for the short block times.

In Chapter two it was found that the block time fitted the log-normal distribution and that it was best classified by season, aircraft type, city-pair and time of day. Intuitively, one would thus be motivated to predict the block time means by equating these as a function of precisely these same variables. However, since the city pairs are fixed, this would inhibit the generalization applicable to city pairs for which no data exists. It would seem more appealing to translate this factor to a distance variable while including an average headwind factor as well. A plausible model is thus the following one:

$$\text{Block time} = \alpha_0 + \alpha_1 \cdot \text{distance} + \alpha_2 \cdot \text{time of day} + \alpha_3 \cdot \text{season} + \alpha_4 \cdot \text{aircraft type} + \alpha_5 \cdot \text{wind effect} + \text{error} \tag{4.1}$$

The block time also depends on the geographic route and direction of the flight and this is accounted for by the inclusion of both the distance and the average headwind factors. The time of day variable is split into one-hour intervals from 7:00 a.m. to midnight. The hours in between are grouped into one additional interval. As for the aircraft type, numerical codes are used as indicators. The season variable takes one of two values depicting the summer or winter season (6 months each).

Motivated by Franke's discussion, model (4.1) could be further refined by taking into account the effect of the months of the year on the block time. One way to do this is to treat the season independent variable as a linear combination of trigonometric functions of time t, (t = 1, 2, .. , 12). Then model (4.1) would become:

$$\text{Block time} = \beta_0 + \beta_1 \cdot \text{distance} + \beta_2 \cdot \text{time of day} + \beta_3 \cdot \sin\left(\frac{2\pi t}{T}\right) + \beta_4 \cdot \cos\left(\frac{2\pi t}{T}\right) + \beta_5 \cdot \text{aircraft type} + \beta_6 \cdot \text{wind factor} + \text{error}, \tag{4.2}$$

where T = 12.

In Franke's prediction procedure, a somewhat different approach is pursued. The following model is assumed:

$$\text{Block time} = \mu_F + \mu_S + \mu_{X_D} + \mu_{X_A} + \mu_H + \text{error} \qquad (4.3)$$

where μ_F is the mean of flight time F, μ_S is the mean of the seasonanal variance S of F, μ_{X_D} and μ_{X_A} are the mean taxi times at departure (X_D) and arrival (X_A) respectively and μ_H is the mean holding time (H) at arrival. To obtain the prediction equation of the above model, Franke considers independent regressions as follows:

(1) $\mu_F = \delta_1 + \delta_2 \cdot \text{distance}$

(2) $\mu_S = \Theta_1 + \Theta_2 \sin \frac{2\pi t}{T} + \Theta_3 \cos \frac{2\pi t}{T}$

where δ_i, $i = 1, 2$ and Θ_j, $j = 1, 2, 3$ are the corresponding regression coefficients, t is the month of the year index ranging from 1 to 12, and T = 12. Let $\hat{\delta}_i$ and $\hat{\Theta}_j$ be their least squares estimates.

The three other variables X_D, X_A, and H are each averaged from historical data, classified by individual station or by stations with similar conditions. These averages serve as the estimates $\hat{\mu}_{X_D}$, $\hat{\mu}_{X_A}$ and $\hat{\mu}_H$ respectively. The prediction equation (4.3) now becomes:

$$\widehat{\text{Mean block time}} = \hat{\delta}_1 + \hat{\Theta}_1 + \hat{\Theta}_2 \sin \frac{2\pi t}{T} + \hat{\Theta}_3 \cos \frac{2\ t}{T} + \hat{\delta}_2 \cdot \text{distance} + \hat{\mu}_{X_D} + \hat{\mu}_{X_A} + \hat{\mu}_H \qquad (4.4)$$

In performing independent estimations of the components of model (4.3), it is conceivable that any intercorrelations between the variables may be omitted. In addition to this, it seems quite difficult to obtain a good measure of the fit for model (4.3) while for the more direct regression approach used in model (4.1), R^2 together with the F-ratio may serve for this purpose.

Returning to our former discussion, we now have a procedure to predict the mean block time (4.1). A relation between μ and σ was also estimated in (3.15). To estimate the parameter μ and σ for a new B_{ij}, we use both of these results together with some lognormal theory as follows below. Let μ, and σ^2 be the mean and variance of the logged block time variable, respectively. Also, let α and β^2 be the corresponding mean and variance of the block time. Since the block time data were found to be lognormal, equations (2.4) and (2.5) are applicable and give the following equation:

$$2 \cdot \ln \alpha = \sigma^2 + 2\mu \qquad (4.5)$$

Model (4.1) permits the prediction of α. Therefore we may use the calculated $\hat{\alpha}$ and replace α in (4.5) by $\hat{\alpha}$ as follows:

$$2 \cdot \ln \hat{\alpha} = \sigma^2 + 2\mu \qquad (4.6)$$

Now, using the estimated linear relationship between μ and σ [c.f. (3.14)],

$$\mu = \hat{\gamma}_1 + \hat{\gamma}_2\sigma \qquad (4.7)$$

we obtain a system of two equations in two unknowns.

Substituting (4.7) in (4.6) we obtain the following quadratic equation in $\hat{\sigma}$:

$$\sigma^2 + 2\hat{\gamma}_2\sigma + 2(\hat{\gamma}_1 - \ln \hat{\alpha}) = 0 \qquad (4.8)$$

Its solution is given by:

$$\sigma = -\hat{\gamma}_2 \pm \left(\hat{\gamma}_2^2 - 2(\hat{\gamma}_1 - \ln \hat{\alpha})\right)^{\frac{1}{2}} . \qquad (4.9)$$

In certain cases, (4.9) may give complex, negative or two positive roots. In the first two cases, the roots are not acceptable, no prediction can be made and the analyst must either investigate the goodness of the prediction models or use other methods of prediction (for instance, $\tilde{\mu} = \ln \hat{\alpha}$ and $\tilde{\sigma} = \frac{\tilde{\mu} - \hat{\gamma}_1}{\hat{\gamma}_2}$). If (4.9) gives two positive roots one can investigate their sizes and perhaps reject a "bad" solution using empirical criteria such as $0 < \sigma < 1$.

Using the estimated values of γ_1 and γ_2 with (4.9), no complex root was obtained for a range of α from 20 to 600 minutes. This range gives the limits of block times for the present flight legs of the Air Canada Schedule.

Hence, using our estimates α from the prediction model (4.1), together with the estimated relation between μ and σ in (3.14) allows, using lognormal theory, the prediction of $\hat{\mu}$ and $\hat{\sigma}$ for a new B_{ij}. The estimate $\hat{\mu}$ is obtained by substituting $\hat{\sigma}$ from (4.9) into (4.7).

Some data were recorded in an attempt to use model (4.1). For this exercise, the data collected were for one season only (summer) and thus the season independent variable in (4.1) was therefore equated to zero. The second column of the Table 4.1 below gives the average block times for the city pairs, aircraft types and times of the day shown on the other columns of the table. These averages, calculated for the months of July and August 1973, served as the dependent variable in solving model (4.1).

Table 4.1

City Pair	Avg.Block time	Aircraft type	Time of day	Distance	Wind Effect
ANU -BDA	145.63	2	5	927	-2
ANU -YUL	266.22	1	6	1803	-7
ANU -YYZ	275.29	1	5	1833	-8
ANU -YYZ	273.95	2	6	1833	1
BDA -ANU	150.38	2	11	927	1
BDA -YUL	148.64	1	4	892	-15
BDA -YYZ	160.72	1	8	978	-19
BDA -YYZ	161.75	2	9	978	-19
BGI -YUL	305.17	2	11	2075	-5
BGI -YYZ	310.38	2	9	2088	-7
BRU -YUL	454.33	1	7	2997	-32

The time of day represents the hourly intervals: 1 for 7:00 - 7:59 hrs, 2 for 8:00 - 8:59 hrs, ..., 17 for 23:00 - 23:59 hrs, and 18 for 0:00 - 6:59 hrs. The distance (in great circle miles) and the wind effect were both obtained from the World Enroute Winds tables, published by the Boeing Company. The two aircraft types 1 and 2 represent the DC8 and D8S aircraft respectively.

Using the above data, model (4.1) was applied and gave the following results:

$$\text{Mean block} = 12.2665 + 2.0667 \times \text{Aircraft type} - .0392 \times \text{time of day} + .139639 \times \text{distance} - .584665 \times \text{wind effect.} \quad (4.10)$$

The F-ratio obtained for the above regression has 4 and 6 degrees of freedom, is $F_{4,6} = 1760.7$ which is highly significant. The $R^2 = .99915$

which is of course very high and supports the goodness of fit of the line. As we might expect the block time varies positively with the distance as is indicated by the estimated regression coefficient. Also as expected, the distance and the wind factor account for most of the variance. The t-values for these are 77.9 and -3.9 respectively which are both significant at $\alpha = .01$. When regressing the block time variable with the latter two independent variables, a good regression is also obtained:

$$\text{Mean block} = 15.7979 + .1396267 \times \text{distance} - .5251184 \times \text{wind effect.} \tag{4.11}$$

Here, the R^2 is .99908 It is believed that the aircraft type and time of day influences would be greater in (4.10) with a larger sample of observations.

When the regression (4.1) and the relation (3.14) are performed on a complete, representative set of data, the corresponding estimated regression coefficients may then be used to estimate $\hat{\mu}$ and $\hat{\sigma}$ as in (4.6). Table 4.2 below gives the values calculated by (4.10) and compares them with their corresponding observed values. As may be observed the percent relative errors tabulated below are quite small.

Using this set-up for predicting μ and σ, we are making some underlying assumptions. First, the space of the lognormal parameters is being restricted to a line which is estimated by (3.14). Also, α, the mean block time is assumed to be a linear function of some independent variables. In both cases we have obtained large R^2 and F-ratio's which support the goodness of fit of the lines. Therefore, it is felt that these assumptions are somewhat justified.

TABLE 4.2

Observed B_{ij} (minutes)	Calculated B_{ij} (minutes)	% error ((Calc.-Obs.) ÷ Obs.) x 100
145.63	146.82	.816%
266.22	269.96	1.405
275.29	274.77	−.188
273.95	271.54	−.880
150.38	144.83	−3.691
148.64	147.50	−.716
160.72	161.70	.607
161.75	163.72	1.220
305.17	308.64	1.138
310.38	311.71	.427
454.33	451.27	−.674

Alternatively, it may have been possible to pursue the prediction of $\hat{\mu}$ and $\hat{\sigma}$ by other approaches. For instance, we may elect to relate α and β^2 to other independent variables in a way similar to the prediction model for α in (4.1). When such estimates $\hat{\alpha}$ and $\hat{\beta}^2$ are obtained, we may solve for $\hat{\mu}$ and $\hat{\sigma}$ with the aid of the theoretical relations (2.4) and (2.5). We may also attempt to relate μ and σ^2 to some independent variables. No matter which prediction approach is selected, some assumptions on the B_{ij}, other than that of lognormality must be made. Thus, in any case, the problem becomes one of testing the validity of the assumptions, as we have done above.

APPENDIX A

RETRIEVED DATA FROM MONTH-LESS-CREW TAPE

(Per flight-leg)

A. Scheduled data.

1. Scheduled originating date
   - year
   - month
   - day
2. Scheduled day of year (1-365)
3. Day of week (1-7)
4. Flight serial number
5. Leg sequence number
6. Scheduled times
   - scheduled departure time and departure time zone
   - scheduled arrival time and arrival time zone
7. Stations
   - departure station
   - arrival station
8. Service Type
9. Scheduled leg-miles
10. Scheduled block minutes
11. Scheduled departure and arrival dates
    - year
    - month
    - day
12. Scheduled departure time for originating station of flight
13. Scheduled origination/termination code

B. Actual data.

1. Actual departure date
   - year
   - month
   - day

2. Actual times
   - time out
   - time off
   - time on
   - time in

3. Station delay minutes

4. Actual aircraft code

5. Aircraft serial number

6. Actual flight minutes

7. Actual block minutes

8. Actual leg miles

9. Irregularities
   - leg type
   - dupe code
   - irregular reason code

AIR CANADA FILE DESCRIPTION
OF MONTH-LESS-CREW TAPES

1. Flight Code.

   Identifies type of flight; codes are:

   SPACE - Scheduled Flight
   0 - Revenue Extra Section
   1 - Charter
   2 - Excess on Scheduled & Extra Section
   3 - Operational Extra Sections
   4 - Ferry
   5 - Test
   6 - Courtesy and Publicity
   7 - Familiarization Flights
   8 - Training Flights
   9 - Competency Flights

2. Scheduled Originating Date - Year, month, day.

   Date on which flight is scheduled to depart from originating station.

3. Flight Number.

   Flight designator assigned to a specific flight.

4. Scheduled Numeric Day of Year.

   Numeric day of year on which flight is scheduled to depart from originating station.

5. Numeric Day of Week.

   Numeric day of week on which flight is scheduled to depart from station, based on Scheduled Departure Date. Monday = day 1.

6. Leg Sequence Number.

   Numbering of flight legs in operating sequence within flight number.

7. Departure Station.

   From Station alphabetic designator.

8. Scheduled Departure Time.

   Scheduled time out of station in minutes.

9. Departure Station Time Zone.

Numeric Time Zone code of departure station.

10. Arrival Station.

To station alphabetic designator.

11. Scheduled Arrival Time.

Scheduled time into station in minutes.

12. Arrival Station Time Zone.

Numeric time zone code of arrival station.

13. Service Type Code.

Numeric service code. N. American = 1 Atlantic = 2 and Southern = 3.

14. Service Transfer Code.

Numeric service transfer code. Codes are:

1 = All domestic N. American flights
2 = International N. American flights
3 = Atlantic flights
4 = Southern flights

NOTE: Unlike Service code which identifies the service to which each flight leg is applicable, the service transfer code indicates the Service to which the entire flight (all flight legs) is chargeable.

15. Strata.

Strata number to which a flight leg applies.

16. Route Number.

Route number to which a flight leg applies.

17. Direction Code.

Direction in which a flight leg is operating.

1 = East/South
2 = West/North

18. Log Number.

Numeric code assigned to a flight leg regardless of direction.

19. Scheduled Leg Miles.

Flight leg mileage scheduled.

20. Scheduled Block Minutes.

Scheduled Elapsed time from scheduled time out to scheduled time in.

21. Scheduled Aircraft Code.

Alpha-numeric code of aircraft type scheduled to operate flight leg.

22. Installed Seats - First Class.

Number of first class seats installed on aircraft scheduled to operate the flight leg.

23. Installed Seats - Economy.

Number of economy seats installed on aircraft scheduled to operate the flight leg.

24. From Station Region Code.

New Customer Service Region Codes. Codes are:

Code	Region
3	U.S.
4	Southern
5	Western
6	Central
7	Eastern
8	European

25. To Station Region Code.

Same as item 24 only pertains to arrival station.

26. Flight Movement Load Flight Code.

Indicates Flight Code shown on Flight Routing File which could be changed (when writing out a new record) as a result of the Dupe code indicated on the Flight Routing File.

27. Flight Movement Load Dupe Code.

Dupe code indicated on Flight Routing File. Codes are:

Code	Explanation
1 & 2	Used when a change of equipment occurs at an unscheduled point.
3	Indicates a diversion. (Excess flying).
4	Indicates a Charter flight.
5	Indicates that an Extra Section flight leg is non-revenue.

28. Scheduled Departure Date - Year, month, day.

Date on which flight leg is scheduled to depart from departure station.

29. Scheduled Arrival Date - Year, month, day.

Date on which flight ~~leg is~~ scheduled to arrive at arrival station.

30. Day Code.

Indicates the number of days difference between the Scheduled Originating Date and the Scheduled Departure Date. e.g., if a flight were scheduled to originate on the first and the Scheduled Departure Date was the third, the Day Code would read '2'.

31. Scheduled Departure Time - Originating Station.

Scheduled Departure time of Originating Station in minutes.

32. Current Week Code.

The most current week (period) on the MTH-LES-CREW file is indicated as a zero in this field, previous week(s) are shown as a one.

33. Leg Type.

Code which indicates whether an irregularity has occurred on the flight leg. Codes are:

Code	Explanation
SPACE	No Irregularity.
1	Irregularity - Shows flight leg as it actually operated.
3	Irregularity - Shows flight leg as it should have operated.

34. Dupe Code.

Refer to item 27.

35. Irregular Reason Code.

Reason for Irregularity into or out of station. Codes are:

Code	Reason
1X	Cancelled.
2X	No stop.
3X	Landed (Unscheduled stop).
5X	Originating (Unscheduled origination).
6X	Terminating (Unscheduled termination).
7X	Returned (attempt).

36. Scheduled Origination.

Scheduled Originating station of Flight, indicated by a 1.

37. Scheduled Termination.

Scheduled terminating station of Flight, indicated by a 1.

38. Departure Delay Reason.

Reason Code for Delay out of Departure Station. This code will appear only when a station delay has occurred.

39. Departure Delay Minutes.

Number of minutes flight was delayed out of departure station, based on comparison of actual time out to scheduled time out.

40. Arrival Delay Reason.

Not used.

41. Arrival Delay Minutes.

Number of minutes flight was late arriving at station, based on comparison of actual time in to Scheduled time in.

42. Station Delay Minutes.

Delay minutes chargeable to Departure Station. Arrived at by subtracting the Arrival Delay minutes from the Departure Delay minutes. If the Arrival Delay minutes are greater than zero the Departure Delay minutes then become the Station Delay minutes.

43. Enroute Delay Minutes.

Number of minutes flight leg was delayed in flight. Arrived at by subtracting the Departure Delay Minutes from the Arrival Delay minutes at the next scheduled downline station.

44. Originating Departure Code.

Denotes a downline origination of a flight Signified by a 1.

45. Actual Departure Date - Year, month, day.

Date on which flight actually departed from Departure Station.

46. Actual Leg Times.

Actual times out of ramp, off ground, on ground, and in-ramp, indicated in minutes.

47. Actual Aircraft Code.

Alpha-numeric code of aircraft type actually operating flight leg.

48. Actual Aircraft Number.

Air Canada number of Aircraft actually operating flight leg.

49. Actual Flight Minutes.

Actual elapsed time from time off to time on.

50. Actual Block Minutes.

Actual elapsed time from time out to time in.

51. Scheduled Block Minutes Actual Aircraft.

Scheduled elapsed time from scheduled time out to scheduled time in.

52. Actual Leg Miles.

Flight leg mileage actually flown.

53. From Station Collator.

Air Canada numeric station identifier of the Departure station.

54. Maximum Payload Space Available.

Space available for Revenue Payload indicated in pounds.

55. Fuel Boarded.

Quantity of Fuel boarded indicated in either gallons or litres.

56. Mail Boarded.

Total pounds of Mail boarded.

57. Express Boarded.

Total pounds of Express boarded.

58. Freight Boarded.

Total pounds of Freight boarded.

59. Comat Boarded.

Total pounds of Comat boarded.

60. Mail Carried.

Total pounds of Mail carried.

61. Express Carried.

Total pounds of Express carried.

62. Freight Carried.

Total pounds of Freight carried.

63. Comat Carried.

Total pounds of Comat carried.

64. Seats Available-First Class.

First Class seats available for sale.

65. Seats Available-Economy Class.

Economy Class seats available for sale.

66. Total Passengers Carried.

Total passengers carried, First and Economy, including CON's and POS as indicated on the Flight Load Message.

67. Contingent Passengers.

Contingent Passengers carried.

68. Passengers Boarded-First Class.

Total First Class passengers boarded, includes Revenue and POS (NOC/AOG).

69. Passengers Boarded-Economy Class.

Total Economy Class passengers boarded, includes Revenue and POS (NOC/AOG).

70. Passengers Carried-Reservec First Class.

Total First Class passengers carried, includes Revenue and POS (NOC/AOG).

71. Passengers Carried-Reservec Economy Class.

Total Economy Class passengers carried, includes Revenue and POS (NOC/AOG).

72. Passengers Carried-First Class.

First Class Revenue passengers carried, includes adjustments.

73. Passengers Carried-Economy Class.

Economy Class Revenue passengers carried, includes adjustments.

74. NOC/AOG Passengers Carried-First Class.

First Class NOC/AOG (POS) passengers carried.

75. NOC/AOG Passengers Carried-Economy Class.

Economy Class NOC/AOG (POS) passengers carried.

APPENDIX B

This Appendix gives the results obtained from the BMD statistical package (program BMD02R) which performs a simple linear regression of the logged block time means on their standard deviations. The table of residuals is also given.

SUB-PROBLM 1

DEPENDENT VARIABLE 1
MAXIMUM NUMBER OF STEPS 4
F-LEVEL FOR INCLUSION 0.010000
F-LEVEL FOR DELETION 0.005000
TOLERANCE LEVEL 0.001000

STEP NUMBER 1
VARIABLE ENTERED 2

MULTIPLE R 0.8117
STD. ERROR OF EST. 0.4915

ANALYSIS OF VARIANCE

	DF	SUM OF SQUARES	MEAN SQUARE	F RATIO
REGRESSION	1	30.807	30.807	127.509
RESIDUAL	66	15.946	0.242	

VARIABLES IN EQUATION

VARIABLE	COEFFICIENT	STD. ERROR	F TO REMOVE
(CONSTANT	5.98801)		
2	-20.71715	1.83467	127.5092 (2)

VARIABLES NOT IN EQUATION

VARIABLE	PARTIAL CORR.	TOLERANCE	F TO ENTER

F-LEVEL OR TOLERANCE INSUFFICIENT FOR FURTHER COMPUTATION

LIST OF RESIDUALS

CASE NUMBER	Y X(1)	Y COMPUTED	RESIDUAL	X(2)
1	6.0829	5.2990	0.7839	0.0333
2	6.0776	5.3106	0.7670	0.0327
3	6.3054	5.3433	0.9621	0.0311
4	6.1336	5.4094	0.7242	0.0279
5	3.5525	3.6563	-0.1038	0.1125
6	3.5694	3.0482	0.5212	0.1419
7	5.2799	5.0427	0.2362	0.0456
8	5.2089	5.0698	0.1991	0.0443
9	5.1519	5.1649	-0.0130	0.0397
10	3.6740	4.1112	-0.4372	0.0906
11	4.7644	4.5838	0.1806	0.0676
12	4.7706	4.6725	0.0981	0.0635
13	4.8075	4.2534	0.5541	0.0837
14	4.7362	4.6965	0.0397	0.0623
15	4.3378	4.8492	-0.5114	0.0550
16	4.4404	4.4667	-0.0263	0.0734
17	6.1014	5.3536	0.7478	0.0306
18	4.5929	4.7850	-0.1921	0.0581
19	4.1754	3.7354	0.4400	0.1087
20	4.2208	3.0963	1.1245	0.1396
21	5.8768	5.3520	0.5248	0.0307
22	5.6345	5.3029	0.3316	0.0331
23	6.0604	5.2774	0.7830	0.0343
24	3.8118	4.0775	-0.2657	0.0922
25	4.8177	3.6492	1.1685	0.1129
26	4.2727	3.5256	0.7471	0.1189
27	5.9575	5.1473	0.8102	0.0406
28	5.0598	5.0110	0.0488	0.0472
29	4.2377	4.7968	-0.5591	0.0575
30	4.2316	4.8724	-0.6408	0.0536
31	4.3322	4.9505	-0.6183	0.0501
32	4.3535	5.0715	-0.7180	0.0442
33	3.7311	3.6571	0.0740	0.1125
34	3.7162	3.5256	0.1906	0.1189
35	3.6094	3.4943	0.1151	0.1204
36	3.5600	3.8744	-0.3144	0.1020
37	4.4746	4.9722	-0.4976	0.0490
38	4.4660	4.6288	-0.1628	0.0656
39	4.4476	5.0234	-0.5758	0.0466
40	4.4382	4.5610	-0.1228	0.0689
41	4.4419	4.5243	-0.0824	0.0706
42	4.8480	4.8326	0.0154	0.0558
43	4.8666	4.9171	-0.0505	0.0517
44	4.8414	5.0744	-0.2330	0.0441
45	3.6941	3.8688	-0.1748	0.1023
46	4.4425	5.0387	-0.5962	0.0453
47	3.5664	4.1318	-0.5654	0.0896
48	3.4760	3.9037	-0.4277	0.1006
49	3.5333	3.7361	-0.2028	0.1067
50	3.3042	3.2697	0.0345	0.1312
51	3.4262	3.8825	-0.4563	0.1016
52	3.4939	3.7433	-0.2494	0.1083
53	3.6096	3.9165	-0.3069	0.1000
54	3.4375	4.0077	-0.5702	0.0956
55	3.4844	3.1525	0.3319	0.1369

CASE NUMBER	Y X(1)	Y COMPUTED	RESIDUAL	X(2)
56	3.4970	4.1332	-0.6362	0.0895
57	3.4868	3.5583	-0.0715	0.1173
58	3.4929	3.4207	0.0722	0.1239
59	4.0327	4.4562	-0.4235	0.0759
60	4.0394	4.4394	-0.4000	0.0747
61	5.6127	5.4960	0.1167	0.0237
62	5.0008	5.1494	-0.1486	0.0405
63	5.0785	4.9704	0.1081	0.0491
64	5.0848	4.9196	0.1652	0.0516
65	4.0819	4.4280	-0.3461	0.0753
66	3.8923	4.2532	-0.3609	0.0837
67	3.8772	4.9339	-1.0567	0.0509
68	4.3060	4.1960	0.1100	0.0865

FINISH CARD ENCOUNTERED
PROGRAM TERMINATED

APPENDIX C

Application of the model on two cycles extracted from the summer 1973 Air Canada Schedule (D9S Aircraft).

TABLE 1

Cycle A:

FLIGHT 346

YAM	YYZ	YOW	YUL
07:00	07:55-08:25	09:15-09:35	10:05-

FLIGHT 227

YUL	YQG	YWG	YYC	YVR
-11:25	12:55-13:25	15:40-16:20	18:15-18:35	20:25

Cycle B:

FLIGHT 271

YWG	YQR	YYC	YVR
06:00	07:00-07:20	08:30-08:50	10:05-

FLIGHT 280 (YVR – YQR – YWG), FLIGHT 289 (YWG – YQR)

YVR	YQR	YWG	YQR
-11:25	13:15-13:35	14:30-15:30	16:30-

FLIGHT 228

YQR	YWG	YOW
-17:15	18:10-18:35	20:55

TABLE 2

Estimates of block time parameters

City Pair	Section of day ***	$\hat{\mu}$	$\hat{\sigma}$
YAM - YYZ	2	4.03362	0.057130
YOW - YUL	3	3.47600	0.100610
YQG - YWG	4	4.88817	0.040620
YQR - YWG	4	3.99912	0.062467
YQR - YYC	2	4.24126	0.053950
YUL - YQG	3	4.44474	0.044830
YVR - YQR	3	4.70895	0.051106
YWG - YOW	5	4.90770	0.045269
YWG - YQR	2	4.10333	0.066746
YWG - YQR	5	4.07327	0.055489
YYC - YVR	2	4.31898	0.067968
YYC - YVR	5	4.29096	0.055610
YYZ - YOW	2	3.84623	0.075578

TABLE 3

Estimates of transit time parameters (all services Domestic)

Station	Distance indices*	$\hat{\mu}$	$\hat{\sigma}$
YOW	6 - 6	2.96083	0.356370
YQG	1 - 1	3.44872	0.257650
YQR	6 - 6	2.96594	0.184573
YVR	6 - 1	3.56371	0.187682
YWG	1 - 1	3.19808	0.265860
YWG	6 - 1	3.18499	0.201312
YYC	1 - 6	3.12048	0.21036
YYC	6 - 6	3.05369	0.208036
YYZ	6 - 6	3.40492	0.126960

TABLE 4

Estimates of turn time parameters (all services Domestic).

Station	$\hat{\mu}$	$\hat{\sigma}$
YQR	3.04112	0.25120
YUL	3.63962	0.29239
YVR	3.69887	0.25943
YWG	3.62386	0.29128

TABLE 5

Station Delay frequency distribution

YOW (3)		YUL (3)		YVR (3)		YYZ (2)	
MINS.	FREQ.	MINS.	FREQ.	MINS.	FREQ.	MINS.	FREQ.
0	55	0	53	0	32	0	40
1	10	1	3	1	3	1	7
2	8	2	11	2	11	2	8
3	9	3	10	3	9	3	8
4	7	4	8	4	7	4	7
5	11	5	7	5	10	5	8
6	9	6	2	6	5	6	1
7	5	7	4	7	4	7	4
8	4	8	3	8	5	8	4
9	2	9	6	10	3	9	2
10	4	11	4	11	3	10	3
11	2	13	2	12	4	11	1
12	3	15	5	13	2	12	3
13	1	18	3	14	3	13	1
14	1	22	3	16	3	15	1
15	2	24	2	19	2	18	1
17	2	29	4	21	3	20	2
20	1	32	2	28	2	22	1
23	1	33	1	30	2	24	2
24	1	35	2	31	1	29	2
26	1	40	2	32	2	30	1
40	2	60	2	35	2	40	1
45	2	63	2	38	1	60	1
				40	2	68	1
				43	1	75	1
				50	[illegible]		

TABLE 5 (Continued)

YWG (2)

MINS.	FREQ.
0	33
1	6
2	9
3	5
4	1
5	1
6	2
7	2
8	1
10	1
20	1

YWG (5)

MINS.	FREQ.
0	26
1	6
2	2
3	6
4	3
5	6
6	1
8	1
10	2
12	2
13	1
14	1
18	1
23	1
30	2
35	1

YYC (5)

MINS.	FREQ.
0	32
1	9
2	9
3	9
4	8
5	9
6	6
7	3
8	1
9	1
11	1
13	1
14	1
16	1
19	1
25	1

YYC (2)

MINS.	FREQ.
0	36
1	4
2	4
3	8
4	6
5	8
6	4
7	5
8	5
9	1
10	1
11	2
15	2
16	1
20	1
22	1
25	1
48	1

YAM (2)***

MINS.	FREQ.
0	18
5	3
6	1
8	1
12	2
13	1
15	1
19	1
28	1
43	1
48	1

YQG (4)

MINS.	FREQ.
0	11
1	1
2	6
4	3
5	5
7	1
9	1
10	1
11	1
25	1
30	1

YQR (2)

MINS.	FREQ.
0	31
1	3
2	6
3	8
4	5
5	4
6	2
8	2
9	2

YQR (4)

MINS.	FREQ.
0	20
1	4
2	4
3	6
4	9
5	9
6	2
7	1
9	1
10	2
11	2
28	1
40	1

* Distance indices: 6 = 0 - 500 miles
1 = 500 - 1000 miles

** Section of day: 1 = 0 - 5:59 hrs, 2 = 6:00 - 8:59 hrs,
3 = 9:00 - 11:59 hrs, 4 = 12:00 - 14:59 hrs,
5 = 15:00 - 19:59 hrs, 6 = 20:00 - 24:00 hrs.

*** (.) Denotes the section of day.

ON TIME PERFORMANCE PROBABILITIES

DOS SUMMER 1973

D A = DEPARTURE / ARRIVAL DELAY.
L = DUE TO LEG.
PL = DUE TO PREVIOUS LEG.
C = DUE TO CYCLE.

CYCLE YAM-YYZ-YOW-YUL-YOG-YWG-YYC-YVR-YEG-YWG-

FLIGHT NO.	LEG	D A	DUE TO	MAXIMUM DELAY MINUTES 0	5	10	15	30	60
346	YAM-YYZ	D	L	0.600	0.700	0.767	0.867	0.933	1.000
		A	L	0.381	0.887	0.995	1.000	1.000	1.000
346	YYZ-YOW	D	L	0.364	0.709	0.836	0.891	0.964	1.000
		D	PL	0.402	0.768	0.951	0.995	1.000	1.000
		D	C	0.146	0.545	0.795	0.887	0.964	1.000
		A	L	0.842	0.988	1.000	1.000	1.000	1.000
346	YOW-YUL	D	L	0.385	0.699	0.867	0.930	0.972	1.000
		D	PL	0.687	0.845	0.929	0.969	0.999	1.000
		D	C	0.154	0.415	0.643	0.779	0.927	1.000
		A	L	0.181	0.741	0.977	1.000	1.000	1.000
227	YUL-YOG	D	L	0.376	0.652	0.759	0.937	0.922	0.979
		D	PL	0.993	0.997	0.999	1.000	1.000	1.000
		D	C	0.257	0.640	0.759	0.837	0.922	0.979
		A	L	0.912	0.995	1.000	1.000	1.000	1.000
227	YOG-YWG	D	L	0.344	0.813	0.906	0.938	1.000	1.000
		D	PL	0.651	0.811	0.909	0.959	0.999	1.000
		D	C	0.147	0.475	0.639	0.742	0.926	1.000
		A	L	0.696	0.920	0.988	1.000	1.000	1.000
227	YWG-YYC	D	L	0.419	0.790	0.855	0.919	0.954	1.000
		D	PL	0.958	0.989	0.997	1.000	1.000	1.000
		D	C	0.305	0.648	0.773	0.905	0.984	1.000
		A	L	0.593	0.849	0.964	0.994	1.000	1.000
227	YYC-YVR	D	L	0.344	0.817	0.935	0.968	1.000	1.000
		D	PL	0.367	0.621	0.822	0.935	1.000	1.000
		D	C	0.069	0.317	0.546	0.720	1.000	1.000
		A	L	0.725	0.960	0.998	1.000	1.000	1.000

ON TIME PERFORMANCE
PROBABILITIES

DOT SUMMER 1973

D A = DEPARTURE ARRIVAL DELAY.
L = DUE TO LEG.
PL = DUE TO PREVIOUS LEG.
C = DUE TO CYCLE.

CYCLE YWG-YOR-YYC-YVR-YOR-YWG-YOR-YWG-YOW-

				MAXIMUM DELAY MINUTES					
FLIGHT NO.	LEG	D A	DUE TO	0	5	10	15	30	60
271	YWG-YOR	D	L	0.522	0.887	0.984	0.984	1.000	1.000
		A	L	0.496	0.881	0.989	1.000	1.000	1.000
271	YOR-YYC	D	L	0.500	0.919	1.000	1.000	1.000	1.000
		D	PL	0.523	0.827	0.962	0.996	1.000	1.000
		D	C	0.261	0.760	1.000	1.000	1.000	1.000
		A	L	0.500	0.901	0.994	1.000	1.000	1.000
271	YYC-YVR	D	L	0.396	0.725	0.901	0.945	0.989	1.000
		D	PL	0.470	0.775	0.935	0.987	1.000	1.000
		D	C	0.109	0.430	0.841	0.945	0.989	1.000
		A	L	0.452	0.798	0.958	0.995	1.000	1.000
280	YVR-YOR	D	L	0.259	0.591	0.718	0.815	0.911	1.000
		D	PL	0.996	0.999	1.000	1.000	1.000	1.000
		D	C	0.252	0.573	0.713	0.812	0.911	1.000
		A	L	0.469	0.785	0.947	0.992	1.000	1.000
280	YOR-YWG	D	L	0.323	0.939	0.935	0.968	0.964	1.000
		D	PL	0.492	0.761	0.921	0.982	1.000	1.000
		D	C	0.079	0.375	0.591	0.736	0.899	1.000
		A	L	0.609	0.951	0.998	1.000	1.000	1.000
289	YWG-YOR	D	L	0.419	0.790	0.855	0.919	0.984	1.000
		D	PL	0.949	0.972	0.996	0.993	1.000	1.000
		D	C	0.314	0.646	0.746	0.843	0.934	1.000
		A	L	0.702	0.975	1.000	1.000	1.000	1.000
228	YOR-YWG	D	L	0.532	0.742	0.839	0.903	0.919	1.000
		D	PL	1.000	1.000	1.000	1.000	1.000	1.000
		D	C	0.492	0.712	0.824	0.898	0.919	1.000
		A	L	0.494	0.919	0.996	1.000	1.000	1.000
228	YWG-YOW	D	L	0.419	0.790	0.855	0.919	0.954	1.000
		D	PL	0.616	0.963	0.966	0.995	1.000	1.000
		D	C	0.172	0.507	0.674	0.797	0.905	0.947
		A	L	0.796	0.945	0.991	1.000	1.000	1.000

APPENDIX D

This appendix contains the listings of some subroutines required by model (1.6).

The first of these, DSCRLN performs the discretization of a lognormal variable $X \frown \Lambda(\mu, \sigma^2)$. Here, μ and σ are given by DMU and SDEV respectively. The resulting discrete variable has a discrete distribution p(t), t = DT, DT + 1, ..., DT + N - 1.

To compute the numerical convolution of two discrete random variables, X_1 and X_2, the subroutine CONV requires the corresponding discrete distributions $p_1(t)$, t = D1, D1 + 1, ..., D1 + N1 - 1 and $p_2(t)$, t = D1, D2 + 1, ..., D2 + N2 - 1. The distribution of the resulting variable is $p_3(t)$, t = D3, D3 + 1, ..., D3 + N3 - 1.

The VMAX routine computes the distribution of the maximum of two variables. The notation is similar to that of the CONV routine.

```
200 SUBROUTINE DISPLN(DMU,SDEV,P,DT,N)
220 DIMENSION P(1)
240 PB(X)=.5*ERRF(.707107*X)
260 DT=DMU-3.*SDEV
280 UT=DMU+3.*SDEV
300 DTL=EXP(DT)
320 UTL=EXP(UT)
340 ID=DTL
360 IU=UTL+.5
380 DT=FLOAT(ID)
400 T1=DT-1.
420 N=IU-ID
440 XX=(ALOG(T1+1.5)-DMU)/SDEV
460 P(1)=PB(XX)
480 DO 1 I=2,N
500 XY=XX
520 XX=(ALOG(T1+I+.5)-DMU)/SDEV
540 P(I)=PB(XX)-PB(XY)
560 IF(XX*XY.LT.0.)P(I)=.5+P(I)
580 1 CONTINUE
600 N=N+1
620 XX=(ALOG(T1+N-.5)-DMU)/SDEV
640 P(N)=.5-PB(XX)
660 RETURN
680 END
```

```
SUBROUTINE CONV(P1,D1,N1,P2,D2,N2,P3,D3,N3)
DIMENSION P1(1),P2(200),P3(200)
D3=D1+D2
N3=N1+N2
SUM=0.
PTOT=0.
J=0
DO 2 KK=1,N3
K1=MAX1(1,KK-N2+1)
K2=MIN1(KK,N1)
SU=0.
DO 1 K=K1,K2
1 SU=SU+P1(K)*P2(KK-K+1)
SUM=SUM+SU
IF(SUM.LT..001)GO TO 2
IF(SUM.GT..999)GO TO 4
J=J+1
P3(J)=SU
PTOT=PTOT+SU
IF(J.GT.1)GO TO 2
D3=D3+KK-1
2 CONTINUE
4 N3=J
SU=0.
NN3=N3-1
DO 5 L=1,NN3
P3(L)=P3(L)/PTOT
5 SU=SU+P3(L)
P3(N3)=AMAX1(0.,(1.-SU))
IF(N3.LT.150)GO TO 6
CALL FIXSIZ(P3,D3,N3,150)
6 RETURN
END
SUBROUTINE FIXSIZ(P1,D1,N1,NFIX)
DIMENSION P1(1),P2(200)
NDIF=N1-NFIX
IF(NDIF.LE.0)GO TO 1
L1=MAX0(NDIF/3,1)
A=P1(L1)
SU=0.
L2=N1-L1
IZ=1
DO 2 I=L1,N1
IF(P1(I).LT.A)GO TO 4
IF(IZ.GT.NFIX+5)GO TO 4
SU=SU+P1(I)
IZ=I-L1+1
2 P2(IZ)=P1(I)
4 IZ=IZ-1
SU1=0.
DO 3 I=1,IZ
P1(I)=P2(I)/SU
3 SU1=SU1+P1(I)
N1=IZ+1
SU=SU1
P1(N1)=AMAX1((1.-SU),0.)
D1=D1+L1-1
1 RETURN
END
```

```
4420 SUBROUTINE VMAX(P1,D1,N1,P2,D2,N2,P3,D3,N3)
4430 DIMENSION P1(1),P2(200),P3(200)
4440 D3=AMAX1(D1,D2)
4450 M1=IFIX(D1)
4460 M2=IFIX(D2)
4470 N=MIN0((M1+N1),(M2+N2))-1
4480 M=MAX0(M1,M2)
4490 N=N-M+1
4500 MDEV1=M-M1
4510 MDEV2=M-M2
4520 SU1=0.
4525 IF(N.LT.1)GO TO 10
4530 L1=MDEV1
4540 L2=MDEV2+1
4550 IF(L1.LT.1)GO TO 1
4560 DO 2 I=1,L1
4570 2 SU1=SU1+P1(I)
4580 1 SU2=0.
4590 DO 3 I=1,L2
4600 3 SU2=SU2+P2(I)
4610 P3(1)=P1(L1+1)*SU2+P2(L2)*SU1
4620 DO 4 J=2,N
4630 L1=L1+1
4640 L2=L2+1
4650 SU1=SU1+P1(L1)
4660 SU2=SU2+P2(L2)
4670 4 P3(J)=P1(L1+1)*SU2+P2(L2)*SU1
4680 N=N+1
4682 GO TO 11
4683 10 N=1
4684 IF(MDEV2.LT.1)GO TO 5
4686 GO TO 12
4687 11 CONTINUE
4690 IF((N-N1).GT.0)GO TO 5
4700 IF((N-N2).GT.0)GO TO 9
4705 12 CONTINUE
4710 DO 6 I=N,N1
4720 6 P3(I)=P1(I)
4730 N3=N1
4740 GO TO 7
4750 5 DO 8 I=N,N2
4760 8 P3(I)=P2(I)
4770 N3=N2
4780 GO TO 7
4790 9 N3=N
4800 7 RETURN
4810 END

READY
```

REFERENCES AND BIBLIOGRAPHY

1. Aircraft Codes for Types (1973). Report No. 929, Air Canada, Montreal.

2. Aircraft Handling Standards (1972). Report No. 210, Air Canada, Montreal.

3. Aitchison, J. and Brown, J.A.C. (1957). *The Lognormal Distribution, with Special Reference to its Uses in Economics*. University of Cambridge Press.

4. Bindler-Gaspard, D. (1964). Forecasting the Punctuality of a Route with Intermediate Stops. *Proceedings of the Fourth AGIFORS Symposium*, S.159f, Montbazon.

5. B.M.D. - Biomedical Computer Programs. (1973). University of California Press.

6. Block to Block Times and Great Circle Miles. (Sept. 1973). Report No. 828, Air Canada, Montreal.

7. Dequesnay, H. (1961). Simulation of a Caravelle Fleet Operation. *Proceedings of the First AGIFORS Symposium.*

8. Duncanson, A.V. (1974). Regularity Analysis with a Cycle-Simulation Model. Goran Stael von Holstein, Administrative Development, Scandinavian Airlines System. *Presented at the 14th AGIFORS Symposium*, Davos, Switzerland.

9. Etschmaier, M.M. and Roshstein, M. (1973). Estimating the Punctuality Rate Inherent in an Airline Schedule. Technical Report No. 19, Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261.

10. Franke, P. (1972). *Die Analyse der Zuverlassigkeit von Flugplanen.* Doctoral Dissertation, Technical University Darmstadt, Darmstadt, Germany.

11. Franke, P. (1973). Analysis and Improvement of Schedule Punctuality. Technical Report, Deutsche Lufthansa, Germany.

12. Harris, B. (1966). *Theory of Probability.* Addison Wesley.

13. Jackson, K.F. and Smith J.U.M. (1963). A Simulation Model of Airline Operations. *Proceedings of the Third AGIFORS Symposium.*

14. Kendall, M.G. and Stuart A. (1969). *The Advanced Theory of Statistics. Distribution Theory.* Vol. I, Third Edition, Charles Griffin & Company Limited, London.

15. Kendall, M.G. and Stuart A. (1973). *The Advanced Theory of Statistics. Inference and Relationshop.* Vol. II, Third Edition, Charles Griffin & Company Limited, London.

16. Lee, A.M. and Fearnley, J. (1963). Two Problems in Pre-Testing Proposed Flight Schedules by Means of Simulation. *Proceedings of the Third AGIFORS Symposium.*

17. Lilliefors, H.W. (1967). On the Kolmogorov-Smirnov Test for Normality with mean and variance unknown. *Journal of the American Statistical Association,* 62, 399 - 402.

18. Month-Less-Crew File Description (1973), Air Canada, Winnipeg.

19. Neal, M.B. and Hansen, E.M. (March 1967). World Enroute Winds. Commercial Airplane Division, The Boeing Company, Renton, Washington.

20. Oliver, R.M. (1962). Delays to Aircraft Serviced by the Glide-Path. *Operations Research.* Vol. 13, S.201f.

21. Operational Planning Review (1972). Project No. 123, Operational Research Division, Air Canada, Montreal.

22. Operational Research Report (January 1971). Project No. 105 (70/1), Operational Research Division, Air Canada, Montreal.

23. OPSIM Utilization (Operations Simulation). (1967). Project No. 69/4, Operational Research Division, Air Canada, Montreal.

24. Rao, C.R. (1965). *Linear Statistical Inference and its Application.* John Wiley & Sons, Inc., New York.

25. System Survey - Station Services - Central Region (1970). Report No. 1A, Air Canada, Montreal

26. Tobin, N. and Butfield, T. (1970). O.R. Models and their Implementation in the Airline's Planning Cycle. *Proceedings of the Tenth AGIFORS Symposium.*