NATIONAL BUREAU OF STANDARDS REPORT

10111

ANALYSIS OF A CAPACITY CONCEPT

FOR

RUNWAY AND FINAL-APPROACH PATH AIRSPACE

Inter-Agency Agreement DoT FA69-WAI-166
FAA Project Number 187-601-01R
SRDS Report Number RD-69-47



U.S. DEPARTMENT OF COMMERCE

NATIONAL BUREAU OF STANDARDS

AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO

NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards [1] was established by an act of Congress March 3, 1901. Today, in addition to serving as the Nation's central measurement laboratory, the Bureau is a principal focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. To this end the Bureau conducts research and provides central national services in four broad program areas. These are: (1) basic measurements and standards, (2) materials measurements and standards, (3) technological measurements and standards, and (4) transfer of technology.

The Bureau comprises the Institute for Basic Standards, the Institute for Materials Research, the Institute for Applied Technology, the Center for Radiation Research, the Center for Computer Sciences and Technology, and the Office for Information Programs.

THE INSTITUTE FOR BASIC STANDARDS provides the central basis within the United States of a complete and consistent system of physical measurement; coordinates that system with measurement systems of other nations; and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. The Institute consists of an Office of Measurement Services and the following technical divisions:

Applied Mathematics—Electricity—Metrology—Mechanics—Heat—Atomic and Molecular Physics—Radio Physics [2]—Radio Engineering [2]—Time and Frequency [2]—Astrophysics [2]—Cryogenics.[2]

THE INSTITUTE FOR MATERIALS RESEARCH conducts materials research leading to improved methods of measurement standards, and data on the properties of well-characterized materials needed by industry, commerce, educational institutions, and Government; develops, produces, and distributes standard reference materials; relates the physical and chemical properties of materials to their behavior and their interaction with their environments; and provides advisory and research services to other Government agencies. The Institute consists of an Office of Standard Reference Materials and the following divisions:

Analytical Chemistry—Polymers—Metallurgy—Inorganic Materials—Physical Chemistry.

THE INSTITUTE FOR APPLIED TECHNOLOGY provides technical services to promote the use of available technology and to facilitate technological innovation in industry and Government; cooperates with public and private organizations in the development of technological standards, and test methodologies; and provides advisory and research services for Federal, state, and local government agencies. The Institute consists of the following technical divisions and offices:

Engineering Standards—Weights and Measures — Invention and Innovation — Vehicle Systems Research—Product Evaluation—Building Research—Instrument Shops—Measurement Engineering—Electronic Technology—Technical Analysis.

THE CENTER FOR RADIATION RESEARCH engages in research, measurement, and application of radiation to the solution of Bureau mission problems and the problems of other agencies and institutions. The Center consists of the following divisions:

Reactor Radiation—Linac Radiation—Nuclear Radiation—Applied Radiation.

THE CENTER FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides technical services designed to aid Government agencies in the selection, acquisition, and effective use of automatic data processing equipment; and serves as the principal focus for the development of Federal standards for automatic data processing equipment, techniques, and computer languages. The Center consists of the following offices and divisions:

Information Processing Standards—Computer Information — Computer Services — Systems Development—Information Processing Technology.

THE OFFICE FOR INFORMATION PROGRAMS promotes optimum dissemination and accessibility of scientific information generated within NBS and other agencies of the Federal government; promotes the development of the National Standard Reference Data System and a system of information analysis centers dealing with the broader aspects of the National Measurement System, and provides appropriate services to ensure that the NBS staff has optimum accessibility to the scientific information of the world. The Office consists of the following organizational units:

Office of Standard Reference Data—Clearinghouse for Federal Scientific and Technical Information [3]—Office of Technical Information and Publications—Library—Office of Public Information—Office of International Relations.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D.C. 20234.
[2] Located at Boulder, Colorado 80302.
[3] Located at 5285 Port Royal Road, Springfield, Virginia 22151.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT		NBS REPORT
4314403	November 1969	10111

ANALYSIS OF A CAPACITY CONCEPT

FOR

RUNWAY AND FINAL-APPROACH PATH AIRSPACE

Inter-Agency Agreement DoT FA69-WAI-166
FAA Project Number 187-601-01R
SRDS Report Number RD-69-47

Technical Report
to the
Federal Aviation Administration
Systems Analysis Division
Systems Research and Development Service

by
Technical Analysis Division and
Applied Mathematics Division

IMPORTANT NOTICE

NATIONAL BUREAU OF S
for use within the Government
and review. For this reason,
whole or in part, is not autl
Bureau of Standards, Washin
the Report has been specifica

Approved for public release by the Director of the National Institute of Standards and Technology (NIST) on October 9, 2015.

ress accounting documents intended
is subjected to additional evaluation
re listing of this Report, either in
the Office of the Director, National
, by the Government agency for which
l copies for its own use.



U.S. DEPARTMENT OF COMMERCE

NATIONAL BUREAU OF STANDARDS

ABSTRACT

A "maximum throughput rate" concept, for the capacity of a runway and its associated final-approach path airspace, is developed as a possible alternative for some purposes to the present concept embodied in the Airport Capacity Handbook which is based on a "tolerable average delay" level. The new concept is shown to be representable by a simple mathematical formula. In the context of a stream of IFR landings, it is shown to have other properties useful in an operational setting, in particular to have potential value in connection with cost-effectiveness analyses of proposed changes in ATC equipment or procedures. Illustrative numerical calculations and parametric sensitivity analyses are included. A comparison with the ideas and numerical values in the Handbook is carried out. Suggestions for natural extensions of the present brief study are formulated. Technical appendices include peripheral studies of capacity-increasing techniques for deviating from a first-come first-served treatment of arriving aircraft.

TABLE OF CONTENTS

		Page
1.	INTRODUCTION AND SUMMARY	1
2.	THE CAPACITY CONCEPT	7
	2.1 PRELIMINARIES	7
	2.2 INFORMAL DESCRIPTION	9
	2.3 SYNOPSIS OF RESULTS	16
	2.4 MATHEMATICAL FORMULATION AND ANALYSIS	22
3.	APPLICATION TO IFR LANDINGS	33
	3.1 PRELIMINARIES	33
	3.2 TREATMENT OF TIEUP TIMES	35
	3.3 TREATMENT OF BUFFER TIMES	40
	3.4 SYNTHESIS AND COMMENTS	42
4.	NUMERICAL ILLUSTRATIONS	46
	4.1 DATA SETS EMPLOYED	46
	4.2 DISCUSSION OF RESULTS	52
	4.3 COMPARISON WITH SIMULATION OUTPUTS	61
5.	RELATION TO "HANDBOOK" VALUES	66
	5.1 PRELIMINARIES	66
	5.2 A BASIS FOR "MODIFIED HANDBOOK" VALUES	71
	5.3 COMPARISON WITH HANDBOOK CONCEPT	79
6.	POSSIBLE NEXT STEPS	87
7.	REFERENCES	90

TABLE OF CONTENTS (CONT.)

Page

APPENDIX A: PROOF OF EQUATION (2.6) 92

APPENDIX B: JUSTIFICATION OF EQUATION (2.7) 96

APPENDIX C: ON SEQUENCING OF ACCEPTANCES 102

APPENDIX D: MORE ON SEQUENCING 105

APPENDIX E: A DELAY CONCEPT 117

APPENDIX F: A SUPPORTING SIMULATION 126

APPENDIX G: EXTENSION TO A DUAL-USE RUNWAY 135

Figure 2.3.1: DECOMPOSITION OF INTERTOUCHDOWN TIME 19

Figure 2.4.1: SEQUENCE OF INTERSERVICE INTERVALS 22

Figure 3.3.1: TRIANGULAR DISTRIBUTION ON [-R,R] 41

Figure 3.3.2: UNIFORM DISTRIBUTION ON [-R,R] 41

Figure 4.2.1: SENSITIVITY TO CHANGES IN FINAL APPROACH DISTANCE . . 54

Figure 4.2.2: CAPACITY AS A FUNCTION OF POPULATION MIX 56

Figure 4.2.3: SENSITIVITY TO CHANGES IN R 58

Figure 4.2.4: EFFECTS OF R ON EFFICIENCY MEASURE 59

Figure 4.3.1: ANALYTICAL VS. SIMULATION RESULTS (UNIFORM DISTRIBUTION) 63

Figure 4.3.2: ANALYTICAL VS. SIMULATION RESULTS (TRIANGULAR DISTRIBUTION) 64

Figure F.1: SOME SIMULATION OUTPUTS 127

Figure F.2: LISTING OF SIMULATION PROGRAM 133

TABLE OF CONTENTS (CONT.)

		Page
TABLE 4.1.1:	FINAL-APPROACH PATH LENGTHS AND MINIMUM SEPARATION DISTANCES	47
TABLE 4.1.2:	DISTRIBUTIONS OF ERROR IN INTERTOUCHDOWN TIME	47
TABLE 4.1.3:	THRESHOLD PROBABILITIES OF VIOLATION	47
TABLE 4.1.4:	MIXES OF AIRCRAFT TYPES	47
TABLE 4.1.5:	FINAL-APPROACH SPEED AND RUNWAY OCCUPANCY TIME	47
TABLE 5.2.1:	ESTIMATES OF AVERAGE HOLDING TIMES BY AIRCRAFT TYPE	74
TABLE 5.2.2:	PREDICTED VS. EXTRACTED HANDBOOK VALUES	75
TABLE 5.3.1:	COMPARISON WITH VALUES FROM HANDBOOK	82
TABLE 5.3.2:	COMPARISON WITH MODIFIED HANDBOOK VALUES USING INTERTOUCHDOWN TIMES	82
TABLE 5.3.3:	COMPARISON WITH MODIFIED HANDBOOK VALUES USING RUNWAY OCCUPANCY TIMES	82
TABLE D.1:	PENALTY COSTS FOR SEQUENCING EXAMPLE	110
TABLE D.2:	FRACTIONAL PENALTY-REDUCING SOLUTION	110
TABLE F.1:	SIMULATION APPROXIMATIONS OF $K=c^{1/2}$	129

1. INTRODUCTION AND SUMMARY

The currently employed concept, for measuring the "capacity" of an airport's runway system and the associated final-approach airspace, is that set forth in the Airport Capacity Handbook [1][(1)] and based on a number of mathematical and empirical studies carried out by the Airborne Instruments Laboratory (e.g. [2], [3]). In connection with the FAA's Continuing Study of Air Traffic Control System Capacity and Demand[(2)], it appeared advisable to examine and re-evaluate the assumptions on which this concept was based, and as appropriate to initiate the exploration of alternative concepts.

The present report documents one such technical exploration, carried out at the National Bureau of Standards[(3)] by members of the Bureau's Technical Analysis Division (Simulation Group) and Applied Mathematics Division (Operations Research Section). Our assignment had as its theoretical core the task of developing, and bringing to a point permitting numerical application, a mathematical model of a "capacity" concept with the following features:

(a) Arrivals represented by perturbations to a deterministic process, rather than highly random (Poisson distributed) inputs.

(b) Known individual service times, rather than a constant service time.

---

(1) Numbers in square braces indicate references in Section 7 of this report.

(2) FAA Subprogram 187-601, assigned to the Systems Analysis Division, Systems Research and Development Service.

(3) Under Inter-Agency Agreement DoT FA69 WAI-166, PR No. WA5I-9-0629.

The study was undertaken on a short-term[4] "best effort" basis. This had several implications, recognized in advance by the FAA and the study team, for the course of the work. One pertained to technical scope; the new capacity concept was to be developed in detail only for the case of a single runway, a "pure" situation in which attention could focus on the concept's essential features without distracting complications.[5] Other implications were that only readily available data could be employed (i.e., no data-gathering or data-assimilation tasks were included), and that orientation for the study's staff would necessarily be quite limited. Under these circumstances the study was of course critically dependent on close cooperative liaison with the FAA, and we would be remiss in failing to acknowledge here both the constant helpfulness and the substantive contributions of our project monitor at the FAA, Mr. S. P. E. Price.

The body of the report is organized as follows. Section 2 describes the underlying ideas of the new concept, in the general context of "capacity of a service facility serving several customer types." The essential technical difficulty stems from the fact that the allowable interval, between providing service for the first of two consecutive customers and providing service for the second, depends upon the types of both.[6] This complication is overcome by drawing on some relatively recent developments in the mathematical theory of random processes. The result is a simple closed-form formula for the facility's capacity, in

---

[4] Roughly two calendar months.

[5] Prior to operational use of the concept, its development would of course have to be extended to encompass interactions between neighboring or intersecting runways.

[6] E.g., the required distance separation between consecutive aircraft in the final approach of IFR landing translates into a time separation which depends on the approach speeds of both A/C.

terms of (i) the relative proportions of the different customer types in the mix of customers to be served, (ii) for each pair of customer types, the nominal desired interval between provision of service to two consecutive customers of these respective types, and (iii) the precision with which such desired intervals can be achieved. See eq. (2.17), p.31.

This approach appears applicable, as it stands, to situations in which a runway is being used either for arrivals (landings) only, or for departures (takeoffs) only. In Section 3, the application process is illustrated by carrying it out in detail for a stream of IFR landings. The resulting specialized formulas[7] permit one to study the effects, on capacity, of technological or operational improvements which for example might permit reductions in runway occupancy times or minimum separation criteria (e.g., the traditional "3-mile rule"), or might achieve higher precision in the measurement and control processes involved in the landing operation. The capability for such analyses seems quite important, if a capacity concept is to be readily usable for cost-effectiveness analyses of possible new equipment or procedures.

Section 4 continues the illustrative treatment of IFR landings, presenting numerical values of the new capacities calculated for a number of sets of input data supplied by the FAA. These sets exhibit variations in

(a) the mix of aircraft types in the arriving stream,

(b) the length of the final approach path,

---

[7] See eqs. (3.25-26), p. 42.

(c) the prescribed minimum separation distance,

(d) the "just tolerated" probability of violating the separation criterion, and

(e) the probability distribution of deviations from a desired inter-touchdown time interval.

The results are displayed in several graphical formats to help communicate the capacity measure's sensitivity to the various parameters involved.

In Section 5, some of these capacity values are compared with the values given in the Airport Capacity Handbook for the corresponding cases. Our information, on the full set of assumptions governing the results in the Handbook, is not sufficiently explicit to leave us entirely confident of the strict comparability of the two sets of values; on the other hand, since the Handbook is intended to be employed "in the field" without further explanatory material, acceptance of its values for use in the comparison seemed appropriate. In addition, "modified Handbook values" intended to be better-suited for the comparison were derived. The calculated new capacity values were found to differ systematically and significantly from the Handbook values (both modified and unmodified).

The present document of course constitutes only an initial exploration of the new capacity concept. Section 6 sketches some of the next steps that would be involved in bringing the concept closer to operational status. Such further development may well be useful to the FAA in connection with its Continuing Study of Air Traffic Control System Capacity and Demand.

Apart from a bibliography, the remainder of the report consists of technical appendices. Some of these (A and B) contain detailed mathematical justifications of formulas given in the main text. One (Appendix F) describes a fast-time simulation performed to check the adequacy of an approximation entering the analytical treatment. Two (C and D) contain mathematical studies, stimulated by but somewhat peripheral to the main theme of the project; they deal with possible strategies for replacing a first-come-first-served policy toward arrivals with some sort of "optimal sequencing."

The two appendices not mentioned yet contain discussions of topics quite germane to the central purposes of the present study. One (Appendix G) deals with the extension of the capacity concept to the case of a runway used for both takeoffs and landings. Such an extension is evidently of considerable practical and theoretical importance; only the beginnings of a treatment could be developed during the time available, and this topic probably tops the list of significant next steps in pursuing the present line of research.

The other appendix (E) takes up the development of a delay concept to be associated with the new capacity concept. What seems of special interest here is that the attempt led (indeed, drove) us to take into account the "capacity" of the next higher layer in the Air Traffic Control System. This association, of a quality-of-service measure (delay) at one level of the ATC system with a capacity measure one level higher, is suggestive of possible similar relationships at other elements and levels of the system.

We conclude this introduction by listing the members of the project staff[8]:

Simulation Group (Technical Analysis Division):

M. J. Aronoff (Project Manager)

M. D. Maltese

J. T. McQueen

W. Steele

M. Wanger (Consultant)

Operations Research Section (Applied Math Division):

A. J. Goldman

W. A. Horn

J. Levy

M. H. Pearl

[8] Almost all participated on a part-time basis.

2. THE CAPACITY CONCEPT

2.1 Preliminaries

Before the description of the new capacity concept is begun, some preliminary remarks by way of orientation may be helpful.

The first point to be addressed is the whole idea of "alternative capacity concepts." For contrast, consider the length of a rod. One can propose different methods and instruments for measuring this length, and discuss the relative precisions and accuracies of each, but such discussions refer to alternative approaches to the measurement of a well-defined physical property of the rod [1], not to "alternative concepts" of length! Why then is it sensible to consider alternative concepts for the capacity of a runway [2]?

The answer, of course, is that "capacity" is not a "well-defined physical property" of the runway. A particular capacity concept is an attempt to capture the essentials of a particular intuitive notion, in a quantitative form suitable for certain uses. Studies which begin with different intuitive notions, or are directed toward different goals, may well find different capacity concepts appropriate. (The situation is analogous to that in engineering economics, where the appropriate definition of "plant capacity" can be strongly dependent on the use to

---

[1] Assuming specified environmental conditions, and that the context is not a level of detail at which surface irregularities and the like become significant.

[2] Throughout the report, we shall use "runway" as an abbreviation for the awkwardly long phrase "runway and its final-approach path airspace."

which such "capacity numbers" are to be put.)

This leads to the second point to be stressed here. The capacity concept embodied in the Airport Capacity Handbook expresses the following intuitive notion:

> maximum traffic rate which can be accommodated without average delay reaching an unacceptable level.

On the other hand, the capacity concept to be described below is based on a quite different intuitive notion, of the "maximum throughput rate" variety; it is not directly associated with any quality-of-service indicator such as average delay [3], the study of such indicators (in particular, the evaluation of their levels as more or less acceptable) being regarded as admitting and deserving attention somewhat apart from "capacity" considerations. We must therefore explicitly ask the reader, who through familiarity with or continuing use of the Handbook has developed an automatic association of "capacity" with "delay," to suspend this association of ideas in examining the new capacity model.

---

[3] A tentative treatment of a "delay" concept compatible with the new capacity concept is given in Appendix E.

2.2 Informal Description

The new capacity concept will now be described, in general terms relating to a facility which serves a stream of customers of several types.[3a] Adopting this initial generality of language has two advantages. First, it promotes the recognition of possibly useful analogies between airport runways and other specific examples of service facilities; these may for example provide clues toward helpful literature references. Second, it places in sharper focus those more special features which characterize the airport runway situation.

As noted earlier, the intuitive notion underlying the capacity concept is of a "maximum throughput" type. We therefore list two factors which are major influences on a facility's throughput rate:

(a) Customer availability. Since customers cannot be served before they arrive, the throughput rate would be degraded by idle intervals in which no customers are at hand.

(b) Tieup times. Very roughly, this refers to how long a customer "ties up" the facility, i.e., how fast the facility can service that customer and get on to the next one. Clearly, a reduction in such times (due, say, to improved equipment or operational procedures) will tend to increase the facility's throughput rate.

Before leaving this subsection, we shall have to come back for a more searching examination of this "tieup time" notion. But for the moment, let us return to the "maximum throughput rate" notion and ask with respect to which of the two preceding factors the "maximization"

[3a] Unless otherwise specified, customers are always assumed served in order of arrival.

should take place. This depends upon the anticipated uses of the capacity concept. The concept under study here is intended to prove useful in cost-effectiveness analyses and comparisons of possible changes in equipment or procedures, changes which might reduce tieup times. Thus the concept should not involve some prior optimization with respect to tieup times, for this would "wash out" the capability to use the concept in comparing approaches to the improvement of this factor.

On the other hand, the factor of customer availability is a natural candidate for the "maximization." While idle periods at the facility affect its actual throughput rate in a given situation, they are irrelevent to its potential for throughput under heavier workloads, and hence to its "capacity" as normally conceived. The "maximum" aspect of our concept will therefore be expressed in the following assumption of continuous demand[4]:

> As each customer's service is completed, another customer is at hand.

With this point clarified, the following informal description of the capacity concept can be given:

> the average mean throughput rate over a prolonged period of continuous demand.

Here "throughput" has the familiar meaning of "number of customers served," but other terms in the description require explanation. First,

---

[4] This is consistent with the treatment in [4] (p. 67), in which "capacity" is considered only for situations satisfying the analog of the continuous demand assumption.

a "prolonged" period is one which is long relative to the maximum tieup time for a single customer. If the capacity concept is to be useful, it should be numerically insensitive to the exact duration of the period, and the range of "prolonged periods" should include durations which are typical rather than unreasonably long relative to peak load conditions at the facility[5].

Second, although the terms "average and "mean" are generally used interchangeably, their employment above involves a deliberate distinction. The mean throughput rate of the facility during a period of duration T is the number of customers served during that period, divided by T. There is some inexactness here, as to whether or not to count in customers whose service is in progress at the beginning or end of the period. This, however, involves an uncertainty of at most 1 or 2 customers in the count, an uncertainty which is negligible since the period in question is a prolonged one.

For the applications of interest here, the tieup time for a customer can depend on his type, and perhaps on the type of the next customer as well. Thus the mean throughput rate during a period of time will depend on the exact sequence of types exhibited by the customers served during that period. Clearly a capacity concept based on such detailed specifications, of a sequence of types almost certain to change from day to day, would be of little practical use. What is done, therefore, is (in effect) to average the mean throughput rate over all possible sequences

---

[5] That is, we would not want a runway capacity concept which applied only to periods of several hours' continuous demand.

of those customer types liable to use the facility during the period; the "weights" (probabilities) involved in the averaging operation are of course based on the relative proportions of the various customer types in the "mix" which the facility serves. This explains the usage of "average." Of course the possibility of frequent large deviations from the average must be looked into.

Because of the assumption of continuous demand, the facility's capacity (as described above) is limited only by the tieup times associated with the successive customers it serves. This "tieup time" notion, which is so critical to the capacity concept, must therefore be examined with some care; this will occupy the next part of our informal discussion.

One initial point is that a customer can tie up the facility for longer than the "processing time" physically required to provide his service. For example, some sort of clean-up may be required before another customer can be served. On a production line, a "setup time" after, and in addition to, the processing time may be needed if the next customer (production batch) is of a different type from the current one. This is a good example for present purposes, since it illustrates how a customer's tieup time might depend in part on the type of the <u>next</u> customer. At any rate, the principal idea is that tieup times will in general be longer than processing times; rather, they correspond to what are called "holding times" in the queuing literature stimulated by communications-network considerations.

Another aspect of tieup times is that they may not be determined solely by physical laws (in the simplest case, that two bodies can't occupy the same space at the same time), but also by man-made regulations. Safety considerations, for example, may lead to rules inhibiting the next customer in approaching the facility while the current customer is being served, or even from having followed the current customer too closely toward the facility. For the runway case, this regulatory element figures very strongly. The distinction is important, because it is physically possible for constraints expressing man-made rules (rather than physical laws) to be violated; such violations may typically be inadvertent, infrequent, and slight, but they can in principle occur.

The notion of tieup time will now be made more concrete, but at the cost of introducing two more ideas. One is that of an identification point in the use of the facility by a customer; this is merely some clear-cut stage in the service process which can be defined in a uniform way for all customers. It might for example be defined to be the start of processing or the end (for landing aircraft, these correspond respectively to initiation of final approach, and to turnoff from the runway). We have chosen it in later sections to be a clearly distinguishable intermediate point (namely, touchdown).

The second new idea is that of the interservice interval associated with a customer. This is simply the time interval between the moment at which the customer's identification point occurs, and the corresponding moment for the next customer. The length (duration) of this interval will

be called the customer's interservice time. The "identification point" idea was introduced just in order to provide definite moments between which to measure the interservice time.

We can now define

tieup time = required minimum value of interservice time.[6]

The word "required" is intended to recall the largely "regulatory" nature of the tieup time, and the fact that violations of the requirement are physically possible. Because of this possibility, it is necessary to distinguish clearly among

tieup time

(actual) interservice time

desired interservice time.

Ideally, the second and third of these would coincide with the first; the capacity resulting in this case might be called the ideal capacity of the facility. In practice, however, an interservice interval cannot be made to assume a desired (aimed-at) duration exactly. Deviations from perfect accuracy may be due to any or all of a variety of reasons: inevitable residues of imprecision in knowledge as to when the current service period began, as to the exact duration between start-of-service and identification point, etc.; inability to bring the new customer into "servicing position" at precisely an intended moment; and so forth.

---

[6] Thus, tieup time (more precisely, the associated interval) corresponds to "service time" as treated in [4], p. 24.

As a consequence of this inevitably less-than-perfect accuracy, choosing the desired interservice time to coincide with the tieup time may be too risky, creating an unacceptably high probability that the actual interservice time will turn out to be less than the tieup time -- i.e., that a violation will occur. We shall therefore assume that the desired interservice time which the facility's operators aim at achieving, is systematically somewhat larger than the required minimum level[7] (the tieup time). The increment will be called the buffer time; it is a "safety margin" added to the tieup time in order to reduce the probability of violation. With this addition, a violation will only occur when the random error in achieving the desired interservice time is so large (negative) as to more than offset the buffer time; such extreme values of the random error should be relatively rare.

The ratio of the "actual" capacity (which involves the buffer times) to the ideal capacity defined above, provides one measure of the facility's efficiency[7a]. This measure of course has only a limited scope of relevance; it does not pertain to possible improvements (in equipment, procedures, or whatever) or possible regulation changes which might reduce tieup times, but only to the effect on capacity of the imperfections in measurement and control which make the buffer times necessary.

---

[7] This idea is taken from [5]. There is no implication that the increase need be officially sanctioned, or even deliberately set and consciously recognized by the operators themselves; see however [8], p. 29, para. 1.

[7a] This has little relation to the usage of the same term in [4] (pp. 154-155); the latter in fact comes close to our concept of "capacity."

2.3 Synopsis of Results

The preceding informal discussion can be summarized as follows:

First, the facility's "capacity" under the proposed concept is defined as its average mean throughput rate over a prolonged interval of continuous demand. Here the "average" is in effect one over all possible sequences of customer types which might arise in the stream being served, the necessary "weights" or probabilities for the average being derived from the proportions of different customer types among the population of facility users. The term "continuous demand" refers to a heavy-traffic condition (the only situation in which a capacity concept is of much interest), represented by the assumption that at the completion of each service there is another customer at hand to be served.

Second, throughput rate (and hence capacity) is limited by two sets of factors:

(a) One set consists of tieup times, which measure how long each customer has to "tie up" the facility. These depend on the quality of the technology and operational procedures employed, which determines how rapidly a customer can be served and how much time need be spent on other necessary activities (cleanup, setup, etc.). They also depend on regulations which prescribe minimum separations between consecutive customers; since such regulations are not physical laws, their occasional violation is physically possible. Specifically, tieup times enter as minimum required interservice-interval durations.

(b) The second set of limiting factors consists of buffer times. They arise because the service process cannot avoid some degree of random error in seeking to achieve a specified interservice time. Specifically, the buffer times are increments which are added to the tieup times to obtain desired (aimed-at) interservice time spacings large enough that attempts to achieve them, despite random errors, have only a very low probability of violating the tieup time regulations. Thus the necessary sizes of the buffers depend on how tightly the distribution of errors is "packed" around zero, i.e., on the precision available in the landing and landing-control process.

In the next subsection 2.4, a simple mathematical formula to express the capacity concept will be derived. Then, in Section 3, this formula will be specialized to the case of a stream of IFR landings at a runway. The resulting specialized formula is the basis for the numerical work reported in Sections 4 and 5.

The derivations of formulas, in 2.4 and Section 3, necessarily involve material which is relatively detailed and technical. Some readers may prefer to omit this material, going at once to the more concrete illustrations in Section 4. In the rest of the present subsection, we therefore provide a synopsis of the results of subsection 2.4 and Section 3, to provide a direct bridge to Section 4. The results are of course stated without proof, and no mention is made of the various alternatives and generalizations taken up in the detailed text.

The situation under consideration is a continuous-demand stream of IFR landings at a single runway. Interservice times are measured, for definiteness, between successive touchdowns.

Suppose first, for simplicity, that only two aircraft types are involved: A and B. These types are present, in the mix of aircraft using the runway, in certain relative proportions p_A and p_B (positive numbers summing to 1). Let

r_{AB} = tieup time (required minimum interservice time[7b]) for a type A aircraft if followed by a type B aircraft;

r_{BA}, r_{AA} and r_{BB} are defined analogously. Also, let

b_{AB} = buffer time for a type A aircraft if followed by a type B aircraft.

b_{BA}, b_{AA} and b_{BB} are defined analogously.

Consider a touchdown chosen at random. The probability that it is an "AB-touchdown", i.e., involves a type A aircraft followed by a type B aircraft, is given[7c] by the product $p_A p_B$. In this case the (actual) intertouchdown time, denoted I_{AB}, is equal to the desired (aimed-at) intertouchdown time plus a random (unbiased) error term, e. Furthermore, as described earlier, the desired intertouchdown time is the sum of tieup time and buffer time. Therefore

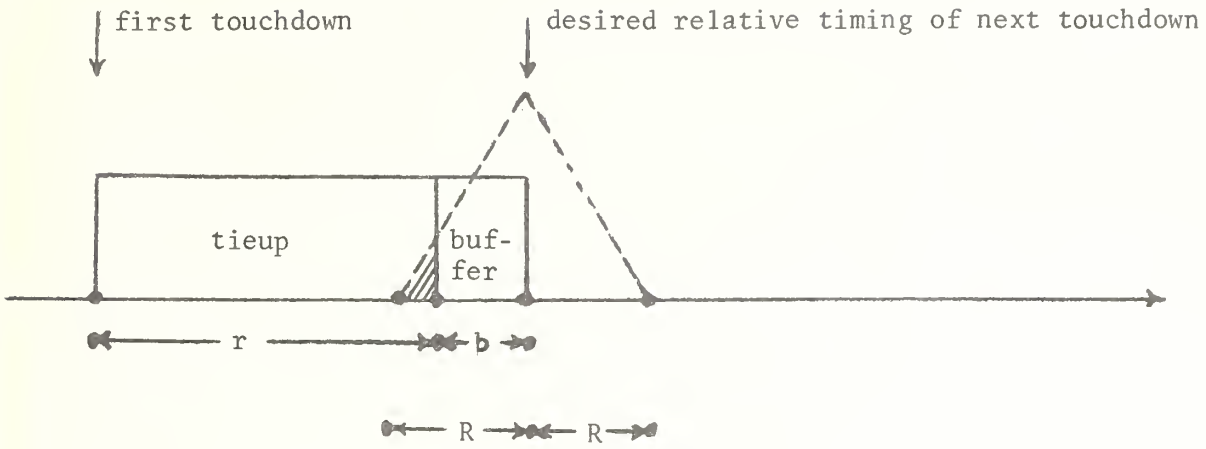
$$I_{AB} = r_{AB} + b_{AB} + e.$$

See figure 2.3.1, in which the random error e is assumed to follow a symmetric triangular distribution ranging from (-R) to (+R), where R is a measure of control precision.

[7b] Minimum allowable time implied by separation regulations and physical laws.

[7c] Recall the first-come-first-served assumption.

Figure 2.3.1: Decomposition of Intertouchdown Time (see Explanation)



Explanation: Dotted "curve" shows probability distribution of the random error in achieving desired intertouchdown time. Area both under this curve and preceding end of tieup interval (see shaded triangle) gives probability of violation; it would be increased to 0.5 if there were no buffer time, i.e., if the dotted distribution were shifted to be centered at the end of the tieup interval.

The average value of I_{AB} is therefore

$$(I_{AB})_{av} = r_{AB} + b_{AB}.$$

Similarly, AA - touchdowns occur with probability $p_A p_A = (p_A)^2$, and give rise to an average intertouchdown time of

$$(I_{AA})_{av} = r_{AA} + b_{AA},$$

with like results for the other two cases (BA-touchdowns and BB-touchdowns). The average length of a random touchdown is therefore given by the expression

$$P_A P_B (r_{AB} + b_{AB}) + P_A P_A (r_{AA} + b_{AA})$$

$$+ P_B P_A (r_{BA} + b_{BA}) + P_B P_B (r_{BB} + b_{BB}),$$

obtained by summing the average touchdown times for the four cases, each "weighted" by the probability of that case.

It is shown in subsection 2.4 that the capacity is given by the reciprocal of the average length of a random touchdown, i.e. the reciprocal of the expression above. For more than two aircraft types, this formula generalizes to

$$C = 1/\Sigma_{ij} P_i P_j (r_{ij} + b_{ij})$$

where the sum is over all ordered pairs (i,j) of aircraft types.

Now it is necessary to specify how the numerical values of tieup times (r's) and buffer times (b's) are to be determined. For an AB-touchdown, these would be r_{AB} and b_{AB}.

The tieup time r_{AB} is found as the larger of two quantities, corresponding to the two regulations to be observed. One of these quantities is the typical runway occupancy time for type A aircraft; its presence corresponds to the requirement that the next aircraft (here, type B) not touch down until the current one (here, type A) has cleared the runway. The second quantity is a straightforward translation, into a separation of touchdown times, of the minimum distance separation

imposed along the final approach path. This translation of course involves the final approach speeds of the two aircraft (with due attention to whether the gap between them is opening or closing during the approach), the size (S) of the required distance separation -- e.g., 3 n. mi. --, and the length (L) of the final approach path.

The buffer time b_{AB} is chosen, as described earlier, to reduce the probability of violation to a realistically "endurable" threshold level. This permits its calculation (details in Section 3) in terms of (i) the threshold level (p_v) and (ii) the probability distribution of the random error in attaining a desired intertouchdown interval. For our illustrative work, this distribution was taken to range from (-R) to (+R), and to be either triangular (as in Figure 2.3.1) or uniform; thus it is characterized by the single datum R.

This completes the synopsis. For the next subsection, we revert to the "service facility" level of generality.

2.4 Mathematical Formulation and Analysis

To begin the mathematical formulation of the material described informally in subsection 2.2, we consider a stream of customers[7d]

$$C_1, C_2, \ldots, C_n, \ldots$$

for the facility. For simplicity, the moment at which the interservice interval of C_1 begins is taken as the origin of the time axis, $t = 0$. Let

$$I_1, I_2, \ldots, I_n, \ldots$$

designate the corresponding sequence of lengths of (actual) interservice intervals, e.g., the interservice period of C_3 begins at time $t = I_1 + I_2$.

For any interval of time $[0,T]$ starting at $t = 0$, let

$N(T)$ = number of interservice intervals in the period $[0,T]$, so that

$N(T)/T$ = mean throughput rate in $[0,T]$.

$N(T)$ is explicitly defined by the conditions

$$I_1 + I_2 + \ldots + I_{N(T)} \leq T < I_1 + I_2 + \ldots + I_{N(T)} + I_{N(T)+1}. \qquad (2.1)$$

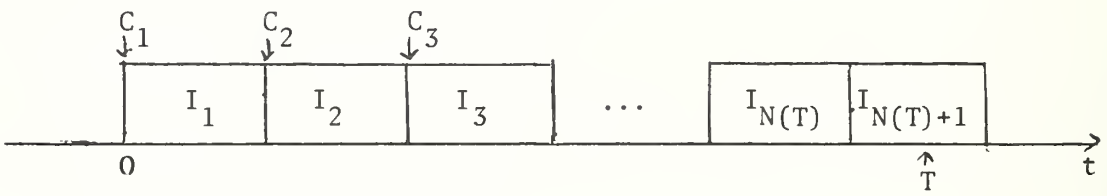


Figure 2.4.1: Sequence of Interservice Intervals

[7d] The symbol "C" without subscript is reserved to mean "capacity."

We can delimit the mean throughput if we have estimates of the largest and smallest interservice intervals that would ever occur under continuous demand, i.e., quantities I_{max} and I_{min} such that

$$I_{min} \leq I_n \leq I_{max} \qquad \text{(all n)}.$$

For then it follows readily from (2.1) that

$$(1/I_{max}) - (1/T) \leq N(T)/T \leq (1/I_{min}); \qquad (2.2)$$

for large T, the term (1/T) on the left can be ignored, leading to

$$(1/I_{max}) \leq N(T)/T \leq (1/I_{min}).$$

Now the interservice interval I_n has been assumed to depend on what types of customers C_n and C_{n+1} are, and the sequence of successive types is to be regarded as randomized rather than specified. Thus each I_n is a random variable, and so N(T) is a random variable as well, the same holding true for the mean throughput rate N(T)/T. Furthermore, our informal description of the capacity concept identified it as the average value of this last-mentioned random variable, for prolonged intervals [0,T]. Thus, using the customary symbol "E" for the operation of taking the average (or "expected") value of a random variable, we have

$$\text{Capacity} = E\,[N(T)/T] \qquad \text{(large T)}. \qquad (2.3)$$

Since E [N(T)/T] is not perfectly constant for large T, we formulate (2.3) more precisely as

$$\text{Capacity} = \lim_{T \to \infty} E\,[N(T)/T]. \qquad (2.4)$$

Once the intuitive ideas behind the capacity concept have been made precise enough to admit explicit mathematical representation, the questions to be resolved involve problems in mathematical analysis rather than formulation. Specifically:

QUESTION 1. Is there a usefully simple formula for evaluating capacity, in terms of the limit in equation (2.4)?

QUESTION 2. Is the convergence to the limit, as T increases, sufficiently rapid that the limit is an adequate approximation to the result for large but finite T in the practical — interest range of durations? (Cf. footnote 5 of this chapter.)

QUESTION 3. For large T, is the variance of the random variable N(T)/T so small as to indicate that the random variable is adequately represented by its average value?

We proceed next to the analysis of these questions. The mix of customer types can be represented by a set of numbers

p_i = relative proportion of type i customers in the mix.

These p_i's are non-negative and sum to 1; we can interpret them as probabilities

$$p_i = \text{Prob } \{C_n \text{ is of type i}\}.$$

This interpretation requires that the type of each successive customer be regarded as arising "at random" independent of the types of all previous customers, e.g., no deliberate "sequencing by type" is involved.[8]

---

[8] Some topics related to sequencing are taken up in Appendices C and D.

Suppose first, for simplicity, that the probability distribution function of I_n depends only on the type of customer C_n. Let

T_i = average value of random variable I_n, if C_n is of type i;

these averages, "conditional" on the type of C_n, can be combined to yield

$\Sigma_i p_i T_i$ = (unconditional) average value of I_n. (2.5)

The random process is one which starts afresh ("renews itself") with the end of each interservice period; after the n-th period the type of the next customer C_{n+1} arises at random in accordance with the probabilities p_i, and then the value of I_{n+1} arises at random in accordance with the probability distribution function for I_{n+1} associated with whichever customer type turns up. Such a process is called a renewal process, and there is a fairly extensive technical literature on the subject. In particular, it is known ([6], p. 359) that for large T, the random variable N(T) is approximately normally distributed with mean T/μ where μ is here the quantity in (2.5), and with variance $T\sigma^2/\mu^3$ where σ^2 is the (unconditional) common variance of the random variables I_n.

From this result it follows, first, that[9]

$$\lim_{T \to \infty} E\,[N(T)/T] = 1/\Sigma_i p_i T_i. \tag{2.6}$$

This gives an affirmative answer to Question 1, assuming that the T_i's are readily calculable. Second, it follows that the variance of N(T)/T (T^{-2} times that of N(T)) is for large T approximately proportional to 1/T; since 1/T tends to zero rapidly for large T, an affirmative answer to

---

[9] In Appendix A we give a proof of the next formula which is relatively elementary, involving only Laplace Transform techniques.

Question 3 is also at hand. That the answer to the remaining Question 2 is also favorable, follows from a more general result to be cited later.

Thus, if the length I_n of the n-th interservice interval could be assumed to depend only on the type of the n-th customer C_n, then our analysis could be regarded as successfully concluded, and it would be time to move from the present level of generality to more specific and runway-oriented applications. But in fact, for the type of application motivating this study, the probability distribution of I_n really does depend on the type of the following customer C_{n+1} as well as that of C_n, and so we are not yet done.

When this point in the project was reached, some thought was given to the idea of trying an approximation in which I_n is split into two parts; a first one whose probability distribution depended only on C_n's type, and a second one depending only on C_{n+1}'s type. Then the first part of I_n and the second part of I_{n-1} together form an interval depending only on C_n's type, and so the previous results on renewal processes would be applicable. This idea might have worked out well; it is noted here for its intrinsic interest and because it may suggest analogies useful in other situations. However, it was laid aside when the more compelling approach described next was uncovered.

Let us say that the facility is in "state i" if the current interservice interval involves a customer (C_n) who is of type i. Then the time history of the facility is a random process involving a sequence of transitions from state to state (or sometimes from a state to the same state), these transitions occurring when a new interservice interval begins. The probability distribution of the time interval spent in a state, before the next transition, depends on the pair of states between which the transition occurs.

Such a random process is known as a semi-Markov or Markov renewal process. These processes, considerably more general than renewal processes, were introduced more recently and have been studied less extensively. By suitably assembling and specializing results in the literature on this subject[10], we arrive at the following conclusions.

First, with the notation

T_{ij} = average value of τ_n if C_n is of type i and C_{n+1} is of type j,

we have

$$\text{Capacity} = \lim_{T\to\infty} E\,[N(T)/T]$$

$$= 1/\Sigma_{ij}\ p_i\, p_j\, T_{ij}, \tag{2.7}$$

[10] Details are given in Appendix B.

where the summation is over all ordered pairs (i,j) of customer types. This formula, which yields an affirmative solution to Question 1 if the T_{ij}'s are readily calculable from known quantities, can be regarded as the climax of the more theoretical part of the study. Second, the error in approximating E [N(T)/T] for large T by the limit in (2.7), is essentially proportional to 1/T, so that an affirmative answer to Question 2 is indicated[11]. The coefficient of proportionality is also available, as an explicit combination of complicated quantities whose simplification we lacked time to attempt.

As for Question 3, in the time available we were unable[12] to extract from the literature (or to devise ourselves) an analytical proof that the variance of E [N(T)/T] tends rapidly to zero as T increases, although it seemed likely that this variance tends to zero roughly proportionally to 1/T. We were therefore obliged to check Question 3 by a Monte Carlo fast-time simulation for the type of application described in Sections 3 and 4. The results were positive; they are reported in subsection 4.3.

The considerations described at the end of subsection 2.2 will now be brought into play, in order to reach a more specific form of equation (2.7). Suppose customer C_n is of type i, and C_{n+1} is of type j. Let

$$d_{ij} = \text{desired interservice time for } C_n, \quad \text{if } (C_n, C_{n+1}) \text{ are of types } (i,j). \tag{2.8}$$

---

[11] This implies the same result for the renewal-process case discussed earlier.

[12] For a later bulletin, see the portion of Appendix B following equation (B.13).

Then the actual interservice time, the random variable I_n, is given by

$$I_n = d_{ij} + e_n \tag{2.9}$$

where

e_n = random error in achieving a desired interservice interval.

We now make the further assumption that

$$E(e_n) = 0, \tag{2.10}$$

i.e. that there is no systematic bias in the actual interservice intervals as compared with the desired ones. The justification is that such a systematic bias, if it existed, would be detected by the facility's management and cancelled by appropriately altering the desired (aimed-at) interval lengths. From eqs. (2.9) and (2.10) it follows that

$$E(I_n) = d_{ij}.$$

But $E(I_n)$, under the stated assumptions on the types of C_n and C_{n+1}, is what was defined earlier as T_{ij}. Thus eq. (2.7) becomes

$$\text{Capacity} = 1/\Sigma_{ij}\, p_i\, p_j\, d_{ij}. \tag{2.11}$$

Recall that the tieup time for C_n was defined as the smallest size that I_n should have, according to regulations. With the notation

r_{ij} = tieup time for C_n if (C_n, C_{n+1}) are of respective types (i,j),

the probability of violation of the tieup time constraint is given by

p_V = Prob $\{I_n < r_{ij}$, given that (C_n, C_{n+1}) are of types $(i,j)\}$.

Using equation (2.9), and omitting the "given that" clause for brevity, we have

$$p_V = \text{Prob } \{d_{ij} + e_n < r_{ij}\}$$

or equivalently

$$p_V = \text{Prob } \{e_n < r_{ij} - d_{ij}\}. \tag{2.12}$$

The random error e_n has some probability distribution, which might conceivably depend on the types of C_n and C_{n+1}. We therefore introduce the cumulative distribution functions

$$F_{ij}(x) = \text{Prob } \{e_n < x, \text{ given that } (C_n, C_{n+1}) \text{ are of types } (i,j)\}. \tag{2.13}$$

Thus (2.12) takes the form

$$p_V = F_{ij}(r_{ij} - d_{ij}). \tag{2.14}$$

We now adopt the viewpoint that p_V is not a derived quantity, but rather a (small) realistically endurable probability of violation. The equation (2.14) can be solved[13] for d_{ij}; if $F_{ij}^{-1}(p_V)$ denotes the smallest value of x for which $F_{ij}(x) = p_V$, then

---

[13] Assuming F_{ij} is continuous.

$$d_{ij} = r_{ij} - F_{ij}^{-1}(p_v). \tag{2.15}$$

Since p_v is small, $F_{ij}^{-1}(p_v)$ will be negative, and the quantity

$$b_{ij} = -F_{ij}^{-1}(p_v) \tag{2.16}$$

represents the "buffer" or "safety margin" mentioned earlier (see pp.17,19). The consequence $d_{ij} = r_{ij} + b_{ij}$ of (2.15-16), on substitution in (2.11), yields

$$\boxed{\text{Capacity} = 1/\Sigma_{ij}\, p_i\, p_j\, (r_{ij} + b_{ij}),} \tag{2.17}$$

where r_{ij} and b_{ij} are defined below (2.11) and in (2.16) respectively.

[Of course, more sophisticated models for treating the random error can be introduced if they seem appropriate. For example, suppose[14] that e_n might be better represented as a random multiple (sometimes negative) of the desired interservice time to which it is the random error. Denoting the multiplier by y, and its cumulative distribution function by G, we have

$$\begin{aligned} p_v &= \text{Prob}\{ d_{ij} + e_n < r_{ij} \} \\ &= \text{Prob}\{ d_{ij}(1 + y) < r_{ij} \} \\ &= \text{Prob}\{ y < (r_{ij}/d_{ij}) - 1 \} \\ &= G(r_{ij}/d_{ij} - 1). \end{aligned}$$

Solving for d_{ij} yields

$$d_{ij} = r_{ij}/[1 + G^{-1}(p_v)]$$

rather than (2.15), and thus

$$b_{ij} = d_{ij} - r_{ij} = [-G^{-1}(p_v)]\, r_{ij}/[1 + G^{-1}(p_v)] \tag{2.18}$$

rather than (2.16).]

[14] This is intended as a hypothetical illustration.

This completes the theoretical analysis. The results appear clearly applicable to a runway serving either a stream of landing aircraft, or a stream of departing aircraft. To bear out this contention, the preceding material is applied in some detail to the "IFR landings" case in the following sections of this report.

Further investigation would be required to extend this material to interacting runways, or to a runway serving both arrivals and departures. The latter topic is taken up in a preliminary way in Appendix G.

3. APPLICATION TO IFR LANDINGS

3.1 Preliminaries

In the previous section, a capacity concept for a class of facilities serving a multi-type stream of customers was introduced and analyzed. A relatively simple formula for it was derived, namely

$$\text{Capacity} = 1/\Sigma_{ij}\, p_i p_j\, (r_{ij} + b_{ij}) \qquad (3.1)$$

where the summation is over all ordered pairs (i,j) of customer types, and

p_i = relative proportion of customer "mix" which is of type i,

r_{ij} = minimum time which <u>should</u> elapse between providing service to a customer of type i, and providing service to the next customer if of type j,

b_{ij} = "buffer" time added to r_{ij} as a safety margin, because an intended interservice spacing cannot be achieved with perfect precision; it reduces the probability of violating the constraint expressed by r_{ij} to a prescribed level p_v.

In the present section, our aim is to specialize equation (3.1), and hence the capacity concept, to the case of a runway handling a stream of IFR landings. This specialized formula provides the basis for the specific numerical illustrations presented in Section 4.

The "facility" in question is clearly the runway together with its final-approach path airspace. The "customers" are evidently the aircraft in the landing stream. Several definitions of "service" might be possible;

the choice among them is immaterial so long as the chosen one is used in a careful and consistent way. Here we shall regard the interservice period of an aircraft as beginning witb its touchdown, and ending with the touchdown of the next aircraft.

The "input data" p_i in the capacity formula clearly represent the mix of different aircraft types in the arriving scheme. It remains to derive appropriate formulas for the other quantities appearing in equation (3.1), namely the "tieup times" r_{ij} and the "buffer times" b_{ij}. These two tasks are carried out in subsections 3.2 and 3.3 respectively.[1] The results are assembled, together with some comments, in the concluding subsection 3.4.

---

[1] Our treatment of these topics is generally similar to that in [5], though differing in several specifics.

3.2 Treatment of Tieup Times

For IFR landings on a runway, the minimum period which should elapse between the touchdown of one aircraft and that of the next, is governed by two considerations. First, the second aircraft of the pair should not touch down until the service of the first is complete. With the notations

T' = time of touchdown for first A/C of pair,

T'' = time of touchdown for second A/C of pair,

OT' = runway occupancy time for first A/C of pair,

this condition reads

$$T'' - T' \geq OT'. \tag{3.2}$$

Second, for IFR landings there is a minimum distance separation constraint on the two aircraft as they share the final approach path (e.g., the "3-mile rule"). This requirement is expressed by a datum

S = prescribed minimum distance separation between two A/C on final approach path, (3.3)

and translates (as will be seen below) into a time separation

SEP = minimum time separation, corresponding to S, between touchdowns for the two A/C. (3.4)

Thus, in addition to (3.3), we have

$$T'' - T' \geq SEP. \tag{3.5}$$

Combining (3.2) and (3.5) yields the result:

$$\text{tieup time for first A/C in pair} = \max (OT', SEP). \tag{3.6}$$

To make equation (3.6) useful, the translation from the distance separation S to the time separation SEP must be carried out explicitly. Transition from a distance to a time duration of course requires some sort of velocity data; in the present instance these are

$$v' = \text{final approach speed of first A/C,}$$
$$v'' = \text{final approach speed of second A/C.} \qquad (3.7)$$

The same idealization as in [5] will be adopted, namely that these speeds can be treated as constant over the final approach path; a more delicate analysis could of course employ variable velocity-time profiles, if they could be satisfactorily specified. In addition to v' and v'', there is an additional element of the situation's geometry which must be known; this is

$$L = \text{length of final approach path.} \qquad (3.8)$$

There are two cases to be considered:

CASE 1. Suppose $v' \leq v''$. Then the second aircraft is overtaking the first, so that the gap between them is closing as both proceed down the final approach path. The aircraft are therefore closest together at the last moment at which they share the approach path, namely at the time (T') when the first member of the pair touches down. If the distance separation criterion (S) is satisfied at this moment, it will also be satisfied at all prior times when both aircraft are on the final approach path.

With the temporary notation

χ = distance between the A/C when the first one touches down.

we see that the second aircraft will touch down χ/v'' later than the first, i.e.

$$T'' - T' = \chi/v''.$$

The distance separation condition, which reads $\chi \geq S$, is therefore equivalent to

$$T'' - T' \geq S/v''.$$

In other words

$$\boxed{\text{SEP} = S/v'' \qquad (\text{if } v' \leq v'').} \tag{3.9}$$

CASE 2. Now suppose $v' > v''$. Then the gap between the two aircraft is <u>opening</u> as they share the final approach path. The aircraft are therefore closest together at the moment when they <u>first</u> share the path, namely at the moment when the second aircraft begins its final approach (passes the outer marker).[2] Suppose this occurs at time t, and now adopt the temporary notation

χ = distance between the two A/C at time t.

Then the second aircraft, before touching down, must cover the full length (L) of the final approach path; it will therefore touch down at time

$$T'' = t + (L/v''). \tag{3.10}$$

---

[2] An alternative is noted below.

The first aircraft has only the distance $L - \chi$ along the approach path to cover; it will therefore touch down at time

$$T' = t + (L - \chi)/v'. \tag{3.11}$$

From (3.10) and (3.11), we have

$$T'' - T' = (\chi/v') + L\,(1/v'' - 1/v').$$

The distance separation criterion, which reads $\chi \geq S$, is therefore equivalent to

$$T'' - T' \geq (S/v') + L\,(1/v'' - 1/v').$$

In other words,

$$\boxed{\text{SEP} = (S/v') + L\,(1/v'' - 1/v') \qquad (\text{if } v' > v'').} \tag{3.12}$$

Finally, combining (3.6), (3.9) and (3.11) yields, for the tieup time of the first aircraft of the pair,

$$\text{tieup time} = \max\,(OT', S/v'') \qquad (v' \leq v'') \tag{3.13}$$

in CASE 1, and

$$\text{tieup time} = \max\,(OT', S/v' + L\,(1/v'' - 1/v')) \qquad (v' > v'') \tag{3.14}$$

in CASE 2.

Two technical points about the preceding analysis should be noted before going on. First, it has of course been assumed that $S < L$, i.e. that it is allowable for two aircraft simultaneously to be in the final

approach phase of their landings. Second, a different approach to CASE 2 (opening gap) is adopted in [5]. The viewpoint there is that as the two aircraft approach the outer marker, they would be altitude separated, with the second (slower) one higher. The distance separation criterion is then imposed when the first aircraft begins its final approach, on the assumption that by then the second aircraft has already begun its approach from the higher altitude. This situation of course admits an analysis similar to the one given above, though resulting in a slightly different formula for the tieup time. Our information on this point is that the postulated altitude separation is not so universal a practice as to command inclusion in the present illustrative application of the capacity concept.

3.3 Treatment of Buffer Times [3]

Recall from equations (2.9), (2.15) and (2.16) that the actual interval between the touchdowns of the two successive aircraft discussed above is given by

(tieup time) + b + e

where b is the buffer time we wish to represent, and e is a random error in achieving a desired time separation of the two touchdowns. The error e was defined so as to have average value 0, and the desired buffer time was calculable from (2.16) as

$$b = - F^{-1} (p_v) \tag{3.15}$$

where F is the cumulative distribution function of e, and p_v is the probability of violation.

Without an explicit investigation of error sources and their statistical variation, nothing definitive can be said about the proper choice of the distribution function F. Some considerations on this topic appear in [5], [7] and [8]. In particular ([8], p.30), error patterns that appeared to be normally distributed have been observed. For the present illustrative purposes, we shall work with two mathematically simple distributions which represent sharply different assumptions about the tendency of errors to cluster about the zero (no-error) point. Each of these distributions depends on a single parameter, which also serves as a measure of precision for the landing and landing-control process.

[3] Based on (2.16) rather than the alternative (2.18).

The probability density (or frequency) functions for these two distributions are shown in Figures 3.3.1 and 3.3.2 (whose vertical scales differ). The first is the triangular distribution over the finite interval [-R, R], where R measures the spread of errors. This is an analytically more tractable substitute for the normal distribution, and also lacks the sometimes embarrassing infinite "tails" of the latter. The second is the uniform distribution over the same interval [-R, R]; it is much more pessimistic, allowing for no peaking at 0 whatever.

For these two distributions, under the natural assumption $p_v < 1/2$, equation (3.15) will be shown to yield

$$b = R\,[1 - (2p_v)^{1/2}] \qquad \text{(triangular)}, \qquad (3.16)$$

$$b = R\,[1 - 2p_v] \qquad \text{(uniform)}. \qquad (3.17)$$

Figure 3.3.1: Triangular Distribution on [-R, R].

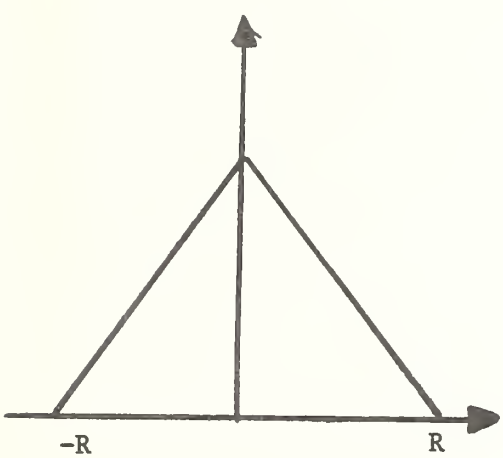
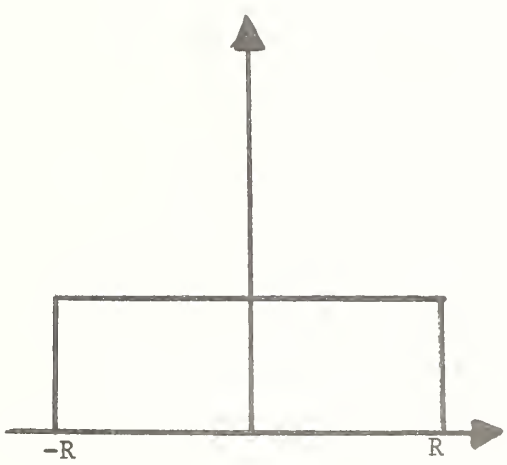


Figure 3.3.2: Uniform Distribution on [-R, R].



The details for Figure 3.3.2 will be given first. The area of the rectangle, like the area under any probability density curve, must be unity. Since the rectangle's base has length 2R, its height must be 1/2R. Thus the equation of the probability density function is

$$f(x) = 1/2R \qquad (-R \leq x \leq R),$$

and so that of the cumulative density function is

$$F(x) = \int_{-R}^{x} f(e)\, de = (x + R)/2R \qquad (-R \leq x \leq R).$$

The equation $F(x) = p_v$ therefore leads to

$$x = F^{-1}(p_v) = R\,[2p_v - 1].$$

so that

$$b = -F^{-1}(p_v) = R\,[1 - 2p_v],$$

corroborating (3.16).

For Figure 3.3.1, the area must again be unity; since the base of the triangle is 2R, its altitude is 1/R. Thus the left-hand leg of the triangle joins the points (-R,0) and (0,1/R); its equation (that of the probability density function) is therefore

$$f(x) = \frac{(1/R)}{R}\,(x + R) \qquad (-R \leq x \leq R),$$

so that the cumulative density function is

$$F(x) = \int_{-R}^{x} f(e)\, de = (x + R)^2/2R^2 \qquad (-R \leq x \leq R).$$

The equation $F(x) = p_v$ leads to

$$x = F^{-1}(p_v) = R\,[(2p_v)^{1/2} - 1],$$

so that

$$b = -F^{-1}(p_v) = R\,[1 - (2p_v)^{1/2}],$$

corroborating (3.17).

3.4 Synthesis and Comments

It is relatively routine to assemble the results of the last two subsections. For this purpose, let

$$OT_i = \text{runway occupancy time for type i A/C,} \tag{3.18}$$

$$v_i = \text{final approach speed for type i A/C.} \tag{3.19}$$

Then (3.13) and (3.14) yield

$$r_{ij} = \max \{OT_i, S/v_j\} \qquad (v_i \leq v_j), \tag{3.20}$$

$$r_{ij} = \max \{OT_i, S/v_i + L\,(1/v_j - 1/v_i)\} \qquad (v_i > v_j). \tag{3.21}$$

Next, by (3.16) and (3.17),

$$b_{ij} = R\,[1 - (2p_v)^{1/2}] \qquad \text{(triangular)}, \tag{3.22}$$

$$b_{ij} = R\,[1 - 2p_v] \qquad \text{(uniform)}. \tag{3.23}$$

When the last four formulas are substituted into equation (3.1), the result is somewhat simplified because b_{ij} as given above is not dependent on the pair (i,j). This result, with notation

$$C = \text{capacity,} \tag{3.24}$$

takes the form

$$\boxed{C = 1/\{\Sigma_{ij} p_i p_j r_{ij} + R\,[1 - (2p_v)^{1/2}]\}} \tag{3.25}$$

for the triangular error distribution, and

$$\boxed{C = 1/\{\Sigma_{ij} p_i p_j r_{ij} + R\,[1 - 2p_v]\}} \tag{3.26}$$

for the uniform distribution, with r_{ij} given by (3.20) and (3.21).

Numerical calculations based on these formulas will be presented in the next section. However, a few comments about the formulas seem in order at this point.

First, the formulas are sufficiently simple to permit easy derivation of equations for sensitivity coefficients, i.e., partial derivatives with respect to each of the quantities entering the capacity measure. (These derivatives exist except in the unlikely coincidence of a tie between the quantities competing for the maximum in (3.20) and (3.21).) Thus much of the appropriate sensitivity analysis, with respect to changes in (for example)

the separation distance (S)

the violation probability (p_v)

the control-precision measure (R)

the runway occupancy times (OT_i)

the mix of aircraft types (p_i)

can be carried out analytically, rather than purely numerically as in Section 4.

Second, these formulas (3.25) and (3.26) are based on the mathematical model of random errors leading to eq. (2.16). If for example the approach leading to (2.18) were used, the result would be

$$C = [1 + G^{-1}(p_v)]/\Sigma_{ij} p_i p_j r_{ij}. \tag{3.27}$$

In this case, the efficiency of the landing process, as measured by the ratio of its capacities with and without the need for buffer times,

turns out to depend on p_v alone.

$$C/C_{ideal} = 1 + G^{-1}(p_v) < 1. \tag{3.28}$$

Third, three generalizations can be handled with essentially no extra effort, except that for the first two b_{ij} will really depend on the pair (i,j) so that (3.1) rather than (3.25 - 26) must be used: We can replace the error range R by a set of R_{ij}'s; this may be advantageous in working toward a more sophisticated error model. We can generalize the threshold violation probability p_v to $(p_v)_{ij}$'s; this is desirable since in principle one might prefer a "safety policy" sensitive to considerations (such as number of passengers) which may typically differ among aircraft types. Similarly, the minimum desired separation distance S can be generalized to a set of S_{ij}'s.[4]

This leads naturally to the fourth comment. In the previous development, nothing specific had to be said about what classification of the arriving aircraft into "types" was to be employed. Implicit in equations (3.19) and (3.20), however, is a first restriction on which classification schemes are admissible.[5] The requirement for OT_i and v_i

---

[4] Incidentally, the treatment of S and p_v as independent parameters seems somewhat odd; one would expect both to be related to some "safety criterion", perhaps with a possibility of tradeoffs.

[5] It was for this reason that the notations OT_i, v_i, t_{ij}, b_{ij} were avoided in subsections 3.2 and 3.3.

to be meaningful, clearly, is that runway occupancy times (OT) and final approach speeds (v) for the aircraft of any one type be adequately representable by single nominal numerical values. This phrase "adequately representable" is not as burdensome as it might appear; for example it does not require that OT and v be essentially constant within each type, a condition that would lead to an undesirable proliferation of types. Rather, it requires that a reasonably representative nominal value for each $r_{ij} + b_{ij}$ be assignable.

In the present version, given by eqs. (3.25) and (3.26), this stipulation restricts only the r_{ij}'s. A glance at the formulas (3.20-21) for these tieup times reveals that some degree of sophistication, in finding nominal values for each class, may be called for. For example the v's appear only in the form 1/v, so that averaging of 1/v's rather than of v's is indicated.[6] Moreover, even the use of average OT's and 1/v's is not rigorous, since these quantities enter the tieup time via a nonlinear mathematical operator ("max"). This point cannot be pursued further here, but seems to merit further analysis; we note only (a) that the general difficulty in question seems to be equally pertinent to the capacity concept embodied in the Handbook, and (b) that an aircraft classification scheme developed for other purposes may not be especially suitable for "capacity" analysis.

---

[6] The average of 1/v is not in general the reciprocal of the average of v; the bias is systematic, in the direction $[1/v]_{avg} \geq 1/v_{avg}$.

4. NUMERICAL ILLUSTRATIONS

4.1 Data Sets Employed

In Section 3, the general capacity concept formulated in Section 2 was applied to the case of a single runway serving a stream of IFR landings. In particular, the general formula (2.17) derived for this concept was specialized (in a way based on further "modeling" assumptions) to a specific version appropriate to this application, and given in equations (3.25) and (3.26).

To obtain concrete illustrative numerical results[1] for examination, this formula was exercised to calculate "capacity" values for a number of sets of input data provided mainly by the FAA. These results are discussed in subsection 4.2, where they are displayed in several graphical formats to help communicate the capacity measure's sensitivity to the various parameters involved. In subsection 4.3 some of the results are compared with the outputs of a fast-time Monte Carlo simulation, and excellent agreement is found. A further comparison, this one with capacities given in or derived from the Airport Capacity Handbook [1], is carried out in Section 5.

The remainder of this subsection is devoted to specifying the data sets employed in calculating the capacity values. We will list the quantities which enter the calculations, and present the numerical values assigned to each.

First, there are two geometrical quantities

[1] We emphasize "illustrative," because an emphasis on data considerations quite beyond this study's scope would be needed to produce numbers sufficiently reliable for operational use.

Table 4.1.1: Final-Approach Path Lengths (L) and Minimum Separation Distances (S), in N. Mi.

L :	6	8	10
S :	2	3	4

Table 4.1.2: Distributions over [-R, R] of Error in Intertouchdown Time.

Form :	triangular (symm.), uniform
R(sec.) :	15, 20, 25, 30, 35, 40, 45

Table 4.1.3: Threshold Probabilities of Violation

p_V :	0.01 , 0.05

Table 4.1.4: Mixes of Aircraft Types

Mix	p_A	p_B	p_C	p_D	p_E
I(Fast)	0.6	0.2	0.2	0.0	0.0
II(Med.)	0.2	0.4	0.2	0.2	0.0
III(Slow)	0.0	0.3	0.3	0.2	0.2

Table 4.1.5: Final-Approach Speed (v, in knots) and Runway Occupancy Time (OT, in sec.) by Aircraft Type.

Type :	A	B	C	D	E
v :	165	150	135	120	105
OT :	59	52	45	38	31

L = length of final approach path (n. mi.),

S = required minimum distance separation along final approach patn (n. mi.).

For these, the value-sets given in Table 4.1.1 were presented by the FAA as covering the range of interest. Since L and S are to be varied independently, the Table represents 9 combinations in all; the "nominal" combination among these is taken to be

L = 8 n. mi. ; S = 3 n. mi.

The next two factors to be considered describe the probability distribution of the random error in achieving a desired intertouchdown interval. This distribution is assumed to be symmetric and centered at 0, and to depend on a single parameter determining its "spread." The two factors are

R = half-range of distribution (i.e., the errors range over the interval [-R, R]),

F = (functional) form of distribution.

A traditional mathematical representation of random errors is that of the normal (Gaussian) distribution; it is used for example in [5], and some empirical backup for this choice exists ([8], p. 30). On the other hand, that distribution is a little awkward to work with analytically, and its infinite "tails" are not really appropriate. We therefore employed, as one functional form, the more tractable triangular distribution, which has a satisfactory resemblance to the normal for our present illustrative purposes. As a deliberately pessimistic alternative, the uniform distribution was used. (These distributions are sketched in Figures 3.3.1 and 3.3.2.)

Data to guide the choice of R-values were not initially at hand. However our project monitor (Mr. S. P. E. Price of the FAA) obtained, from one of the authors of [8], information showing deviations of between (-8) sec. and roughly 52 sec. between scheduled and actual inter-aircraft arrival intervals. Writing this as 22 ± 30 sec., we regard the 22 seconds as a "buffer time" rather than a systematic unintended bias; this interpretation is supported by [8], p. 29, para. 1. Thus we are led to take

$$R = 30 \text{ sec.}$$

as nominal value.

This however involves considerable uncertainty. Some intervals in [8] are described (p. 28) as recorded "at the runway boundary"; the seriousness of the difference between errors in such intervals and those in intertouchdown times is not clear to us. The "mean errors" reported ([8], p. 29) are much smaller than 22 sec., apparently inconsistent with our use of a symmetric error distribution. Using the observed range of a random variable as an estimate of its theoretical range is of course not sound practice, but time did not permit setting up a proper estimation based on the distribution of observed errors. We tried to allow in our calculations for some of these uncertainties, by permitting R to vary quite widely as shown in Table 4.1.2. This is not really adequate, however; quite clearly a more intensive effort to determine suitable error distributions would be an important part of further work along the lines of this report.

The next parameter to be considered is the threshold probability (p_v) of a violation. Values of 1% and 5%, which appeared in [5], were approved for use here. The value 0% of course suggests itself as of special interest; however, because the error distributions we employ are relatively "short-tailed," the results for 0% would not differ greatly from those for 1%.

The three mixes of aircraft types shown in Table 4.1.4 were provided by the FAA. Ranging from "fast" to "slow" in the average final-approach speed involved, they represent a progression from a mix representative of a large airport to one more typical of a smaller facility. The classification into "types" is that employed in the Airport Capacity Handbook. In the Table, p_i denotes the proportion of type i aircraft in the mix.

Counting up possible combinations of data at this point, we find 108 for each of the 7 R-values. This had two implications for our work. First, although the formulas are simple enough that hand computation for a small number of cases is perfectly feasible, the number of cases involved here is large enough to call for a computer program. Accordingly, a straightforward FORTRAN code was written and employed for the calculations. Second, further proliferation of cases would be discouraged; the results could be computed quickly enough, but time for their thoughtful interpretation and meaningful summarization (in this report) was lacking.

This decision affected our treatment of the remaining two sets of input data, namely the typical final approach speeds (v_i) and runway occupancy times (OT_i) for the various aircraft types. The particular

numerical values given in Table 4.1.5 were provided by the FAA. These are only "representative" sets of values, however, and so sensitivity analyses with respect to changes in them should be made. Such analyses were not performed, for the reason just given, and so remain as unfinished business for any possible continuation of this work.

4.2 Discussion of Results

In this subsection we present and discuss three groupings of calculated capacities, and one grouping of values of the "efficiency" measure defined at the end of subsection 2.2. There are three types of comments to be made concerning these results. One type is simply confirmatory and qualitative; it involves checking that as some parameter varies, the calculated values of capacity or efficiency do indeed change in the direction indicated by common-sense. The second type examines these changes quantitatively, pointing out the sensitivity of the calculated quantity to changes in the various input data. The third type notes, in those cases where common-sense fails to indicate a "proper" direction of change, which direction actually occurred. All three types arise below.

We begin with Figure 4.2.1[(2)], in which R=30 sec., and the population mix (II) is held fixed. There are a number of observations to be made:

(a) For the triangular distribution (believed more realistic than the uniform), and the nominal value S=3 n. mi., the capacity (C) varies from 32.5 to 36.5 aircraft/hr. For population mixes I and III, the corresponding ranges are 35.3-38.7 and 29.0-32.5 aircraft/hr. These results appear quite compatible with the range of values (31.0-37.0) reported in [8], p. 22. Thus our procedure appears to be producing numbers "in the right ball-park."

---

(2) To avoid misleading first impressions, note that the ordinate scale of this and other figures does not start at 0.

(b) C is systematically higher for the triangular error distribution than for the uniform distribution. This is as it should be, the latter assumption not crediting the landing process with any ability to make small errors more frequent than larger ones. The variation amounts to 5% or more in many cases.

(c) Since capacity can presumably be increased at the cost of greater indifference to the regulations[3], it is not surprising that C increases with p_v. This is more pronounced for the triangular distribution than for the uniform. The difference can be traced back to equations (3.16-17): the formula for C contains p_v in the form $(2p_v)^{1/2}$ in the triangular case and $(2p_v)$ in the uniform case, and for small p_v the first form varies much more rapidly than does the second.

(d) As would be expected, capacity increases when the minimum distance separation (S) on the final-approach path can be reduced. The sensitivity of C to S is quite marked; e.g. for L=8 n. mi., p_v=0.01 and the triangular error distribution we have

$$C = 27.1, 33.0, 41.2 \text{ aircraft/hr.}$$

corresponding respectively to

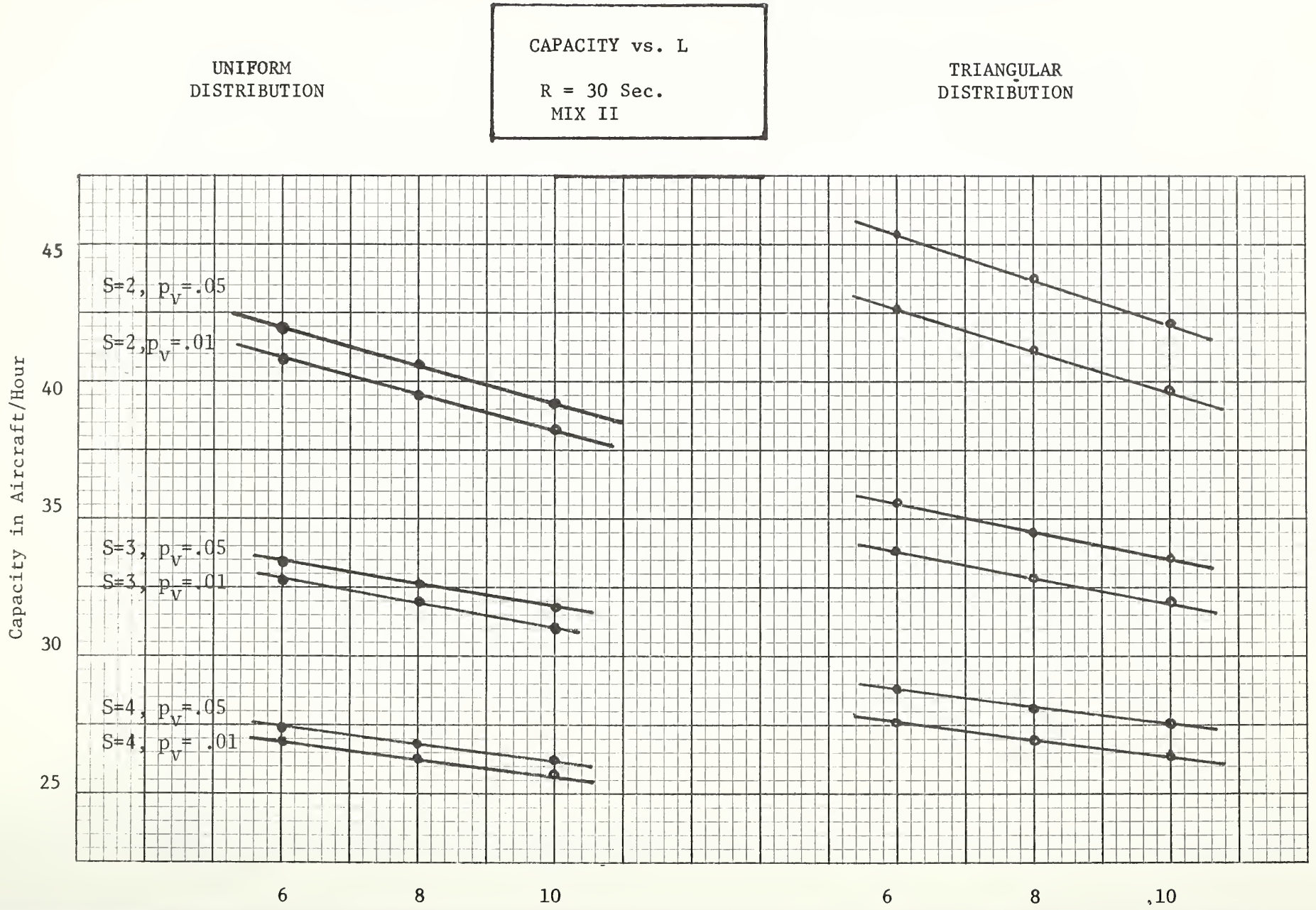
$$S = 4, 3, 2 \text{ n. mi.}$$

(e) C is consistently a decreasing function of the length L of the final-approach path. This too is as it should be: a

---

(3) Our model imposes no penalty to throughput rate for a violation; incorporating such a feature might be desirable.

Figure 4.2.1 - Sensitivity to Changes in Final Approach Distance



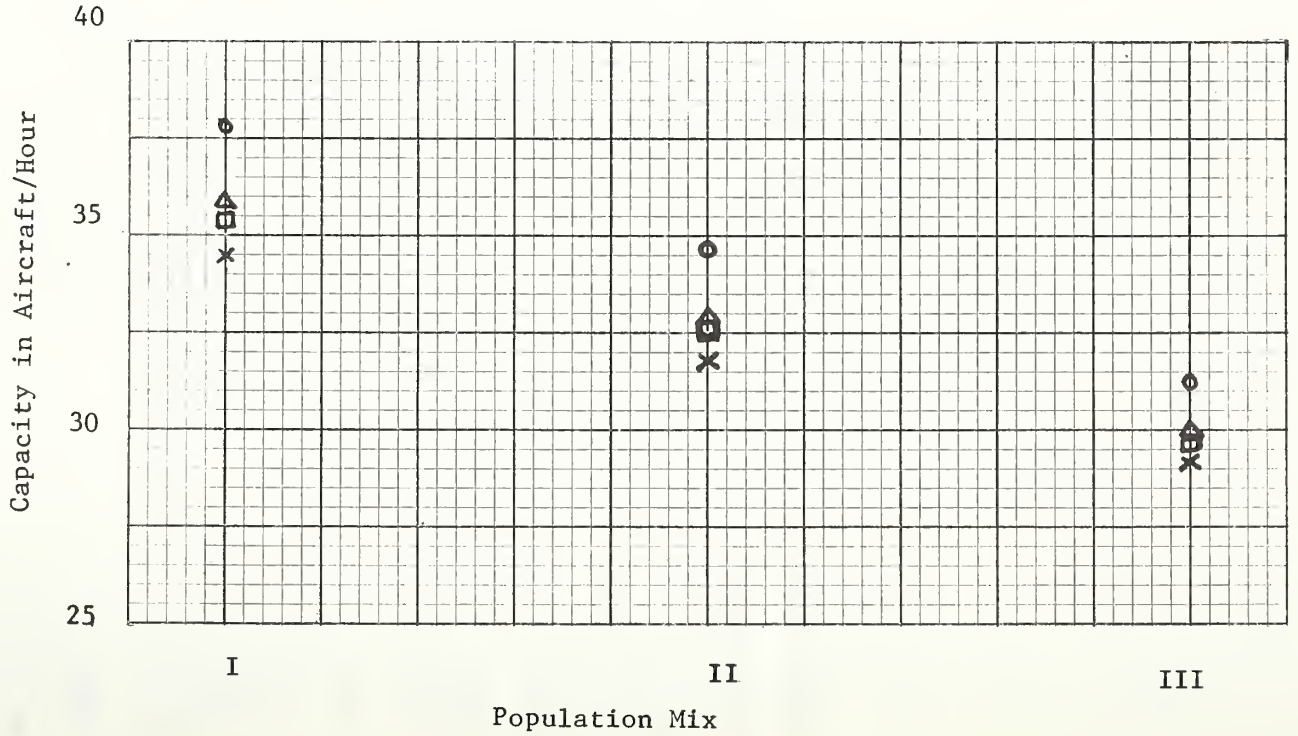
longer path permits an "opening gap" between two successive aircraft to open even wider, with consequent loss in throughput rate, whereas the "continuous demand" assumption assures that "closing gaps" will finally close to a separation S regardless of L, so that they are not shortened further by an increase in L. What is somewhat surprising is the nearly linear variation of C with L; it was not evident that our data combinations would be on a linear portion of the generally non-linear C vs. L curves.

We turn now to Figure 4.2.2, designed to display the influence of population mix on capacity. Clearly C decreases, as one moves from the "fast" mix through the"medium" one to the "slow." Now aircraft making faster approaches should indeed be able to achieve more landings per unit time --- the minimum distance separation translates into a smaller required time separation --- unless the constraint associated with runway occupancy prevents realizing the potentially greater throughput from the faster mix. The seriousness of this last possibility is accentuated by the fact that the fast-approach types have longer runway occupancy times; see Table 4.1.5. Thus there are two opposing tendencies involved in how C will vary with population mix, and without calculation it was not clear which would win out. From the direction of variation shown in the Figure, it follows that tieup times are in fact (for our data) generally determined by the final-approach separation criterion rather than by runway occupancy considerations. This observation will gain in significance when a comparison with Handbook values is made in Section 5; see there the discussion of Table 5.3.1.

Figure 4.2.2 - Capacity as a Function of Population Mix

CONSTANTS
R = 30 Sec.
L = 8 N.M.
S = 3 N.M.

DISTRIBUTION	p_v	SYMBOL
TRIANGULAR	.05	○
TRIANGULAR	.01	△
UNIFORM	.05	□
UNIFORM	.01	×



The observation also provides an exceptionally sharp example of how such a model can prove useful in cost-effectiveness analyses: under the stated circumstances, investments (in equipment, R. and D. or whatever) whose principal aim is to reduce runway occupancy times are clearly ill-advised if increasing capacity is the major goal!

Figure 4.2.3 was prepared to show how capacity (C) depends on the measure (R) of error spread. The variation of C-values is considerable, roughly 20% for Mix II for example, so that pinning down realistic values of R more tightly is indeed important. The functional dependence of C upon R can be seen from equations (3.25-26) on p.42 to have the form

$$C = 1/(A+BR)$$

where A and B are positive constants (relative to R) with $B < 1$, but near 1 when p_v is small. It follows[3a] that as R is reduced, successive reductions have <u>increasing</u> marginal benefit (in upping the value of C). This contrasts with the more common and less promising "<u>decreasing</u> marginal benefits" situation. Thus, purely from the "effectiveness" side of the cost-effectiveness ledger, attempts aimed at reducing R in order to increase C seem to merit a tentative label of "high-payoff." (This of course says nothing of technical feasibility and cost considerations.)

For R=0, C assumes the "ideal capacity" value C_{ideal} corresponding to perfectly precise control of intertouchdown times. Figure 4.2.4 repeats the information of the previous figure in a "normalized" form, using as ordinate not C but rather the efficiency measure

$$E = C/C_{ideal}$$

[3a] Since $\partial C/\partial R = -B/(A+BR)^2$, which increases in absolute magnitude as R decreases.

Figure 4.2.3 - Sensitivity to Changes in R

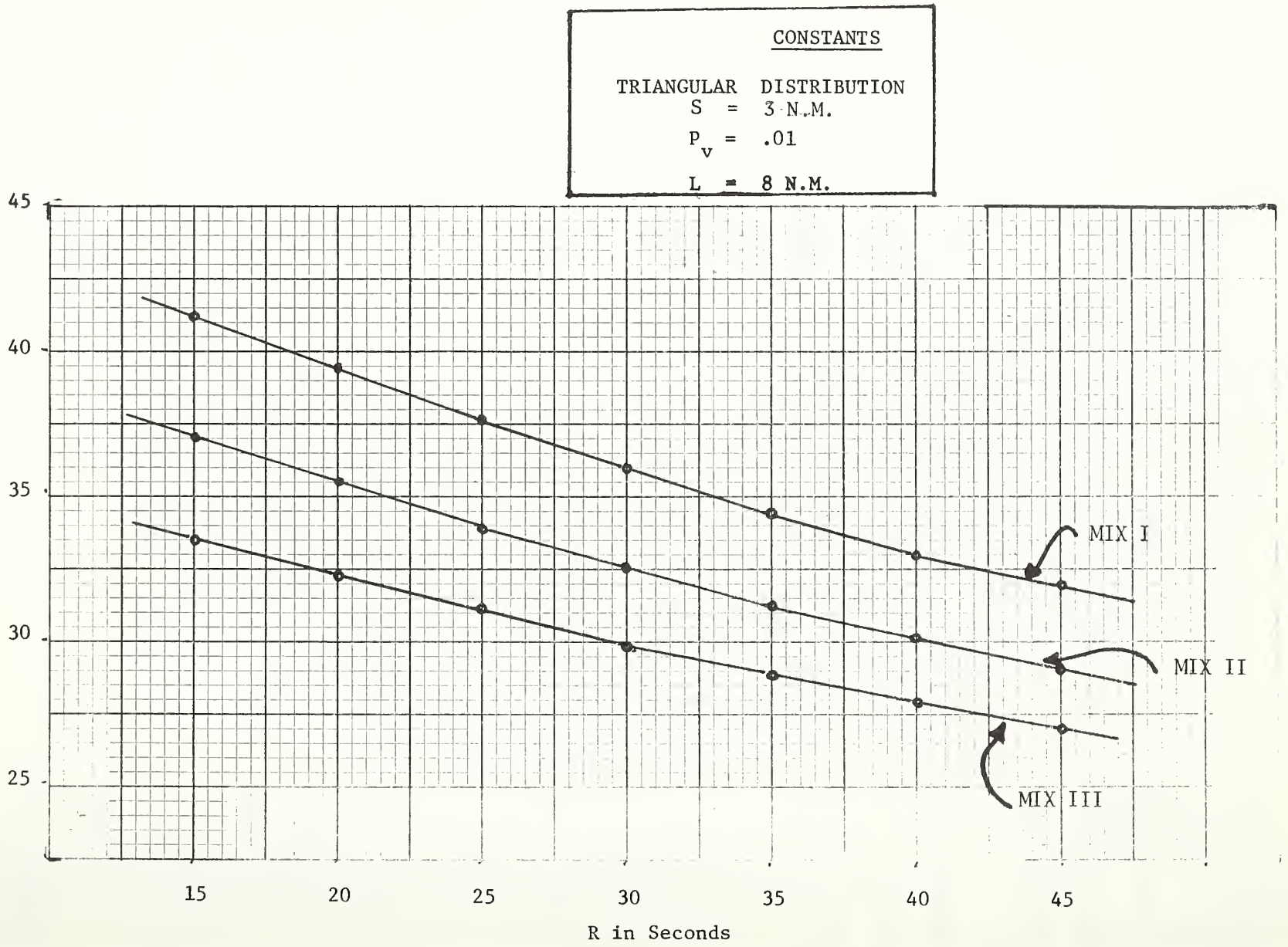
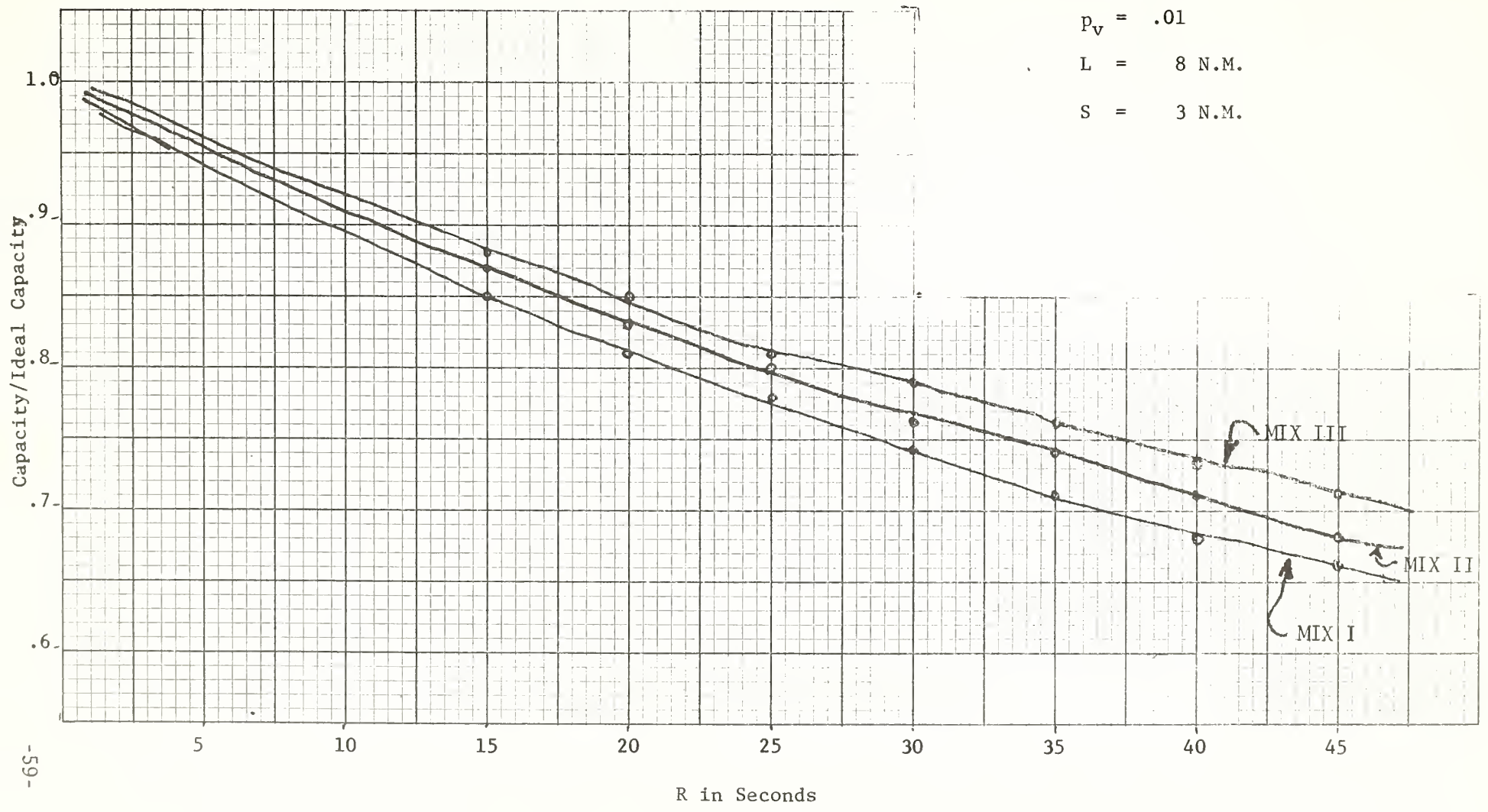


Figure 4.2.4 - Effects of R on Efficiency Measure



suggested in subsection 2.2. This measure increases to the value 1 as R is reduced to 0. As the Figure shows, the rate of increase in E per unit decrease in R is considerable throughout at least the lower half of the R-range considered, again illustrating the potential from such reductions.

On comparing Figures 4.2.3 and 4.2.4, it is interesting to note that the rank order of population mixes reverses when E rather than C is the criterion. This must mean that for slower mixes, buffer times are less significant relative to tieup times than for faster mixes. Thus average tieup times must be less for slower mixes; since runway occupancy times are higher, we conclude as before that "final-approach separation" rather than "runway occupancy" must be dominant in determining tieup times.

Various other arrangements of our results could of course also have been plotted and discussed, but it is hoped that those given above suffice to illustrate the kinds of analysis that are possible.

4.3 Comparison with Simulation Outputs

Our capacity concept was defined, informally in subsection 2.2 and more precisely in subsection 2.4, as the average mean throughput rate of the facility for a prolonged period of continuous demand. If the concept is to be useful relative to a peak load interval at a runway during which one can meaningfully speak of a single population mix, then a period of at most one or two hours had better be sufficiently "prolonged" for the concept to apply. Moreover, the concept is defined as an "average"; its potential usefulness is therefore much enhanced if one can show that the associated random variable has a small variance, thus guaranteeing (e.g., by Tchebychev's Inequality; see [6], Vol. I, p. 183) that large fluctuations from the average are very rare.

These points were explicitly addressed in subsection 2.4. Let T be the duration of the "prolonged period," and N(T) --- a random variable --- the number of customers served (aircraft landed) during that period. The "average" in question is that of the random variable N(T)/T. Now it is shown in the second part of Appendix B that for large T, the variance of N(T)/T is approximately proportional to 1/T, and so declines rapidly toward 0. It is also shown that the average, which depends on T, differs from its limit as $T \to \infty$ by a quantity which is approximately proportional to 1/T for large T, and so also declines rapidly to 0, implying that the average will soon (as T increases) "settle down" to be virtually independent of T.

These arguments appear to provide assurance on the two points raised above, and such arguments have generally proved to be reliable guides in applying similar mathematical concepts to normal real-world situations. However,

two slim possibilities for trouble still exist. One of them can be identified by the double appearance of the phrase "for large T" in the last paragraph; it could conceivably be the case that 1-2 hours is not "large" enough for the clause following the phrase to apply. The second danger is associated with the double appearance of the phrase "proportional to 1/T"; we did not have time to try to estimate the associated constants of proportionality, and just possibly they might be so large that their quotient by a T of 1 or 2 hours would not be so very small at all.

To check on these undesirable possibilities,[4] a fast-time Monte Carlo simulation of the IFR landing stream was designed, programmed, and run. For each of several of the data combinations presented in subsection 4.1, a sample of 100 simulated 2-hour periods was generated. For the first

half-hour

hour

hour-and-a half

2 hours

of this 2-hour period, we recorded the associated sample average (to compare with the limit as $T \to \infty$ of the theoretical average), and the sample standard deviation.

The results shown in Figures 4.3.1 and 4.3.2 (for R=30 sec.) are quite representative. The two curves delimit a band, of width (± 1 sample standard

---

(4) And on whether the variance did appear to be converging to 0; this was not yet known at the time. Incidentally, the bounds on N(T)/T provided by equation (2.2) were calculated in our numerical work, and in general did not confine N(T)/T so tightly as to rule out significant fluctuations from the average.

Figure 4.3.1 - Analytical vs. Simulation Results

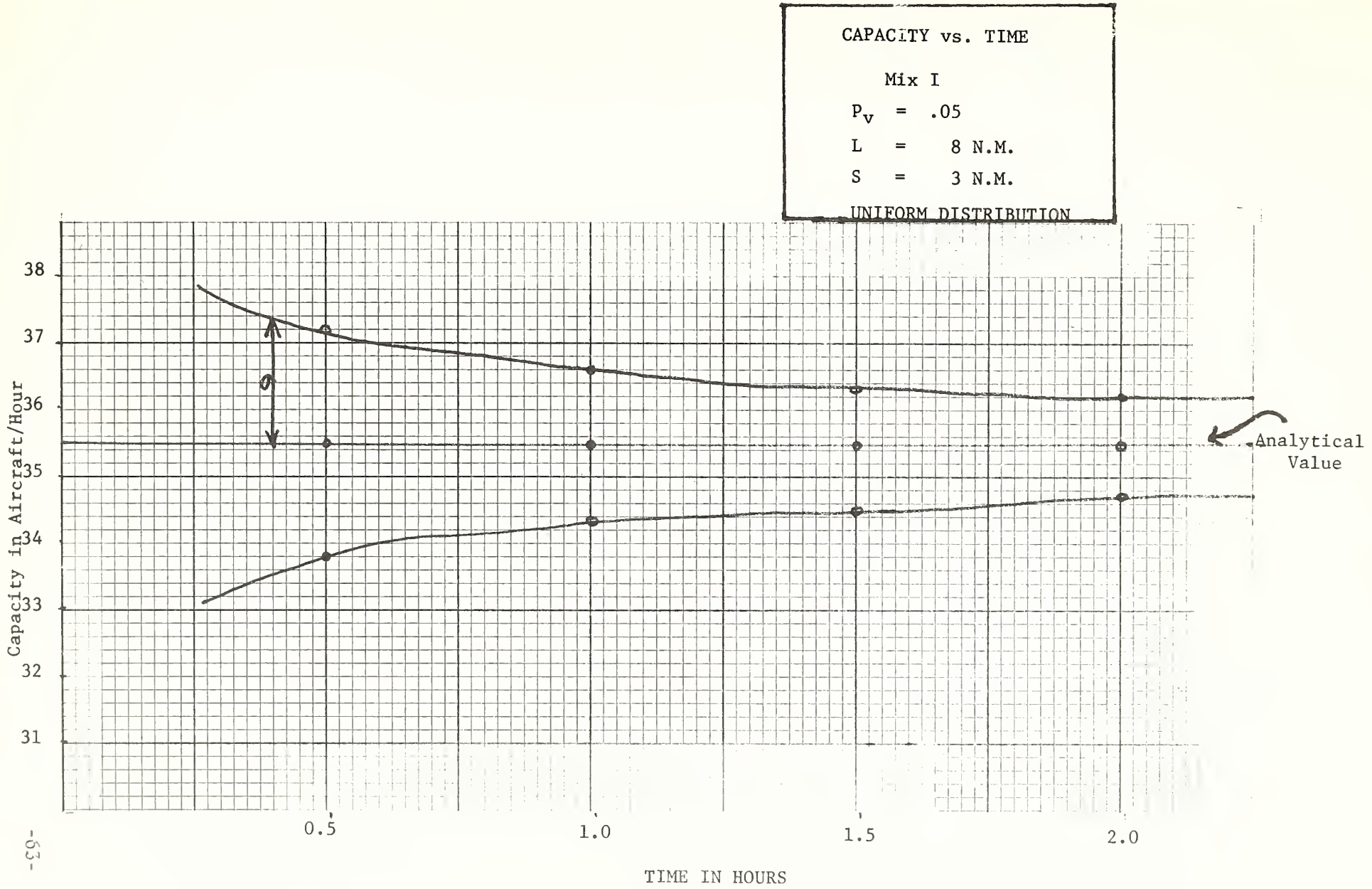
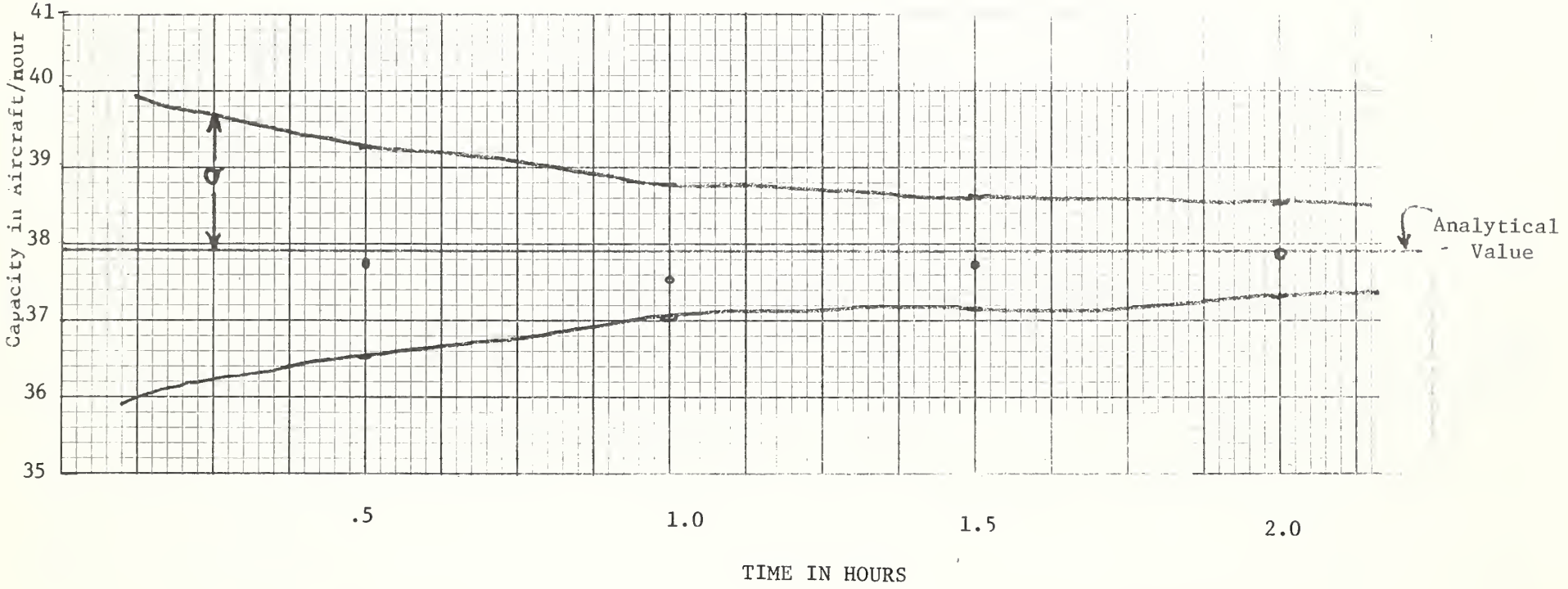


Figure 4.3.2 - Analytical vs. Simulation Results



CAPACITY VS. TIME

MIX I

p_v = .05

L = 8 N.M.

S = 3 N.M.

TRIANGULAR DISTRIBUTION

deviation), around the capacity value calculated from the recommended formula. We see that for the triangular-distribution case, as early as T=0.5 hr., the sample average differs from the theoretical value by less than 0.5 aircraft/hr., while the sample standard deviation is only about 1.4 aircraft/hr. and drops below 1.0 aircraft/hr. at T=1 hr. For the uniform distribution, the standard deviations are just a bit larger,while the settling-down of the average is even more rapid.

This establishes that the conceivable difficulties mentioned above should not in fact arise. Further specifics are given in Appendix F. Note, incidentally, that the sample averages all lie on the same side of the theoretical value; this is consistent with the fact that the constant B, in eq. (B.14), has a definite sign.

5. RELATION TO "HANDBOOK" VALUES

5.1 Preliminaries

One of our specific assignments was to compare the new capacity concept with that embodied in the Airport Capacity Handbook [1].(1)

So far as underlying ideas are concerned, a basic difference between the two concepts is that pointed out in subsection 2.1: the Handbook's "capacities" represent maximum traffic rates which can be handled without average delay exceeding a prescribed level, whereas the present concept is based on a "maximum throughput rate" idea not explicitly limited by delay considerations. On these grounds alone, the present concept should assign a higher capacity number to a given situation than would the Handbook.

At a more technical level, the Handbook's results are based on a mathematical model which assumes a Poisson distribution of "customer" arrival times. We understand that this assumption has been questioned as being too pessimistic, since arrivals for the facility in question (here, a runway and its final-approach path airspace) have been subjected to previous control actions tending to regularize their flow; thus a higher-order Erlang distribution has been suggested as a possibly more appropriate mathematical representation than the Poisson. At any rate, any really probabilistic treatment of arrival times admits the possibility of occasional idle periods for the facility, whereas our concept is defined in terms of a

---

(1) This reference is the first edition of the Handbook; the second edition, dated 6/69, became available too late for use in this study.

"continuous demand" scenario which rules out such periods as not properly relevant to a "capacity" measure. This difference, too, will tend to make our values higher than the Handbook's.

We turn now to a numerical comparison of the present concept's capacity numbers with those given by the Handbook, for the case of a single runway handling a stream of IFR landings. Our capacity numbers are taken from the material in Section 4, using the data and parameter values given in subsection 4.1.[2] How the comparable "Handbook values" were obtained will be discussed here and in subsection 5.2, before the comparative numbers are presented and discussed in subsection 5.3.

The need to discuss how "Handbook values" for the comparison were determined may at first seem absurd; one would think the only possible answer is "we looked them up in the Handbook." Moreover, the look-up procedure for IFR landings at a runway is quite simple:

(a) First one consults Figure 16-1 (p.16-2 of the Handbook). This gives a curve for average delay (which we will denote $\bar{D}$) versus λ, the arrival rate parameter in the underlying Poisson-distribution model. The Handbook recommends $\bar{D}$=4 min. as the appropriate "tolerable level", and has λ=21.7 landings/ hr. explicitly marked on its Figure 16-1 as the corresponding nominal value of "capacity."

---

[2] The half-range R, of the random error in achieving a desired intertouchdown spacing, was held fixed at a "nominal" level R=30 sec.

(b) Figure 16-1, however, is based on some single nominal mix of aircraft types among the users of the facility. The remainder of the Handbook's Chapter XVI, Figure 16-2 through 16-10, gives multiplicative correction factors to be applied to the basic λ =21.7 to obtain the capacity for the "actual" population mix at hand.

This straightforward look-up process was in fact employed to obtain one of the sets of "Handbook values" which appears in the comparison of subsection 5.3. However, we had two misgivings about it. The less important of the two is that the procedure is somewhat laborious and liable to human error; if a larger set of comparisons were required in the future, a computer program based on the mathematical models underlying the Handbook (or at least its Chapter XVI) would clearly be more suitable than a visual-manual lookup from graphs.

The second misgiving is more serious. For a proper comparison of the Handbook's values with ours, the two concepts should be applied (as nearly as possible) to the same sets of cases. However, the Handbook's material permits variation only of the mix of aircraft types and not of a number of quantities used in our concept which appear to be relevant:[3]

---

[3] We would not expect the parameters R_V and R of our mathematical model to be directly relevant for the Handbook's capacity concept. Insensitivity to these factors, as well as the four listed above, appears to reduce the usefulness of the Handbook's capacity information for cost-effectiveness analyses at a higher level than airport design.

length (L) of final-approach path
minimum distance separation (S) required along path
final approach speeds (v_i) by A/C type
runway occupancy times (OT_i) by A/C type.

Presumably, nominal values of these quantities are implicit in the Handbook's contents, but these values might not be compatible with those (in subsection 4.1) used in evaluating this report's capacity concept. In particular, the data used in Section 4 are based on experience more recent than the period (1963-65) mentioned in the title of the relevant Handbook chapter.[4]

Thus, to ensure proper comparability capability we again needed to know the mathematical models underlying the Handbook's graphs, so that our data could be entered into these models to determine what capacities the Handbook's concept would yield for the same cases evaluated in our Section 4. At this point, we cannot avoid some critical comments concerning the documentation (at least, that available to us) of the methodology underlying the Handbook. This documentation was found to be essentially impenetrable, despite the high motivation and technical competence of the staff members examining it. Some sections were more yielding than others, but it was simply not possible for us to extract a clear checkable statement of the mathematical models and calculations used in obtaining the Handbook's curves[5], even for the especially simple

---

[4] On the other hand, Chapter XVII of the Handbook takes up IFR operations for the period 1970-1975; possible use of this chapter is taken up in subsection 5.3.

[5] We understand that others have had the same experience.

case of a single runway handling a stream of IFR landings. (Such documentation does not even permit the reader, with a reasonable effort, to satisfy himself as to the correctness of the reported work. More satisfactory supporting documents may in fact exist, but the time-frame of our study precluded an intensive search.)

We are therefore unable to be fully certain that any of the "Handbook values" employed in subsection 5.3 are simultaneously (i) truly comparable with those given by the present concept and (ii) true representatives of what the Handbook's methods would produce. As one set of "Handbook values" in the comparison, we employ capacity values extracted from Chapter XVI by look-up; their use seems acceptable since the Handbook is after all intended to be employed "as is" without supplementary explanatory material. In addition, some detective work was performed to "estimate" the model used in the Handbook's Chapter XVI; this work, which provides a basis for calculating "modified Handbook values" possibly better-suited for the comparisons to be made, is reported in subsection 5.2. Some readers may prefer to skip over this next subsection, which is relatively technical in content.

5.2 A Basis for "Modified Handbook" Values

The text of a 1965 professional-society presentation by an FAA staff member describes ([15],p.1) the underlying model, for arrivals only at a single runway, as a "simple Poisson-fed queue with first-come-first-served discipline". The term "simple" in this queuing context has two possible interpretations as regards the probabilistic distribution of holding times: either constant ($= 1/\mu$), or exponential (with average $= 1/\mu$). The formulas for average delay in these two cases are standard ones ([9],p.347):

$$\bar{D} = \lambda/2\mu(\mu-\lambda) \qquad \text{(constant)}, \qquad (5.1)$$

$$\bar{D} = \lambda/\mu(\mu-\lambda) \qquad \text{(exponential)}. \qquad (5.2)$$

Although Figure 1 of [15] contains (5.1) rather than (5.2), implying that constant rather than exponential holding times were assumed, both possibilities were maintained through this phase of the analysis.

The capacity concept of the Handbook requires setting[5a]

$$\bar{D} = \bar{D}_o = 4 \text{ min.}$$

and solving for the corresponding value of λ (the capacity).

(5a) That is, $\bar{D}_o$ will stand for 4 minutes (expressed in appropriate units) in all the formulas to follow.

This yields

$$\lambda = 2\mu^2 \bar{D}_o / (1 + 2\mu \bar{D}_o) \qquad \text{(constant)}, \qquad (5.3)$$

$$\lambda = \mu^2 \bar{D}_o / (1 + \mu \bar{D}_o) \qquad \text{(exponential)}. \qquad (5.4)$$

For present purposes, however, it is more relevant to solve for the average holding time $1/\mu$ in equations (5.1) and (5.2) with $\bar{D} = \bar{D}_o$:

$$1/\mu = [\bar{D}_o^2 + (2\bar{D}_o/\lambda)]^{1/2} - \bar{D}_o \qquad \text{(constant)} \qquad (5.5)$$

$$1/\mu = [(\bar{D}_o/2)^2 + (\bar{D}_o/\lambda)]^{1/2} - \bar{D}_o/2 \qquad \text{(exponential)} \qquad (5.6)$$

In using the model for predicting capacity (λ) by (5.3) or (5.4), a definite value of μ must be employed. The Handbook values of λ depend on the mix of aircraft types (the p_i); since $\bar{D}_o$ is constant, μ must depend on the mix. We now assume that in the method underlying the Handbook, this dependence is of the form

$$1/\mu = \Sigma_i p_i (1/\mu_i) \qquad (5.7)$$

where $1/\mu_i$ is a model parameter representing the average holding time for type i aircraft; thus $1/\mu$ is a general average holding time.

If the $1/\mu_i$ were known, then (5.7) together with (5.3) or (5.4) represents the conjectured Handbook model employed in its Chapter XVI. It only remains to estimate the $1/\mu_i$, which are initially unknown to us.

For this purpose, we can regard (5.7) as giving $1/\mu$ as a linear form in the variables p_i, with unknown coefficients $1/\mu_i$ which are to be estimated. Such estimation becomes a standard least-squares linear regression problem (for which computer codes are readily available), given the appropriate data consisting of an adequate number of p_i-combinations (aircraft-type mixes) together with the value of $1/\mu$ corresponding to each. The value of $1/\mu$ corresponding to a given population-mix can in turn be found by first using look-up in the Handbook's Chapter XVI to determine a value of capacity ($\hat{\lambda}$), and then determining $1/\mu$ from (5.5) or (5.6).

This procedure was followed, using a set of 15 different population mixes. The resulting estimates are shown in Table 5.2.1; the closeness of the R^2 statistic[5b] to 1, and the smallness of the estimate standard deviations (about 1% of the estimates), provide some assurance that an approach like that sketched above really does underlie the Handbook's "IFR landings, single runway" curves. In addition, eq. (5.7) is consistent with the use of averaging procedures elsewhere in the Handbook, e.g. in calculating nominal "runway ratings" ([1], p. C-2, right-hand column). Note that neither column in Table 5.2.1 has its entries arranged in order of size; we do not know whether or not this is cause for disturbance.

[5b] Here and on the next page only, R denotes a multiple correlation coefficient (a measure of goodness-of-fit of a proposed formula to a set of data); this should not be confused with the usage of "R" in the bulk of the report as a measure of spread in intertouchdown time errors.

As a further check, the estimated values of the $1/\mu_i$'s were used together with equations (5.7) and (5.3-4) to "predict" the Handbook values of capacity for three additional mixes of aircraft types (not among the 15 mixes used in the estimation). Table 5.2.2 shows that the predicted values are in fairly good agreement with

Table 5.2.1: Estimated Values of Average Holding Times ($1/\mu_i$) by Aircraft Type.(*)

	Constant(**)	Exponential(***)
$1/\mu_A$:	171.5	145.3
$1/\mu_B$:	112.9	98.8
$1/\mu_C$:	115.9	101.3
$1/\mu_{D+E}$:	134.3	116.0

(*) Each column based on 15 observations.

(**) In seconds per A/C. $R^2 = 0.9818$.

(***) In seconds per A/C. $R^2 = 0.9829$.

Table 5.2.2: Predicted vs. Extracted Values of Handbook Capacities[(a)] for Three Mixes

Mix	p_A	p_B	p_C	p_{D+E}	Const.[(aa)]	Exp.[(aaa)]	Extracted
I	0.6	0.2	0.2	0.0	18.4	18.5	18.2
II	0.2	0.4	0.2	0.2	22.0	20.0	21.7
III	0.0	0.3	0.3	0.4	23.9	23.4	23.4

those extracted from the Handbook's curves.

In connection with this predictive process, it is convenient to rewrite equations (5.3) and (5.4) in terms of $1/\mu$ (the quantity given by (5.7)) rather than μ. The new versions are

$$\lambda = 2\,\bar{D}_o \,/\, (1/\mu)[2\,\bar{D}_o + 1/\mu] \qquad \text{(constant)}, \qquad (5.8)$$

$$\lambda = \bar{D}_o \,/\, (1/\mu)[\,\bar{D}_o + 1/\mu] \qquad \text{(exponential)}. \qquad (5.9)$$

The contents of this subsection so far, can be summarized as follows: We have made a plausible case that the mathematical formulas

---

(a) In A/C per hour.

(aa) Predicted values assuming constant holding-time model.

(aaa) Predicted values using exponential holding-time model.

underlying Chapter XVI of the Handbook are equivalent to the procedure of calculating a mix-dependent $1/\mu$ from eq. (5.7), and then computing the capacity (λ) from (5.8) or (5.9). Since eq. (5.1) rather than (5.2) appears in [15], it is probably (5.8) rather than (5.9) which should be used. Moreover, the numerical values of the $1/\mu_i$'s used for the 1963-65 curves in Chapter XVI should be approximately those in Table 5.2.1.

It follows that "modified Handbook values" of λ, probably more suitable for comparison with numerical values of this report's capacity concept than are those extracted from the Handbook's Chapter XVI, can be obtained by using the procedure described in the last paragraph. However, the $1/\mu_i$'s of Table 5.2.1 should be replaced with values of average holding time (by aircraft type) consistent with our data in Section 4. This will be tried in the next subsection.

The remainder of the present subsection consists of two comments of a more technical nature, which some readers may prefer to omit. The first of these concerns the approach (involving linear least-squares regression) used above to estimate values of the $1/\mu_i$ from the curves in the Handbook's Chapter XVI. This approach was such that $1/\mu$ played the role of "dependent variable" in the regression. Since we are "fitting" a model to be used in calculating values of λ, it would be theoretically preferable to perform parameter estimation with λ

serving as dependent variable. The appropriate functional form is that obtained by substituting equation (5.7) into (5.8) and (5.9). Note that the resultant functional form is nonlinear (specifically, reciprocal quadratic) in the parameters $1/\mu_i$ to be estimated; thus the theoretically preferable procedure requires use of a nonlinear regression algorithm (e.g., [16]). Such an algorithm was therefore applied, and proved to yield only negligible changes from the estimates found by linear regression.

The second technical comment concerns the use(reported in [15]) of equation (5.1), apparently in conjunction with equation (5.7), in producing the Handbook's curves in Chapter XVI. Unfortunately this procedure is simply not correct; eq. (5.1) applies only to constant holding times, whereas $1/\mu$ as given by (5.7) is an average holding time representing a probabilistic mixture of constant holding times, but not itself a constant holding time. Equation (5.1) should be replaced by ([9],p.345):

$$\bar{D} = \{\lambda / 2\mu (\mu-\lambda)\} \{1 + s^2\mu^2\} \qquad (5.10)$$

where s^2 is the variance of holding time, given (when constant holding times $1/\mu_i$ are assumed for each aircraft type) by

$$s^2 = \Sigma_i \, p_i \, [\, (1/\mu_i) - (1/\mu)]^2 \, . \qquad (5.11)$$

Thus (5.8) should be corrected by replacing $\bar{D}_o$ with $\bar{D}_o \, (1 + s^2\mu^2)$. We have not attempted to investigate the numerical seriousness of this flaw.

5.3 Comparison with Handbook Concept

With the preceding explanations and preparations complete, we can begin the actual comparison of "capacity values" (presented in Section 4) for this report's concept with corresponding values of the concept embodied in the Handbook. Since the two concepts are based on different intuitive notions of "capacity", the meaningfulness of such numerical comparisons should not be over-estimated.

Four sets of comparisons will be made. These represent four different possibilities for obtaining "Handbook values" properly comparable with those obtained for the present concept. The first two sets involve values directly extracted from the Handbook; the third and fourth involve "modified Handbook values" based on the ideas in subsection 5.2.

The first and second sets of comparisons are presented in Table 5.3.1. Two columns of values are given for the capacity concept presented in this report; they correspond to the assumptions of a triangular or (more pessimistically) a uniform distribution for the random error in achieving a desired intertouchdown interval. Each entry in these two columns consists of (i) a range of values, followed by (ii) a single number. The range of values arises because the formula for this report's concept depends explicitly on a number of factors not employed in extracting values from the Handbook; thus a single "Handbook case" cor-

responds to a whole family of the cases considered in Section 4, and hence to a range of capacity-values. The single number in the entry is one extracted (for definiteness) from this range, and corresponds to the specific conditions

$L = 8$ n. mi.,

$S = 3$ n. mi.,

$p_v = 0.01$.

The third column in Table 5.3.1 gives capacity values taken from Chapter XVI of the Handbook. As noted earlier, these refer to the time period 1963-65, which casts doubt on their comparibility with the material in the preceding columns. Therefore, in the fourth column of Table 5-3-1, we present the analogous values extracted from the Handbook's Chapter XVII, which deals with IFR operations in the period 1970-75. These numbers are presented subject to the following reservations:

(a) They are based on the Airborne Instrument Laboratory's projections (specified at least in part in [2]) on the effects by 1970 of improved traffic-control techniques and equipment. These projections were made no later than 1963 (the date of [1] and [2]), and so represent a forecast roughly 7 years into the future. They may therefore be proving considerably over-optimistic or over-pessimistic,[6] and no doubt have been updated.

---

[6] This is not a criticism, but merely an acknowledgment of the difficulties attending medium -range and long-range forecasting.

(b) Chapter XVII deals mainly with "mixed" IFR operations, i.e., both landings and takeoffs. While "landings only" is certainly a special case of this, it is an extreme case, and the curves in Chapter XVII may be based on approximations whose quality deteriorates for extreme cases.

(c) From the instructions for Chapter XVII (see Example 1, p. 17-1 of [1]), it appears that this chapter's "capacity value" for a single runway serving a stream of landings is obtained either from Figure 17-6, or from the smaller of the readings from Figures 17-1 and 17-6. The uncertainty on this point causes no trouble here, since for the particular mixes under examination both alternatives give the same result. What is disturbing is that the capacity values in each of these two figures depends only upon the proportion of Class A aircraft in the mix. In addition, the "bending backwards" of the curve in Figure 17-6 causes concern as to whether the methodology is consistent with that employed in Chapter XVI.

Examination of Table 5.3.1 reveals two salient points. First, the values for the proposed concept are systematically and distinctly greater than those taken from Chapter XVI of the Handbook. (The same is not true for the values from Chapter XVII, but for the reasons given above not much significance is attached to this.)

Second, the values of the proposed capacity concept diminish as one goes from faster to slower mixes, whereas the Handbook values increase.

Table 5.3.1: Comparison with Values Extracted from Handbook. (A/C per hour)

MIX	PROPOSED CONCEPT		HANDBOOK EXTRACTS	
	Triang.	Uniform	Chap. XVI	Chap. XVII
I (Fast)	29.0-46.2 (35.9)	28.1-42.7 (34.7)	18.2	36.5
II (Med.)	26.5-45.4 (33.0)	25.8-41.9 (31.9)	21.7	38.0
III (Slow)	24.0-42.3 (30.0)	23.4-39.3 (29.1)	23.4	39.2

Table 5.3.2: Comparison of Proposed-Concept Values (C) with Modified Handbook Values (λ) using Intertouchdown Times as Holding Times. (A/C per hour)

C:	20	25	30	35	40	45	50	55	60
λ (const.):	11.5	19.2	24.0	28.8	33.7	38.6	43.5	48.4	53.3
λ (exp.) :	11.4	15.6	20.0	24.5	29.1	33.8	38.5	43.2	48.0

Table 5.3.3: Comparison with Modified Handbook Values using Runway Occupancy Times as Holding Times. (A/C per hour)

MIX	PROPOSED CONCEPT		MODIFIED HANDBOOK	
	Triang.	Uniform	Const.	Exp.
I (Fast)	29.0-46.2 (35.9)	28.1-42.7 (34.7)	62.9	57.4
II (Med.)	26.5-45.4 (33.0)	25.8-41.9 (31.9)	66.4	60.7
III (Slow)	24.0-42.3 (30.0)	23.4-39.3 (29.1)	77.0	71.2

This divergence can be explained in part by the following line of reasoning:

On the one hand, in passing from aircraft types with higher final approach speeds to those making slower approaches, runway occupancy times typically diminish (see Table 4.1.5), and this tends to increase capacity for the slower mixes. On the other hand, diminished speeds will increase the time separation into which the minimum distance-separation requirement translates, and this tends to decrease capacity. Which of these two opposing tendencies wins out is largely determined by the relative frequencies with which runway occupancy and distance separation (respectively) determine tieup time; see equations (3.13) and (3.14). This last sentence is couched in the language of the present report; in connection with the Handbook, we must instead speak less concretely of the relative influences of runway occupancy times and of distance separation criteria on the "holding time" concept(7) entered into the underlying queuing model.

Examination of intermediate results, in the calculations reported in Section 4, reveals that tieup time was determined by "separation" rather than "runway occupancy" except for some cases with S = 2. Thus the decrease in our capacity values, for slower mixes of aircraft types, is explained. Evidently runway occupancy times play a more influential role in the "holding times" entering the Handbook's mathematical model.(8)

It is interesting that the discrepancy probably derives from the "holding time" concept in the work supporting the Handbook. The treatment of holding times could after all be varied while retaining the essence

---

(7) In the queuing-literature sense of tying up a facility, not the ATC sense of "holding."

(8) An alternative possibility, that the Handbook assumes a smaller required distance separation, is ruled out by p. 3-9 of [2].

of the delay-based capacity concept, and so possibly the divergence could be removed. It might also be removed by suitable changes in the calculated values of the proposed concept, due say to variations from the (nondefinitive) data on runway occupancy times and final approach speeds used in Section 4. But the decreasing trend down each of the first two columns of Table 5.3.1 seems too clear-cut to be reversed by reasonable perturbations of the input data.

We turn now to comparisons with "modified Handbook values." In subsection 5.2, evidence was assembled that such values could be produced by one of the formulas

$$\lambda = 2\,\overline{D}_o/(1/\mu)[2\,\overline{D}_o + 1/\mu] \qquad \text{(constant)}, \tag{5.12}$$

$$\lambda = \overline{D}_o/(1/\mu)[\overline{D}_o + 1/\mu] \qquad \text{(exponential)}, \tag{5.13}$$

where the parenthetical term refers to the assumed distribution of holding times in the underlying queuing model, $\overline{D}_o = 4$ min., and $1/\mu$ is a mix-dependent average holding time given by

$$1/\mu = \Sigma_i\, p_i\, (1/\mu_i) \tag{5.14}$$

with $1/\mu_i$ the average holding time for type i aircraft.

A set of modified Handbook (capacity) values is obtained by assigning appropriate values to the $1/\mu_i$. Without a clear understanding of the holding-time concept

used in the Handbook's implicit mathematical model, one cannot be certain how such values should be assigned. However, two alternatives will be examined. One of them should produce results quite properly comparable with those of Section 4 for the proposed capacity concept, but at the probable cost of doing real violence to the "holding time" idea embodied in the Handbook's Chapter XVI. The second is believed to hew more closely to the ideas underlying the Handbook, but its results are less comparable with those of Section 4.

Specifically, the first idea is to equate $1/\mu_i$ to the average inter-touchdown time for type i aircraft (i.e., between the touchdown of such an aircraft and that of the next aircraft in the stream). In the context of Sections 2 and 3, this becomes

$$1/\mu_i = \Sigma_j \, p_j \, (r_{ij} + b_{ij}),$$

and so equation (5.14) yields

$$1/\mu = \Sigma_{ij} \, p_i \, p_j \, (r_{ij} + b_{ij}).$$

Now comparison with equation (2.17) yields

$$C = \mu \qquad (5.15)$$

where C is the capacity concept proposed in this report. Thus equations (5.12) and (5.13) --- see the more convenient forms (5.3) and (5.4) --- become

$$\lambda = 2 \, \overline{D}_o \, C^2/(1 + 2 \, \overline{D}_o \, C) \qquad \text{(constant)}, \qquad (5.16)$$

$$\lambda = \overline{D}_o \, C^2/(1 + \overline{D}_o \, C) \qquad \text{(exponential)}. \qquad (5.17)$$

This gives an explicit algebraic conversion from values (C), of the capacity concept used in this report, to "associated" modified Handbook (λ) values. With the temporary notation

$$A = 2\,\overline{D}_o \text{ in (5.16)}; \quad A = \overline{D}_o \text{ in (5.17)},$$

the preceding relations become

$$\lambda/C = AC/(1 + AC) = 1 - 1/(1 + AC). \tag{5.18}$$

From this it is apparent that $\lambda < C$, and that λ/C increases from 0 toward 1 as C increases from 0. In particular, λ increases with C so that the type of discrepancy noted in connection with Table 5.3.1 cannot occur; this is not surprising since the "holding time" concept underlying λ has been bodily replaced by that used in our calculation of C-values. A numerical comparison is given in Table 5.3.2; the second row is more likely to be relevant than the third.

A second alternative is to equate the $1/\mu_i$'s to the runway occupancy times OT_i. In view of the above discussion of Table 5.3.1, this was thought to reflect more closely the holding-time concept used in deriving the Handbook's curves. The resulting comparison, using the OT_i-values in Table 4.1.5, is given in Table 5.3.3. We see that the "modified Handbook values" obtained look considerably too high. Thus the holding times used in the Handbook probably include more than just runway occupancy.

6. POSSIBLE NEXT STEPS

The present report of course constitutes only an initial exploration of the new capacity concept. In this Section, we sketch some of the next steps that would be involved in investigating the concept more fully and bringing it closer to operational status. Such further development may well be useful to the FAA in connection with its Continuing Study of Air Traffic Control System Capacity and Demand.

The steps to be described fall into three groups. One pertains to extending the scope of the concept. The second involves more intensive application and exploitation of the theoretical progress already made than was possible within the present study's time span. The third concerns attempts to derive sharper information on certain technical points arising during the study.

In the first and most ambitious category, five tasks suggest themselves:

(a) Extend the concept to deal with a runway serving both arrivals and departures.

(b) Extend the concept to deal with interactions between runways.

(c) Extend the concept to a "weighted capacity" notion, in which "throughput" is measured in a way sensitive (for example) to the number of passengers served and not merely to the number of aircraft.

(d) Modify the mathematical model to incorporate some penalty in capacity associated with violations.

(e) Continue study of an associated delay concept (see Appendix E).

These five tasks are too much to tackle together in a small-scale effort. Item (a) seems to deserve top priority, and progress on it might well have spillover usefulness to the subsequent study of (b). We believe that task (c) requires only modest additional theoretical development beyond what has already been done, but are not quite so confident concerning (d). On balance, it would seem that (a) and (c) represent a reasonable pair of targets for the "extension" portion of a follow-on study; further work on (e) also seems called for.

There are several possibilities in the "further applications" category:

(a) Perform sensitivity analyses with respect to variations in the runway occupancy times and final approach speeds attributed to various aircraft types.

(b) Examine the effects of introducing violation - probability thresholds and distance separation rules which depend on the types of aircraft involved.

(c) Derive analytical formulas for sensitivity coefficients.

(d) Examine alternative and better-founded models for the distribution of errors in achieving desired intertouchdown times.

(e) Apply the methods to a stream of IFR takeoffs, and perhaps to VFR operations as well.

(f) Redo the comparison with the Handbook's concept, using the second edition of the Handbook.

Of these tasks, only (d) appears to be of fundamental significance, and only (d) and (e) promise any real difficulty. The remaining items

should in principle require only relatively minor efforts, so that those of them considered to hold sufficient interest might as well be undertaken.

We have identified three tasks of the "derive sharper information" variety. The first two are aimed at reducing the incentive for auxiliary simulations like that reported in subsection 4.3:

(a) Recall that the capacity measure, over a prolonged period of duration T, differs from its limit as $T \to \infty$ by a term essentially proportional to 1/T. Analytically determine or estimate the constant of proportionality.

(b) Recall that the mean throughput rate, over a prolonged period of duration T, has a variance essentially proportional to 1/T. Analytically determine or estimate the constant of proportionality.

(c) Investigate the effects of the fact that approach speeds and runway occupancy times vary within an aircraft "type."

In addition to the three categories of tasks listed above, we mention also the possible investigation of capacity enhancement through sequencing of customers in other than "first-come-first-served" order. Such study might begin along the lines of Appendices C and D, e.g. by investigating what capacity increases are theoretically possible using the idea proposed in Appendix C.

7. REFERENCES[1]

1. Airport Capacity Handbook.* Airborne Instruments Laboratory, 6/63.

2. Operational Evaluation of Airport Runway Design and Capacity.* Airborne Instruments Laboratory, 1/63. E. N. Hooten et al.

3. Airport Runway and Taxiway Design.* Airborne Instruments Laboratory, 7/60. M. A. Warskow et al.

4. Staff Study: Air Traffic Control System Capacity and Demand (Report #1 of a Continuing Study). FAA Systems Analysis Division, Systems Research and Development Service, 9/68 Working Draft. S. P. E. Price.

5. (Untitled manuscript.) MITRE, 6/9/69 Draft. R. M. Harris.

6. Introduction to Probability Theory and its Applications, Vol. II. J. Wiley and Sons, 1966. W. Feller.

7. Probabilities of Missed Approach for Random vs. Controlled Arrivals.* Control Technology, Inc., 4/63. R. V. Morse and T. W. Connolly.

8. Development and Application of a Terminal Spacing System. FAA (NAFEC) Report NA-67-3, 12/67. D. A. Martin and F. M. Willett, Jr.

9. Mathematical Methods of Operations Research. J. Wiley and Sons, 1959. T. L. Saaty.

10. The Laplace Transform. Princeton U. Press, 1946. D. V. Widder.

11. Finite Markov Chains. Van Nostrand, 1960. J. Kemeny and J. Snell.

12. Mathematical Theory of Reliability. J. Wiley and Sons, 1965. R. E. Barlow and F. Proschan.

13. Introduction to Linear Programming. McGraw-Hill, 1960. W. W. Garvin.

14. Limiting Covariances in Markov Renewal Processes. Berkeley Operations Research Center Report, 7/22/64. W. S. Jewell.

15. An Airport Capacity Handbook: Results of a FAA Research Program in Mathematical Models and Techniques for Determining Airport Capacity. Paper D3b at 27th National Meeting, Operations Research Society of America, 5/7/65. O. Shapiro.

16. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. J. Soc. Ind. Appl. Math. 11 (1963), pp. 431-441. D. W. Marquardt.

---

[1] Titles followed by an asterisk indicate contractor reports to the FAA.

17. Industrial Scheduling. Prentice-Hall, 1963. J. F. Muth and G. L. Thompson (ed.).

18. Theory of Scheduling. Addison-Wesley, 1967. R. W. Conway, W. L. Maxwell and L. W. Miller.

19. Theory of Stochastic Processes. J. Wiley and Sons, 1965. D. R. Cox and H. D. Miller.

20. Stochastic Service Systems. J. Wiley and Sons, 1962. J. Riordan.

21. Some Vistas of Modern Mathematics. U.of Kentucky Press, 1968. R. Bellman.

22. Theory of Linear Economic Models. McGraw-Hill, 1960. D. Gale.

23. Priority Rules for Runway Use. Operations Research 12 (1964), pp. 941-950. G. Pestalozzi.

24. Some Problems in Traffic Delay. Operations Research 10 (1962), pp. 74-104. G. H. Weiss and A. A. Maradudin.

25. Distribution of Gaps and Blocks in a Traffic Stream. Operations Research 10 (1962), pp. 197-217. R. M. Oliver.

26. Waiting for a Gap in Traffic. Berkeley Operations Research Center Report, 6/15/61. W. S. Jewell.

27. Airport Runway Capacity and Delays, I. Berkeley Operations Research Center Report, 8/28/62. R. M. Oliver.

*p.25

APPENDIX A: PROOF OF EQUATION (2.6)*

In this appendix we give a relatively elementary proof (using only Laplace transforms) of equation (2.6) of the main text[1]. The situation being analyzed involves a sequence $I_1, I_2, \ldots$ of random interservice times. The probability distribution of each I_n is assumed (a deliberate oversimplification) to depend only on the type of customer C_n, but this type is in turn picked from a probability distribution (given by the p_i's) which is the same for all n. Thus, if I_n has cumulative distribution function H_i when C_n is of type i, then its unconditional probability distribution is $H = \Sigma_i p_i H_i$. In particular its (unconditional) average, which here will be denoted μ, is given by equation (2.5) of the text.

The formula to be proved is

$$\lim_{T \to \infty} E[N(T)/T] = 1/\mu , \tag{A.1}$$

where the random variable $N(T)$, the number of customers served during the period $[0,T]$, is defined by

$$I_1 + \ldots + I_{N(T)} \leq T < I_1 + \ldots + I_{N(T)} + I_{N(T)+1} . \tag{A.2}$$

By the definition of "average value" for a discrete random variable $N(T)$,

$$E[N(T)] = \sum_{m=0}^{\infty} m \, \text{Prob}\{N(T) = m\}$$

$$= \sum_{m=0}^{\infty} m \, [\text{Prob}\{N(T) \geq m\} - \text{Prob}\{N(T) \geq m+1\}]$$

---

(1) The treatment is formal, omitting some technical details required for mathematical rigor.

$$= \sum_{m=0}^{\infty} m \text{ Prob}\{N(T) \geq m\}$$

$$- \sum_{m=0}^{\infty} [(m+1)-1] \text{ Prob}\{N(T) \geq m+1\}$$

$$= \sum_{m=0}^{\infty} m \text{ Prob } \{N(T) \geq m\}$$

$$- \sum_{m=0}^{\infty} (m+1) \text{ Prob}\{N(T) \geq m+1\}$$

$$+ \sum_{m=0}^{\infty} \text{Prob}\{N(T) \geq m+1\} .$$

Replacing the "dummy variable" m in the first sum by $m+1$ shows that the second sum cancels the first except for the latter's term for $m = 0$, which is zero anyway. Making the same replacement in the third term, we are left with

$$E[N(T)] = \sum_{m=1}^{\infty} \text{Prob}\{N(T) \geq m\} . \tag{A.3}$$

Now by (A.2), $N(T) \geq m$ holds if and only if

$$I_1 + \ldots + I_m \leq T . \tag{A.4}$$

Since $I_1, \ldots, I_m$ are independent random variables, with cumulative distribution function H, the distribution function of the sum $I_1 + \ldots + I_m$ is the m-fold convolution H^{*m} of H. Thus (A.3) becomes

$$E[N(T)] = \sum_{m=1}^{\infty} (H^{*m})(T) . \tag{A.5}$$

We now take Laplace transforms of both sides of equation (A.5). Let $\bar{E}(s)$ and $\bar{H}(s)$ denote the Laplace transforms of $E[N(T)]$ and $H(T)$ respectively. Since the Laplace transform converts convolutions into products, the transform of function $(H^{*m})(T)$ is the power $[\bar{H}(s)]^m$. Hence

$$\bar{E}(s) = \sum_{m=1}^{\infty} [\bar{H}(s)]^m .$$

The right-hand side is a geometric series[2] which can be summed explicitly, yielding

$$\bar{E}(s) = \bar{H}(s)/[1-\bar{H}(s)] . \qquad (A.6)$$

Next, expand $\bar{H}(s)$ in a power series[3] about s=0:

$$\bar{H}(s) = \bar{H}(0) + s\bar{H}'(0) + o(s) \qquad (A.7)$$

where "o(s)" here represents terms which vanish at least as fast as s^2 as $s \to 0$. Since

$$\bar{H}(s) = \int_0^{\infty} \exp(-sT)dH(T) ,$$

we have

$$\bar{H}(0) = \int_0^{\infty} dH(T) = 1$$

(because H is a cumulative distribution function), and

$$d\bar{H}/ds = - \int_0^{\infty} T \exp(-sT)dH(T)$$

so that

$$\bar{H}'(0) = - \int_0^{\infty} T \, dH(T) = - \mu .$$

---

[2] Convergent for $s > 0$, since the exponential is < 1 in the integral expression for $\bar{H}(s)$ given below.

[3] Convergence is needed only near $s = 0$.

Substituting these results into (A.6) yields[4]

$$\bar{E}(s) = [1-\mu s+o(s)]/[\mu s+o(s)]$$
$$= (1/\mu s) + o(s) \ . \qquad (A.8)$$

Now the "Tauberian theory" of the Laplace transform permits one to infer the limiting behavior as $T \to \infty$ of the original function, here $E[N(T)]$, from the limiting behavior as $s \to 0$ of its Laplace transform, here $\bar{E}(s)$. Since $\bar{E}(s)$ behaves like $1/\mu s$ as $s \to 0$, it follows[5] that $E[N(T)]$ behaves like T/μ as $T \to \infty$. From this the desired result (A.1) follows.

The preceding material has been presented to make it conveniently available to interested readers who have not previously encountered it; it is not an original contribution.

---

[4] Here different appearances of "o(s)" can refer to different functions of s with the property described on p.94.

[5] See [10], p.192.

*p.27

APPENDIX B: JUSTIFICATION OF EQUATION (2.7)*

In this appendix we provide background material leading up to equation (2.7) of the main text. Recall that the runway is considered to be in "state i" during the interservice interval of a customer of type i. Let

$$p_{ij} = \text{Prob}\{C_{n+1} \text{ is of type } j \text{ , given that } C_n \text{ is of type } i\} . \tag{B.1}$$

Then the sequence of types of arriving customers constitutes a type of random process known as a Markov chain, with (p_{ij}) as its table of transition probabilities. A process such as those to be considered here admits a unique set of positive numbers x_i such that

$$\begin{aligned} \Sigma_i x_i p_{ij} &= x_j \qquad (\text{all } j) , \\ \Sigma_i x_i &= 1 , \end{aligned} \tag{B.2}$$

and these numbers have a number of properties[1] permitting one to regard x_i as the long-term probability of being in state i .

Our situation has still more mathematical structure, however. This involves the sequence of interservice times I_n. If customers C_n and C_{n+1} are of respective types i and j , then I_n is a random variable whose cumulative distribution function will be denoted H_{ij} , and whose average value will be denoted T_{ij} . It will be convenient to set

$$T_i = \Sigma_j p_{ij} T_{ij} , \tag{B.3}$$

which can be interpreted as the average interservice time for a customer of type i.

---

[1] See [11].

Our concern is with the random variable $N(T)$. To study it, it is convenient to introduce the "initial conditions"

$$a_i = \text{Prob}\{C_1 \text{ is of type } i\} \tag{B.4}$$

and the auxiliary random variables

$$N_{ij}(T) = \text{number of type } j \text{ customers served during } [0,T], \text{ given that } C_1 \text{ is of type } i .$$

Then the average number of customers of <u>all</u> types served during $[0,T]$, given that C_1 was of type i, is

$$E[\Sigma_j N_{ij}(T)] = \Sigma_j E[N_{ij}(T)] .$$

Averaging over the possible types of C_1, we find that the (unconditional) average value of the total number of customers served during $[0,T]$, i.e. of $N(T)$, is

$$E[N(T)] = \Sigma_i a_i \Sigma_j E[N_{ij}(T)] . \tag{B.5}$$

To apply this, we first obtain a formula for $E[N_{ij}(T)]$. Consider, as a random variable, the time <u>from</u> the initiation of service to a type i customer <u>to</u> the next moment at which service to a type j customer is started.[1a] Let ℓ_{ij} denote the average value of this random variable, and $\ell_{ij}^{(2)}$ its second moment. Both ℓ_{ij} and $\ell_{jj}^{(2)}$ appear in the formula for $E[N_{ij}(T)]$, and so we must be able to evaluate them.

It can be shown ([12], pp. 132-3, eqs. (2.7-8)) that

$$\ell_{jj} = (1/x_j)\ \Sigma_k x_k T_k$$

[1a] Other customers of types different from j may have intervened.

where the x's and T_k's are those of (B.2) and (B.3), and that

$$\ell_{ij} = \Sigma_{k\neq j} p_{ik}\ell_{kj} + T_i \ . \tag{B.7}$$

If the x's and T_k's are known, then the ℓ_{jj} can be found from (B.6); for each j, (B.7) gives a set of linear equations which can in general be solved for the ℓ_{ij} with $i\neq j$.

T_i was the average value of a random variable with distribution $\Sigma_j p_{ij}H_{ij}$; let $T_i^{(2)}$ denote the second moment of such a random variable. Then ([12], p.134, eq.(2.10)),

$$\ell_{jj}^{(2)} = (1/x_j)[\Sigma_k x_k T_k^{(2)} + 2\Sigma_{k\neq j}\ell_{kj}\Sigma_r x_r p_{rk} T_{rk}] \ . \tag{B.8}$$

Finally, the formula for $E[N_{ij}(T)]$ is given, in terms of $\ell_{jj}^{(2)}$ and the ℓ_{ij}'s, by[2] ([12], p.137, Theorem 2.10)

$$E[N_{ij}(T)] = T/\ell_{jj} + [\ell_{jj}^{(2)} - 2\ell_{jj}\ell_{ij}]/2\ell_{jj}^2 + o(1) \tag{B.9}$$

where "$o(1)$" denotes a function of T which tends to 0 as $T \to \infty$. With the abbreviation

$$B_{ij} = [\ell_{jj}^{(2)} - 2\ell_{jj}\ell_{ij}]/2\ell_{jj}^2 \ , \tag{B.10}$$

this yields

$$E[N_{ij}(T)/T] = 1/\ell_{jj} + B_{ij}/T + o(1/T) \tag{B.11}$$

---

(2) The formula in [12] contains an extra term allowing for certain periodic phenomena not relevant here.

where "$o(1/T)$" denotes a function of T which tends to 0, as $T \to \infty$, so fast that $o(1/T)/(1/T) \to 0$.

Substitution into (B.5), and division by T, leads to

$$E[N(T)/T] = \Sigma_i a_i \Sigma_j [1/\ell_{jj} + B_{ij}/T] + o(1/T) . \tag{B.12}$$

Now set

$$B = \Sigma_{ij} a_i B_{ij} , \tag{B.13}$$

use the fact $\Sigma_i a_i = 1$, and use (B.2) and (B.6) to substitute for ℓ_{jj} in (B.12). The result is

$$E[N(T)/T] = 1/\Sigma_k x_k T_k + B/T + o(1/T) . \tag{B.14}$$

This is the main result. It asserts that

$$\lim_{T \to \infty} E[N(T)/T] = 1/\Sigma_k x_k T_k , \tag{B.15}$$

and that for large T, since $B/T \gg o(1/T)$, the error in replacing $E[N(T)/T]$ by the limit is essentially B/T. Using (B.3), we can rewrite (B.15) as

$$\lim_{T \to \infty} E[N(T)/T] = 1/\Sigma_{ij} x_i p_{ij} T_{ij} . \tag{B.16}$$

In the case of concern to us, the probability that a given customer is of type j is assumed to depend only on the general population mix, not on the type of the previous customer. Mathematically, we have p_j's

rather than p_{ij}'s. By direct substitution into equations (B.2), we find that in this case the x's are given by the p's themselves. Thus (B.16) becomes

$$\lim_{T \to \infty} E[N(T)/T] = 1/\Sigma_{ij} p_i p_j T_{ij} \ , \tag{B.17}$$

reproducing eq.(2.7) of the main text.

In this case we also have $a_i = p_i$, so that (B.13) becomes

$$B = \Sigma_{ij} p_i B_{ij} \ .$$

Using (B.10) and the fact that $\Sigma_i p_i = 1$, we obtain

$$B = \Sigma_j [\ell_{jj}^{(2)} - 2\ell_{jj} \Sigma_i p_i \ell_{ij}]/2\ell_{jj}^2 \ . \tag{B.18}$$

Since all $x_i = p_i$, it may be possible to simplify (B.18) considerably further using (B.6-8), but there has not been time to pursue this possibility for getting sharper information on the error term in (B.14).

So far the discussion has dealt with the behavior of average value $E[N(T)/T]$ for large T . We are also concerned, however, with the behavior of the <u>variance</u> $Var[N(T)/T]$ of this random variable; see "QUESTION 3" in subsection 2.4. For this purpose, set

$$N_i(T) = \Sigma_j N_{ij}(T) \tag{B.19}$$

and observe that

$$Var[N(T)] = \Sigma_i a_i Var[N_i(T)] \ . \tag{B.20}$$

By (B.19),

$$\mathrm{Var}[N_i(T)] = \Sigma_{jk}\mathrm{Cov}[N_{ij}(T),\ N_{ik}(T)] \ . \tag{B.21}$$

We employ the results[3] of [14],p.9. These results are given explicitly only for the case corresponding to the presence of just two customer types, but it is stated in [14] that the more general case displays the same behavior. Specifically, there exist constants c_{jk} such that

$$\lim_{T \to \infty} \{\mathrm{Cov}[N_{ij}(T),\ N_{ik}(T)]/T\} = c_{jk}. \tag{B.22}$$

It follows from (B.21) that

$$\lim_{T \to \infty} \{\mathrm{Var}[N_i(T)]/T\} = \Sigma_{jk}c_{jk} = c \ . \tag{B.23}$$

From (B.20) and the fact $\Sigma_i a_i = 1$, we have

$$\lim_{T \to \infty} \{\mathrm{Var}[N(T)]/T\} = c \ . \tag{B.24}$$

Since $\mathrm{Var}[N(T)/T] = (1/T^2)\mathrm{Var}[N(T)]$, we see that for large T , $\mathrm{Var}[N(T)/T]$ behaves like c/T . Further study may permit estimation of $\underline{c}$.

---

(3) We are grateful to Professor R. Pyke (University of Washington) for directing us to this unpublished report.

APPENDIX C. ON SEQUENCING OF ACCEPTANCES

The bulk of Appendix B dealt with a situation more general than that treated in the main text. Specifically, the probability distribution of the "next customer type" was permitted to depend on the current customer type, rather than only the customer mix. Thus we had

$$p_{ij} = \text{Prob}\{C_{n+1} \text{ is of type } j \text{ , given that } C_n \text{ is of type } i\} . \tag{C.1}$$

The resultant formula for capacity (equation (B.17)) was found to be

$$C = 1/\Sigma_{ij} x_i p_{ij} T_{ij} \tag{C.2}$$

where the positive x_i give the unique solution of the system

$$\begin{aligned} \Sigma_i x_i p_{ij} &= x_j \qquad \text{(all } j\text{)} , \\ \Sigma_i x_i &= 1 . \end{aligned} \tag{C.3}$$

Only after this derivation, in Appendix B , was the specialization $p_{ij} = p_j$ made. Its implication is that no attempt at sequencing customers of different types is made. With appropriate changes in technology and operating procedures, however, such sequencing might be worth consideration. We wish therefore to present one formulation (but not a solution!) of a type of optimal sequencing problem which may merit study in such a context.

The "independent" variables of this optimization are the p_{ij}'s of (C.1), which represent the control policy. The auxiliary variables are the x_i's. The objective is to maximize the capacity C as given by (C.2), or equivalently, to

$$\text{minimize } \Sigma_{ij} x_i p_{ij} T_{ij} \ . \qquad (C.4)$$

The variables must satisfy three sets of constraints. One consists of the relations (C.3), which determine the x's in terms of the p_{ij}'s. A second set consists of

$$p_{ij} \geq 0 \qquad \text{(all } i,j\text{)}, \qquad (C.5)$$

$$\Sigma_j p_{ij} = 1 \qquad \text{(all } i\text{)}, \qquad (C.6)$$

$$\Sigma_i p_i p_{ij} = p_j \qquad \text{(all } j\text{)}. \qquad (C.7)$$

expressing the requirement that the p_{ij}'s do constitute a set of "transition probabilities" compatible with the <u>given</u> population mix prescribed by the p_j's. The third set, which cannot be written in closed form, restrain the p_{ij}'s to be such as to admit a unique and positive solution (x's) to the system (C.3) of linear equations.

It is apparent, by comparison of (C.3) and (C.7), that all $x_i = p_i$ so that (C.4) becomes

$$\text{minimize } \Sigma_{ij} p_{ij} (p_i T_{ij}) \ . \qquad (C.8)$$

Thus, <u>if</u> the above-mentioned third set of constraints could be ignored, we would have a linear programming problem of a special form called the <u>generalized transportation problem</u>, which has known solution methods more computationally effective than those for linear programs in general(1).

However, it is not apparent that every (or even any) optimal solution of the generalized transportation problem will satisfy the third set of constraints, so that further investigation is needed. Moreover, a full mathematical formulation of the sequencing problem should include some representation of the <u>costs</u> of sequencing control (both monetary, and in delays to some customers) as well as its benefits (in increasing capacity).

---

(1) See Chapter 10 of [13].

ADDENDUM: The uniquely reversible change of variables, from p_{ij} to $y_{ij} = p_i p_{ij}$, transforms the generalized transportation problem into

$$\text{minimize } \Sigma_{ij}\, T_{ij}\, y_{ij}$$

subject to

$$y_{ij} \geq 0 \qquad \text{(all i,j)},$$

$$\Sigma_j\, y_{ij} = p_i \qquad \text{(all i)},$$

$$\Sigma_i\, y_{ij} = p_j \qquad \text{(all j)}.$$

The non-zero coefficients in the constraints are now all unity, i.e., the problem has been transformed into an "ordinary" transportation problem for which still simpler and more readily available solution algorithms exist. Here T_{ij} plays the role of a "unit cost" per pair of consecutive landings consisting of a type i aircraft followed by one of type j; y_{ij} is the relative frequency of such landing-pairs among all consecutive landing-pairs.

APPENDIX D: MORE ON SEQUENCING

At present, landings at runways are typically handled on a first-come-first-served basis. Technological advances, however, may permit enhancing the capacity of this and other ATC System levels by adopting some more sophisticated "sequencing" procedure.[1/] This notion receives considerable attention in [4], with its explicit recognition of "selection algorithms" (p.64) and its careful distinction (pp.68-69) between (i) "maximum capacity" and (ii) the smaller "normative capacity" achievable by a first-come-first-served policy.[2/]

The mathematical-logical justification for useful sequencing procedures are likely to resemble or be based upon material in the standard references on this field, e.g. [17] and [18]. Appendix C's content, while closely related to the body of the report, stands somewhat apart from the class of problems typically considered in this body of technical literature. The material that follows is much more in the tradition of "sequencing" research, but makes relatively little contact with the balance of our text despite having been directly stimulated by the present study. This is quite understandable, since our report's capacity concept as presently developed is a "normative capacity" (see the last paragraph) rather than one defined so as to allow for the potential benefits of deliberate deviations from "first-come-first-served."

[1/] Conversely, the desire to achieve such a capability may influence the planning of technological R. and D.

[2/] In this connection, we may note the (non-technical) lead paper in [17], describing why operating staff are often practically unaware that sequencing problems are significant in their activities.

We will be concerned with two of the many possible mathematical formulations of the following problem: If a facility can serve only one customer at a time, in what order should it accept the customers presented to it? The word "should" will be interpreted here as referring to some "penalty function" to be minimized. This function is assumed to be the sum of penalties (possibly zero) relating to individual customers and the times at which they are served.

An important question in the formulation of such mathematical models is the nature of the information available to the "sequencer" --- how much is known about the customers, and when is this knowledge available? We shall deal here only with the simplest and most favorable case: before the first arrival of a customer, all relevant information about all customers (arrival time, length of service required, penalty-related data) is assumed known, exactly rather than probabilistically. The natural directions for generalization are obvious.

In the first formulation, time is treated as divided into discrete equal-length periods. The model involves a set of decision variables.

$x_{nt} = 1$ if customer n begins service in time period t,

$x_{nt} = 0$ otherwise.

There are three sets of constraints. The first,

$$x_{nt} = 0 \text{ or } 1 \qquad \text{(all n, t)}, \qquad \text{(D.1)}$$

expresses the dichotomous nature of each decision. The second,

$$\Sigma_t x_{nt} = 1 \qquad \text{(all n)} \qquad \text{(D.2)}$$

expresses the fact that each customer (n) must begin receiving service in one and only one of the time periods.

The third set of constraints involves a set of problem data, namely[3/]

d(n) = required duration of service to customer n.

Consider any time period t. If customer n begins service in time period s, this will tie up the facility during period t if and only if $s \leq t \leq s + d(n) - 1$, or equivalently

$$\max (1, t + 1 - d(n)) = r(n,t) \leq s \leq t.$$

Thus the requirement that the facility can be used by at most one customer at a time during any period (t) takes the form

$$\Sigma_n \Sigma^t_{s=r(n,t)} x_{ns} \leq 1. \qquad (D.3)$$

Since we are interested in "good" sequencing of customers, the mathematical model is naturally an optimizing one. Specifically, we seek to

$$\text{minimize } \Sigma_{nt} c_{nt} x_{nt}, \qquad (D.4)$$

subject to the constraints (D.1), (D.2), (D.3) given above. In the situations we have in mind, for fixed n the "penalty" coefficient c_{nt} is 0 for small values of t (customer n is being served early enough), but then increases with t (lateness penalties). This is however not essential; if for example e(n) is the earliest possible time at which customer n can be available for service, then this condition can be expressed by

$$c_{nt} = \infty \qquad \text{for } t < e(n),$$

[3/] For some of what follows, d(n) can be generalized to d(n,t), i.e. to service durations which depend on when service begins. Each d(n) is assumed integral.

though an attractive alternative is an extra constraint

$$\Sigma_{t < e(n)} x_{nt} = 0 \qquad \text{(all n)}.$$

What we have now is an integer linear programming problem (ILP) in the constrained variables x_{nt}. Its mathematical structure is sufficiently special that it might admit an especially efficient solution method (further study of this possibility is indicated), but for the moment the only apparent solution methods are those used for ILP's in general. For problems involving moderate to large numbers of variables and/or constraints, these methods are somewhat unpredictable in their ability to reach a solution in a reasonable amount of computer time. If we deal either with numerous customers (many n-values), or with a fine subdivision of the time scale (many t-values), we will in fact have a large problem. Therefore some experimentation, to see how the methods behave for this particular class of problems, is called for.

This difficulty would not arise if we had a "continuous" rather than an "integer" linear programming problem, since available solution methods for the former category can handle quite large problems without strain. Some important classes of ILP's have the property that their continuous analogs possess optimal solutions in which all variables have integer values. The pleasant consequence is that the more reliable solution methods for the continuous case can be used to solve such ILP's. We therefore investigated whether the problems at hand might fall into this category. Unfortunately the answer turned out to be negative, as is shown by the counter-example given in Tables E.1 - E.2 and discussed below.[4/]

[4/] Simpler counter-examples might exist.

The example involves 4 customers and 10 time periods. Table E.1 presents the associated penalty coefficients; c_{nt} is the entry in row n and column t. Customer n is assumed to require n time periods of service, i.e. d(n) = n.

We first show that the minimum value for the ILP is ≥ 2. Note that there are 10 time periods in which to accomplish Σ_n d(n) = 10 units of service; hence the facility can never be idle. To achieve a cost < 2, at most one of the following four statements can be violated:

(a) Customer 1 begins service before period 10, hence ends it before period 10.

(b) Customer 2 begins service before period 6, hence ends it before period 7.

(c) Customer 3 begins service before period 7, hence ends it before period 9.

(d) Customer 4 begins service before period 6, hence ends it before period 9.

If all but possibly (a) are true, at most customer 1's single unit of service is left to complete with periods 9 and 10 still left, leading to an idle period. If all but possibly (b) are true, then customers 3 and 4 have received their 7 units of service during periods 1 - 8; if customer 1 used the remaining one of these 8 periods (it might have been in period 1 or 4 or 5 or 8) then customer 2 must begin in period 9, with a cost of at least 7, while if customer 2 started service in period 8 (the only other possibility), the cost is at least 4.

Table D.1: Penalty Costs (c_{nt}) for Sequencing Example

	t:	1	2	3	4	5	6	7	8	9	10
n =	1	0	0	0	0	0	0	0	0	0	1
	2	0	0	0	0	0	1	2	4	7	11
	3	0	0	0	0	0	0	1	2	4	7
	4	0	0	0	0	0	1	2	4	7	11

Table D.2: Fractional Solution (x_{nt}) Reducing Total Penalty to 5/3

n = 1	0	0	0	0	0	0	0	0	1/3	2/3
2	1/3	0	1/3	0	1/3	0	0	0	0	0
3	1/3	0	0	0	0	0	1/3	1/3	0	0
4	1/3	0	0	1/3	1/3	0	0	0	0	0

Now suppose all but possibly (c) are true. Then all customers except the third have completed service before period 10, hence customer 3 receives service during this last period (since the facility cannot be idle), hence customer 3 began service in period 8 yielding a cost of at least 2. Similarly, if all but possibly (d) are true, then period 10 finds customer 4 still receiving service which must have begun in period 7, yielding a cost of at least 2.

The minimum value for the ILP is actually 2, as is seen for example by setting

$$x_{31} = x_{24} = x_{16} = x_{47} = 1 \qquad (x_{nt} = 0 \text{ otherwise}).$$

To prove the point, it now suffices to exhibit a (necessarily fractional) solution to the continuous analog which achieves a value < 2. The continuous analog is obtained by replacing (D.1) with $0 \leq x_{nt} \leq 1$, and a fractional solution to it which yields a total penalty 5/3 is given in Table E.2.

We turn now to the second formulation. It involves much more restrictive assumptions, but yields quite specific instructions as to an optimal sequencing policy. The restrictive assumptions are that all customers have the same priority (i.e., penalty) function, differing only in timing, and that all customers require the same duration of service. Further research, to see what conclusions can be derived when these conditions are weakened, seems worthwhile.

Specifically, we have a known availability time a_n for each customer n, and also a non-decreasing continuous function P with P(0) = 0. The

penalty associated with serving customer n at time t is taken as

$$p_n(t) = P(t - a_n).$$

Let t_n be the (variable) time at which customer n is served, and let $\vec{t}$ be the vector with n-th component t_n. Then the problem is to choose $\vec{t}$ so as to minimize the total penalty

$$S(\vec{t}) = \Sigma_n\, p_n(t_n) = \Sigma_n\, P(t_n - a_n), \qquad (D.5)$$

subject to three sets of constraints. The first set,

$$t_n \geq a_n \qquad \text{(all n)}, \qquad (D.6)$$

simply asserts that a customer cannot be served before he is available. The second asserts that the facility can serve at most one customer at a time; in terms of the common duration (d) of each customer's service, this reads

$$|t_m - t_n| \geq d \qquad (m \neq n). \qquad (D.7)$$

The third set expresses the requirement that all customers receive and complete service during a specified time interval, say [0, T]:

$$0 \leq t_n < t_n + d \leq T \qquad \text{(all n)}. \qquad (D.8)$$

Deferring the question of whether constraints (D.6) through (D.8) are consistent, we assume consistency and discuss the problem of finding an optimal (S-minimizing) service schedule. The solution will depend on a further assumption, namely that the function P is convex (i.e., is never under-estimated by linear interpolation). This has the significance that increasing delay carries non-decreasing marginal penalties.

Let N be the number of customers. It is convenient to index them in order of availability, so that

$$a_1 \leq a_2 \leq \cdots \leq a_N. \tag{D.9}$$

In terms of this ordering, we claim that the following simple rules yield an optimal policy:

Rule 1 : At time a_1, serve customer 1.

Rule 2 : At the conclusion of a service, if there are customers available, serve the one with lowest index. If there are none, but not all customers are yet served, serve the first one to become available (breaking ties, say, by giving preference to lower indices).

This "first-come-first-served" policy serves customer 1 at time $t_1^* = a_1$, and the other customers at times t_n^* determined recursively by

$$t_{n+1}^* = \max \{a_{n+1}, t_n^* + d\}. \tag{D.10}$$

To prove that it is optimal, we begin with any optimal schedule $\vec{t}$, and show that $\vec{t}$ can be transformed into the schedule $\vec{t}^*$ given by the two rules in a finite sequence of steps none of which increases the value of the function S to be minimized. Each step operates on a "current schedule" $\vec{s}$ ($\vec{t}$ initially, and then the output of the preceding step) to produce a new schedule which will obviously satisfy all the constraints (D.6-8) because the current schedule did.

The steps are of two types. The first type simply "tightens up" schedule $\vec{s}$, eliminating slack periods but not changing the order in which customers are served. In terms of the current schedule $\vec{s}$ with its components s_n, this step can be formally described as follows:

Find the smallest s_n such that $s_n > a_n$ and either (a) $s_n = \min_m s_m$, or (b) $s_n > \min^* \{s_m + d\}$ where the "min*" is a minimum taken over all m such that $s_m < s_m$. In situation (a), change s_n to a_n. In situation (b), change s_n to $\min^* \{s_m + d\}$.

When no such s_n exists, so that no further application of a Type 1 step is possible, the schedule has been "tightened up" as much as possible subject to the current order of service. In applying the second kind of step, the current schedule $\vec{s}$ is assumed to have been already "tightened" by Type 1 steps.

A Type 2 step rectifies a situation in which service is begun to a customer other than the one with smallest a_n among those available at the time. The formal description is as follows:

Find the smallest s_n such that $s_m > s_n$ for some $m < n$. With such an m, form a new schedule $\vec{s}^1$ from $\vec{s}$ by interchanging components s_m and s_n.

Note that $m < n$ implies $a_m \leq a_n \leq s_n$, so that in the new schedule customer m is not served before he is available. To show that the objective function is not increased, we calculate

$$\Delta S = S(\vec{s}) - S(\vec{s}^1)$$

$$= [p_n(s_n) + p_m(s_m)] - [p_n(s_m) + p_m(s_n)]$$

$$= [P(s_n - a_n) + P(s_m - a_m)] - [P(s_m - a_n) + P(s_n - a_m)].$$

Setting

$$x = s_n - a_n$$

$$y = s_m - s_n > 0,$$

$$z = a_n - a_m \geq 0,$$

we have

$$\Delta S = [P(x) + P(x + y + z)] - [P(x + y) + P(x + z)]. \qquad (D.11)$$

Since function P is never under-estimated by linear interpolation,

$$P(x + y) = P\left(\frac{z}{y+z} \cdot x + \frac{y}{y+z} \cdot (x + y + z)\right)$$

$$\leq \frac{z}{y+z} P(x) + \frac{y}{y+z} P(x + y + z),$$

and similarly

$$P(x + z) \leq \frac{y}{y+z} P(x) + \frac{z}{y+z} P(x + y + z).$$

Adding the last two inequalities gives

$$P(x + y) + P(x + z) \leq P(x) + P(x + y + z),$$

which by comparison with (D.11) shows that $\Delta S \geq 0$ as desired.

The over-all process of passing from $\vec{t}$ to $\vec{t}^*$ can now be described as follows: Apply Type 1 steps as long as possible, then a Type 2 step, then Type 1 steps as long as possible, etc. The process must terminate, since each sequence of consecutive Type 1 steps contains no more than N steps, while each Type 2 step brings the order in which customers are served closer (i.e., fewer "inversions") to the order $\{1, 2, \ldots, N\}$. This order is clearly achieved by the last Type 2 step. Note that the process is a conceptual one used only for <u>proof</u> purposes; for <u>calculating</u> the optimal schedule one sets $t_1^* = a_1$ and employs the recursive formula (D.10).

It only remains to consider whether the constraints (D.6-8) are consistent. If they are satisfied by at least one $\vec{t}$ (optimal or not), the above construction will lead from $\vec{t}$ to $\vec{t}^*$, which must also satisfy

the constraints; conversely the constraints are surely consistent if $\vec{t}^*$ satisfies them. Thus consistency can be checked by computing $\vec{t}^*$ as described above, and seeing whether it obeys the constraints; this is simply a question of whether $t_N^* + d \leq T$.

In conclusion, we emphasize that the two specific mathematical models presented above are illustrative and exploratory; the directions in which they might have to be modified and generalized to be useful in an ATC setting are not yet known to us.

APPENDIX E : A DELAY CONCEPT

As emphasized in subsection 2.2, this report's capacity concept differs in a basic way from that in the Airport Capacity Handbook, in that it is not bound up with some notion of a tolerable level of "delay." We can be somewhat more explicit. In a textbook-type queuing situation, there is no real problem in defining "delay." A customer arrives and is served at definite times (in general, known only probabilistically in advance); his delay is the duration of the period between these two events. The "continuous demand" assumption of Section 2, however, only requires that a customer be on hand whenever the facility is no longer tied up by the previous customer. Nothing is said about how much earlier that "on-hand" customer was in fact available to be served had no others been ahead of him. One of the essential ingredients for discussing "delay," namely a specification of when customers arrive, is therefore missing.

This separation of capacity from delay has its advantages. It permits capacity-focused studies to proceed without bogging down in questions as to what delay levels are "acceptable." Even more significantly, it avoids the complications stemming from the fact that the ATC System and its elements are far from "textbook" in their complexity, and do not provide clear and compellingly "right" ways to define "delay." These complications are communicated very vividly in [4]; some of them can be roughly described as follows:

(a) It does not take much sophistication to suggest that delay, as a "quality of service" concept, might well be measured not from the moment of a customer's "arrival" (which might by chance be early), but rather from the moment of some

scheduled or anticipated initiation of service. How should these be defined in an ATC setting? Flight plans are altered en route; should "scheduled or anticipated" refer to the expectations at takeoff, or after the last fight-plan updating prior to reaching the terminal area, or at some intermediate time?

(b) During its flight, an aircraft is served by a number of elements of the ATC System. These may be regarded as service facilities, but in some cases they can best "serve" a customer by increasing his "local" delay, rather than by passing him as promptly as possible into a soon-to-dissipate high-congestion situation further "down stream."(1) How should a delay concept be formulated so as to reflect such a multi-facility situation, with its overtones of potential conflict of interest among facilities in case each is rated solely by the delays which customers incur at its hands?

(c) It is an economic truism that in situations of competition for a "scarce resource" (e.g. airspace, or prime treatment from the ATC System), better treatment for some of the competing entities normally implies worse treatment for others. Thus for example "total system-wide delay" is not in itself adequate as a measure of performance; a proper analysis should also consider the equitability of the way in which this total is distributed among various "customers." How should a delay concept be defined so as to capture this aspect of the situation?

---

(1) That is, performing some portion of a flight somewhat less quickly than possible may be preferable if it avoids a severe slowdown later on.

We have just given good reasons why the absence of delay considerations from our capacity concept might be considered cause for rejoicing. Still, "delay" is so important in the ATC context as a measure of service quality, as to compel at least one exploration of what type of demand concept might reasonably be matched with the capacity concept developed in the main text.

The setting is that of Sections 3-5, namely a single runway (plus its final-approach path airspace) serving a stream of landings. As noted earlier, a treatment of delay will require some representation of the times of customers' "arrivals." Arrivals where? We will make two attempts at answering this question, find that both run afoul of our need to preserve mathematical simplicity in this initial exploration of delay concepts, and then settle for a compromise.

(a) From the view of actual or imagined risk it is at least plausible that "total time airborne" is what's to be minimized for a given flight. This suggests that a customer's "arrival," the moment from which his "delay" is measured in the standard queuing language, should be regarded as occurring as early as possible in the flight, say as soon as "enroute" status is reached. The trouble with this approach becomes apparent if we consider two flights departing for New York, one from Los Angeles at 10 a.m. and the second from Boston at 10:30. The Los Angeles flight would be counted as the first of the two to "arrive" at New York for service, though of course the second would typically be served first at New York. Thus the "far-out arrivals" approach would lead to severe violation of the assumption of a "first-come-first-served" service policy, and the mathematical analysis would become much more difficult than is tolerable at this stage of our study.

(b) Since our capacity concept is firmly associated with the runway and its immediate vicinity, there is a natural inclination (opposite to that in (a)) to define "arrival" so that the same will be true of the associated delay concept. A natural "close-in" definition would be that of arrival in the holding pattern. The flow of entries to the holding pattern, however, may be too well regulated by previous ATC controls to permit description as "Poisson arrivals" --- and this assumption that the Poisson distribution can be used is also important to the simplicity of the analysis.

In what follows, therefore, we shall regard "arrival" as occurring at a suitable point intermediate between the two extremes described above: sufficiently "far out" from the runway that the aircrafts' diverse origins and takeoff times combine to yield a pattern of arrival times that is acceptably Poisson-like, yet sufficiently "close in" that the first-come-first-served assumption is a reasonable approximation. Perhaps the "terminal boundary" will do. In a longer-term study with less stress on initial mathematical simplicity, one of the two extremes might possibly prove preferable to this compromise.

We next consider how to express, for present purposes, the assumption of continuous demand used in the capacity concept. Its statement there (subsection 2.2) was that at the end of each service period, another customer is invariably at hand. This condition cannot be achieved in a setting of Poisson arrivals; to retain the simplicity of the Poisson distribution, some other mathematical representation of the same intuitive

idea is needed. For this purpose, consider the following random variable: The time, from a completion of a service period only to find no "next customer" at hand, to the next occurrence of such a situation. Our modified statement of the "continuous demand" assumption is that the average value of this random variable be infinite[1a].

The mathematical formulation of this requirement proceeds as follows. We are dealing with a Poisson-fed queue at a facility operating under a first-come-first-served policy. Let λ denote the average arrival rate. For simplicity, it will be assumed that the "service time" for an aircraft[2] has a probability distribution depending on the type of that aircraft only (and not on the type of the next aircraft). If $1/\mu_i$ denotes the average service time for type i aircraft, and p_i is the proportion of type i aircraft in the customer "mix," then

$$1/\mu = \Sigma_i \, p_i \, (1/\mu_i) \qquad \text{(E.1)}$$

is the mean service time. The requirement stated at the end of the last paragraph can be expressed (see [19], pp. 115-117) as $\lambda = \mu$. Unfortunately, the delay concept under these conditions is quite degenerate. since average delay is given ([9], p. 345) by

$$\bar{D} = \{\lambda/2\mu(\mu - \lambda)\} \, (1 + s^2\mu^2) \qquad \text{(E.2)}$$

where s^2 is the variance of service time, setting $\lambda = \mu$ leads to an infinite average delay.

(1a) Roughly speaking, instead of assuming such situations cannot happen at all, we assume that on the average their occurrences are "infinitely far apart;" this seems to express the same intuitive idea.

(2) Here "service time" has the same meaning as did "holding time" in Section 5.

This outcome is neither physically reasonable (in the "runway" context) nor useful for our aims. We are therefore led to reexamine the physical reasonableness of the assumptions underlying the queuing model. The weak point appears to be a hypothesis so customary that it was not even stated explicitly before: that there is no bound to the length of the queue.

In fact, there must be a finite practical limit to what the airspace and ATC capabilities of the terminal area can handle. We interpret this as a finite value of

h = maximum queue length

= maximum number of A/C which can be held in the terminal area.

When the queue is "full" (has h members), new arrivals are abstractly considered to be "turned away"; we will not try here to think through how best to interpret this phrase[3/] in terms of diversion to other terminals and/or restrictions on original schedules. What seems worth noting, for its possible analogy-value in other studies of capacity and delay in an ATC setting, is the way in which the study of "delay" at one level (the runway and final-approach airspace) of the ATC System hierarchy has driven us to bring in the "capacity" at the next higher level (the terminal area).

With a finite maximum queue length (h), the average delay is at most h times the average service-time, and so is certainly finite. However,

---

3/ In the runway situation it typically represents a balking by the customer at an unacceptable wait, not a rejection by the facility.

the simple condition $\lambda = \mu$ no longer expresses exactly the previous formulation of the condition of continuous demand. To analyze the situation further, let $\rho = \lambda/\mu$ and let I be the long-term probability that the "facility" is idle. In terms of h, ρ, we have

$$I = (1 - \rho)/(1 - \rho^{h+1}) \qquad (E.2)$$

for the simple case of exponentially distributed service times ([20], p.73, formula for p_n with n = 0). As ρ increases to 1 (i.e., λ increases to μ), I approaches the nonzero limit 1/(h+1).

In general, one would expect that making $I \to 0$ (to express "continuous demand") would require taking $\rho \to \infty$ (i.e., $\lambda \to \infty$, for fixed μ). And one would further expect, in this case, that those customers who do join the queue almost invariably find h-1 others ahead of them and thus must suffer an average delay

$$\bar{D} = (h - 1)/\mu. \qquad (E.3)$$

This can be verified for exponential service times (and probably more generally, though we have not yet tried to do so), using the results on p.73 of [20]: The limiting probability that an arriving customer finds the queue less than full and so enters it ($(1-p_k)$ in the notation of the reference), is

$$(1 - \rho^{h})/(1 - \rho^{h+1}) \qquad (E.4)$$

and the mean delay for all customers (assuming 0 delay for one not joining the queue[4/]) is

---

[4/] This does not really jibe with the language on p.73 of [20], but appears to check with the mathematics; time was lacking to resolve this point definitively.

$$I\,\rho^2\,[1 - h\,\rho^{h-1} + (h-1)\,\rho^h]/(\mu\,\rho)(1-\rho)^2. \qquad \text{(E.5)}$$

The average delay for those customers who do join the queue is then obtained as the quotient of (E.5) by (E.4), with equation (E.2) used to substitute for I:

$$\bar{D} = (1/\mu)\,\rho\,[1 - h\,\rho^{h+1} + (h-1)\,\rho^h]/(1-\rho)\,(1-\rho^h). \qquad \text{(E.6)}$$

It can be shown that the limit of this quotient as $\rho \to \infty$ is indeed given by (E.3).

If the sole source of an aircraft's "delay", once within the terminal area, is that associated with queuing-up for use of the (sole) final-approach airspace and runway, then our capacity measure $C = \mu$ (see eq. (5.15)) so that equation (E.3) gives the compact expression

$$\bar{D} = (h-1)/C \qquad \text{(E.7)}$$

for average delay as a quotient of capacities at two levels of the ATC system. It remains to see how generally this holds.

Much of the previous material has involved an attempt to force the delay concept to include the "continuous demand" assumption already present in the capacity concept. The necessity for such an inclusion seems less evident for delay than for capacity, however. It may be best simply to accept $\bar{D}$ as a function of ρ (i.e. of arrival rate λ, for fixed μ), to be calculated using equation (E.6). Interest would naturally concentrate on situations with $\lambda > \mu$. The relationship with our capacity concept would enter via $C = \mu$, or some more general $1/\mu = (1/C) + \delta$ where δ is an average delay within the terminal area due to factors other than queuing for the runway.

Further study of these ideas would include developing the analog of eq. (E.6) for more general service-time distributions, in particular those implicit in the main text. Our study of this whole topic has been so brief, however, that further thought may well reveal a superior conceptual approach.

APPENDIX F: A SUPPORTING SIMULATION

In Appendix B a somewhat complicated analysis was given of the behavior, for large T, of the average and variance of the mean throughput rate of a facility over a time period [0,T]. The "facility" of interest here consists of a runway (and its final-approach path airspace) serving a stream of IFR landings. As a check on the analysis, and (more significantly) to discover what constitutes "large T" to which the analytical results apply, a model to simulate the runway over a period of time was desired.

A fast-time, Monte-Carlo simulation was developed, coded in FORTRAN (a sample listing follows), and operated for selected data combinations from subsection 4.1. In particular, each run was made for a specific population mix, error distribution type and range of distribution. Within each run the probability of violation (p_v), minimum separation distance between aircraft (S), and final approach distance (L) were allowed to take on all values mentioned in Tables 4.1.1 and 4.1.3.

For each combination of p_v, S and L, 100 simulations were made of a two hour period, with "the number of aircraft serviced so far" recorded after each half hour. This yielded four arrays, with 100 sample points in each. For each array, the mean and standard deviation (σ) of the sample points were calculated and printed. (Illustrative outputs for one set of cases are given in Figure F.1.)

Figure F.1 : Some Simulation Outputs

THE POPULATION MIX USED FOR THIS RUN WAS .6-A .2-B .2-C .0-D .0-E

THE VALUE OF R USED IN THIS RUN WAS 30. SECONDS WITH A TRIANGULAR DISTI

L	S	PV	MEAN .5	STD. DEV. .5	MEAN 1.0	STD. DEV. 1.0	MEAN 1.5	STD. DEV. 1.5	MEAN 2.0	STD. DEV. 2.0
6.	2.	.05	46.22	1.61	46.21	1.05	46.18	.82	46.19	.66
6.	2.	.01	43.25	1.54	43.31	.97	43.33	.72	43.32	.54
6.	3.	.05	38.72	1.31	38.71	.87	38.70	.71	38.68	.58
6.	3.	.01	36.64	1.27	36.53	.92	36.56	.72	36.60	.55
6.	4.	.05	31.21	1.05	31.20	.75	31.20	.56	31.21	.47
6.	4.	.01	29.97	1.00	29.90	.69	29.87	.53	29.87	.45
8.	2.	.05	45.23	1.63	45.23	1.13	45.21	.88	45.21	.71
8.	2.	.01	42.49	1.41	42.46	.91	42.47	.77	42.45	.66
8.	3.	.05	37.75	1.39	37.57	.90	37.71	.77	37.67	.68
8.	3.	.01	35.82	1.40	35.88	.95	35.89	.72	35.91	.56
8.	4.	.05	30.73	1.13	30.75	.72	30.70	.53	30.73	.46
8.	4.	.01	29.13	1.10	29.16	.78	29.36	.67	29.40	.48
0.	2.	.05	44.22	1.50	44.20	1.13	44.12	.87	44.14	.69
0.	2.	.01	41.52	1.51	41.50	1.04	41.46	.75	41.52	.64
0.	3.	.05	37.22	1.63	37.03	1.06	37.08	.90	37.07	.73
0.	3.	.01	35.18	1.44	35.20	.94	35.27	.73	35.22	.62
0.	4.	.05	30.32	1.16	30.31	.77	30.25	.64	30.21	.51
0.	4.	.01	29.06	1.17	28.94	.78	28.99	.63	28.91	.48

The resulting sample means were used to verify that the analytical results are indeed applicable to relatively short time periods; this is discussed in subsection 4.3.

The resulting values of σ were also used to see whether the time periods in question were long enough for the theoretical result

$$\sigma^2 \simeq c\ T, \qquad \text{(F.1)}$$

expressed in eq. (B.24), to take hold. If this relationship did indeed apply, then the quantity

$$K = \sigma/T^{1/2}$$

should be approximately constant (equal to $c^{1/2}$). Values of K for T = 0.5, 1.0, 1.5 and 2.0 hrs. were calculated for several sets of the simulation outputs in Figure F.1, and are shown in Table F.1. In that Table, K(T) designates the value of $\sigma/T^{1/2}$ for the simulated time period [0,T].

TABLE F.1 : Simulation Approximations of $K = c^{1/2}$.

L	S	p_v	DISTN.	K(.5)	K(1)	K(1.5)	K(2)
8	3	.05	UNIFORM	.42	1.19	1.15	1.07
8	3	.05	TRIANGULAR	.35	.90	.95	.96
10	4	.01	UNIFORM	.31	.98	.93	.89
10	4	.01	TRIANGULAR	.29	.78	.77	.68
6	3	.05	UNIFORM	.41	1.15	1.05	1.08
6	3	.05	TRIANGULAR	.33	.87	.87	.82

These results indicate that the value of K at 0.5 hrs. is quite misleading as to its limiting value for large T; moreover the "approach to the limit" is not evident for T = 1.5 hrs. It may occur by T = 2 hrs., but this cannot be ascertained without calculations for more values of T between 1.5 and 2 hrs. Unfortunately time did not permit either such intermediate calculations, or going on to higher values of T.

To avoid possible misunderstanding, we note that these findings do not limit the usefulness of the analytical results; as noted in subsection 4.3, σ is "small enough" for T = 0.5 hrs., and certainly for T = 1 hr. The findings merely illuminate the theoretical question of the range of validity for equation (F.1).

We conclude this Appendix with a listing (Figure F.2) of the simulation program. This is preceded by a glossary of the symbols appearing in the listing. There are some divergences from the notation in the main text, namely

$O\,T \rightarrow RO\,T$, $\quad L \rightarrow X$

$S \rightarrow Y$, $\quad r \rightarrow SERV$.

The listing is given for the triangular distribution case only.

Glossary of Simulation Terms

V(I) = Final approach velocity of i-th A/C, defined in knots but internally converted to N.M./sec.

ROT(I) = Runway occupancy time of i-th A/C, in sec.

X(I) = Length of i-th chosen value of final approach path, in N.M.

Y(I) = Length of ith chosen value of minimum separation distance, in N.M.

SERV(I,J) = Tieup time for A/C of type i followed by A/C of type j, in sec.

PV(I) = Probability of violation, with PV(1) = .05 and PV(2) = .01.

PMIX(I) = Percentage of the population mix of type i.

RANDNO(ARG_1, ARG_2) = A function generating a uniformly distributed pseudo random number between 0 and .9999, using the power residual method: $RANDNO = \frac{Mod_{10000}(3ARG_1 + ARG_2)}{10000}$

BUFFER(I) = Added separation time corresponding to PV(I) in sec.

R = Half-range of the error distribution, in sec.

ISEED1, ISEED2 = Seeds of random number generator.

TIMINT(I) = Endpoint of i-th time period for which samples are to be taken, in sec. (1/2, 1, 1 1/2, 2 hrs.)

QTYLND(I,J) = Number of A/C which have landed before the end of the i-th period, in the j-th sample.

AVF(I) = Mean of 100 samples for i-th time period, in A/C per hour.

STDDEV(I) = Standard Deviation of 100 samples for i-th time period, in A/C per hour.

IPER = Index of current period.

ITER = Index of current iteration.

ENDTIM = Endpoint of current time period, in sec.

NOLAND = Number of landings finished at current clock time.

TIME = Current clock time.

VAR = A triangularly distributed random variable with mean 0 and range 2R.

ACTIM = Time required to land current A/C.

Figure F.2 : Listing of Simulation Program

```
      DIMENSION V(5),ROT(5),X(3),Y(3),PV(2),SERV(5,5),PMIX(5),BUFFER(2)
     1    ,QTYLND(4,100),AVG(4),STODEV(4),TIMINT(4),NUMBRS(100)
      DATA V/165.,150.,135.,120.,105./
     1    ,ROT/59.,52.,45.,38.,31./    ,R/30./   ,TIMINT/0.5,1.0,1.5,2.0/
     2    ,PV/0.05,0.01/     ,X/6.0,8.0,10.0/     ,Y/2.0,3.0,4.0/
     3    ,PMIX/0.6,0.2,0.2,0.0,0.0/     ,ISEED1/3113/     ,ISEED2/2157/
C ------  THIS FUNCTION GIVES A RANDOM NUMBER BETWEEN 0 AND 1 ----------
C ------  USING THE  POWER RESIDUAL METHOD
      RANDNO(I,J)=MOD(3*I+J,10000)/10000.
C -------------------------------------------------------------------
      WRITE(6,780)PMIX
  780 FORMAT('1    THE POPULATION MIX USED FOR THIS RUN WAS ',F3.1,'-A',
     1  F3.1,'-B',F3.1,'-C',F3.1,'-D',F3.1,'-E'/)
      DO1I=2,5
    1 PMIX(I)=PMIX(I-1)+PMIX(I)
      DO2I=1,5
    2 V(I)=V(I)/3600.0
      BUFFER(1)=R*(1.0-SQRT(2.0*PV(1)))
      BUFFER(2)=R*(1.0-SQRT(2.0*PV(2)))
      DO500IX=1,3
      DO500IY=1,3
      ITEST=0
      DO10I=1,5
      DO10J=1,5
      IF(V(I).GT.V(J))S=X(IX)/V(J)-(X(IX)-Y(IY))/V(I)
      IF(V(I).LE.V(J))S=Y(IY)/V(J)
   10 SERV(I,J)=AMAX1(ROT(I),S)
      IF(ITEST.EQ.1) WRITE(6,781)((SERV(I,J),J=1,5),I=1,5)
  781 FORMAT(//5F10.3)
      DO450 IPV=1,2
      DO440 ITER=1,100
      IPER=1
      ENDTIM=1800.
   40 TIME=0.0
      NOLAND=0
      JTYPE=0
   51 RNDNUM=RANDNO(ISEED1,ISEED2)
      ISEED1=RNDNUM*10000
      ITYPE=JTYPE
      JTYPE=0
   50 JTYPE=JTYPE+1
      IF(RNDNUM.GT.PMIX(JTYPE))GOTO50
      IF(ITYPE.EQ.0)GOTO51
   52 RNDNUM=RANDNO(ISEED1,ISEED2)
      ISEED1=RNDNUM*10000
      RNDNM2=RANDNO(ISEED1,ISEED2)
      ISEED1=RNDNM2*10000
      VAR=(RNDNUM+RNDNM2-1.)*R
      ACTIME=SERV(ITYPE,JTYPE)+BUFFER(IPV)+VAR
      TIME=TIME+ACTIME
      IF(TIME.GE.ENDTIM)GOTO60
      NOLAND=NOLAND+1
      GOTO51
   60 QTYLND(IPER,ITER)=NOLAND+(ENDTIM+ACTIME-TIME)/ACTIME
```

```
      NOLAND=NOLAND+1
      IPER=IPER+1
      ENDTIM=ENDTIM+1800.
      IF(IPER.LE.4)GOTO51
  440 CONTINUE
      IF(ITEST.EQ.1) WRITE(6,782)QTYLND
  782 FORMAT(4F20.3)
      DO448IPER=1,4
      SUM=0.0
      DO441I=1,100
  441 SUM=SUM+QTYLND(IPER,I)
      AVG(IPER)=SUM/100.
      SUM1=0.0
      DO442I=1,100
  442 SUM1=SUM1+(AVG(IPER)-QTYLND(IPER,I))**2
      SIGSQD=SUM1/100.
  448 STDDEV(IPER)=SQRT(SIGSQD)
      DO 449 IPER=1,4
      AVG(IPER)=AVG(IPER)/TIMINT(IPER)
  449 STDDEV(IPER)=STDDEV(IPER)/TIMINT(IPER)
      WRITE(6,789)X(IX),Y(IY),PV(IPV),BUFFER(IPV), (TIMINT(I),AVG(I),
     1   STDDEV(I),I=1,4)
  789 FORMAT(///'  FOR THE CASE OF X=',F3.0,' , Y=',F3.0,' AND PV=',F4.2
     1   ,' (WHICH MEANS A BUFFER OF ',F3.0,' ) THE RESULTS FOLLOW.'/
     2   4(/20X,'FOR',F5.2,' HOUR(S) THE MEAN ARRIVAL RATE WAS',F8.3,
     3   ' WITH A STANDARD DEVIATION OF',F6.3))
      ITEST=0
  450 CONTINUE
  500 CONTINUE
      STOP
      END
```

APPENDIX G: EXTENSION TO A DUAL-USE RUNWAY

The capacity concept developed and applied in the body of this report dealt with a facility serving a single stream of customers, e.g. a runway serving a stream of IFR landings. An important next step, in bringing this approach nearer to practical usefulness, is clearly its extension to the case of a runway serving both arrivals (landings) and departures (take-offs) --- in a broader context, to facilities which serve a pair of customer-streams. The present appendix records our extremely tentative and exploratory thoughts on this topic.

As in the single-use case, the initial problem is not that of deriving a mathematical formula for a well-defined property of the runway. Rather, it is that of deciding what intuitive notion we are seeking to express in a quantitative way.(1)

It seems reasonable that the desired "capacity" should be descriptive of something like "maximum throughput rate" for departures, arrivals, or both, under some appropriate assumptions about the "service discipline" which sets priorities for the processing of the two streams. And once the "something like" and "some appropriate" are pinned down, it seems likely that translation first into a rigorous description and then into a mathematical formula can indeed be carried out using suitable generalizations (which may present technical difficulties) of the techniques employed in Sections 2 and 3. This technical development, involving the various "separation rules" involved, will not be taken up here; our aim is to get a start on the "pinning down" of the basic ideas involved.

---

(1) "It comes as a bit of a shock when one first realizes that the real problem of science is not so much 'what is the answer?' as 'what is the question?'" [21], p.89.

For a simple symbolic notation, let

A = throughput rate for arrivals,

D = throughput rate for departures.

Going back to fundamentals, we can safely say that a major role of a "capacity" concept is to distinguish those numerical pairs (A, D) which the facility can turn out from those it isn't capable of producing. This can for example be accomplished (at least in principle) by listing the "feasible" combinations (A, D), or by giving an algebraic or geometric description of the corresponding "feasible region" in the first quadrant of a plane with (A, D) as Cartesian coordinates.

However, "capacity" suggests not only the notion of "feasible" (i.e., attainable), but also the notion of "feasible and unimprovable," as in the qualifying adjective of "maximum throughput rate." To capture this additional idea, we fix attention not on the whole feasible region, but rather its "upper frontier," the set of feasible points (A, D) which would become infeasible were either coordinate increased further. Economists, using a similar diagram to discuss an economy or a plant which turns out two products, term such points efficient[2]; we will therefore refer to the upper frontier as the efficient curve for the facility ("production possibility curve" would be the technical phrase). Its points represent combinations of throughput rates for landings and takeoffs respectively, such that neither can be increased without detriment to the other; such limitations arise because the same "scarce resource" (use of the runway and its environs) is required in the production of both "goods."

[2] See [22], p. 307.

At this point our capacity concept involves not a single number ("scalar") but rather a vector with an A-component and a D-component; and moreover not just one such two-dimensional vector, but rather the infinite set of them which comprise the efficient curve. On the basis of our very limited consideration to date of the dual-use case, it does not seem possible to obtain a more definitive kind of "answer" without introducing additional considerations or assumptions (see the next paragraph). Thus it presently appears that one of the analytical tasks for future work, perhaps a principal one, is to determine the efficient curve. Its equation might be written

$$E(A, D) = 0 \qquad (G.1)$$

or perhaps in the solved-for-D form

$$D = e(A) \qquad (G.2)$$

giving the maximum throughput rate for departures when A is the specified throughput level for arrivals. The functions E in (G.1) and e in (G.2) would contain as parameters the various quantities characterizing the situation treated, e.g.: relative proportions of different aircraft types among arrivals and departures; relevant speeds, lengths and times; required separation distances and/or times; violation probability thresholds and associated error distributions like those in the body of this report.

In search of possible clues to what sort of "additional considerations" might be introduced to single out a definite point on the efficient curve, we might consider how the economist treats the analogous problem when considering a plant which can turn out two products in various quantity

combinations. His (quite natural) approach is to postulate that plant management will elect to operate at that point on the efficient curve[3] which maximizes the function

$$P(A, D) = \text{net profit from production and sale of the output mix } (A, D). \qquad (G.3)$$

That is, $P(A, D)$ is to be maximized subject to the constraint given by eq. (G.1). Under the simplifying assumptions that the two products are sold in independent markets, and that production costs can be assigned unambiguously to one or the other of the products (i.e., "joint costs" are negligible), the simpler functional form

$$P(A, D) = P_1(A) + P_2(D) \qquad (G.4)$$

results; with the further assumption of constant per unit profits in each market (P_A and P_D), we have the further simplification

$$P(A, D) = P_A A + P_D D. \qquad (G.5)$$

Although "profit" is presumably not the appropriate criterion for our runway situation, the last paragraph does suggest that the selection of a single point from the efficient curve must involve the optimization of some "scoring function" analogous to $P(A, D)$; its optimized value might then serve as a single "capacity number." There might be some inclination to choose as scoring function the total operation rate A+D (summing components is certainly one way to convert a vector into a scalar!), but no real justification is presently apparent for this choice, especially since its compatibility with the preference actually accorded arrivals is dubious.

[3] We omit discussing the "more production is more profitable" condition needed to ensure that the chosen operating point will be on the efficient curve rather than interior to the feasible region.

A similar idea, which singles out a point on the efficient curve though not yielding a scalar "capacity," is the minimization of some penalty function over that curve. The sort of function which suggests itself most naturally is the sum of two terms, one for each of the two streams; each term is the average cost of delay for that stream, perhaps obtained as the product of average delay by a constant (stream-specific) cost per unit delay.[4] This approach seems to incorporate many realistic aspects of the situation, yet raises many difficulties: the complications (noted in Appendix E) associated with the use of "delay" concepts; the need to specify demand patterns before aggregate delay effects become meaningful[5]; the fact that "capacity," if taken as the minimizing point on the efficient curve, can depend on cost factors not necessarily related to the runway and its surroundings (e.g., fuel costs and crew wage rates).

With matters in this inconclusive state, we shift to a discussion of what "scenario" should be employed in determining the efficient curve (i.e., what are the "appropriate assumptions" alluded to in the third paragraph of this appendix). Since our interest is capacity-focused, some version of the assumption of continuous demand (see subsection 2.2) should surely be imposed; as a minimum, whenever the facility is free to handle a customer, a customer from at least one of the two streams should be at hand.

Should the "continuous demand" condition be applied to either stream separately? Here we must take account of the current requirement that arrivals take priority over departures.[6] If this rule is absolute, and

[4] Cf. [23].

[5] This is analogous to the need to know supply-demand curves in the A-market and D-market before a profit function P(A, D) can be determined.

[6] Except of course for emergencies, e.g., a Coast Guard plane taking off on a rescue mission.

if "continuous demand" for arrivals obtains, the "production rate" for departures will clearly be zero.[7] This outcome being unacceptable, we must assume either that the priority-for-arrivals rule is not absolute, or that the traffic load of arrivals is less than continuous-demand, or both.

Consider first whether our scenario should treat the priority-for-arrivals rule as less than absolute. This raises the question of whether our goal is fidelity to current official rules, or fidelity to actual practice (in which "rules" may become "guidelines" to be tempered by the professional judgement of controllers), or the ability to encompass possibly quite different future situations to which it might be desired to apply the capacity concept. There are clearly subtleties involved; for example a departure may have been fitted into a "gap" in the stream of arrivals, but did the gap in fact arise out of "permissiveness", a deliberate failure to close up the arrival stream as tightly as possible? We will have to seek further information and guidance on these points.

Representing an absolute rigidly-followed rule in a mathematical model may present technical difficulties, but is at least unambiguous. If the priority-for-arrivals rule is not of this nature, how should degrees and kinds of deviations from it be modeled? Presumably the decision criteria (whether articulated or not) for exceptions to the rule are not of some mathematically convenient random nature, but rather refer to how long and long-suffering is the queue of aircraft waiting to depart. While

[7] We assume that departures cannot be interspersed between arrivals subject to both priority and "continuous demand" for the latter, but have not checked out this plausible impression.

further study is required, contemplation of these potential difficulties yields a fairly strong predisposition to treat the priority rule as "hard" should this prove a factually acceptable option.

Given the prospective treatment of the priority rule as rigid, we are obliged to turn to the second alternative posed above, that of dropping the "continuous demand" assumption on the stream of aircraft arriving for landings. This stream will therefore exhibit gaps, which (if long enough) can accommodate takeoffs. It seems reasonable to impose the continuous-demand assumption on the departure stream when determining the function in (G.2), so that the remaining major problems involve (a) arriving at a mathematical description of the stream of arrivals, and (b) deriving from this description the necessary information on the frequency of gaps of various sizes in this stream.

Regarding (b), there is a modest literature on "gapology" developed mainly in connection with road traffic, e.g. a pedestrian seeking to cross a traffic stream, or a vehicle from a side road seeking to enter or cross such a stream. We cite here only [24] - [26], as well as the runway-oriented [27]; a careful study of the potential usefulness of these papers and the further references they contain, as well as later work on the topic, had to be postponed for a possible follow-on project.

Concerning (a), one might hypothesize Poisson arrivals as in the work underlying the Handbook [1] (see for example [2], p. 8-1), but this would be subject to the criticisms noted earlier on p. 66. The use of less tractable probability distributions will surely complicate the mathematics,

but a quick appraisal of the references cited above offers hope that these difficulties can be coped with.

This winds up the account of our present ideas on extending the capacity concept to a runway used both for landings and for takeoffs. Quite obviously, our progress on this topic has not reached the point (jubilantly announced on p. 24 when attained in connection with the single-stream runway) at which conceptual problems give way to mathematical problems. However, considerable conceptual structuring and clarification have been accomplished, and initially preferred lines for further work have been identified: these include seeking formulas or numerical solution methods for determining the efficient curve under the assumptions of strict priority for landings and of continuous demand for departures.

We conclude by recalling, for contrast and comparison, the approach adopted in the Handbook [1]. On the technical side, both streams are regarded as Poisson-generated ([2], p. 8-1; [15], p. 2) and the priority-of-arrivals rule is assumed to hold strictly ([15]; pp. 1-2). The conceptual aspects can be inferred from sample exercises (Examples 1-3, [1], pp. 17-1, 2): using the mix of aircraft types (specifically, the Class A proportion), a value of A+D is read from Figure 17.1. The user must supply the desired A/D ratio by which to convert this (A+D)-value into a pair (A, D) with the indicated sum. However, this point (A, D) may not be attainable. Figure 17.6 is entered with the mix of A/C types to obtain a value A*, and $A_C=\min(A, A^*)$ is taken as the "arrival capacity." Define D_C by $A_C/D_C=A/D$;

then apparently A_c+D_c is taken as the total capacity. Average delays for both streams are obtained by entering Figure 17.7 using A+D, A_c+D_c, A_c and A.

The look-up in Figure 17.1 (together with the user-supplied ratio A/D) can be interpreted as yielding a point (A, D) which may fail to be in the feasible region; in that case it is replaced by the point (A_c, D_c) in which the efficient curve meets the line joining (A, D) to the origin (0, 0). The rationale for this graphical construction, which involves maintenance of the original A/D ratio, is not evident.


DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS