A MODEL FOR INVESTIGATING THE RELIABILITY OF TRAIN-TO-TRAIN CONNECTIONS IN RAILROAD FREIGHT YARDS

by

PETER ALEXANDER KERR

B.S.C.E., Union College
(1969)

Submitted in partial fulfillment

of the requirements for the degree of

Master of Science

at the

Massachusetts Institute of Technology

September, 1973

Signature redacted

Signature of Author..
Department of Civil Engineering
September, 1973

Signature redacted

Certified by..
Thesis Supervisor

Signature redacted

Accepted by...
Chairman, Departmental Committee on Graduate Students
of the Department of Civil Engineering

ABSTRACT

A MODEL FOR INVESTIGATING THE RELIABILITY OF TRAIN-TO-TRAIN CONNECTIONS IN RAILROAD FREIGHT YARDS

by

PETER ALEXANDER KERR

Submitted to the Department of Civil Engineering on 25 August 1973 in partial fulfillment of the requirements for the degree of Master of Science.

This thesis forms part of a research case study which tested the findings of previous research into railroad reliability. This mathematical model of operation of a classification yard was designed to predict, using two submodels, the probability of making a connection and the mean time for cars in that connection - two parameters chosen to represent yard performance. The model form involved both linear and logit functions. Significant variables included the time between arrival and departure, and the number of cars involved. Calibrated models were used to investigate typical connections at each of two yards, specific changes in operation at one yard, and the changes to the mean time under increased frequency of operation.

Three major conclusions were drawn:

1) There is a range of yard time between the arrival of a car on train A and the departure of that car on train B for which the mean time in the yard is constant.

2) The time available between arrival and departure is the most important factor in determining whether or not a car will make a connection.

3) Under increased frequency of operation, the improvements in mean time exhibit decreasing returns to scale.

Thesis Supervisor: Joseph M. Sussman

Title: Associate Professor of Civil Engineering

ACKNOWLEDGEMENTS

The author would like to acknowledge the support of

Carl D. Martland who kept this work on the right track,

Joseph M. Sussman who, as supervisor, insured that the work was technically correct, and

Miss Jean McBeth who did the typing.

The work on this thesis was supported as part of a research effort sponsored by the Federal Railway Administration (Contract # DOT-FR-10006).

TABLE OF CONTENTS

Page

Title........ 1

Abstract........ 2

Acknowledgements........ 3

Table of Contents........ 4

Index to Figures........ 8

Index to Tables........ 9

Chapter 1. Introduction........ 11

1.1 Background........ 11

1.1.1 Modal Choice........ 11

1.1.2 Previous Studies........ 13

1.2 Purpose of Thesis........ 16

1.3 Outline of Chapter Contents........ 17

Chapter 2. Basis of the Model........ 21

2.1 Introduction........ 21

2.2 How Rail Freight Moves........ 21

2.2.1 Origin to Destination Trip........ 21

2.2.2 Classification Yard Operations........ 23

2.3 Individual Train-to-Train Connection Performance........ 29

2.4 Summary of the Chapter........ 33

Chapter 3. Development and Calibration of the Model........ 35

3.1 Introduction........ 35

3.2 Variables........ 35

TABLE OF CONTENTS
(continued)

Page

3.2.1 Dependent Variable.................................... 36

3.2.2 Independent Variables.................................. 36

3.2.2.1 Continuous Variables Included in the Model 36

3.2.2.2 Discrete Variables Included in the Model.. 38

3.2.2.3 Variables Not Used in the Model........... 39

3.3 Form of the Model.. 40

3.3.1 Probability Submodel................................... 40

3.3.1.1 Assumptions About the Submodel............ 40

3.3.1.2 Previous Findings......................... 41

3.3.1.3 Functional Form of the Probability Submodel.................................. 43

3.3.2 Mean Time Submodel.................................... 44

3.4 Data Used.. 44

3.4.1 Limitations of the Data................................ 45

3.4.2 Observation of A Connection............................ 47

3.5 Calibration of the Probability Submodel for Two Yards..... 48

3.5.1 Validity of the Calibrated Submodels................... 51

3.5.2 Application of the Probability Submodel to Typical Connections at Each Yard.............................. 55

3.6 Mean Time Submodel....................................... 60

3.7 Interpretation of the Model.............................. 66

3.7.1 Implications for Railroad Operations................... 66

3.7.2 Time of Day Influences................................. 72

3.7.3 Differences Between the Models for Each Yard....... 73

TABLE OF CONTENTS
(continued)

Page

3.8 Summary of Chapter 75

Chapter 4. Applications of the Model 77

4.1 Introduction 77

4.2 Application of the Model in a Case Study 77

4.2.1 Background 77

4.2.2 Possible Changes 78

4.2.3 Evaluation Procedure 80

4.2.4 Results of Evaluation 81

4.2.5 Implementation 81

4.2.6 Summary 88

4.3 Use of the Model to Evaluate Increased Frequency of Operation 88

4.3.1 Introduction 88

4.3.2 Modifications to the Current Models 88

4.3.3 Analysis and Interpretation 92

4.3.4 Conclusion 96

4.4 Summary of Chapter 99

Chapter 5. Summary and Conclusions 100

5.1 Summary 100

5.2 Conclusions 103

5.3 Areas for Extension to This Model 106

TABLE OF CONTENTS
(continued)

Page

Appendices..110

Appendix A - Calibration Method..............................111

Appendix B - Logit Curve Calibration.........................116

Appendix C - Analysis of the Mean Time Submodel by Differential Calculus..........................117

Appendix D - Discussion of Possible Interdependence of the Model Variables..............................119

Bibliography..120

INDEX TO FIGURES

Figure		Page
1-1	Structure of the Model	18
2-1	Journey of a Freight Car	22
2-2	Yard Time Parameters	26
2-3	Sample Yard Time Distributions	30
3-1	Car Moving on Schedule vs. Available Yard Time	42
3-2	Sample Probability Curves at Yard A	58
3-3	Sample Probability Curves at Yard D	61
3-4	Mean Time Curves at Yard A	64
3-5	Mean Time Curves at Yard D	65
4-1	Mean Time, One Train Per Day	90
4-2	Mean Time, More Than One Train Per Day	91
4-3	Mean Time Curves as Frequency is Changed, Yard A	93
4-4	Mean Time Curves as Frequency is Changed, Yard D	94
4-5	Decreasing Returns to Scale as Frequency is Increased	97
4-6	Mean Time - Cost Trade-Off (Hypothetical)	98

INDEX TO TABLES

Table		Page
1-1	Decline of Railroad Freight Share	12
1-2	Examples of N-Day-%	15
3-1	Characteristics of Yards A and D	50
3-2	Regression Results	52
3-3	Probability Submodel Equations	53
3-4	Correlation Matrices	54
3-5	Corrected Coefficients for the Time of Day Variables	56
3-6	Changes to Independent Variables Needed to Increase PMAKE by 10% at Yard A	59
3-7	Changes to Independent Variables Needed to Increase PMAKE by 10% at Yard D	62
3-8	Changes in Probability as Available Yard Time is Increased	70
3-9	Time of Day Variables as Indicators of Congestion	74
4-1	Trains Selected for the Case Study	79
4-2	Evaluation of Having Train 1931 Depart at 0300	82
4-3	Evaluation of Having Train 1041 Arrive at 2100	83
4-4	Evaluation of Having Train 1621 Depart at 1400	84
4-5	Evaluation of Having Train 1721 Arrive at 1930	85
4-6	Evaluation of Having Train 1521 Arrive at 0930	86
4-7	Summary of Operational Impacts	87
4-8	Summary of Mean Time Curves as Frequency is Changed	95
5-1	Increased Range of Constant Mean Time and Decreasing Total Range of Mean Time	105
5-2	Decreasing Returns to Scale of Mean Time Improvement Under Increasing Frequency	107

INDEX TO TABLES
(continued)

Table | Page

A-1 Sensitivity/Search Results at Yard A........................ 114

A-2 Sensitivity/Search Results at Yard D........................ 115

Chapter 1

Introduction

1.1 Background

1.1.1 Model Choice

A firm with goods to move considers several factors when choosing a mode of shipment. These factors include price, time, reliability, loss and damage, and availability of special equipment. With consideration of these factors, over the past several years shippers have been choosing other forms of transport over rail with increasing frequency (Table 1-1). To regain their customers railways must improve on the areas that affect choice.

Over the past few years railroads have been improving the availability of special equipment,[1] working to improve loss and damage,[2] and decreasing time, where possible, by the use of run-through trains and unit trains. Railroads are not able to fully control the rates charged, as the Interstate Commerce Commission has the power to set rates in order to maintain competition between transport modes. The last factor, reliability, defined as the consistency in the length of time necessary for a car to travel

---

1. "Box Car Problems Meet Shippers at the Door;" Railway Age, 30 Apr. 1973, Vol. 174, No. 8.
2. "L and D Hits the Downward Trail;" Railway Age, 28 May 1973, Vol. 174, No. 10.

Table 1-1

Decline of Railroad Freight Share

Year	% total ton-miles				
	Rail	Highway	Water	Pipeline	Other
1945	67.3	6.5	13.9	12.3	0.0
1950	56.2	16.5	15.4	12.0	0.0
1955	49.5	17.5	17.0	15.9	0.0
1960	44.1	21.7	16.8	17.4	0.1
1965	43.3	21.9	16.0	18.7	0.1
1970	39.7	21.3	16.5	22.3	0.2

Source: American Trucking Trends
1972

between an origin and a destination, remains an area open for improvement and research.[3]

1.1.2 Previous Studies

Research into the area of reliability has focussed on two areas, the impact to the shipper/consignee of poor reailbility and the operations and policies of the railroads themselves that affect reliability. Ainsworth[4] investigated the additional inventory costs to firms caused by poor reliability. He found that the backup inventory needed increases with the trip time through the network, the standard deviation of the trip time, and the accuracy of a firm's own prediction of need. A firm, given equal shipping costs, will choose the mode that will, in its opinion give the best combination of transit time and reliability.

Martland[5], Reid and O'Doherty[6], Folk[7], and Belovarac and Kneafsey[8], investigated the railroads themselves for areas of operation and policy that affect reliability. Martland reviewed parameters to measure reliability of line haul trips from shipper to consignee through the network. Three major parameters were considered, "on-time

---

3. "How to Make Yards Work for Us," Modern Railroads, July 1973, Vol. 28, No. 7.
4. D. Ainsworth, "Implications of Inconsistent Rail Services,: Proceedings 1972. Transportation Research Forum, Vol. XIII, No. 1, 1972.
5. C. D. Martland, Rail Trip Time Reliability: Evaluation of Performance Measures and Analysis of Trip Time Data (M.I.T., Cambridge, 1972).
6. R. Reid, J. O'Doherty, J. Sussman, and A. S. Lang, The Impact of Classification Yard Performance on Rail Trip Time Reliability (M.I.T., Cambridge, 1972).
7. J. F. Folk, Models for Investigating Rail Trip Time Reliability (M.I.T., Cambridge, 1972).
8. Belovarac and J. T. Kneafsey, Determinants of Line Haul Reliability, (M.I.T., Cambridge, 1972).

performance," the N-day-%, and variance of the trip time distribution. The N-day-% is defined as the highest percentage of cars arriving on N consecutive days. (Table 1-2). The N-day-% was chosen over the other two, as "on-time performance" changes with the definition of "on time" and the N-day-% is least affected by the skewness of the distributions, which tend to have long tails on the side greater than the mean. Using this parameter Martland examined railroad data to find origin to destination pairs with poor reliability (i.e., with low N-day-%'s). A further examination of these pairs indicated that the principle area where the poor reliability and the ensuing delay occured was the classification yard with the line haul portion second.

Reid and O'Doherty specifically investigated the classification yard to find what caused the delays and lack of reliability Martland noted. The primary cause was found to be missed connections (i.e. having to wait for a second train). Several factors were found to affect whether or not cars will make a connection. These included cancellations of trains, late arrivals, and a hold/no hold policy. Folk specifically examined, through the use of a simulation, how a hold/no hold policy affects network reliability. A no hold policy was defined as having a train depart as close to the schedule as possible even if more cars for that train are due in the yard just after the train leaves. A hold policy, on the other hand, would wait as long as possible to see if more cars will arrive. If not enough arrive, the train will be cancelled. Folk's simulation suggested that a policy in between strict hold or strict no hold would yield the best reliability for through cars at yards.

Table 1-2

Examples of N-day-%

O-D pair	trip time distribution (days) 1	2	3	4	5	6	6+	2-day-%	3-day-%
1	10%	50	30	0	0	10	0	80	90
2	0	20	30	30	10	10	0	60	80
3	0	0	30	40	30	0	0	70	100
4	0	10	20	20	20	10	20	40	60

Belovarac and Kneafsey investigated the secondary source of delay, the line haul portion of an origin to destination trip. They found that the principle causes of poor reliability were variances in actual running time, actual departure time, and length of stops at intermediate points.

Up to this point the research has been diagnostic. The area and causes of poor reliability have been pin-pointed, but solutions have not. This thesis proposes a model that will allow the impacts on yard performance of changes in network operation to be predicted and evaluated. Implicit in the prediction of the yard impacts is the assumption that improving yard performance will also improve the O-D trip performance as seen by a shipper and as measured by the N-day-%.

1.2 Purpose of Thesis

Past research has found that a primary cause of poor reliability in car movement occurs due to missed connections at freight yards. This thesis proposes a two-part model for predicting the performance of through cars involved in train-to-train connections at railroad freight yards as a function of operating and schedule policies. Through the use of this model, specific policy alternatives for changing railroad operations to improve reliability in freight yards may be investigated. The indices chosen to measure the performance of cars are:

1. the probability of making a train-to-train connection (p(MAKE)), defined as the probability that a car will leave of today's train and not have to wait for tomorrow's.

2. the mean time spent in the yard for all cars making a certain connection.

The first part of the model (first submodel) predicts the probability of making a train-to-train connection based on the operating characteristics of the two trains and the number and type of the cars involved. As p(MAKE) approaches zero, the connection is less reliable. As p(MAKE) nears one, the connection becomes more reliable. The value of p(MAKE) found in the first submodel is an input for the second part of the model.

The second part of the model (second submodel) uses the value of p(MAKE) from the first part to predict the mean time spent in the yard by cars making that connection. A desirable goal is to reduce the mean yard time (or wait) and, at the same time, to increase reliability. This will improve the total transportation service seen by the shipper/consignee. Figure 1-1 shows the relationship of this submodel to the first submodel.

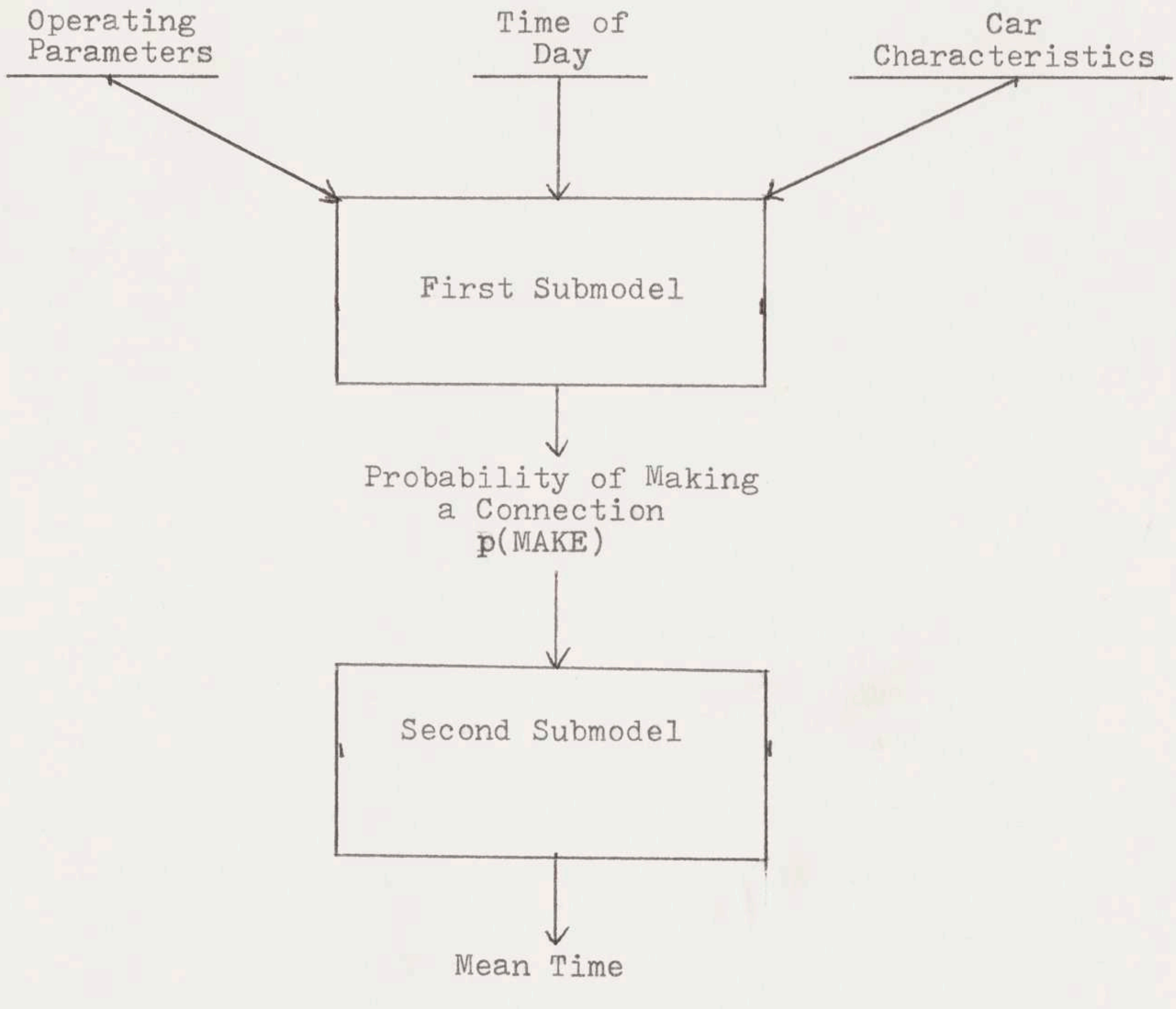
The model proposed in this thesis, then, predicts what might happen to the probability of making (or missing) a connection and to the mean yard time as a railroad changes its operating policies and schedule in the yard or on the line haul portion of a trip in an effort to improve reliability. The prediction of what might occur is the principal application of the model.

1.3 Outline of Chapter Contents

In this chapter the background and introduction to the model proposed in this thesis was presented. One of the characteristics of transportation service a shipper/consignee looks for is high reliability. Because the reliability of railroad service is low, research has been done into the impacts and causes of poor rail

Figure 1-1

Structure of the Model



reliability. This research has been diagnostic, and solutions have not been proposed. The two-part model to be presented was developed to predict what might happen to performance at a freight yard, as measured by the probability of making a connection and the mean time spent at a yard, when changes are made in an attempt to improve reliability.

In Chapter 2 the basis of the model will be presented. Both the line haul and the yard portions of a trip will be examined to determine where delays occur. Based on past research, the causes of delays and missed connections in freight yards will be fully discussed. In the second section the causes of the missed connections will be reviewed again from the view point of specific cars in a specific train-to-train connection rather than general car movement through a yard.

Chapter 3 will present the development of the model form and its calibration on data from two yards. The first section discusses the variables that represent the various causes of delay in a freight yard. In later sections, the form of each of the submodels will be fully developed, the calibration method will be presented, and the source of data and its affect on the model will be mentioned. The calibration of the model for two yards will be carried out with the equation for the probability of making a connection from the first submodel used as the basis for the equation for the mean time of the second submodel. An analysis by differential calculus will be performed on the second submodel to see if for some set of conditions there might exist a minimum mean time. A discussion of the results and interpretation of the calibrations will be presented.

Chapter 4 will present and develop two applications of the model. These are the actual use in a case study and the investigation of improvements in the mean time when the frequency of train operation is changed.

Finally, in Chapter 5 the thesis will be summarized, conclusions will be drawn, and areas for future research will be presented.

Chapter 2

Basis of the Model

2.1 Introduction

In this chapter the underlying basis for the proposed model will be examined. The basis of the model is formed by the previously mentioned research into the reliability of railroad freight movement both on a network and a classification yard scale. The previous research has focused on factors affecting car or train performance and reliability both through freight yards and on the line haul portion of a trip in the aggregate. In the course of this previous research, it was found that the principal reason cars experienced delays and decreased reliability was the missing of connections in freight yards due to cancellations, a hold/no hold policy, late arrivals, congestion, yard policies, and other factors.

The model proposed in this thesis will examine the effect of these various factors on performance of connections between specific train pairs. How these factors affect individual connections, as measured by the probability of making a connection, will be the subject of the second major section of this chapter.

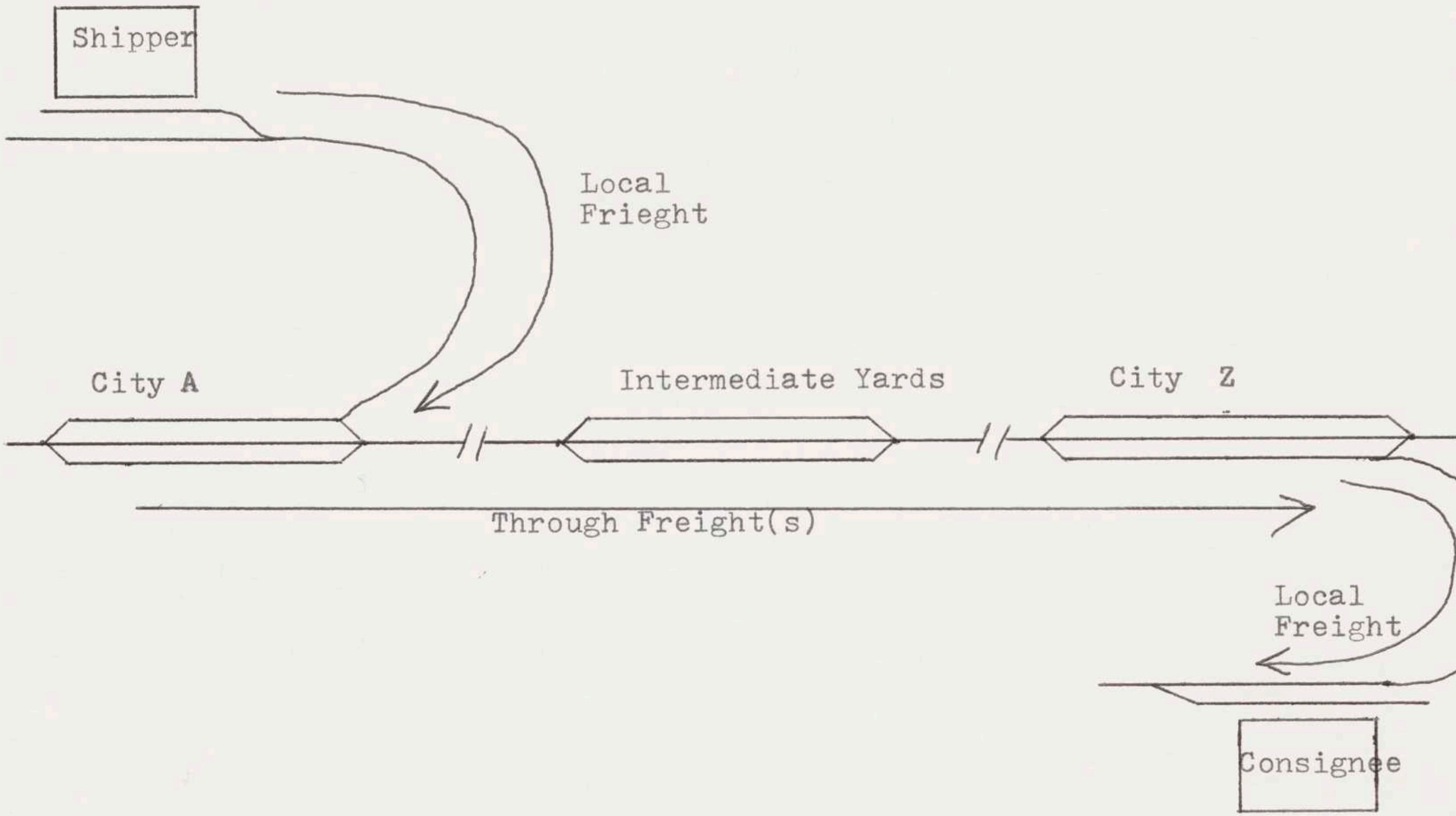
2.2 How Fail Freight Moves

2.2.1 Origin to Destination Trip

A freight car starts its journey (Figure 2-1) from shipper to consignee by being located empty and routed to the

Figure 2-1

Journey of a Freight Car



shipper's siding. When the car is loaded, a local freight which typically operates on a loose schedule picks up the car and carries it and any others it might have picked up to a classification yard. At the classification yard the cars from the local train are sorted by destination in preparation for inclusion in a through freight.

The through freight carries this car and others headed in the direction of the next classification yard. At this next and succeeding classification yards the process of sorting and transferring from train to train may be repeated again and again until the car arrives at the consignee's city. At the destination city a second local takes the car to the consignee's rail siding, where the journey ends.

There are many opportunities for delay in this O-D rail trip making procedure. These include breakdown of equipment, weather, poor track conditions, and no train operating on a given day. The previous studies into the reliability of rail operations[1,2] have shown that the classification yard is a primary source of in-transit unreliability and delay. In the next section the classification yard will be looked at in more detail.

2.2.2 Classification Yard Operations

When a car enters a railroad classification yard in order to be transferred from one train to another rather than enter and leave on the same train, several operations are necessary. When the inbound train carrying the car in question enters the receiving area, the

1. Martland; op. cit.
2. Reid, O'Doherty, et al; op. cit.

road engines and caboose are usually removed, and the brakes are released individually on each car. While this is going on, the consist of the train is compared with the advance information from the previous yard or on-line scanners (ACI), and the paperwork for each car is reviewed to check on the proper outbound block it should be assigned to.

When the brakes are released and the paperwork finished, the cars wait until a yard switcher takes the string of cars and either sorts each car individually (a flat yard) or pushes the cut over the hump (hump yard). In either method, when the proper block assignments are finished, the cars wait again for an outbound train in the proper direction.

At some time, either because a scheduled train is being assembled or added to or because there are enough cars in the same general direction to operate a special train, the car in question along with the others in its block will be placed on that outbound train by a yard switcher. The road engines and caboose will be placed at appropriate ends, the brakes will be reset, and the paperwork transferred. Finally, the car will leave the yard for another yard where the entire operation may take place again.

The yard operation discussed above has many areas which could cause missed connections and delays. These include:

1) Cancellation of the outbound train or block

Many delays in a yard are caused by cancellation of either the outbound train or block.[3] The outbound train could have been

---

3. Ibid, p. 19.

cancelled by lack of motive power, too few cars, lack of caboose, lack of crew, or some other cause. Outbound blocks are cancelled because of capacity or length constraints (see (6) below).

2) Late arrival of the inbound train

A second cause of delay is a late arrival of the inbound train. Unless it has priority or the yard is clear, it will often have to wait before the cars can be sorted.

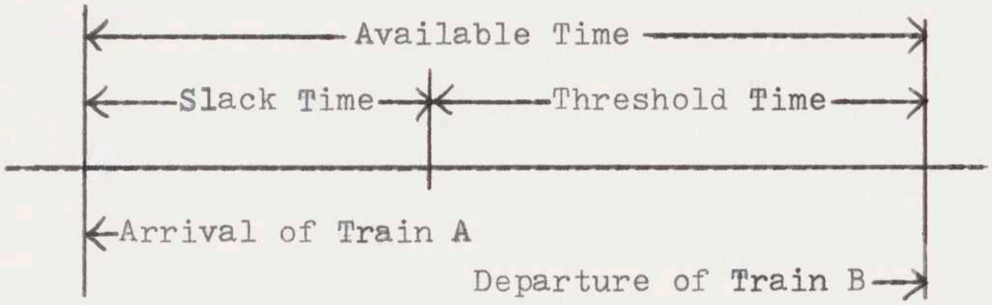
In addition, it is possible that a given actual arrival time is within the threshold time. (See Figure 2-2). The threshold time is the minimum time, even if a car is first priority, between the arrival of a car and its subsequent departure. This minimum time includes sorting time, preparation time for sorting and for departure, outbound make up time, and, depending on the yard, crew and power turnaround requirements. If a car arrives during this period, there is no chance that it will make the connection, as there is not enough time to complete all the operations necessary before the outbound train departs, unless, for some reason, the outbound train is also delayed. One reason for departure delay is a hold policy at a yard.

3) Hold/no hold policy

The hold/no hold policy of the yard also affects the probability of making a connection. If the dispatcher waits for the most tonnage before assembling or sending out a train (a hold policy), incoming cars will have an increased probability of making the connection due to the additional time allowed. A policy that prefers to send out whatever tonnage is available at the scheduled

Figure 2-2

Yard Time Parameters



time rather than wait (no hold), may cause delays, as the threshold time is more strictly observed.

The hold/no hold policy of one yard clearly affects other yards. If train B is held at one yard, it will be delayed arriving at the next yard as train A of an A,B pair. As mentioned above, late arrival of the inbound train is a major cause of delay. Hence, such a policy at a yard or throughout a system would affect network reliability and transit time.

4) Non-uniform handling of cars

Non-uniform handling of cars could be caused by:

a) a priority assigned to a car either by the railroad or external demands.

b) the volume of cars

A high priority car is usually taken first in the sorting procedures. A low priority car waits until all the others have been sorted. The former case means high reliability; the latter, low. A large volume of cars on one connection could interfere with operations at the yard by clogging tracks. Hence, there might be a tendency to move that group of cars first.

5) The time allowed in the yard

It is possible that a given car has too short a scheduled yard time. If the scheduled or available time calls for the inbound cars to arrive during the threshold time, the cars will be delayed, as there is not enough time available to complete the various operations involved in transferring a car from one train to another. As the time allowed increases above the minimum time necessary, additional time is available to make up for late arrivals, yard congestion, etc.

Hence, the probability of making the connection increases. At some point there will be no further increase in probability, as more than enough time is allowed to compensate for the unreliability of arrival and any congestion.

6) Tonnage and length constraints

Various reasons -- grades, curves, poor motive power, and others -- place a limit on the length and weight of a train. Those cars unable to be carried on a train because of such limits are said to miss the connection due to cancellation caused by tonnage and length constraints.

Tonnage and length constraints are also related to the hold/no hold policy. A policy which favors short quick trains will cause a no hold yard policy, as the minimum tonnage or number of cars needed to operate a train can be reached quickly. Conversely, a long or heavy train policy will cause a preference to hold trains, as cars can be added up to an absolute maximum.

7) Congestion

When a car enters the receiving area of a yard there may be other cars ahead of it waiting to be sorted. When the switcher doing the sorting finally gets to this car and the others with it, it may be too late for it to be added to the outbound train. This connection is said to have been missed due to congestion.

8) A cyclic operating policy of the yard

In some cases[4] a yard may switch or hump cars in a cyclical manner, alternating one direction with another. If a train, on

---

4. _Ibid_, p. 34.

schedule or not, arrives during the wrong part of the cycle, it must wait until the cycle shifts, even if connections may be missed.

9) Other miscellaneous causes

Several other miscellaneous causes may force a missed connection. These include repairs, miss-routes, and "no-bills." These occurrences tend to be random and unpredictable.

These several factors influence whether or not all cars, in general, will or will not make a connection. In the next section how these factors affect connections between specific train pairs will be examined.

2.3 Individual Train-to-Train Connection Performance

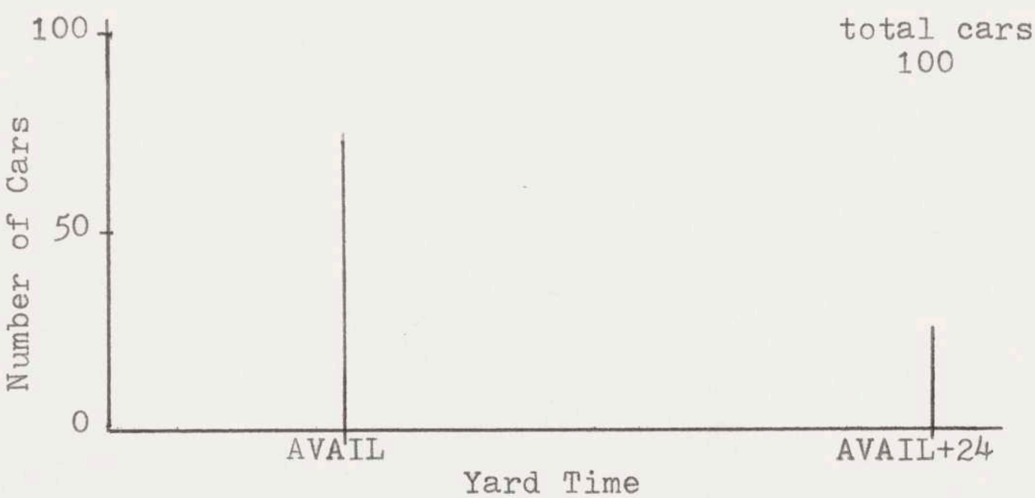
The performance of a connection in this model is measured by the probability of making a connection and the mean time of cars making that connection. Which cars make the connection can be shown through the use of yard time distributions for cars in any specific connection. A car that misses the connection for some reason has a long yard time, while one that makes it has a shorter time. In the hypothetical distributions discussed below and shown in Figure 2-3, what caused the missed connection is not always known, as any one of several factors could cause the same pattern to occur. In addition, the delay caused by the missed connection is assumed to be 24 hours (i.e. until tomorrow's train) because examination of railroad blocking and train make-up books indicates that although there might be two or more trains per day bound for a given destination, each one usually carries different types or groups of cars.

A perfect connection would have a distribution of yard

Figure 2-3

Sample Yard Time Distributions

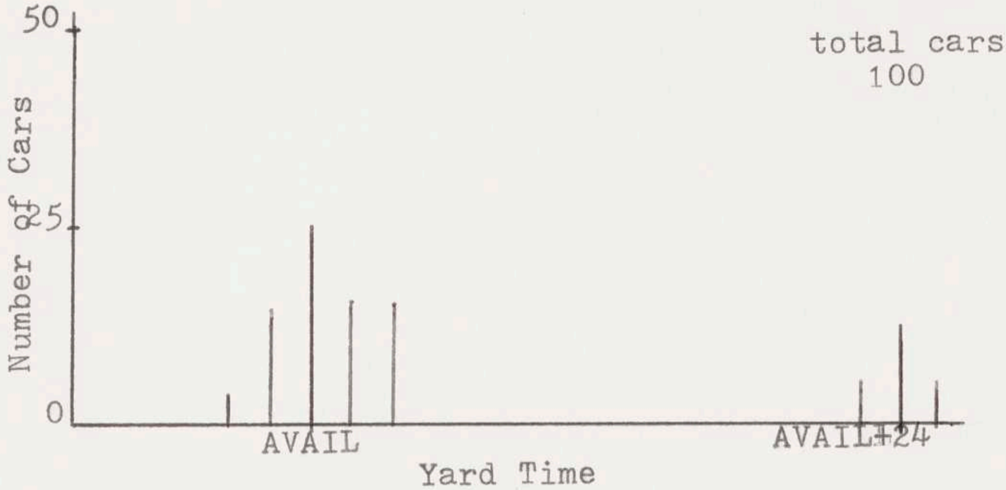
a) Distribution of yard times if all cars make the connection and both trains are on schedule.



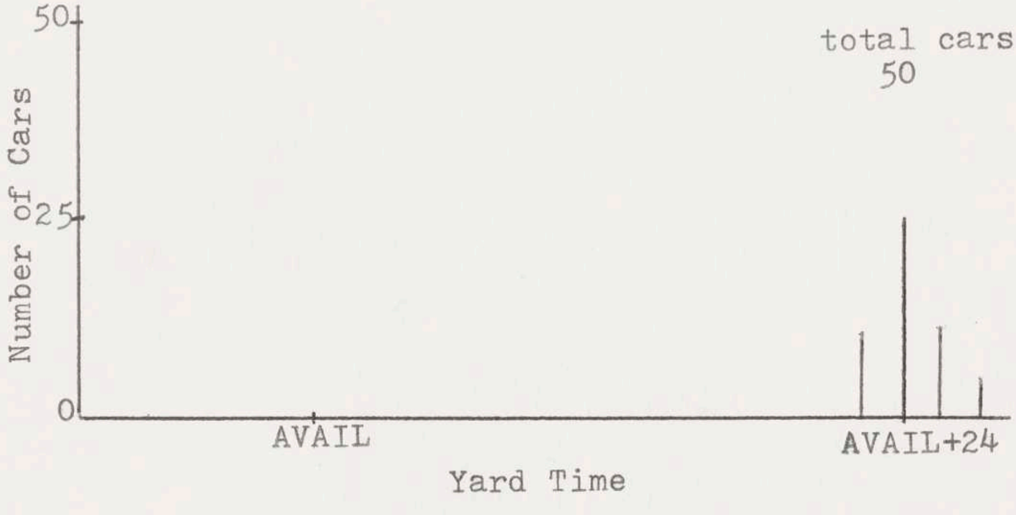
b) Distribution of yard times if train B is cancelled about one time in four.

Figure 2-3

(Con'd)



c) Distribution of yard times if either or both trains A and B run erratically.



d) Distribution of yard times if available yard time is less than the threshold time.

times as shown in Figure 2-3a. All cars make the connection within the time between the arrival of train A and the departure of train B (called the available yard time) and both trains are on schedule.

If cancellations of the outbound train occur, the distribution of yard times will include some cars moving the next day, and, thereby, reduce the number of cars moving on the first try. Figure 2-3b shows an example of a distribution when train B or the block from train A to B is cancelled about 25% of the time but usually runs on schedule otherwise.

The previous examples of distributions were based on perfect schedule adherence. Unreliable operation of trains A or B or both do not necessarily prevent the connection from being made, but instead "spread-out" the yard distribution. Figure 2-3c shows an example of a distribution with cancellations and unreliable operation of trains A and B. The cars in the sub-distribution around the available yard time are considered to have made the connection, even though the distribution is not a single spike.

A scheduled connection time that is too short will appear as a distribution similar to the one in Figure 2-3d. No cars or very few will make the connection in the available yard time. However, all the cars should make the next day's train. On the other hand, a time for a connection that is long because the inbound train is scheduled in too early or because the train arrived during the wrong part of the yard switching cycle would only cause increased available yard time.

The hold/no hold policy also would appear as a distribution similar to the one in Figure 2-3c. A no-hold policy would tend to "tighten" the distribution and it would appear more reliable, while a hold policy would "spread out" the distribution. Of course, as mentioned above, the policy at one yard affects others.

Besides being caused by a hold/no hold policy, a yard time distribution as shown in Figure 2-3c could also occur due to tonnage and length constraints, congestion, or a combination of causes.

Random miscellaneous causes of delay appear as cars moving on the second or third day. This widening of the distribution reduces the number of cars on a given connection that will move within 24 hours plus available yard time.

From the information shown by yard time distributions, an observation of the probability of making a connection can be defined as the number of cars that make the connection on the first day divided by the total number of cars involved. By this definition in Figure 2-3a, the probability of making the connection would be 100% (=1.00); while in Figure 2-3c, it would be 75% (=0.75). Observed values of p(MAKE) computed in this way from actual yard time distributions are the values of the dependent variable in the calibration of the first (probability) submodel carried out in the next chapter.

2.4 Summary of the Chapter

In this chapter the basis of the model was discussed. During an origin to destination trip a freight car may encounter many delays. The primary cause of delays has previously been found to be missed connections at railroad freight yards. Several factors were found to influence whether or not a connection is missed, including cancellation of trains or blocks, schedules, late arrivals, and the hold/no hold policy of the yards passed through.

When observing a specific connection pair, the factors previously mentioned may or may not affect that particular connection. Of the ones that do affect that pair, it is often not possible to tell which one caused the missed connection. Hence, several factors will have to be considered simultaneously. The next chapter develops and calibrates the model to account for these many factors.

Chapter 3

Development and Calibration of the Model

3.1 Introduction

In this chapter the model will be developed and calibrated using variables suggested by the previous sections concerning the basis of the model. The first section will discuss a number of variables that can be used to represent the various factors influencing the probability of making a connection. Both variables that were in fact used and that were not used in the model will be presented. The next section will combine these variables into a mathematical form for each of the submodels which can then be calibrated on the data. The source and type of data on which the model was calibrated will be the subject of the third section. The fourth section will contain information as to

1) The characteristics of each yard
2) The calibration of the equations for the probability of making a connection
3) A brief discussion of the results

After the calibrated probability submodel is presented, the mean time submodel is developed in its own section. The interpretation of the model results as to the implications for railroad policy conclude the chapter.

3.2 Variables

The discussion in Chapter 2 suggests factors which affect connection probability that can be represented by independent variables. This section will examine these independent variables and others which might affect the probability. In addition, the primary dependent variable, the probability of making a connection, will be discussed.

3.2.1 Dependent Variable

The primary dependent variable, the probability of making a connection, PMAKE (=p(MAKE)), is defined as the number of cars that make a connection on the first try divided by the total number of cars involved in that connection.

3.2.2 Independent Variables

In this section independent variables for the model will be discussed. The discussion will first look at the continuous and discrete variables that were in fact used. Following this, some of the variables that were not used will be mentioned.

3.2.2.1 Continuous Variables Included in the Model

1) The average available yard time (AVAIL)

As found by previous work[1], the time a car has to make a connection has an effect on the probability of making a connection. Either the time between the actual or scheduled arrival and departure times could be used to represent the available yard time. Because arrivals and departures are often not on schedule, the difference between the actual mean arrival of train A and actual departure of train B will be used as the average available yard time (AVAIL).

2) Inbound train reliability (SA or SDARR)

The standard deviation of arrival time is used to measure

---

1. Reid, O'Doherty, et al, op. cit.

the reliability of the inbound train. This measure was chosen because previous research by Belovarac and Kneafsey[2] found that the standard deviation of arrival is related to the departure deviation at the previous yard, and the unreliability of the line haul portion of a car's trip. Hence, the standard deviation of arrival stands, in part, for what has occurred previous to a car's arrival at the yard under investigation.

This measure of arrival reliability was chosen in preference to others which could have been used including the % later than schedule and the N-hour-% (defined similarly to the N-day-%), because it was desired to maintain consistency with the work of Belovarac and Kneafsey.

3) Outbound train reliability (SD or SDDPT)

The standard deviation of departure time was chosen as the measure of outbound train reliability. In this use it is, in part, a surrogate for dispatching policy (hold/no hold). A high standard deviation might indicate that a train is often held for more cars, while a low value indicates that it is not held. Although alternate measures similar to the ones for the SA variable could have been used to measure outbound reliability, this measure was chosen for consistency with SA and AVAIL.

4) The average volume of traffic per day involved in the connection (N)

The importance of a connection is measured by the number of cars that pass from train A to train B in a given day as well as

---

2. Belovarac and Kneafsey, _op_. _cit_., pp. 13, 50.

the priority. If no special situation exists, the hypothesis is that the yard will tend to move large groups of cars in preference to small ones. The results of the model will show if this assumption is reasonable.

5) Average length of the outbound train (L)

The length of the outbound train could affect the probability of making a connection in different ways. If a train runs short due to lack of motive power or runs long with low horsepower available per ton, there will be little additional capacity. If a train runs short due to operating policy, there might be additional capacity.

3.2.2.2 Discrete Variables Included in the Model

1) Load or empty (COND)

This variable is included because railroads make a distinction between loaded and empty cars. If they are handled differently, the model will assign a significant value to the coefficient of this variable.

2) Time of day (TA,TB,TC,TD,TE,TF)

The earlier investigations did not address the effect of time of day on the probability of making a connection. Time of day differences might occur due to the number of crews working, the number of cars in the yard (congestion), or yardmaster policy. The six dummy variables, TA to TF, were chosen to represent various sections of the day during which cars might arrive. (Time of departure could have been used instead.)

Time of Day	
TA	0600-1000
TB	1000-1400
TC	1400-1800
TD	1800-2200
TE	2200-0200
TF	0200-0600

The sections were chosen to correspond with the shifts worked by the railroad from which data were obtained (q.v.). Each connection pair was assigned one of the six discrete variables corresponding to the time period in which a majority of cars arrived.

3) Priority (PRI, PA)

The priority of a car, if known, affects its chance of making a given train-to-train connection. Information from the schedule book and discussions with operating personnel identified two special priorities. A high priority was assigned to connections involving mail/express, TOFC, COFC, and new automobiles. These connections were assigned the variable PRI. A low priority was assigned to a certain type of commodity handled by the railroad. The variable PA marks these connections.

3.2.2.3 Variables Not Used in the Model

1) Tonnage

The tonnage of the block of cars involved in a connection and the tonnage of the outbound train could have been used in place of the number of cars in a connection and the length of the outbound train. Because information on the number of cars and the outbound length was more easily available, tonnage variables were not used.

2) Direction of a train

It is possible that a railroad policy might require cars headed, say west, to be moved in preference to cars headed in other directions at certain yards. Because no policy of this type is known to exist on the railroad involved, direction variables were not used.

3) Class of train

The class of trains involved in a connection, express, through, or local, may affect the probability of cars to make connections. However, in this model the variable PRI is used for connections to express trains and the scheduled locals are handled the same as through trains. Hence, discrete variables for class of train are not needed.

3.3 Form of the Model

3.3.1 Probability Submodel

This section will develop and explain a form for the probability submodel which uses the variables put forth in the previous section.

3.3.1.1 Assumptions about the Submodel

Two assumptions were made about this submodel in order to facilitate its use in a case study (see Chapter 4).

1) The submodel is a linear relationship among the variable functions rather than a product form. A linear relationship permits the magnitude of any proposed change to be calculated directly rather

than through the use of elasticities.

2) The variables are independent.

Given the linearity and independence assumptions, a linearly separable model can be used:

$$f(V_1, V_2, V_3, \ldots V_n) =$$

$$f_1(V_1) + f_2(V_2) + f_3(V_3) + \ldots\ldots + f_n(V_n)$$

This form makes it convenient to show the effect of changes in each variable, independent of the other variables.

3.3.1.2 Previous Findings

The previous work on aggregate car performance can be used as a starting point to find the individual functions within the probability submodel.

Reid and O'Doherty[3] show that with respect to available yard time (AVAIL) the curve of the percentage of cars moving "on schedule" (i.e., making a connection) takes on a non-linear shape. (Figure 3-1) One possible curve that might be useful and has been used in both transportation and econometric analyses is a logit curve.[4,5]

$$PMAKE = 1 / (1 + EXP(-G))$$

$$\text{where } G = b(AVAIL + a)$$

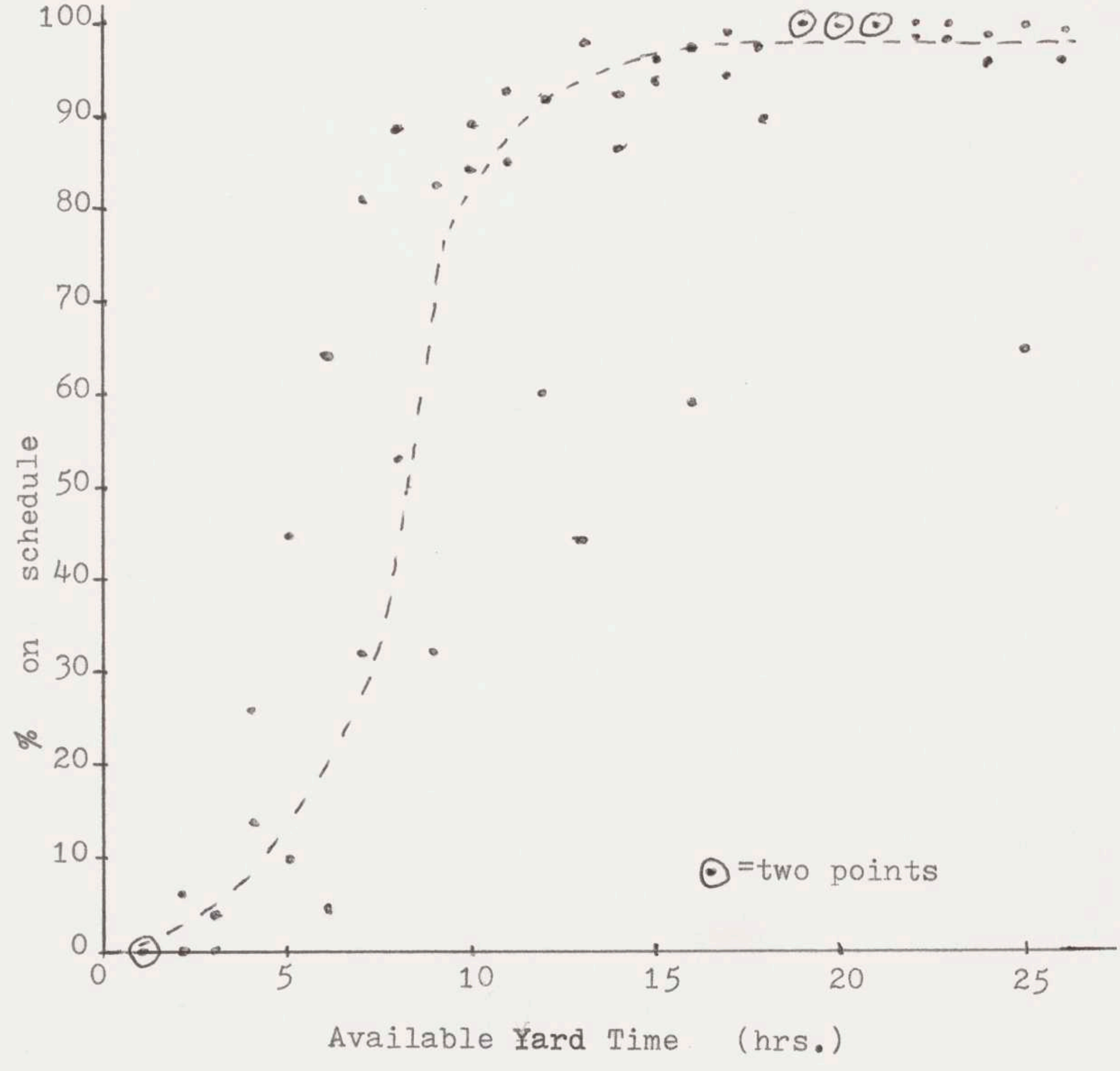
The same work by Reid and O'Doherty[6] also suggests that the

---

3. Reid, O'Doherty, et al, p. 28 ff., Appendix A.
4. See M. Ben-Akiva, Structure of Passenger Travel Demand Models, Ph.D. Thesis, (M.I.T., Cambridge, 1970) for an advanced example.
5. H. Theil, Principles of Econometrics (New York, 1970), pp. 628 ff.
6. O'Doherty, et al, op. cit., p. 20.

Figure 3-1

Cars Moving on Schedule
vs.
Available Yard Time



Source: Reid, O'Doherty, et.al.;
Tables A-XII and A-XVI

mean time in the yard generally decreases as the number of cars involved (N) increases. In the range of interest of this model, N = 2 to 20, the change is nearly linear. For this to happen in the context of this model, the probability of making a connection must decrease as N increases in a linear manner.[7]

The previous works did not directly investigate the effect of inbound or outbound reliability, outbound train length, time of day, or priority on the connection probability. This model will extend the previous work by adding those variables.

3.3.1.3 Functional Form of the Probability Submodel

Based on the previous discussions, the following equation for the probability submodel is reasonable:

$$PMAKE = k_o + k_1 * f_1(AVAIL) + k_2 * f_2(N) + k_3 * f_3(SA) + k_4 * f_4(SD) + - - - - -$$

or

$$PMAKE = k_o + k_1 * f(AVAIL) + f(R)$$

where $f(AVAIL) = 1 / (1 + EXP(-G))$

$G = b(AVAIL + a)$

and all other relationships are assumed to be linear

k_i = regression coefficients

This form is not necessarily the best or the only form the submodel could take. However, this form is linearly separable as well

7. See the following discussion on the form of the mean time submodel.

as allowing the prior information with respect to $f_i(V_i)$ to be used.

3.3.2 Mean Time Submodel

The mean time submodel is based on the assumption that cars that do not make the first outbound train depart on the next train. If trains run daily, this would be tomorrow's train, 24 hours later. The mean time is, then, simply the weighted average yard time of the cars that go today and cars that wait 24 hours.

Mean Time = MT = AVAIL + 24*(1-PMAKE)

PMAKE from the first submodel

Let 1 - PMAKE = PMISS

Mean Time = AVAIL + 24*PMISS

The value of AVAIL that minimizes the mean time can be found by differentiation of MT with respect to AVAIL when PMAKE is calibrated.

3.4 Data Used

Data for the calibration of this model was taken from the daily yard reports of a major U.S. railroad[8] for two weeks in late November-early December, 1972. The reports consisted of records for each of several thousand cars entering several yards on Railroad C. Each record consisted of the car identifier; an inbound train arrival time, day, and number; an outbound train departure time, day, and number; the time spent at the yard; and whether loaded or empty. These data were then sorted by inbound-outbound pair, day and time,

8. The confidentiality of these data was required by that railroad. Hence, this railroad will be known as Railroad C and the two yards as Yards A and D.

loaded or empty. Analyses of these sorts yielded inbound train times and arrival variances, outbound train times and departure variances, and outbound length as well as the distribution of yard times for the loaded and empty cars involved in each individual connection pair.

3.4.1 Limitations of the Data

The data contained three factors that affect the model. First, the assumption was made that cars whose yard times are less than three (3) hours are either 1) arriving and departing on the same train and are, therefore, not really making a connection, or 2) being "block switched" to avoid the delays of the sorting procedure. These last cars would be of an extra high priority. Hence, these cars are not included in the data set. This means that the range of AVAIL, which can be at most 24 hours due to the cyclical nature of railroad operations, should not start at 0 hours, as there would be no record of any cars in the first 3 hours. From the work of Reid and O'Doherty[9] a car has almost no probability of making a connection in less than 3 hours. Hence, 3 hours will be assumed to be the minimum threshold time.[10] This changes the range of values of AVAIL from the spread 0 to 24 hours to the spread 3 to 27 hours, and a value of AVAIL less than 3 hours will be considered AVAIL + 24 hours.

Second, the data does not indicate which train a car should have departed on.[11] Because of this, it is impossible to tell

---

9. See Figure 3-1.
10. This also agrees with the personal observation of this author at Yard D, where the threshold time is approximately 2-3 hours.
11. There are certain special instances where proper train can be ascertained from external information. In those cases corrections were applied.

the total number of cars that should make a connection. What is known is how many cars actually made the connection train A to train B, and through the use of yard time distributions whether or not the connection was made on the first or second (or third or more) try. Hence, an observed value of p(MAKE) has to be considered: the number of cars that make the connection on the first try divided by the total number of cars that actually make the connection (rather than the total number that should) as seen over the two week period.

Third, the observed value of PMAKE is an estimate of the true value, P. Because making a connection will occur with true probability P and not occur with true probability 1-P, making a connection can be considered a Bernoulli (binary, binomial) process. Hence, the mean is always the true value, P. However, the variance of the estimate of P decreases in proportion to the number of observations of occurrence (n).

Variance of mean = P * (1-P) / n

Std. deviation of mean = SQRT (P * (1-P) / n)

Due to the small sample size, the standard deviation of the mean of the average connection is high relative to the range of PMAKE.

Yard A: P = 0.62 ; n = 11.6 for average connection

std. dev. ave. connection = .14

Yard D: P = 0.82 ; n = 11.4 for average connection

std. dev. ave. connection = .11

The standard deviation of the mean of the average connection is close to the lower bound on the standard error of the calibrated model.[12] Hence, the 95% confidence interval on the value of PMAKE predicted by either probability submodel will be at least 50% of the total range of PMAKE.

3.4.2 Observation of a Connection

The processed yard report data contains the following information for each train pair between which cars actually made a connection:

1) The total number of cars that actually made the connection over 14 days.

2) The number of cars that made the connection on the first try.

3) The mean and standard deviation of the arrival time of the inbound train.

4) The mean and standard deviation of the departure time of the outbound train.

5) The average length of the outbound train.

6) Whether the cars are loaded or empty.

7) The number of times the connection took place.

These seven items yielded values for the independent variables, AVAIL, SA, SD, L, N, TA to TF, and COND, and the dependent variable PMAKE, for one train pair. A value of each of these variables together with whether or not a connection pair is priority or contains cars carrying the special commodity (a value for PRI and PA), obtained from external sources, constitutes one observation.

Several observations were excluded from the data used for the calibration of the models. These included -

12. This occurs because the submodel finds an average of the estimators.

1) Connections with less than 14 total cars.

It was decided that a connection that does not involve at least an average of one car per day over the two week period can not be a regular connection, and, therefore, not important for this model.

2) Connections occurring less than four times.

The error inherent in these observations is large relative to even the average error of the entire data. The standard deviation of an individual connection is also calculated using the Bernoulli process, and, if n is less than four, the confidence interval on that observation is almost double that of the average for the entire data. This would destroy the assumption that data has the same variance throughout made in order to use least squares estimation.

3) Connections for which the correct number of times of occurrence could not be determined.

Connections in this category involved trains without a schedule (published or de facto). Trains that usually operate with no schedule include local industry freights, interchange runs, and trains that operate only when enough traffic is available.

The exclusion of these connections places the emphasis of the model on the prediction of what might happen to the reliability of cars passing through the yard under investigation rather than on cars beginning or ending their journey on Railroad C.

3.5 Calibration of the Probability Submodel for Two Yards

The model was calibrated[13] for two hump yards on Railroad C,

---

13. For the calibration method see Appendix A.

Yards A and D. Table 3-1 gives data on the physical and operational characteristics of the two yards.

Observations

Yard A: 221 train pairs involving 11193 cars
Yard D: 225 train pairs involving 13984 cars

Dependent Variable

PMAKE = the probability of a car making a connection from train A to train B

Independent Variables

AVAIL = average available yard time between the arrival of train A and departure of train B (hrs.) (If less than 3, use AVAIL = 24 + AVAIL)

Q = 1 /(1 + EXP(- b(AVAIL + a)))
a and b found by sensitivity procedure (Appendix A)

N = average number of cars per block per day moving from train A to B

L = average length of outbound train B (100's of cars/day)

SA = standard deviation of the arrival of train A (hrs.)

SD = standard deviation of the departure of train B (hrs.)

COND = 1 if loads, 0 if empties

PA = 1 if special commodity cars, 0 otherwise at Yard A
= 0 at Yard D

PRI = 1 if a priority car, 0 otherwise at Yard A
= 0 at Yard D

TA to TF = 1 if cars arrive in time period, 0 otherwise

Variable	Time Period
TA	0600-1000
TB	1000-1400
TC	1400-1800
TD	1800-2200
TE	2200-0200
TF	0200-0600

Table 3-1

Characteristics of Yards A and D

	Yard A	Yard D
Number of Classifications	64	40
Number of Train Arrivals	>34/day	>18/day
Number of Train Departures	>34/day	>18/day
Average Cars/Day Arriving	2900	1900
Distribution of Car Arrivals		
0600 - 1000	14%	15%
1000 - 1400	23	19
1400 - 1800	25	13
1800 - 2200	9	12
2200 - 0200	5	28
0200 - 0600	23	13
Number of Crews (all shifts)	5	4

Source: Operating and schedule data provided by Railroad C.

Regression Coefficients

The regression results are shown on Tables 3-2 and 3-3.[14]

3.5.1 Validity of the Calibrated Submodels

When the probability submodel form was developed, two explicit assumptions were made:

1) The model is a linear relationship.

2) The variables are independent.

The first assumption seems justified, as the residuals were found to be distributed normally with mean zero and, except for one case, lacking any particular pattern which would suggest that either a different functional form (other than linear) or weighting is necessary. The one exception occurred with respect to the variable AVAIL. The residuals had a greater variance at low values of AVAIL, which is what the Bernoulli process suggests.[15] This would indicate that weighting with respect to a function of AVAIL could reduce the standard error. However, weighting was not used in order to keep the model simple.

The second assumption also seems justified, as the correlation coefficients are near zero (Table 3-4). However, there are two

---

14. The regression program contained in the Statistical Package for the Social Sciences (SPSS) on an IBM 370/165 was employed.
15. The variance of a Bernoulli process is greatest when P = .50 for any n, and PMAKE is nearer .50 at low values of AVAIL than high. There might also have been a problem with unequal variances in the observations due to different numbers of cars or occurrences which made up each one. However, the residual pattern of the unweighted submodel did not indicate that weighting due to unequal number of members of the observation group was necessary.

Table 3-2

Regression Results

Variable	Yard A Coeff. (std err)(F)	Yard D Coeff. (std err)(F)
a	8.5	5.4
b	.15	.60
f(AVAIL)=Q	1.00 (.091)(120)	1.00 (.052)(361)
SA	-.021 (.014)(2.2)	-.034 (.010)(10.5)
SD	.060 (.023)(6.8)	-.030 (.013)(5.5)
L	-.315 (.074)(18.3)	.211 (.048)(19.3)
N	.016 (.004)(14.0)	.005 (.002)(4.4)
COND	*	*
PRI	.417 (.076)(30.1)	
PA	-.270 (.045)(36.5)	
TA	*	*
TB	*	-.050 (.028)(3.3)
TC	*	*
TD	.100 (.066)(2.3)	*
TE	.229 (.069)(10.9)	-.037 (.025)(2.2)
TF	*	*
Constant	.055	-.136
R^2	.54	.66
Std. Err.	.254	.158
F	26.3	60.1
Degs. of Freedom	(9 , 201)	(7 , 217)

* = fails significance test at 95% confidence, coefficient assumed = 0.000.

Table 3-3

Probability Submodel Equations

Yard A

$$PMAKE = .055 + (1 / (1 + EXP(-.15(AVAIL - 8.5)))) -.021SA + .060SD - .315L + .016N -.270PA + .417PRI +.100TD + .229TE$$

$$0 \leqq PMAKE \leqq 1$$

Yard D

$$PMAKE = -.136 + (1 / (1 + EXP(-.60(AVAIL - 5.4)))) -.034SA - .030SD + .211L + .005N -.050TB - .037TD$$

$$0 \leqq PMAKE \leqq 1$$

Table 3-4

Correlation Matrices

	AVAIL	SA	SD	L	N	PRI	PA	
AVAIL	1.00	.09	.11	.12	-.07	-.27	.13	
SA		1.00	.11	.05	-.04	-.14	-.23	
SD			1.00	.44	.07	-.40	-.28	
L				1.00	.09	-.52	-.03	Yard A
N					1.00	.00	.17	
PRI						1.00	-.16	
PA							1.00	

	AVAIL	SA	SD	L	N	
AVAIL	1.00	.03	-.10	-.05	-.22	
SA		1.00	-.06	-.16	-.11	
SD			1.00	.58	.12	Yard D
L				1.00	.15	
N					1.00	

areas of possible problems. The first is with the discrete variables TA to TF at both yards, and the second is with certain of the major variables at both yards.

The problem with the discrete variables TA to TF comes from the fact that, as a set, they are correlated with the constant. This correlation can be removed earily by adding the average of their coefficients to the constant and subtracting that average from each coefficient. This was done, and the results are presented on Table 3-5.

The second area of possible lack of independence concerns the variable PRI with the variables SD and L at Yard A and the variable SD with L at both yards. These cases are not considered serious and are discussed in Appendix D.

In addition to the two explicit assumptions, one implicit assumption was made -- the calibrated submodels would be significant. This assumption also seems justified because comparison of the F statistic for each submodel with the value of F for the appropriate degrees of freedom at the 95% or 99% level reveals that the null hypothesis (i.e. the model is insignificant) can be rejected.

3.5.2 Application of the Probability Submodels to Typical Connections at Each Yard

In this section values for the independent variables will be substituted into the submodel equations for each yard in order to examine the effect of changes in various parameters on the probability of making a typical connection. At Yard A the effect of priority will be examined, while at Yard D the effect of changes in operation and block size will be shown.

Table 3-5

Corrected Coefficients for the
Time of Day Variables

Variable	Yard A	Yard D
TA	-.055	.013
TB	-.055	-.037
TC	-.055	.013
TD	.045	.013
TE	.174	-.024
TF	-.055	.013
Constant	.110	-.150

Yard A

The first example for Yard A (Curve A, Figure 3-2) shows a typical connection:

1) Average length of the outbound train is 100 cars. (L = 1)

2) About 5 cars per day make the connection. (N = 5)

3) The standard deviation of arrival is 2.5 hrs. (SA = 2.5)

4) The standard deviation of departure is 2 hrs. (SD = 2)

5) The train arrives about 2000. (TD = 1)

6) The cars have no particular priority or commodity.

Curve B indicates what would happen if, instead of no particular commodity, the cars from the typical connection carry the special commodity identified by the discrete variable PA. These cars are handled with considerably less reliability.

On the other hand if the cars of the typical connection are priority cars being transferred to an express train, they are handled more reliably than the average car. If AVAIL is greater than 11 hours, these cars will move for certain on the next train.

Table 3-6, which is developed from the submodel equation and the example curves, summarizes the impacts of the various independent variables at Yard A.

Yard D

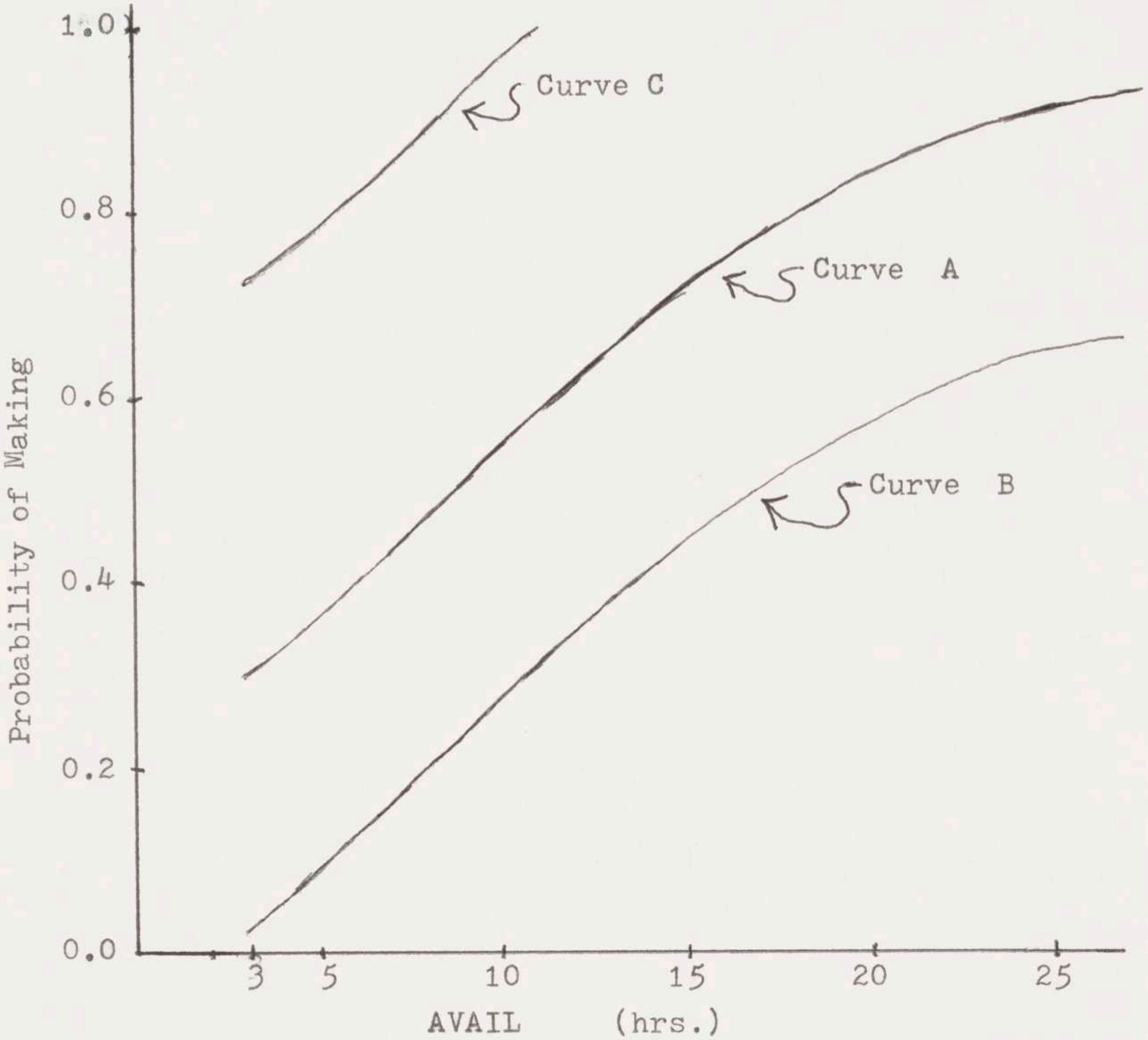
The first example at Yard D (Curve D, Figure 3-3) shows the curve for a typical connection:

1) Average length of the outbound train is 90 cars per day. (L = .90)

2) About 6 cars per day are to make the connection. (N = 6)

Figure 3-2

Sample Probability Curves at Yard A



Curve	L	N	SA	SD	PRI	PA	TD
A	1	5	2.5	2.0	0	0	1
B	1	5	2.5	2.0	0	1	1
C	1	5	2.5	2.0	1	0	1

Table 3-6

Changes to Independent Variables
Needed to Increase PMAKE by 10%
at Yard A

Variable	AVAIL =	3 hr.	9 hr.	15 hr.
AVAIL		+3 hr.	+2.67 hr.	+4 hr.
SA		decrease by 5 hr.[1]		
SD		increase by 1.5 hr.[2]		
L		reduce train length 33 cars		
N		add 8 cars to the block		

[1] This change is not practical, as average SA ≈ 2.5.

[2] This can be accomplished by holding the outbound train more often.

3) Standard deviation of arrival is about 3 hours. (SA = 3)

4) Standard deviation of departure is about 2 hours. (SD = 2)

5) The cars arrive about 2000. (TD = 1)

Curve E shows the effect of doubling the size of the group of cars making a connection (from 6 to 12). The model predicts an increase in yard reliability.

If the arrival and departure reliabilities improve, there is also an increase yard reliability predicted by the model. This increase is shown by curve F.

In a similar fashion to Table 3-6, Table 3-7 summarizes the impacts of the variables at Yard D.

Even though for Yard A, only changes in priority were illustrated, changes similar to those of curves E and F relative to curve D will occur at Yard A because the effects of the change of any one variable can be linearly added to any other. This is a result of the linear separability design of the model.

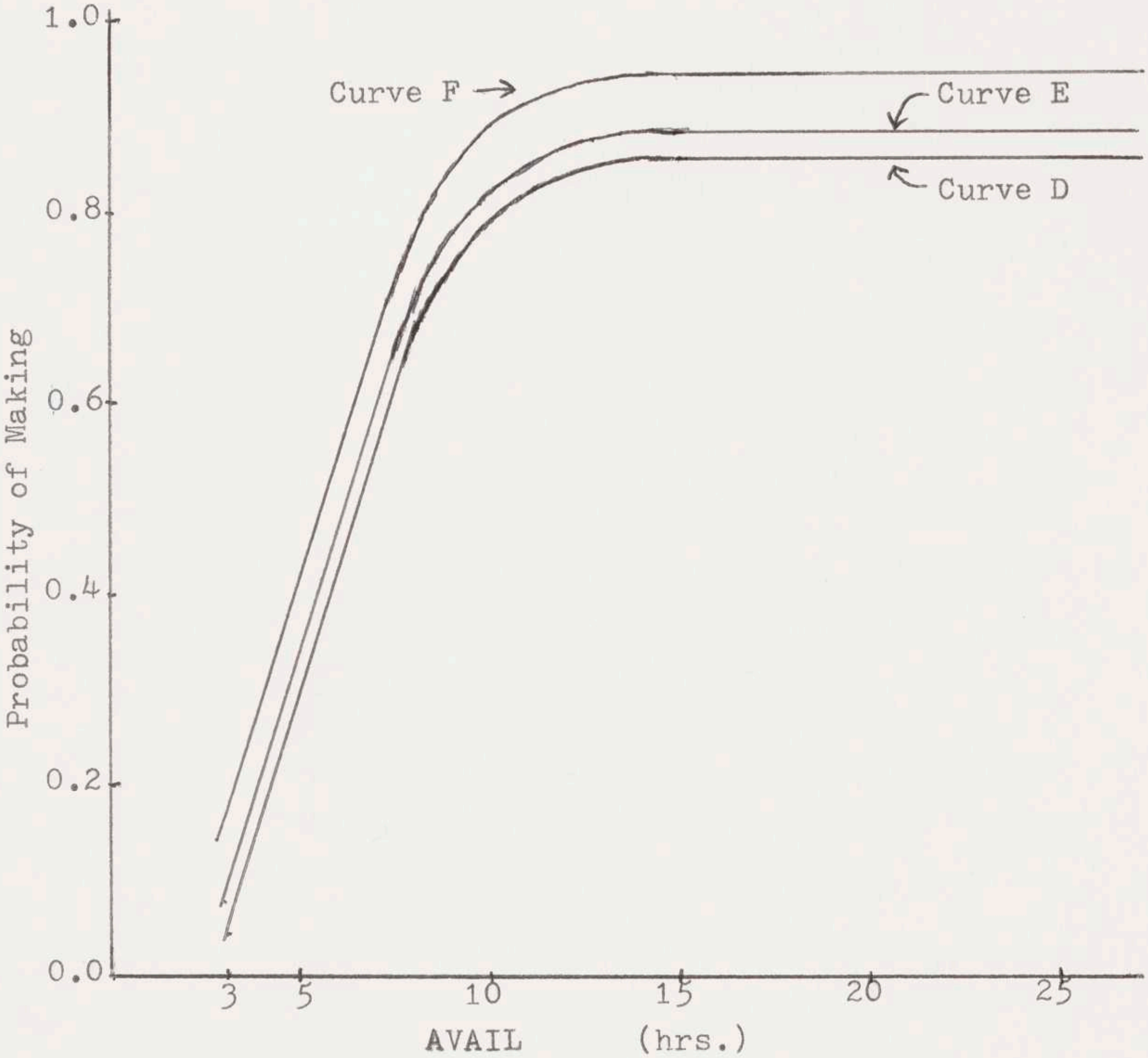
It is apparent that different changes affect the reliability in different ways. However, before discussing trade-offs between changes, the mean time submodel must be presented.

3.6 Mean Time Submodel

The mean time submodel does not require calibration, as it is simply the weighted average or expected value of the time the cars involved in a connection remain in the yard. The assumption is made that the penalty for missing a connection is 24 additional hours wait over the available yard time until tomorrow's train when these cars will leave. So long as train frequency is 1/day, this is not unreasonable,

Figure 3-3

Sample Probability Curves at Yard D



Curve	L	N	SA	SD	TD
D	.9	6	3.0	2.0	1
E	.9	12	3.0	2.0	1
F	.9	6	1.0	1.0	1

Table 3-7

Changes to Independent Variables
Needed to Increase PMAKE by 10%
at Yard D

Variable	AVAIL =	3 hr.	9 hr.	15 hr.
AVAIL		+1 hr.	+6 hr.	-[1]
SA		decrease by 3 hr.[2]		
SD		decrease by 3.3 hr.[2]		
L		allow 50 extra cars per train		
N		add 20 cars to the block		

[1] Can not be done.

[2] Although either change might be accomplished, changing both is more practical.

as any new cars for that connection go to the end of the queue in which the cars that missed are at the front. Therefore:

MT = Mean Yard Time = AVAIL + 24* (1 - p(MAKE))

The value of p(MAKE) is found from the probability submodel and through it the influence of reliability, car characteristics, etc., are brought to bear on the time in the yard.

An important output from the mean time submodel is whether or not a minimum mean time exists for some time allowed in the yard. In this discussion the examples presented in section 3.5.4 will be used to illustrate sample mean time curves.

An analysis by differential calculus (Appendix C) reveals that a maximum or minimum point as defined by the slope of the curve equal to zero does not exist for the parameters of the probability submodel at Yard A. Plotting MT vs. AVAIL for the three examples (A,B,C) used in section 3.5.2 reveals what does occur. (Figure 3-4). The curve passes through a flex point. (Ignoring curve C for the moment). In the range of AVAIL from 5 to 10 hours the curves are "flat." Between 3 and 15 hours, a 12 hour range, the change is less than 2 hours.[16] The lowest point occurs at the boundary, AVAIL = 3.0.

In contrast to Yard A, analysis of the mean time submodel for Yard D reveals the existence of maximum and minimum points as defined by having the slope of the curve equal to zero. Figure 3-5, a plot of MT vs. AVAIL for the three examples (D,E,F) of section 3.5.2 within the range of interest, shows that the value of MT at AVAIL = 9.6 is a minimum. In the 6 hour range between AVAIL = 7.2 to 13.2 hours the difference between MT and the minimum MT is less than 2 hours. In

16. A two hour change will be considered insignificant relative to the entire journey, as two hours is less than 10% of a day.

Figure 3-4

Mean Time Curves at Yard A

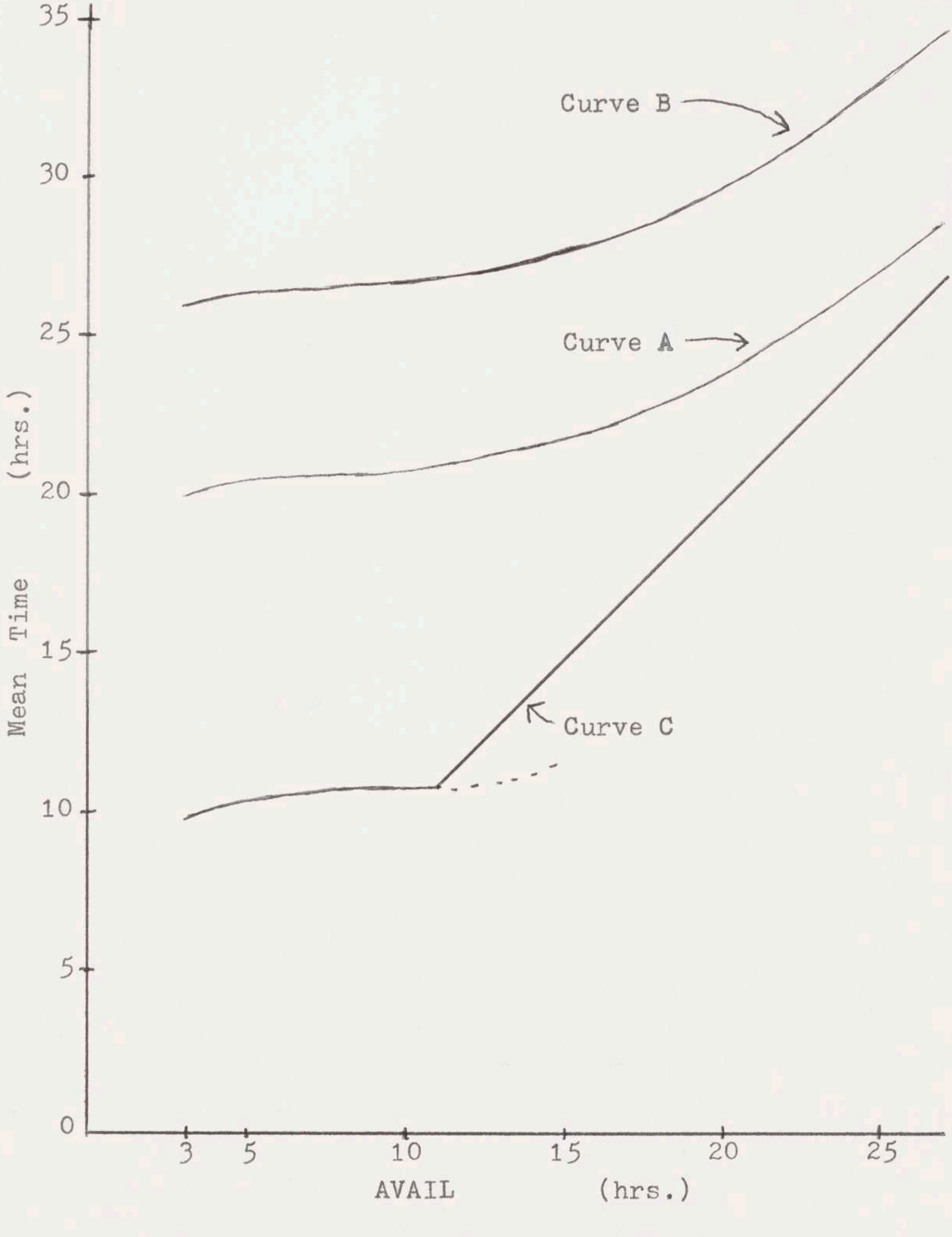
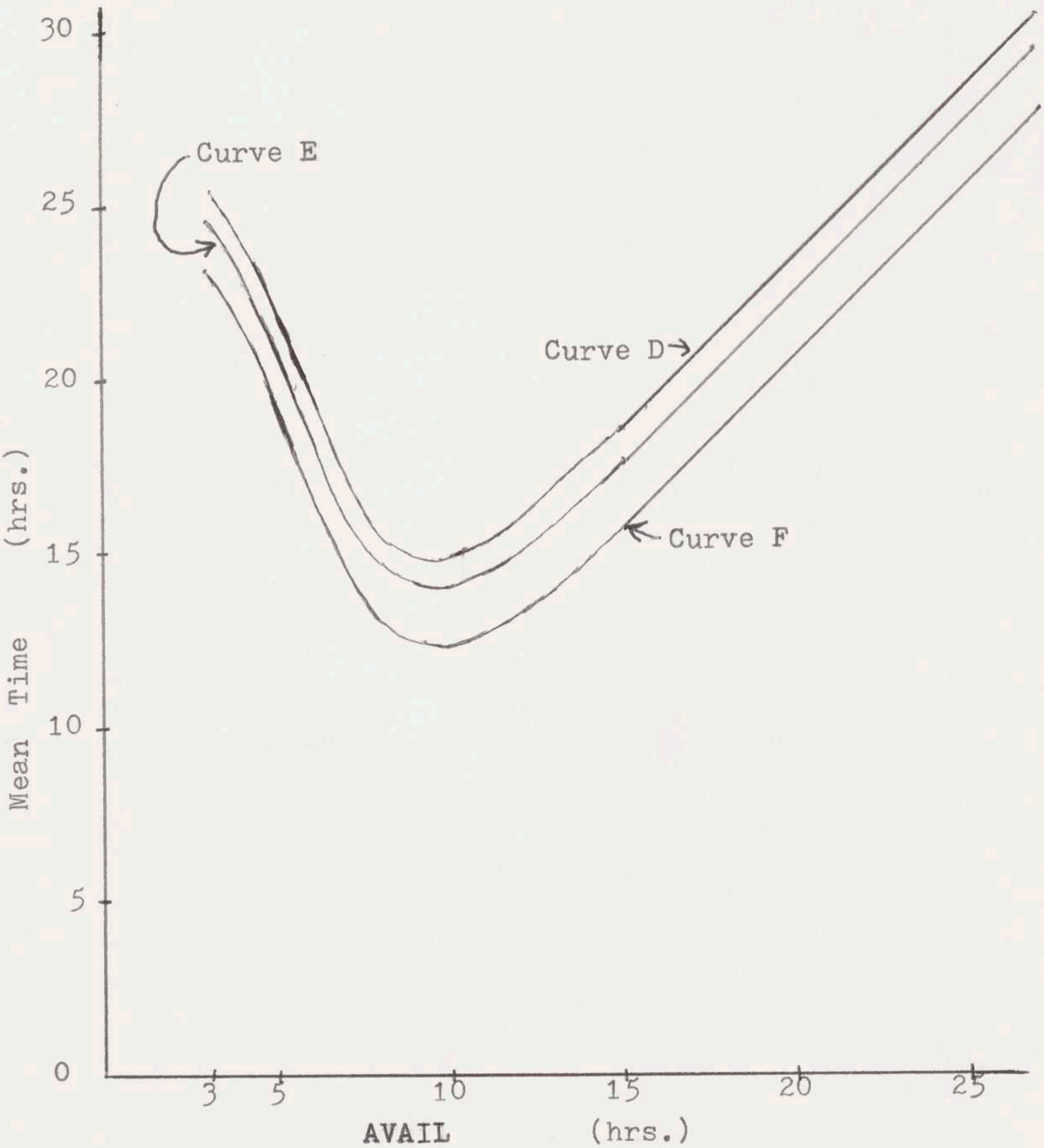


Figure 3-5

Mean Time Curves at Yard D



comparison the range for Yard A was 12 hours.

Within the set of curves for each yard, although the magnitude may change, the shape of the curves does not change for different sets of conditions as long as the first derivative of the probability function is continuous. The first derivative becomes discontinuous when the function tries to cross the upper bound of 1. This occurs in example C. In the range of AVAIL where the probability is bound at 1, the first derivative of PMAKE = 0 and the slope of MT = 1. Hence, MT = AVAIL. However, in the range of AVAIL where the function is not at the boundary, the mean time curves do show the similar shape. The similar shape occurs because the model is constructed with independent first derivatives. Hence, the derivative of any combination of values for PMAKE with respect to AVAIL depends only on the form of f(AVAIL) which is a base curve for each yard.

The implications of these particular forms of the MT curve will be discussed more fully in the following section.

3.7 Interpretation of the Model

3.7.1 Implications for Railroad Operations

The equations of the first submodel express the probability of making a connection at Yards A and D on the basis fo the operating characteristics of the connection pair trains, the time of day, and the specific characteristics of the cars involved. Of the train operating and car characteristics, the most important, for both yards, are available yard time (Q(AVAIL)), volume of traffic (N), and the reliability of arrival (SA) and departure (SD). Combining these characteristics with the results of the mean time submodel reveals several areas where the model suggests improvements and trade-offs in railroad operations. These areas include scheduling policy, line haul operations, and blocking.

1) Scheduling policy

Scheduling policy is the principal area where improvements and trade-offs are suggested by the model. Scheduling policy involves rescheduling trains in an attempt to allow an appropriate value of available yard time for a connection. An appropriate value of available yard time can be found from the mean time submodel, although it may not always be possible to achieve this.

There are two cases of the mean time submodel which occur in the calibrated models. The first case is where a minimum mean time exists for some value of available yard time. Yard D exhibits this behavior. In the other case, there is no minimum mean time as found by having the first derivative equal to zero, but instead, a "flat" spot caused by a flex in the curve, with the lowest value occurring at the boundary.

When a minimum exists in the mean time curve for some value of available yard time (=A, the most efficient yard time), there are two regions of interest, less than and greater than A. If the actual available yard time is less than A, an increase in the available yard time by a schedule change would improve both reliability, as p(MAKE) would increase, and mean yard time, as MT would decrease. If the actual available yard time is greater than A, an increase in available yard time would only improve reliability, as p(MAKE) would increase. This increase, however, would come at the expense of mean yard time. Hence, a good policy for such a yard would be to arrange the connections involving many cars every day in such a way that the available yard time is greater than or equal to the efficient yard time. This would insure the highest reliability for the time spent in the yard. Secondary connections could then be arranged to have available yard times near the most efficient time. Any other or one time connections would be handled as necessary. Following this policy would tend to minimize the wait of cars in the yard and improve reliability, as important connections would be given enough time to occur with high reliability.

When a minimum in the mean time curve does not exist for some value of available yard time, there are also two regions, within the range of the flex and the rest of the curve greater than that range. If the actual yard time is within the range of the flex, the reliability may be increased without greatly changing the mean time. If the actual yard time is in the region beyond the range of the flex, an increase in yard time will increase reliability, but it will occur at the expense of mean time. A scheduling policy in this case would be to arrange as much time as possible within the range of the flex for the major connections and let the others occur as required.

2) Line haul operations

There are three areas where improvements in line haul operation are suggested by the model, consistent operations, schedule adherence, and train speed. Consistent operation is important. At both yards a decrease in the reliability of arrival decreases the probability of making the connection. This decreasing effect on the probability could be as much as 10% for connections involving trains with poor arrival reliability. (Tables 3-2, 3-3, 3-6, and 3-7). The reliability of departure, as indicated by the calibrated probability submodels, affects the probability of making a connection differently at each yard. At Yard A it would seem that trains are often held for cars, as the decrease in reliability, obtained by increasing SD one hour, increases the probability 6%. At Yard D. the opposite occurs - the same increase in SD will decrease the probability 3%.

However, even though the unreliability of departure at Yard A aids significantly the probability of making a connection, it is not good because there is a network involved. A train that leaves erratically at one yard may arrive erratically at the next yard. So although delaying a train at the first yard may aid cars at that yard, it could hinder operation at the next yard by decreasing the arrival time reliability.

In the previous discussion of arrival and departure consistency, no mention was made of schedule adherence, as the standard deviation used in the model was taken around the actual mean arrival or departure time. Analysis of operating data has shown[17] that trains usually arrive and depart late relative to the scheduled time. If trains could arrive earlier or even on time, there could be an increase in available yard time without actually changing the schedule. At values of available yard time less than about 15 hours at Yard A and about 11 hours at Yard D, there are increases in the probability of making a connection for each additional hour of available yard time (Table 3-8) which are greater than any

---

17. J. R. Folk, Some Analysis of Railroad Data (M.I.T., Cambridge, 1972).

Table 3-8

Changes in Probability as
Available Yard Time
Is Increased

AVAIL	Change in Probability[1] Yard A	Yard D
3 hrs.	base	base
4	3.3%	10.9%
5	3.4	13.8
6	3.6	14.7
7	3.7	13.3
8	3.7+	10.2
9	3.8	7.0
10	3.7+	4.3
11	3.7	less than 3.0%
12	3.6	"
13	3.4	"
14	3.3	"
15	3.1	"
16	less than 3,0%	"
⋮		
27	"	"

[1] Change from previous hour.

decrease that would be caused by SA being increased by 1 hour.

For a train to arrive on time or even early, especially if it departs late, train speed might have to be increased. For example, on the 250 mile run from Yard D to Yard A the average train speed is in the order of 15 mph for a running time of about 16.7 hours. If average train speed could be increased to 20 mph, the run time could be decreased to about 12.5 hours. This means cars on a train from Yard D to Yard A could have an extra 4.2 hours of available yard time at Yard A or could make up for a delay at Yard D. At either yard an improvement in reliability occurs without changing the schedule.

If for some reasons a particular train is also running with poor reliability the time gained by faster line haul speeds could offset the effect of the unreliable operation. If SA = 2 hr., $-.02 \times 2 = -.04$. $-.04 + .11$ (a change from 7 to 10 hrs.) $= +.07$, a net improvement at Yard A. Therefore, if trains cannot for some reason operate more reliably, enough time at a yard might compensate for the unreliability and even serve to reduce mean time. This is true especially at low values of available yard time.

3) Blocking

At both yards investigated, as the number of cars per day making the connection from train A to train B increased, the probability of making the connection is also increased. This suggests the following two possibilities:

1) Rather than have trains operate with whatever blocks are ready to go or many small blocks, each train should carry only a few large blocks from yard to yard each day.

2) Skip the yard entirely by using run-through trains. In this case p(MAKE) = 100%, and mean time would be very small.

Whether or not either of these suggestions or any others that might also yield larger blocks are possible or desirable to carry out depends on factors external to this model including yard space and the total amount of traffic in any one direction.

Implicit in these areas of improvement is the supposition that it is possible to carry out any changes suggested. The changing of schedules to allow enough time is very easy and can be accomplished in the short run. Line haul reliability and speed may take longer to accomplish, as additional personnel or equipment may be needed. Departure reliability could be changed at the whim of the yard operator, assuming motive power, crews, and other externals to the model, which very often are not available, become available. Blocking policy is dependent on externals to both the model and, in some cases, to railroads themselves. Hence, although trade-offs exist among these areas, some areas may be easier to effect changes in than others. So even though increased line haul reliability may improve performance, so also will allowing enough yard time which is certainly easier to accomplish.

3.7.2 Time of Day Influences

In the previous discussion of the time of day variables (Section 3.2.2.2), it was suggested that these variables could, in part, stand for congestion as well as any differences in crews or yardmaster policy. If these variables do stand, in part, for congestion, then there should be differences in their coefficients[18] that correspond with the arrival distribution of cars at the yard. As shown by Table 3-9, when traffic is heavier than average in a period, the coefficient is usually less than zero for that period. On the other hand, a light traffic period usually results in a coefficient greater than zero. This indicates that the assumption that time of day variables stand, in part, for congestion as well as any differences in yard crew performance is justified.

The congestion meaning of the time of day variables suggests that there is another factor to consider in any scheduling policy: If there are two major connections to an outbound train, perhaps both inbound trains should not arrive at the same time, as this might cause congestion. Hence, a desirable goal is not only to schedule enough time but to spread traffic evenly throughout the day.

3.7.3 Differences Between the Models for Each Yard

There are several differences between the model results for each yard. A possible reason for these differences is the total number of cars and trains handled including those cars not included in the model which are going/coming from interchange, local industry, and demand operated trains. Yard A handles more total cars and operates more trains than Yard D (Table 3-1).

Additional trains and cars create more options and areas for randomness in the system, as well as more work to be done. At a yard with less work to do, there is more time to plan operations, and, if neither too little nor too much time is allowed, a car has a potential

---

18. Recall that the time of day variables are a set, and if one is significant, they all must be included due to the correlation with the constant.

Table 3-9

Time of Day Variables
as
Indicators of Congestion

Time of Day	Variable	Yard A		Yard D	
		Arriving[1]	Coeff.	Arriving[1]	Coeff.
0600-1000	TA	-3%	-.055	-2%	.013
1000-1400	TB	7	-.055	2	-.037
1400-1800	TC	9	-.055	-4	.013
1800-2200	TD	-8	.045	-5	.013
2200-0200	TE	-11	.174	12	-.024
0200-0600	TF	6	-.055	-3	.013

[1] Difference from mean ($\approx$ 17%)

to move through more quickly, as indicated by the shape of the mean time curves for Yard D. On the other hand, if there is much work to do, planning cannot be as close, cars will be moved as the demands of the moment dictate, and the mean time curve will show a range of times that could be allowed, as happens at Yard A.

3.8 Summary of Chapter

In this chapter a selection of possible independent variables for inclusion in the model was chosen. Some were used. Others such as tonnage and class of inbound train could not be used. With the variables chosen, the form of the model was developed to reflect the influences of these variables.

Data for the calibration were provided by the daily yard reports for two yards on a major U.S. railroad. These data and the processing of them had factors which affected the model, including the range of values and the inherent confidence limits.

The probability submodel calibrated for each of two yards was presented next. Based on these submodels, examples of how the probability of making a typical connection varies with changes in the independent variables were given. The calibrated submodels were next used as the input to the mean time submodel for each yard. It was found that the curve of mean time versus available yard time could take on one of two possible shapes under the assumption of one train per day operation.

The final section discussed the implication of these two mean time curve shapes in conjunction with three areas of improvement and trade-offs in railroad operation indicated by the model:

1) Scheduling policy

2) Line haul operations

3) Blocking policy

At the end of the final section the effect of time of day and the differences in the models for the two yards were considered.

Chapter 4

Applications of the Model

4.1 Introduction

In the previous chapter the model was calibrated for two yards on Railroad C. In this chapter the model will be used to suggest and evaluate alternatives for improving rail freight service at the yards under investigation. The improvements include both increasing the reliability of service and hopefully decreasing the time a car spends waiting at a yard.

The first application of the model to be presented is the use of the model for Yard D in a case study. Here, the model was used to evaluate specific changes in operation to see if there are predicted improvements. A second application of the model will evaluate the impact of changing the frequency of trains operation.

4.2 Application of the Model in a Case Study

4.2.1 Background

In order to test the results of the previous work by Martland, Folk, Reid, and O'Doherty[1] a case study approach was developed. The object of the case study was to examine a specific railroad or part of a railroad for areas where unreliability or poor performance occur, as measured by the indices developed in the previous

---

1. See Chapter 1 of this work.

research. On the basis of data provided by the case study railroad, specific suggestions for changes to improve those areas with alternatives would be developed as a test program. These suggestions were intended

1) to have measurable impacts on operation and reliability which can be predicted

2) to be easy to implement on short notice

3) to produce those impacts with at most a small cost to the railroad or, perhaps, even a small saving.

To meet the requirement that the impacts of any proposed change could be measured and predicted, analytic tools based on previous work had to be devised. The model discussed in this thesis was one of the tools developed to predict the performance of cars at Yards A and D of Railroad C on which the case study was performed.

The requirement that the test program be implemented in a short time frame limited the options for suggested improvements to schedule changes at Yard D.

4.2.2 Possible Changes

All trains at Yard D were considered for possible changes. However, because network effects were involved, the choice was narrowed to those trains operating wholly within that region of Railroad C under close study and for which detailed data was available. In this way, even though the changes were to be made in order to improve operation at Yard D, the effects on other yards in the network could be considered. Out of the remaining trains within the region, five were selected. These trains are listed in Table 4-1. Analysis of the operating data provided by Railroad C indicated that the

Table 4-1

Trains Selected for the Case Study

Train	Average Actual Arrival	Std. Dev. of Arr.	Average Actual Departure	Std. Dev. of Dep.	Average Outbd. Length
1521	0550	3.0 hr	-	-	-
1621	-	-	1220	3.1 hr	115 cars
1721	2335	3.0	-	-	-
1931	-	-	2250	4.5	98
1041	1415	4.1	-	-	-

following recommendations might yield desirable improvements.

1) Have train 1931 depart at 0300 rather than 0000.

2) Have train 1041 arrive at 2100 rather than 2400.

3) Have train 1621 depart at 1400 rather than 1100.

4) Have train 1721 arrive at 1930 rather than 2130.

The following additional change would be evaluated in order to allow for possible changes required to 1521, the opposite number to train 1621.

5) Have train 1521 arrive at 0930 rather than 0630.

When a change such as one of the above is made to a published schedule of operation, it is assumed that the actual operation will follow that change. Even if a train arrives late, it should still be trying to hold to the schedule. Hence, if the schedule is changed so that a train should both leave earlier (or later) and arrive earlier (or later) by the same amount, whatever occurs on the line haul portion of a trip should just occur earlier (or later). What occurs on line haul portion of a trip is not known and, as mentioned in Chapter 3, may not be able to be changed in the short run. This leads to the assumption that only available yard time will change in the short run time frame of the test program, and in the evaluation this was assumed.

4.2.3 Evaluation Procedure

Each of the five suggestions mentioned above was evaluated as follows:

1) Each connection was listed with the number of cars involved over a two-week period (NN) and the actual (observed) mean time, probability, and AVAIL.

2) The model's predictions of the actual probability and mean time were listed.

3) The hoped for new AVAIL was computed. If AVAIL came out greater than 27 hrs. (or less than 3), a correction of 24 hrs. was applied to bring it into the range of the model.

4) On the basis of the new value of AVAIL, the new predictions of both the probability and mean time were computed, using the model for Yard D. As stated above only the value of AVAIL was changed and the values of SD, SDARR, etc., were not changed. It is anticipated that there might be some improvements in these values which could bring additional benefits.

5) To predict one benefit of a change, the hoped for decrease in car hours was computed for each train connecting with the changed train.

6) Finally the weighted average probability and mean time for the actual, predicted actual, and new predicted were computed. It is hoped that the new predicted probability and mean time will be an improvement over the actual.

A change of schedule for a train was determined to be beneficial if there is a predicted decrease in car hours and a predicted increase in the probability of making a connection to or from the changed train relative to the actual. Tables 4-2 to 4-6 show the evaluation procedure carried out for the five recommendations of section 4.2.2.

4.2.4 Results of Evaluation

From the evaluation of each change it can be seen that the suggestions involving trains 1041, 1931, 1621, and 1721 are predicted to cause a net decrease in car-hours. The suggestion for train 1521, on the other hand, could cause a decrease in reliability. This last suggestion was eliminated from further consideration. The impacts of the four beneficial suggestions are summarized in Table 4-7.

4.2.5 Implementation

The four beneficial suggestions were presented, as part of the case study, to Railroad C for possible implementation.

Table 4-2

Evaluation of Having Train 1931 Depart at 0300

				Observed			Predicted[1]		New Prediction			Change in
IB	OB	L/E	NN	AVAIL	p(MAKE)	MT	p(MAKE)	MT	AVAIL	p(MAKE)	MT	Car-Hours
1030	1931	L	77	12.8	.53	24.5	.88	15.8	15.8	.89	18.4	470
1130		L	171	5.6	.48	20.3	.50	17.7	8.6	.84	12.4	1351
1740		L	176	3.6	.20	28.4	.23	22.0	6.6	.65	9.8	3274
1170		L	14	4.7	.00	32.3	.34	20.5	7.7	.74	13.8	258
1951		E	29	13.3	.37	27.3	.74	19.5	16.3	.74	22.5	139
1951		L	30	13.3	.55	24.0	.74	19.5	16.3	.74	22.5	45
1361		E	19	7.9	.50	31.1	.68	15.6	10.9	.82	15.2	302
1561		E	14	15.0	.63	22.4	.80	19.8	18.0	.80	22.8	-6
1561		L	39	15.0	.52	24.3	.81	19.5	18.0	.81	22.5	187
Totals			569									6020
Averages					.39	24.8	.52	19.1		.77	14.5	

[1]Predicted value of observed.

Table 4-3

Evaluation of Having Train 1041 Arrive at 2100

IB	OB	L/E	NN	Observed AVAIL	Observed p(MAKE)	Observed MT	Predicted[1] p(MAKE)	Predicted[1] MT	New Prediction AVAIL	New Prediction p(MAKE)	New Prediction MT	Change in Car-Hours
1041	1030	E	205	13.9	.95	15.2	.93	15.5	7.1	.68	12.4	574
	1840	E	135	20.8	.48	35.6	.90	23.3	14.0	.90	16.4	2592
	1070	E	50	18.3	.96	22.2	.77	23.9	11.5	.74	17.7	225
	1280	E	47	19.5	.98	21.3	.77	25.1	12.7	.76	18.5	132
	1261	E	35	12.9	.71	15.3	.86	16.2	6.1	.38	15.2	4
	1461	L	15	23.0	1.00	22.4	.91	25.1	16.2	.91	18.4	60
	1561	E	28	17.4	.03	50.4	.91	19.5	8.6	.81	15.2	986
	1681	L	150	15.2	.95	15.3	.97	16.0	8.4	.86	11.7	540
	1681	E	99	15.2	.95	16.4	.95	16.5	8.4	.84	12.2	417
Totals			764									5530
Averages					.65	21.2	.91	18.6		.78	14.0	

[1] Predicted values of observed.

Table 4-4

Evaluation of Having Train 1621 Depart at 1400

				Observed			Predicted[1]		New Prediction			Change in
IB	OB	L/E	NN	AVAIL	p(MAKE)	MT	p(MAKE)	MT	AVAIL	p(MAKE)	MT	Car-Hours
1220	1621	L	15	23.6	1.00	24.8	.93	25.4	25.6	.93	27.3	-38
1130		L	28	19.1	1.00	21.2	.97	19.8	21.1	.97	21.8	-13
1860		L	15	16.5	1.00	19.3	.92	18.5	18.5	.92	20.4	-17
1170		L	24	8.4	.92	23.0	.81	13.0	10.4	.90	12.8	245
1951		E	73	26.8	.99	27.2	.90	29.4	4.8	.42	18.7	620
1951		L	33	26.8	.94	31.0	.88	29.7	4.8	.40	19.2	389
1361		E	50	21.4	1.00	23.3	.94	22.8	23.4	.94	24.8	-75
1361		L	118	21.4	1.00	23.0	.97	22.2	23.4	.97	24.1	-130
1561		E	195	4.5	.78	12.2	.45	17.7	6.5	.74	12.7	-97
1561		L	203	4.5	.56	17.3	.46	17.5	6.5	.75	12.5	974
1581		E	88	13.7	.72	20.6	.83	17.8	15.7	.84	19.5	97
1581		L	51	13.7	.74	21.6	.81	18.6	15.7	.82	20.8	82
Totals			888									2036
Averages					.80	19.6	.70	20.1		.78 [2]	17.3	

[1] Predicted values of observed.

[2] Due to accuracy of model and observed data, .78 is considered no change from .80.

Table 4-5

Evaluation of Having Train 1721 Arrive at 1930

				Observed			Predicted[1]		New prediction			Change in
IB	OB	L/E	NN	AVAIL	p(MAKE)	MT	p(MAKE)	MT	AVAIL	p(MAKE)	MT	Car-Hours
1721	1120	L	20	4.4	.05	32.8	.16	24.6	6.4	.45	19.6	264
	1631	L	128	16.7	1.00	19.2	.96	17.7	18.7	.96	19.7	-64
	1851	E	18	15.0	1.00	17.8	.93	16.6	17.0	.93	18.7	-16
	1851	L	48	15.0	1.00	17.8	.94	16.4	17.0	.94	18.4	-16
	1261	E	22	6.5	.50	19.0	.53	17.8	8.5	.73	15.0	88
	1261	L	104	6.5	.37	20.3	.56	17.0	8.5	.76	13.3	728
	1461	E	26	13.6	1.00	14.9	.93	15.3	15.6	.94	17.0	-55
	1461	L	190	13.6	.99	16.1	1.00	13.7	15.6	1.00	15.6	95
	1681	E	16	6.1	.75	18.0	.53	17.4	8.1	.76	13.9	66
	1681	L	27	6.1	.93	13.0	.53	17.4	8.1	.76	13.9	-24
Totals			599									1053
Averages					.82	18.1	.84	16.3		.89	16.4	

[1]Predicted values of observed.

Table 4-6

Evaluation of Having Train 1521 Arrive at 0930

IB	OB	L/E	NN	Observed AVAIL	Observed p(MAKE)	Observed MT	Prediction[1] p(MAKE)	Prediction[1] MT	New Prediction AVAIL	New Prediction p(MAKE)	New Prediction MT	Change in Car-Hours
1521	1120	E	22	22.2	.96	25.4	.84	26.1	19.2	.84	23.0	53
	1631	E	85	10.5	.97	13.8	.93	12.2	7.5	.76	13.3	43
	1631	L	81	10.5	.97	13.4	.93	12.2	7.5	.76	13.3	8
	1851	E	39	8.8	.93	11.9	.87	12.2	5.8	.55	16.6	-183
	1851	L	76	8.8	.91	12.4	.88	11.7	5.8	.56	16.4	-304
	1261	L	55	21.3	1.00	23.5	.88	24.1	18,3	.88	21.1	132
	1261	E	68	21.3	.99	23.5	.88	24.1	18.3	.88	21.1	163
	1461	E	55	7.4	.65	14.1	.76	13.3	4.4	.35	20.0	-325
	1461	L	121	7.4	.82	11.2	.78	12.6	4.4	.37	19.5	-1004
Totals			602									-1417

Averages Evaluation not carried to step 6 due to net decrease in car-hours

[1]Prediction of observed values.

Table 4-7

Summary of Operational Impacts

Suggestion	Observed MT	Prediction of Obs. MT	Predicted New MT	Observed p(MAKE)	Prediction of Obs. p(MAKE)	Predicted New p(MAKE)
Have Train 1931 Leave at 0300	24.8 hr	19.1 hr	14.5 hr	39%	52%	77%
Have Train 1041 Arrive at 2100	21.2	18.6	14.0	65	91	78
Have Train 1621 Leave at 1400	19.6	20.1	17.3	80	70	78 [1]
Have Train 1721 Arrive at 1930	18.1	16.3	16.4	82	84	89

[1] Considered no change from observed value.

Note: For this table MT= average mean time for cars to a departing train or from an arriving train from/to another train.

Recommendations 3 and 4 were put into effect as suggested, while 1 and 2 were modified before use. In the time frame of this thesis the results of the changes were not available. However, data were being collected to see what did occur after the changes were put into effect.

4.2.6 Summary

In this section the model for Yard D was used to evaluate specific schedule changes for implementation in order to predict possible savings in car-hours and reliability increases in the yard. In the following section the model will be used to evaluate the benefits to a wholesale change in policy -- that of more frequent trains.

4.3 Use of the Model to Evaluate Increased Frequency of Operation

4.3.1 Introduction

In the preceding sections and chapters the assumption was made that the penalty for a missed connection was 24 more hours in the yard. This occurred because trains were assumed to operate once per day to each other location. In this section the model will be used to evaluate what happens to cars in a 24-hour period if trains to each location operate more often than once per day. Because it is anticipated that there are benefits to increasing frequency of operation, this section will examine the mean time curves for various frequencies of operation rather than show that there are benefits. The values for the mean time curves will be developed using modified models already calibrated for each of the two yards.

4.3.2 Modifications to the Current Models

In order to predict values for the mean time curves under increased frequencies of operation, three modifications to the current model(s) must be made:

1) Because what the effect of reliability of arrival and departure, block size, train length, and time of day on the probability of making a connection when frequency is changed, is not explicitly known, it will be assumed that the net effect of these variables is zero.

PMAKE = k0 + k1 * f(AVAIL) + f(R)

f(R) = 0 ; k1 = 1

define p(M/t) = k0 + f(AVAIL = t)

and p(S/t) = 1 = p(M/t)

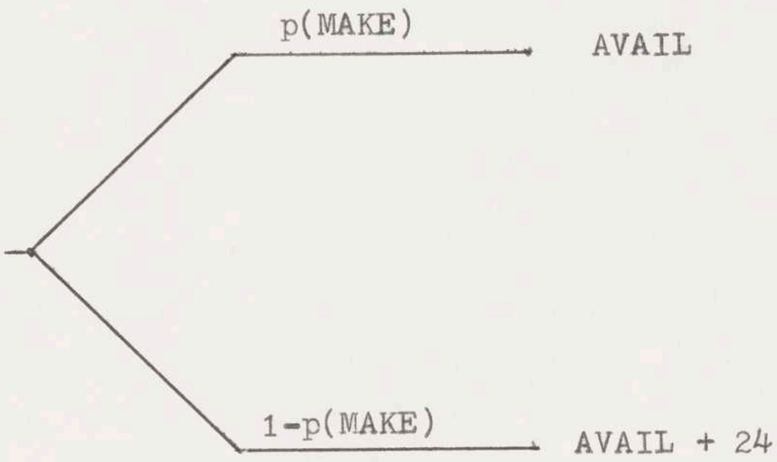
2) A modification of the assumption that a car that misses today will be the first out tomorrow is necessary. The assumption now is that a car that misses today will leave the yard with certainty on the train at AVAIL + 24 but may leave on an earlier one. Because a car may leave on an earlier train, the available yard time is no longer in the range 3 to 27 hours. Instead, it is in the range 3 to 3 + F hours, where F is the new time between trains (F = 24/j, j = # of trains per day).

3) The mean time submodel must be modified also. Figure 4-1 shows the mean time submodel that has been used. Figure 4-2 shows the mean time submodel modified for increased frequency of operation. It is to be noted that the model for one train per day operation is a case of the modified model, when F = 24 (j = 1).

These modifications to the models previously put forward allow easy computation of possible improvements for the cases to be investigated, j = 2, 3, and 4 trains per day. Because the mean time

Figure 4-1

Mean Time, One Train per Day



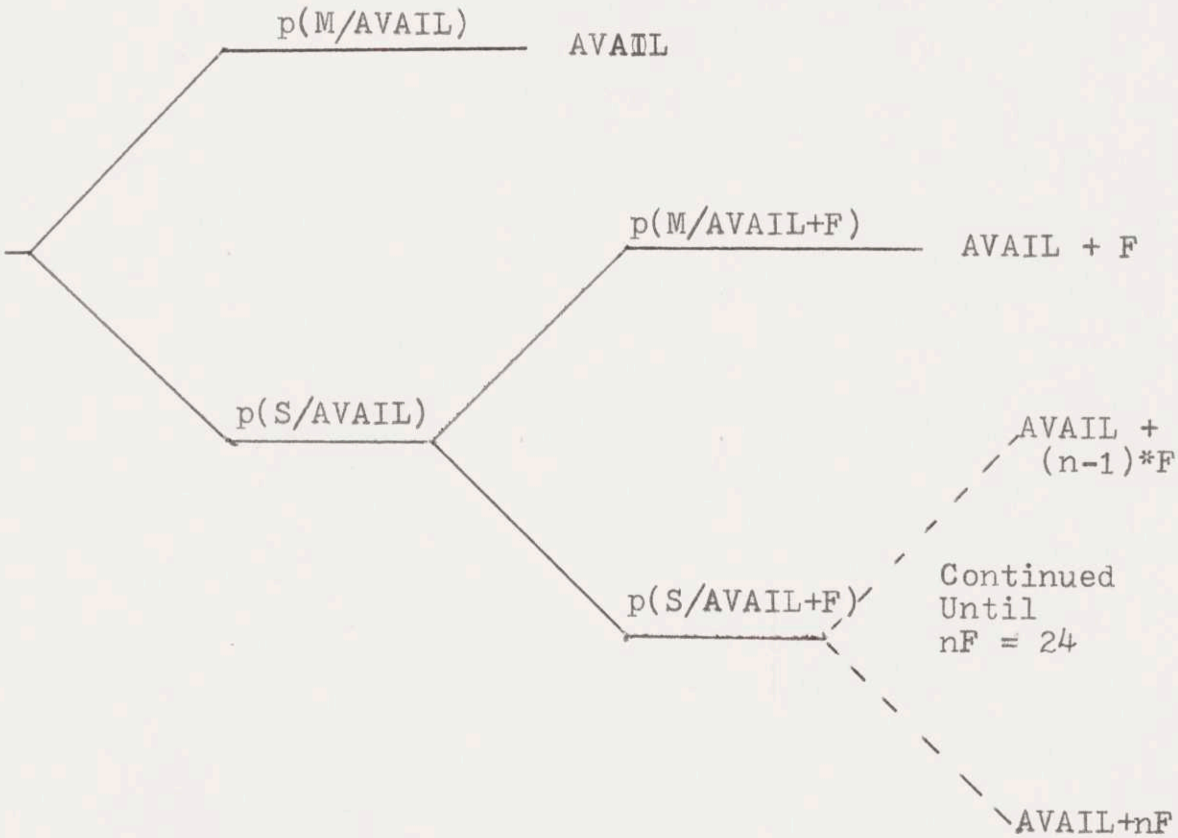
MT = Weighted average yard time

= Expected value of connection "lottery"

= AVAIL + 24 * (1 - p(MAKE))

Figure 4-2

Mean Time, More Than One Train per Day



MT = Weighted average

= Expected value

= AVAIL + p(S/AVAIL)*F + p(S/AVAIL)*p(S/AVAIL+F)*F

+ p(S/AVAIL)* ..... *p(S/AVAIL+nF)*F

model already used is a case of the modified model for j = 1, the mean time values from the unmodified model may be compared directly without having to correct for differences in computational method.

4.3.3 Analysis and Interpretation

Using both probability submodels already calibrated and the appropriate modified mean time model, the mean time curves were computed for frequency = j = 1,2,3, and 4 trains per day. These results are presented on Figures 4-3 (Yard A) and 4-4 (Yard D). The curves are summarized on Table 4-8. As is to be anticipated due to the cyclical nature of operation, the curves repeat themselves over each range of F (= 24/j).

Part of each curve is shown dashed because it is unclear what occurs in each of those periods. What actually happens in that range depends on what occurs in the range 0 to 3 hours. As the data set available from Railroad C did not contain cars on connections made in less than 3 hours, the exact dependency is unknown. However, a "good guess" can be made. As long as there are no "kinks" in the probability submodel from 3 to 27 hours and because time is continuous, it is reasonable to assume that a curve of time vs. time will be "smooth" (having continuous derivatives) and continuous. Hence, in those periods the curve shown dashed is appropriate.

From examination of the curves, two things are apparent - there are decreasing returns to scale and scheduling becomes easier.

The decreasing returns to scale are illustrated on Figure 4-5. For the first additional train per day the change in the lowest or minimum mean time value is about 3 to 5 hours. The second additional

Figure 4-3

Mean Time Curves as Frequency is Changed, Yard A

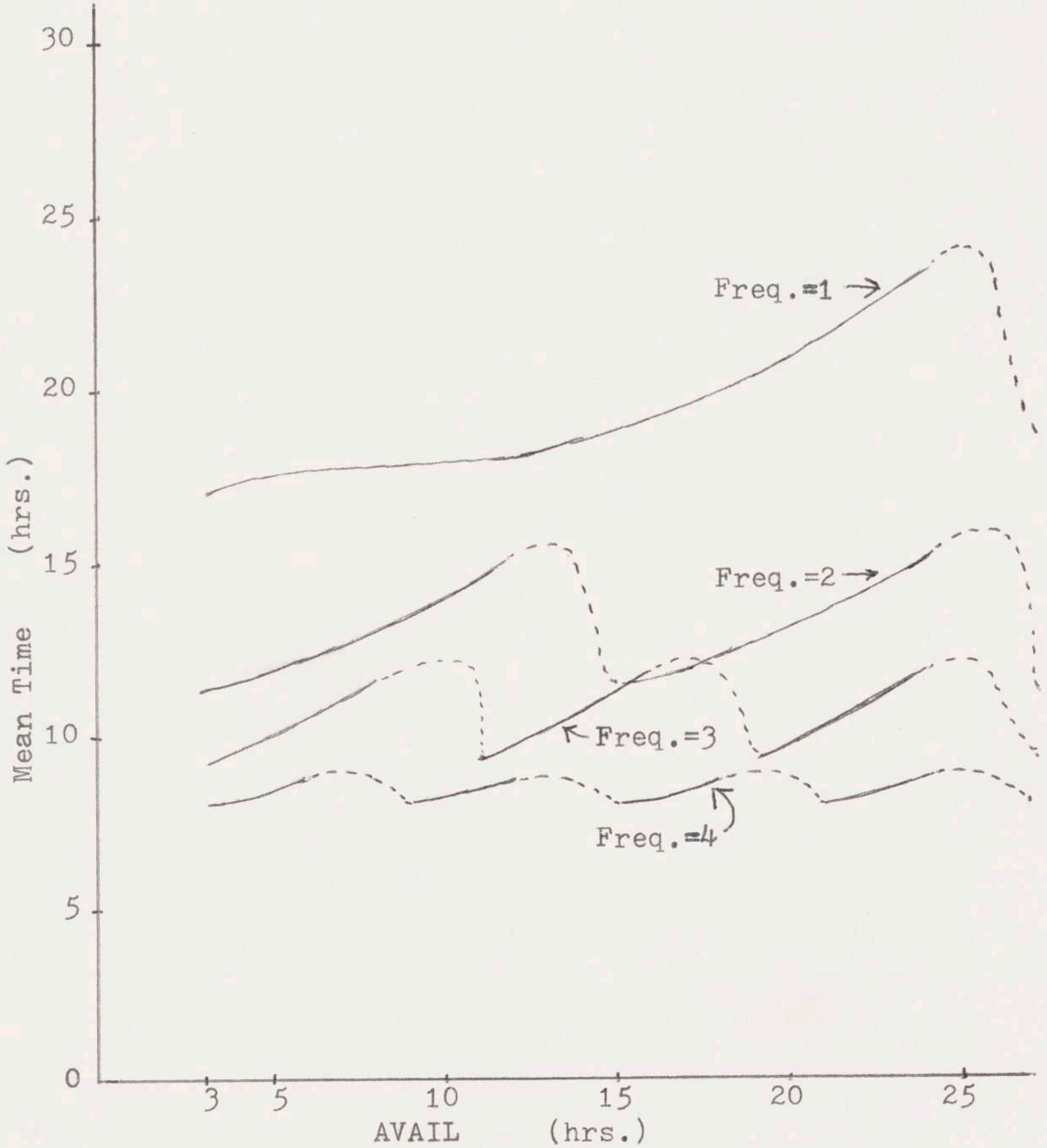


Figure 4-4

Mean Time Curves as
Frequency is Changed,
Yard D

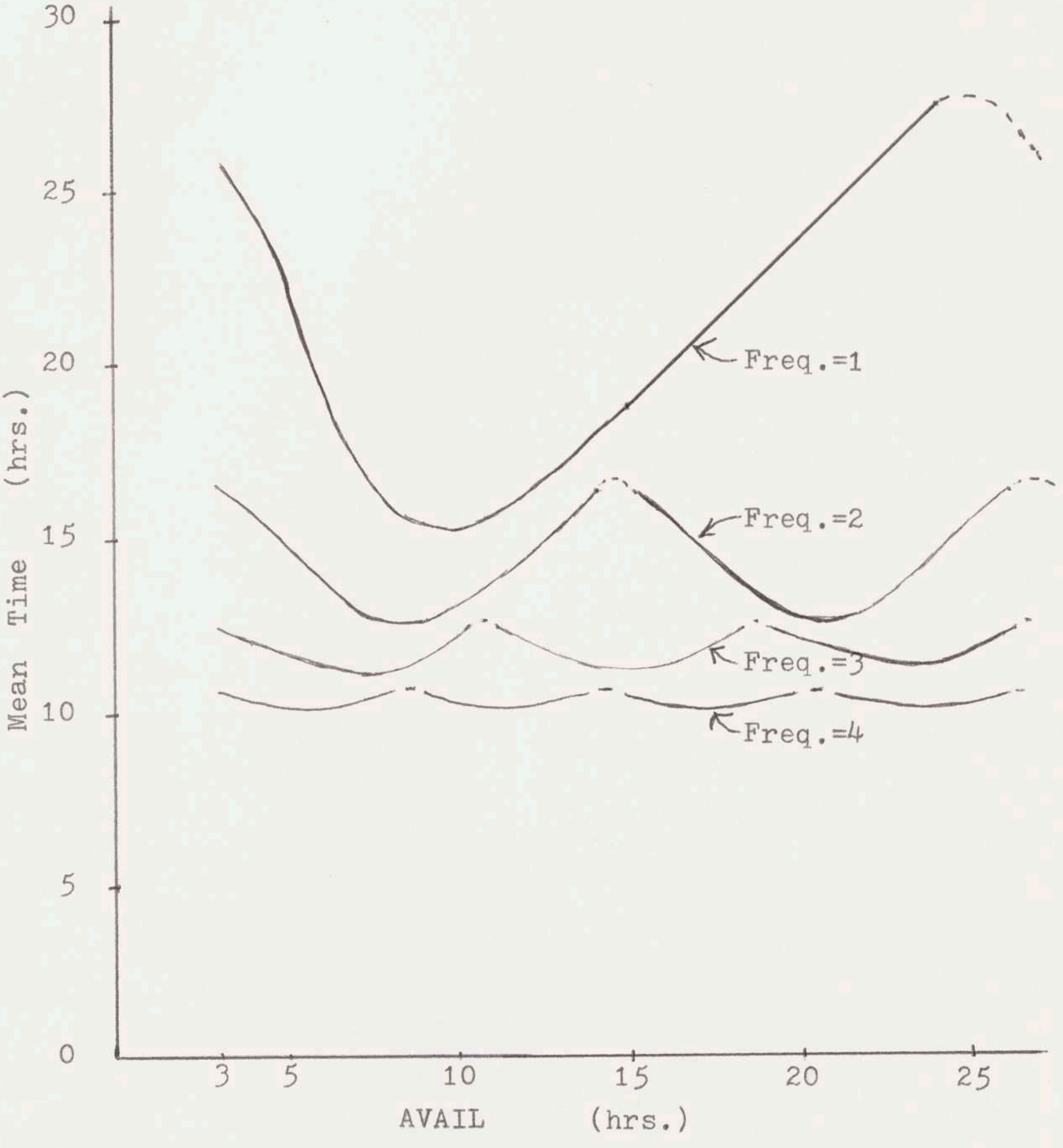


Table 4-8

Summary of Mean Time Curves as Frequency is Changed

	Yard A				Yard D			
Frequency	1	2	3	4	1	2	3	4
Minimum or Lowest Value	17.0	11.1	9.1	8.0	15.2	12.6	11.1	10.1
Change	-	5.9	2.0	1.1	-	2.6	1.5	1.0
Highest Value	26.0	15.4	12.0	9.0	28.0	17.8	12.9	10.7
Change	-	10.6	3.4	3.0	-	10.2	4.9	2.7
Range of AVAIL with Constant MT[1]	12.0	12.0	13.0	24.0	6.0	11.4	24.0	24.0
Change	-	0.0	1.0	11.0	-	5.4	12.6	0.0
Range of MT	9.0	4.1	2.9	1.0	12.8	5.2	1.8	0.6
Change	-	4.9	2.2	1.9	-	7.6	5.8	1.2

[1] Constant MT = less than a 2 hour change from minimum or lowest value

Note: All values are in hours.
All changes are shown as absolute values from previous value.

train causes an additional improvement of only 1.5 to 3 hours. For a third additional train about a 1 hour decrease is indicated. Whether or not these changes are significant enough to justify changing the frequency of operation depends on the cost of implementation of more trains per day and the savings in car usage costs derived from less time on the railroad.

Figure 4-6 compares the curves of Figure 4-5, assuming that they represent the car usage costs, with hypothetical train cost curves, as the determination of actual cost curves for car usage and train operation is beyond the scope of this work. However, even from these hypothetical curves, it is apparent that the cost of train operation is important in selecting how many trains per day to operate. At low cost 3 or 4 trains per day might be "best" for this case; at medium, 2; and at high, only 1, but each individual railroad must make its own decision.

In section 3.7 scheduling policies when a definite minimum or a flex point existed in the mean time curve were discussed. Although both of these shapes are not present in the case of increased frequency, the appropriate policies remain. However, the curves show that in each case scheduling should be easier because 1) the extremes of mean time are reduced and 2) the range in which there is less than a two hour change from the minimum or lowest value of mean time is greater. Hence, if scheduling can not be done as suggested in section 3.7, the increase in mean time under increased frequency will be smaller than it would be in the one train per day case.

4.3.4 Conclusion

In this section the model was used to evaluate changes

Figure 4-5

Decreasing Returns to Scale
as Frequency is Increased

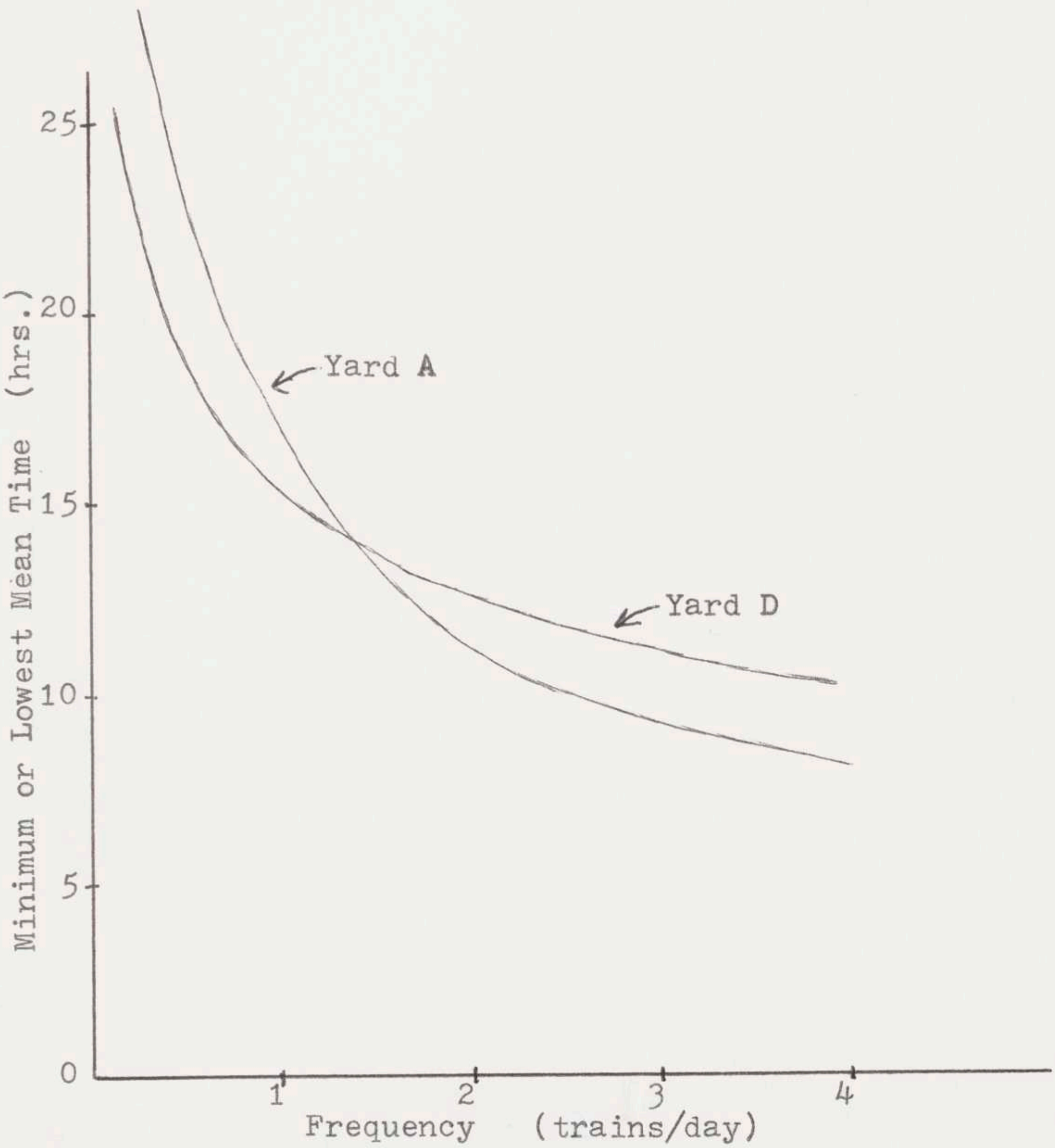
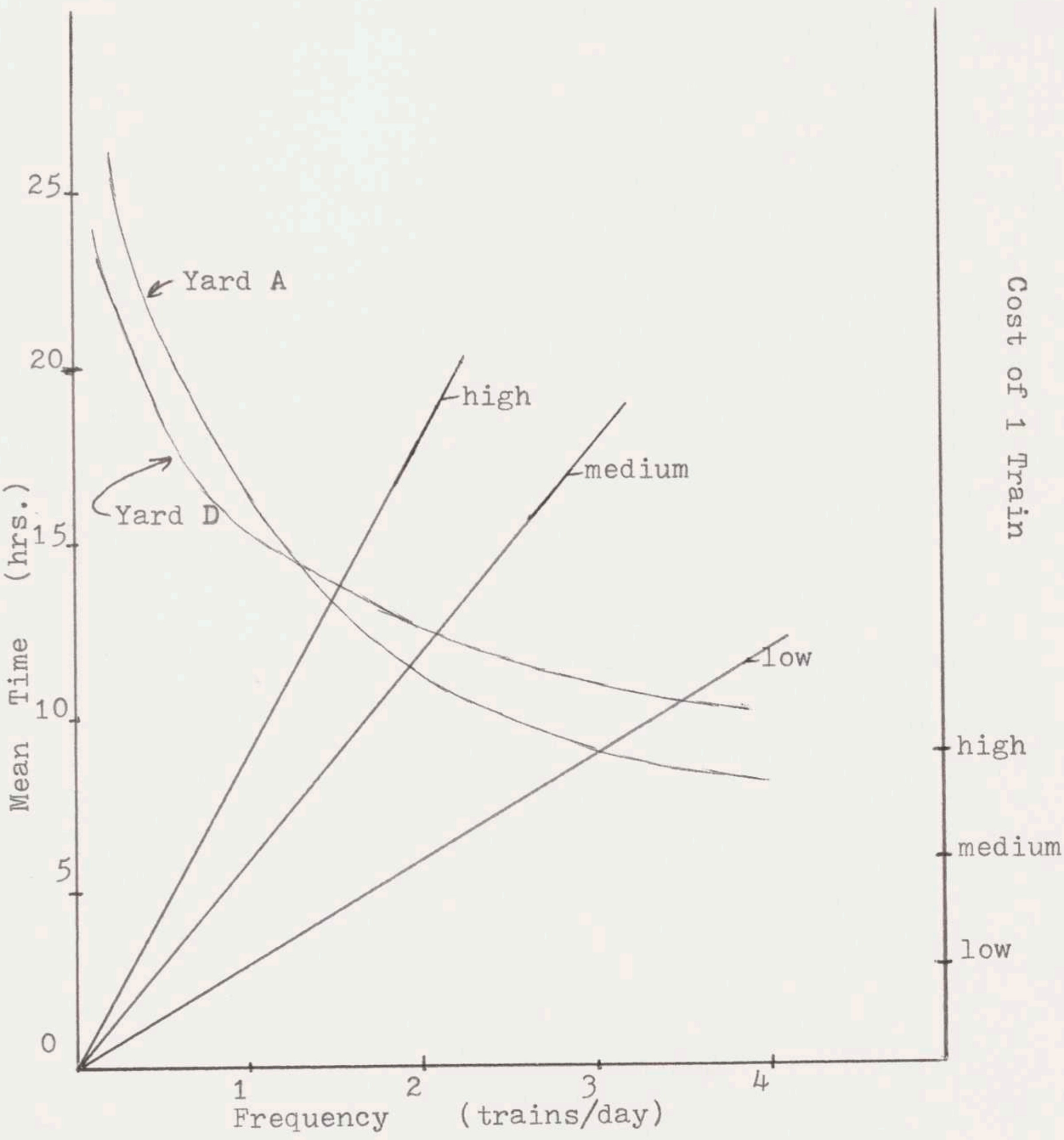


Figure 4-6

Mean Time - Cost Trade-Off
(Hypothetical)



in the mean time at the two yards under investigation if the frequency of train operation is increased. Two conclusions were reached. First, there are decreasing returns to scale with increasing frequency which means a careful mean time - cost trade-off analysis is needed to determine the "best" frequency. Second, scheduling becomes easier, as the range of mean time becomes less.

4.4 Summary of Chapter

In this chapter two applications of the model were presented. The first was to evaluate, as part of a case study, five suggested changes to see if they might be beneficial both in reducing the mean time spent in the yard and increasing reliability. Of the five suggestions four were shown to be potentially beneficial. The fifth would reduce reliability and was rejected. The four useful changes were presented to and implemented by the case study railroad. The second evaluated the changes in mean time under a major change in policy - running more frequent trains. The model had to be modified to do this, but two results were apparent - there are decreasing returns to scale and scheduling becomes easier.

Chapter 5

Summary and Conclusions

5.1 Summary

Based on previous studies which had found that the freight yard was the principal area of poor reliability in the rail network, a model of yard performance containing two submodels was developed as an aid in predicting what might happen to yard performance as changes in operation are made in an effort to improve reliability. Two parameters were chosen to measure yard performance:

1) The probability of making a train-to-train connection which is predicted by the first submodel.

2) The mean time in the yard for all cars making a connection, which is predicted by the second submodel, using the probability value predicted by the first submodel.

The previous research also found several factors which influenced the ability of a car, in general, to make a connection at a yard. These included

1) cancellations of blocks or trains
2) late arrivals
3) hold/no hold policy
4) non-uniform handling due to priority or other cause
5) the time allowed to complete the necessary operations
6) tonnage and length constraints

7) congestion

8) cyclical operating policies

9) other miscellaneous causes including repairs, "no bills," and miss-routes

The effect of these factors on specific connections was illustrated through the use of yard time distributions for cars involved in hypothetical connections. A very reliable connection would show all cars leaving on the first day's train, while less reliable connections would have cars that depart both on the first and second (or more) day's trains. Cars which are miss-routed or are repaired usually appear in the distributions with yard times greater than two days.

Using these distributions, an observation of the dependent variable, the probability of making a connection, was defined as the number of cars that departed on the first day divided by the total number of cars involved.

The independent variables were suggested by the factors that were found to affect connections in general and included:

1) Available yard time

2) Measures of the reliability of arrival and departure

3) Outbound train length

4) Volume of traffic in the connection

5) Indicators of differential handling

6) Time of day

Two assumptions were made concerning how the independent variables entered the model. First, it was assumed that the relationships among the variable functions were linear, and, second, that the variables are independent. These two assumptions allowed the use of a linearly separable form for the probability (or first) submodel.

Review of previous findings revealed that the shape of the curve of probability of making a connection with respect to available yard time is non-linear. A logit function in available yard time was chosen from many possible functions to represent that shape. It was also found from previous research that the probability of making a connection increased linearly as the number of cars involved increased. Because the effect of the other variables on the probability of making a connection had not been previously investigated, linear functions were assumed for these variables.

The mean time (or second) submodel derived its form from the weighted average yard time for all cars in a connection and did not require calibration.

Because the form developed for the probability submodel involved both logit and linear functions, a simple stepwise regression could not be used. Instead, a sensitivity/search procedure was used on data for two yards, A and D. In general, the final results were shown to justify the assumptions of a linear relationship and independence of the variables. Significant variables included the available yard time, the reliability of both arrival and departure, outbound train length, and volume of traffic.

Each of the two submodels was first used separately to investigate typical connections at each yard. From the probability submodel methods of improving the probability of making a connection 10% at each yard were developed. These methods included increasing available yard time, improving arrival reliability, and increasing block size. From the mean time submodel two shapes of the curve of mean time relative to available yard time were found - a convex curve and a curve with a flex. Based on these separate investigations and the parameters of the submodels, the entire model was interpreted from the point of view of railroad operations. Areas of improvement suggested by the model were in the areas of scheduling policy, line haul operations, and blocking policy.

Two applications of the entire model were presented. First, the model for Yard D was used in a case study to evaluate five specific changes in operation intended to improve the probability of making connections and mean time. One of the five was predicted to be a reduction in reliability and was rejected. The others were predicted to have beneficial impacts and were actually proposed to the case study railroad for implementation. Second, the entire model was used to predict values for an analysis of the mean time in the yard as frequency of operation increased. Because, as expected, the mean times dropped with increased frequency, a potential cost saving to a railroad was indicated. Hence, a comparison was made between the decreasing mean times and hypothetical costs of implementation.

5.2 Conclusions

From the work in this thesis several conclusions can be drawn.

1) There is a range of available yard times for which the mean time is constant - a less than two hour change from the minimum or lowest value. At Yard A of this investigation the range is 3 to 15

hours (mean time approximately 18); at Yard D, 7 to 13 hours (mean time approximately 16 hours). The existence of this range of available yard time leads to the second conclusion:

2) The available yard time is the most important factor in determining whether or not a car will make a connection. The reliability of making a connection (as measured by the probability) can be improved as much as 42% at both Yards A and D without changing the mean time simply by schedule modifications to change the available yard time. Because mean time is related to the cost of car usage charged to the railroad, this also means that improving reliability can be accomplished without additional cost. At Yard D the shape of the mean time curve was such that not only was there increased reliability (in four of five cases) but also potential savings from 1053 to 6020 car-hours[1] per two week period, when schedule changes as small as two hours were investigated. A change to no other variable could yield an improvement as large.

3) Under increased frequency the improvement in the mean time exhibits decreasing returns to scale (Table 5-1). Because of the decreasing returns to scale, the "correct" frequency of operation must be determined by careful analysis of the benefits of decreased mean time relative to the cost of implementation. Savings decrease for each additional train; while cost could increase at a constant rate.

4) Scheduling under increased frequency of train operation

---

1. A rough estimate is: 1 car-hour = $.167 = $4.00/24. $4.00 is approximately the average car cost per day.

Table 5-1

Decreasing Returns to Scale of Mean Time Improvements Under Increasing Frequency

Yard	Frequency	Lowest or Minimum Value	Change from Previous
A	1	17.0	-[1]
	2	11.1	5.9
	3	9.1	2.0
	4	8.0	1.1
D	1	15.2	-[1]
	2	12.6	2.6
	3	11.1	1.5
	4	10.1	1.0

[1] Actually infinite

becomes less critical. The model supports this conclusion in two ways. First, the model predicted that the range of available yard time for which the mean time is constant increases (Table 5-2). Second, the "peak" to "valley" range of the mean time curve decreases (Table 5-2). Combining these two means that even if scheduling can not be planned with high reliability and no change of mean time, the "error" would be less.

5) Where possible, larger blocks should be moved instead of many smaller blocks. In both models the probability of making a connection increased as the block size increased. At Yard A this amounted to 1.6% per car; at Yard D, 0.5% per car. However, if the size of every block is increased, the operation of the yard may change, and this model will not apply.

6) There is also some indication that traffic should arrive uniformly throughout the day. This arises from the coefficients of the time of day variables used in the model. At Yard A the range of coefficients is -5.5% to +17.4% (22.9%); while at Yard D it is -3.7% to +1.3% (5%). The range at Yard A is significant, but at Yard D the 5% may not make very much difference.

Because these conclusions bear directly on the day-to-day operations of railroads, the major extension of this work would be to apply the results in the field on a larger scale than the small region of one railroad looked at in the case study (Chapter 4).

5.3 Areas for Extension to This Model

There are several areas for possible future analytic research to extend the model presented in this thesis. These areas include

Table 5-2

Increased Range of Constant Mean Time
and
Decreasing Total Range of Mean Time

Yard	Frequency	Range of Constant[1] MT	Total Range of MT
A	1	12.0	9.0 hr
	2	12.0	4.1
	3	13.0	2.9
	4	24.0	1.0
D	1	6.0	12.8
	2	11.4	5.2
	3	24.0	1.8
	4	24.0	0.6

[1]Hours in 3 to 27 hours of available yard time. Constant MT = change in mean time less than two hours from minimum or lowest value.

1) Could an entire railroad be scheduled based on the probability and mean time submodels of every yard in the railroad involved?

2) What happens if frequency is more than once per day but not at constant intervals throughout the day?

3) What happens if some or all trains run on demand (i.e., no schedule)?

4) What happens during the first 3 hours after a train arrives? Do some cars make connections with an available yard time less than 3 hours? Which ones?

The first area involves calibrating the model presented in this thesis for every yard on a railroad and then,through the use of linear programming techniques and recalibrations based on the revised schedules, find an optimum schedule, if extant, that minimizes the wait of cars in the yards and maximizes the reliability.

The second area involves finding out the chance that a car will arrive within a certain range of available yard time. This is in contrast to the analysis in Chapter 4 where the ranges of available time were equal.

The third area involves finding the probability of a car arrival and the time for X number or more to arrive. This, of course, will invalidate this model, but it may reduce mean time and increase reliability.

The last area involves obtaining records for cars with yard times less than 3 hours, changing the model to include new variables that may explain what occurs in less than 3 hours, and recalibrating.

Some of these suggestions for future research will involve policies not covered in this analysis including per diem, interchange with other railroads, and customer influence, which for the purposes of this analysis were assumed to be negligible, but are present in the real situation.

Appendices

APPENDIX A

Calibration of the Probability Submodels

A.1 Calibration Method

The calibration of the probability submodel required a regression of PMAKE against ko + k1 * (1./(1. + EXP(-b(AVAIL + a)))) + k2 + N + k3 * SA + K4 * SD + k5 * LENGTH + k6 * COND + k7 * PRI + k8 * PA + time of day variables. However, because both linear and logit functions are involved, a simple stepwise linear regression could not be used to calibrate all the coefficients simultaneously. Coefficients a and b must be estimated separately, and a polynomial regression technique employed.

In a polynomial regression new variables may be created which reflect the hypothesized shape, and these new variables are used in a stepwise linear regression. Hence, a new variable Q = 1 /(1 + EXP (-b(AVAIL + a))) was defined for this model. However, values must still be determined for a and b in order to calculate Q. Because there was no prior information as to which values to use, although there was some limit to the range,[1] a technique had to be found to fit this situation - a combination of linear and non-linear forms. Because of the limited range and an easy method of finding a starting point, a search/sensitivity to change procedure was decided upon

---

1. -a less than 27 due to the range of AVAIL, and
   b less than .9 due to computation problems.

out of other possible choices.[2]

A reasonable starting point was to assume, at first, that PMAKE is only a function of AVAIL and use the logit procedure to find a value for a and b.[3,4] This value of a and b would then be used to calculate Q in the full regression. By analysis of the residuals of the first regression, an indication of the first change needed in a or b can be determined. After the first change, the sensitivity/search analysis will be performed by first changing the other coefficient and then alternating to determine if there might be a better set of a and b as indicated by significant improvement in the R^2 and F statistics and by a lower standard error in the full regression.

In performing this sensitivity/search analysis two possibilities could occur:

1) The model is sensitive to changes in both a and b. In this case a point might exist at which R^2 and F are clearly maximized and the standard error minimized for changes in both a and b. If such a point exists, the model using that set of a and b will be accepted as locally optimal.

2) The model is not sensitive to changes in either a or b or both. In this case, a range of points exist which are "best", and an alternate criterion must be used.

The alternate criterion for "best" model arises from the behavior of the model with respect to f(AVAIL). The f(AVAIL)

---

2. Draper and Smith, Applied Regression Analysis (New York, 1966).
3. H. Theil, op. cit., p. 632.
4. Appendix B.

entered into the regression through the use of the variable Q acts as a base curve which the calibration changes to account for the effects of the other variables. If the coefficient of Q is not 1.00, the assumed f(AVAIL) has been modified and is no longer the same initial base curve. Hence, if the model is not sensitive to changes in a or b or both, the combination of a and b that yields a coefficient of Q equal to 1.00 will be used. This means that f(AVAIL) becomes the true base curve.

A.2 Calibration of the Submodels

The sensitivity/search procedure was carried out for the two yards, A and D. For Yard A five series of searches were performed, and representative samples of each series are shown on Table A-1. The first series changed the value of b as suggested by the residuals. Over these searches the model did not exhibit a sensitivity to changes in either a or b. Hence, the alternate criterion was used, and a sixth series of searches was performed to find an a and b which yielded a coefficient of Q equal to 1.00. Such a model was found and will be used.

For Yard D, again the residual analysis suggested changing the value of b for the first series (Table A-2). However, over the remaining two series, it was apparent that the model was not sensitive to changes in a for b = .60. Hence, a fourth series was performed to determine that value of a which gives a coefficient of 1.00 to the variable Q.

Table A-1

Sensitivity/Search Results
at
Yard A

Series	Trial	a	b	k1	R^2	Std. Err.	F [1]
	A0	9.3	.19	.82	.49	.27	25.4
I	A1	9.3	.10	1.30	.54	.255	26.2
	A2	"	.15	.97	.54	.254	26.4
	A3	"	.19	.84	.54	.254	26.3
	A4	"	.40	.59	.52	.261	24.0
II	A5	8.3	.15	1.01	.54	.254	26.3
	A8	10.3	"	.95	.54	.254	26.5
	A10	11.3	"	.92	.54	.254	26.5
	A12	12.3	"	.90	.54	.254	26.4
III	A13	10.8	.10	1.26	.54	.255	26.1
	A15	"	.20	.78	.54	.254	26.7
	A16	"	.25	.69	.54	.254	26.6
IV	A17	9.0	.20	.83	.54	.255	26.0
	A20	12.6	"	.74	.55	.253	26.8
	A24	14.0	"	.73	.54	.254	26.5
V	A25	12.1	.10	1.24	.54	.255	26.0
	A28	"	.30	.61	.55	.253	27.0
VI	A29	8.5	.15	1.00	.54	.254	26.3
	A30	9.3	.14	1.02	.54	.254	26.4
	A31	10.8	.13	1.03	.54	.254	26.3
	A32	11.3	.12	1.09	.54	.254	26.2

[1] F statistic at the 9^{th} step of the regression.

Table A-2

Sensitivity/Search Results
at
Yard D

Series	Trial	a	b	k1	R^2	Std. Err.	F [1]
	D0	4.6	.30	1.17	.59	.18	- [2]
I	D2	4.6	.40	1.22	.65	.160	66.3
	D3	"	.50	1.23	.66	.157	70.1
	D4	"	.60	1.22	.66	.156	85.2
	D5	"	.70	1.22	.66	.157	71.3
	D7	"	.90	1.18	.65	.159	67.9
II	D8	4.0	.60	1.47	.66	.156	85.5
	D10	4.6	"	1.22	.66	.156	85.2
	D11	5.0	"	1.10	.66	.157	85.2
III	D12	4.2	.40	1.33	.65	.159	67.8
	D15	"	.80	1.37	.65	.158	68.5
IV	D16	5.3	.60	1.02	.65	.158	82.8
	D17	5.4	"	1.00	.65	.158	82.3

[1] F statistic at the 5th step of the regression.

[2] Not reportable due to differences in regression for D0.

Appendix B

Logit Curve Calibration

$$T = 1 / (1+EXP(-G)) \quad (1)$$

$$T + T*EXP(-G) = 1 \quad (2)$$

$$T - 1 = -T*EXP(-G) \quad (3)$$

$$(1 - T) / T = EXP(-G) \quad (4)$$

$$ln((1 - T) / T) = -G \quad (5)$$

$$-ln((1 - T) / T) = G \quad (6)$$

$$ln(T / (1 - T)) = G \quad (7)$$

Equation (1) is the logit curve.

Equation (7) is the form in which the logit curve is calibrated.

Appendix C

Analysis of the Mean Time Submodel by Differential Calculus

let x = AVAIL ; k = coeff. of f(AVAIL) ; d = 24.

$$MT = x + d*(1 - k*(1 + EXP(-b(x-a)))^{-1} - f(R)) \qquad (1)$$

$$\frac{\partial f(R)}{\partial x} = 0 \qquad (2a)$$

$$\frac{\partial MT}{\partial x} = 1 - kdbz/(1+z)^2 = 0 \quad \text{for max. or min.} \qquad (2b)$$

$$\text{where } z = EXP(-b(x-a)) \qquad (2c)$$

$$kdbz = (1+z)^2 \qquad (3)$$

$$0 = z^2 + (2-kdb)z + 1 \qquad (4)$$

$$z = \frac{(kdb-2) \pm \sqrt{(2-kdb)^2 - 4}}{2} \qquad (5)$$

$$\ln z = -b(x-a) \qquad z > 0 \qquad (6)$$

$$x-a = -b^{-1}*\ln z \qquad (7)$$

$$x = a - b^{-1}* \ln z \qquad (8)$$

Equation (6) is a combination of equations (2c) and (5).

Yard A: $b = .15$; $a = 8.5$; $k = 1.0$; $d = 24$

$$z = \frac{((.15)(24) - 2) \pm \sqrt{(2 - .15(24))^2 - 4}}{2}$$

$=$ undefined for real roots

$$(2 - .15(24))^2 - 4 < 0$$

Yard D: $b = .60$; $a = 5.4$; $k = 1.0$; $d = 24$

$$z = \frac{((.60)24 - 2) \pm \sqrt{(2 - .60(24))^2 - 4}}{2}$$

$$z = \frac{12.4 \pm \sqrt{(12.4)^2 - 4}}{2}$$

$$z = \frac{12.4 \pm 12.24}{2}$$

$$z = .08 \; , \; 12.32$$

$$\downarrow \qquad \downarrow$$

$$\ln(z) = -2.53 \; , \; 2.51$$

$-2.53 \rightarrow x = 5.4 + 4.2 = 9.6$

$2.51 \rightarrow x = 5.4 - 4.2 = 1.4$

Appendix D

Discussion of Possible Interdependece of the Model Variables

In Chapter 3 the assumption of the independence of the variables was made. When the model was calibrated, four areas of possible interdependence were shown by the correlation matrices:

1) SD with L at Yard A.

2) SD with L at Yard D.

3) PRI with SD at Yard A.

4) PRI with L at Yard A.

All four were related to the hold/no hold policies. If a train was held for more cars, its standard deviation of departure increased, and its length increased due to the additional cars. This would account for the positive correlation of SD with L at both yards. On the other hand, if a train was not held, as an express train would not, the standard deviation would be small and the length may be less.[1] This would account for the correlation of PRI with SD and L.[2] Because the hold/no hold policy, which would have been preferable, could not be a direct input to this model, these correlations were not considered serious.

---

1. Length may also be less because some cars connecting with express trains may have yard times less than 3 hours and not be in the data set.
2. The fact that there might be problems with express trains was one of the reasons the variable PRI was used.

BIBLIOGRAPHY

1. Ainsworth, D, "Implications of Inconsistent Rail Service," Proceedings, 1972, Transportation Research Forum, Vol. XIII, No. 1, 1972.

2. American Trucking Association, American Trucking Trends 1972 (Washington, D. C., 1972).

3. Belovarac, K. and Kneafsey, J., Determination of Line Haul Reliability (M.I.T., Cambridge, R72-38, 1972).

4. Ben-Akiva, M., Structure of Passenger Travel Demand Models (Ph.D. Thesis, M.I.T., Cambridge, 1973).

5. Drake, A., Fundamentals of Applied Probability Theory (New York, 1967).

6. Draper, N. and Smith, Applied Regression Analysis (New York, 1966).

7. Ellison, Frank, Frank Ellison on Model Railroads (New York, 1954).

8. Folk, J., Models for Investigating Rail Trip Time Reliability (M.I.T., Cambridge, R72-40, 1972).

9. Folk, J., Some Analysis of Railroad Data (M.I.T., Cambridge, R72-41, 1972).

10. Hu, T., Econometrics, An Introductory Analysis (Baltimore, 1973).

11. Malinvaud, E., Statistical Methods of Econometrics (Chicago, 1966).

12. Martland, C., Rail Trip Time Reliability: Evaluation of Performance Measures and Analysis of Trip Time Data (M.I.T., Cambridge, R72-37, 1972).

13. O'Doherty, J., Classification Yard Effects in Rail Freight Movement Reliability (C. E. Thesis, M.I.T., Cambridge, January, 1972).

14. Reid, R., Yard Unreliability in Rail Freight Movement (S.M. Thesis, M.I.T., Cambridge, June, 1971).

15. Reid, R., O'Doherty, J., Sussman and Lang, The Impact of Classification Yard Performance on Rail Trip Time Reliability (M.I.T., Cambridge, R72-39, 1972).

16. Theil, H., Principles of Econometrics (New York, 1971).

17. Towers, W., "Yard Switching, Prototype Style," Model Railroader (Milwaukee), Vol. 28, No. 7, July, 1961.

18. Valvanais S., Econometrics (New York, 1959).

19. ________, "Box Car Problems Meet Shippers at the Door," Railway Age (Bristol, Conn.), Vol. 174, No. 8, 30 April 1973.

20. ________, "How to Make Yards Work for Us," Modern Railroads (Chicago), Vol. 28, No. 7, July 1973.

21. ________, "L & D Hits the Downward Trail," Railway Age (Bristol, Conn.), Vol. 174, No. 10, 28 May 1973.