rt No. FRA-OPPD-75-1

PARAMETRIC ANALYSIS OF RAILWAY LINE CAPACITY

AUGUST 1975

FINAL REPORT

Prepared for

U.S. DEPARTMENT OF TRANSPORTATION
FEDERAL RAILROAD ADMINISTRATION
OFFICE OF POLICY AND PROGRAM DEVELOPMENT
WASHINGTON, D.C. 20590

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.
FRA-OPPD-75-1

2. Government Accession No.
--

3. Recipient's Catalog No.
--

4. Title and Subtitle
Parametric Analysis of Railway Line Capacity

5. Report Date
August 1975

6. Performing Organization Code
289-51849-30

7. Author(s)
John C. Prokopy, Richard B. Rubin

8. Performing Organization Report No.
--

9. Performing Organization Name and Address
Peat, Marwick, Mitchell & Co.
1025 Connecticut Avenue, N.W.
Washington, D.C. 20036

10. Work Unit No.
--

11. Contract or Grant No.
DOT-FR-4-5014

12. Sponsoring Agency Name and Address
Economic Programs Division
Federal Railroad Administration,
Department of Transportation
Washington, D.C. 20590

13. Type of Report and Period Covered
Final Report

14. Sponsoring Agency Code
--

15. Supplementary Notes
Contract Technical Officer: William Edson, RA-51

16. Abstract

The problem of analyzing the capability of a rail line to absorb additional traffic has become increasingly important recently as various means of consolidating and expanding usage of the rail systems have been proposed. This report develops a systematic procedure for evaluating the potential capacity of a wide variety of rail line haul facilities under a number of operating conditions and policies.

The procedure is based upon a parametric analysis of a series of rail line cases or operating and physical plant scenarios simulated using a computer train dispatching model. This report first briefly describes the simulation model. Next the structuring of the approach to the analysis is presented along with a description of the cases and the results of the case simulations. Then a parametric analysis of the results is described and procedures for applying the analysis to other conditions are developed concluding with a sample application. Finally, concluding observations are discussed and recommendations for application of the results and for further research are presented.

17. Key Words
Railroad Operations
Simulation Models

18. Distribution Statement
Document is available to the public through the National Technical Information Service, Springfield, Virginia 22151

19. Security Classif. (of this report)
Unclassified

20. Security Classif. (of this page)
--

21. No. of Pages
84

22. Price

NOTICE

NOTICE

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Appendix

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

This report describes a parametric analysis of railway line capacity. The purpose of the project was to perform an analytical examination of key factors affecting railway line capacity in order to:

- determine the relative magnitude of the impact of varying key parameters such as track and signal system configuration and train operating policy in terms of their effect on capacity; and
- develop a parametric model for general application to line capacity analysis.

METHOD

The basic tool for the parametric analysis was a computer train dispatching simulation model developed by Peat, Marwick, Mitchell & Co. (PMM&Co.). The model can be used to simulate operation of a rail line of up to 4 tracks for a specified schedule and set of circumstances. As a part of this project, improvements were made to the model to meet the requirements of the parameters to be analyzed. The train dispatching simulation model was used to simulate several hundred different combinations of track, signal, and train configurations and operating policies.

Various mathematical and statistical analysis techniques were used to analyze the simulation results. The application of these techniques was used to develop the mathematical relationships between various parameters and line capacity.

RESULTS

The following are the most significant results of the project:

- development of the relationship of train delay to number of trains dispatched;
- estimation of sensitivity of average delay per train to various parameters;

- estimation of sensitivity of average delay per train to combinations of parameters;

- development of a meaningful measure of line capacity; and

- development of a set of equations which can be used to estimate capacity of a given line.

A prototypical segment of line 150 miles long was used to develop the relationships of delay in train-hours to the total number of trains dispatched as a function of changes in various parameters (changes in relationships of delay).

The relationships of changes in delay to changes in the parameters were used to develop for each parameter a curve of: $\frac{\text{delay per train}}{\text{trains per day}}$.

Using regression techniques, these relationships were developed as a function of each parameter. The resulting factors were a measure of the relative sensitivity of average delay per train to the various parameters. The measure of sensitivity, similar to the concept of inverse additives used to combine parallel resistances in electronics, was particularly meaningful for parameters, such as siding spacing or average speed, which could easily be represented as continuously variable.

A set of equations was developed to estimate the combined effects of several parameters simultaneously. The equations can be applied to analyze the impact of different line or operating characteristics such as proportion of double-track, signal spacing, train priorities and uniformity of train speeds.

Finally, a measure of line capacity, in terms of maximum permissible delay was developed. This measure was then related to the method of estimating average delay and a set of equations developed to estimate capacity in terms of average delay. Thus, the relative influence of parameters on capacity can be analyzed and tradeoffs of capacity can be performed among various alternative changes in line configuration or operating policy.

CONCLUSIONS

The following conclusions were developed from the parametric analysis. The conclusions relate to:

- the ability of existing rail lines to absorb traffic increases;
- the impact of operating speed on capacity;
- the impact of Centralized Traffic Control (CTC) on capacity for multiple track; and
- the estimation of potential reserve capacity of a line.

The ability of rail lines to absorb considerable increases in traffic without major changes in line or operating characteristics must be questioned. Line capacity was found to be considerably less than widely believed. Capacity is not so much a function of the capability to move trains over a line at all, as it is of the ability to move trains over a line without undue delay. Delays generally exceed acceptable limits before a line will lock up.

The most important parameter in determining capacity, other than the number of tracks, is operating speed. Theoretical capacities for single and double track can only be approached as trains are run at moderately high uniform speeds. The greater the distribution of train speeds, the more the interaction among trains and the greater the delay.

It is interesting to note that CTC on double track is essentially unnecessary to reach theoretical capacity since overtakes would be unnecessary if all trains operated at the same average speed. In fact, CTC may actually increase average delay under normal operations. Its primary usefulness, other than to provide flexibility in the event of track blockage, is to increase the level of service to high priority trains. This improvement in service comes at the cost of increased delay to lower priority trains.

This work can be used as a guide for investigation of alternatives for improving capacity of a line. It can also be used to estimate the potential reserve capacity of a line. Because of its approximate nature, any marginal capacity results should be investigated in depth using detail procedures such as the train dispatching simulation.

RESEARCH RECOMMENDATIONS

There are two general categories of research which should be pursued to develop more information in regard to the parametric relationships of line capacity. One is a continuation in more depth of the present line of research; the other is analysis of topics not covered by this project. Continuation of the present research is required in regard to:

- more analysis of those parameters for which conclusive results were not reached in this study;
- conversion of parameters that were analyzed as discrete variables to continuous relationships; and
- improvement of the procedure for analyzing multiple modifications to the parameters.

New lines of inquiry which are suggested by the present research include:

- impact of increased on-track maintenance time requirements as a function of higher traffic volume; and
- nature of the recovery from major disruption as a function of the percentage utilization of capacity.

I. INTRODUCTION

The utilization of excess capacity of rail lines is widely believed to have potential for solving many kinds of capacity and service problems in transportation. These problems range from providing urban transit rights-of-way to replacing or supplementing long-distance intercity trucking with intermodal trailer-on-flat-car (TOFC) service. The feasibility of these solutions depends upon a careful understanding of rail line capacity.

THE PROBLEM

The subject of main line capacity recently has assumed greater importance because of the potential restructuring of rail service in the northeastern United States and the consequent major shifts in line-haul rail service flows. Problems with tenant or joint use of lines, especially by Amtrak, raise further questions about the precise nature of the interactions between different types of trains and services. The establishment of commuter authorities in many major cities also raises similar problems of joint use and consequent delays.

No generally accepted method for estimating the impacts of train and line charactristics on mainline service has been developed. The interaction among trains on rail lines has historically been examined through manual "stringlines" of actual operations, when available or projected operations when forecasts are being prepared. These stringlines are time-distance charts of train movement and are used to project train interference and the location of bottlenecks. They are tedious to prepare, however, and often only one day's operation is examined for an alternative. The time consuming nature of their careful preparation has also limited the number of options which can be examined. Despite this detailed analysis of operations, railroads are sometimes reluctant to accept the results, because, one day's analysis might not show the range of difficulties which might occur with a slightly different traffic pattern. This study has confirmed that, in fact, day-to-day changes in traffic patterns may have a marked effect on delay for rail lines near capacity.

Rail line capacity, not long ago thought to be far in excess of demand, has recently become much more in short supply than might be imagined. Many railroads have reduced the number of tracks on mainlines, consolidated lines, and downgraded potentially excess main lines. Reductions in speed limits have further reduced capacity. In addition, longer trains may actually reduce capacity, to the extent that handling difficulties and more road failures offset the capacity savings of a reduced number of trains for the same number of cars. Removal or reduction of passenge service, conversely, has added some capacity. While most lines are now generally adequate for the traffic currently handled, any change -- projected growth, proposals to divert traffic, increase in the frequenc of freight services, additions to passenger service -- must be carefull assessed.

PURPOSE OF STUDY

This study attempts to bring into focus the key factors affecting rail way line capacity and to examine their relative impacts on capacity. Th impacts of various track, signal, and train configurations and operatin policies on the capacity of specific types of rail line have been examine using a computerized dispatching simulation model developed by Peat, Marwick, Mitchell & Co. (PMM&Co.). The results of numerous simu lations were analyzed to develop a parametric model for general applic tion.

A secondary purpose of the study was to develop a computer simula tion of train dispatching for use in analyzing in detail specific line alter natives. The parametric analysis can be used to answer policy questio on line capacity and to prepare preliminary analyses of specific situati It is most useful for order of magnitude estimates of the capacity impli cations of proposed changes. The computer model can then be used for detailed comparison of specific alternatives within the range of feasibil indicated by the parametric analysis.

A parametric analysis is a particularly useful tool for analyses involving as many variables as were involved in this study because it provides a coherent means of consolidating results into only one or a few equations. Yet, it still allows the user the flexibility to interpolate or extrapolate those variables which are continuous throughout the range of interest.

DEVELOPING THE PARAMETRIC ANALYSIS

PMM&Co.'s approach to the development of an effective, flexible procedure for the parametric analysis of railway line capacity encompassed the following five basic steps:

1/ Peat, Marwick, Mitchell & Co. Train Dispatching Simulation Model - User's Manual. Prepared for the Federal Railroad Administration, Washington, D.C., March 1975.

- modification of PMM&Co.'s train dispatching simulation (TDS);
- selection of key capacity related parameters;
- determination of procedures for the parametric analysis, including the structure for application of the analysis and a decision tree for determining dominant solutions and insignificant variables;
- evaluation of parameters according to the framework developed above; and
- validation of the model and verification of the accuracy of the parametric analysis.

Modification of TDS

The train dispatching simulation model used in this analysis was developed to study rail line capacity options in a developing country. The model was used in conjunction with a proprietary train performance calculator (TPC) developed by Thomas K. Dyer, Inc. As part of this study, selected TPC runs were prepared by Thomas K. Dyer, Inc. to assist in determining parameters of train performance. The TDS was designed to replicate various aspects of:

- line or track configuration (single and multiple tracks, siding and crossover location, and siding capacity length);
- centralized traffic control and automatic block operation (single and multiple tracks);
- signal system configuration (number of indications and block length);
- temporary service disruptions (train and signal failures, track maintenance requirements);
- dispatching policies (train frequency and size, priorities of service, traffic peaking patterns); and
- train performance (running times, stopping and starting delays, and slowdowns for crossovers).

Selection of Key Parameters

Possible parameters were examined and from this examination the following parameters were chosen as being most important for inclusion in this study:

- speed limits,
- distribution of train speeds,
- siding spacing (single track),
- distribution of siding spacing (single track),
- siding capacity (number of trains per siding),
- siding length vs. train length (single track),
- signal spacing,
- proportion of multiple track,
- crossover spacing (multiple track),
- train power,
- train weight,
- train priorities,
- traffic imbalance,
- traffic peaking patterns, and
- incidence of disruption.

Determination of Procedures for Analysis

Several approaches to designing the parametric analysis were considered. After some refinements based on preliminary results of the analysis, an approach was chosen which was based on parameterizing variations from typical base case rail lines rather than estimating capacity for various idealized lines. It was felt that the use of prototypical lines would be more useful to railroad analysts and would provide greater accuracy in real-world analyses. Decisions on which parameters and what range of values to examine were based on the sensitivity of line capacity observed during early tests of parameters.

Evaluation of Parameters

As evaluation of the parameters proceeded, preliminary analyses of results were made. These preliminary analyses guided further exploration among the possible ranges and combinations of parameters. As the results of many simulations were examined, a number of patterns emerged. Several methods of consolidating the results were tested. The method chosen provides not only the best overall fit of those tested, but also has a relatively high intuitive appeal.

Validation of the Model

Finally, both the model and the parametric analysis methodology were validated against each other and against both real-world data and manual redispatches. The results in all cases were generally satisfactory.

TERMINOLOGY

For purposes of modeling rail line characteristics, certain terms have been adopted. These terms, as used in the parametric analysis, are briefly defined here.

block - a section of track which may be occupied by only one train at a time. Blocks are used to control train separation, and occupancy is regulated either by the dispatcher, an operator at a station, or an automatic signal system.

class - the type of train as defined by its performance characteristics (not as normally defined in railroad terms).

interlocking - any connection between two main line tracks including the transition between single and multiple track.

segment - the section of track between two stations; may contain one or more parallel tracks and must contain at least one signal or train separation block.

siding - a track at a station used for trains to meet, overtake, or perform switching.

station - any point on a rail line where track configuration changes.

yard - several sidings at one station.

REPORT ORGANIZATION

Section II of this report describes the Train Dispatching Simulation model used in developing the parametric analysis. Section III describes the development of the parametric analysis and presents graphs of single variant results. Section IV presents a procedure for applying the results of the parametric analysis to a wide variety of rail lines and includes the sample application. Section V discusses some result and presents some possible areas of further exploration. The reader who is not interested in the derivation or application of the parametric analysis need read only Section V to obtain the conclusions reached from this research. The appendix contains time-distance diagrams (stringlines) of the line operations of the base cases used in the parametric analysis.

II. THE TRAIN DISPATCHING SIMULATION MODEL

METHODOLOGY

The train dispatching simulation model is an event-based simulation that establishes a table of next events for all trains simulated. The model processes each event in time sequence, altering the status of train and track conditions according to the type of event and generating a new next event for each train. This approach appears to be far superior to a time-based logic in all but the most extreme volumes of trains, since relatively few events occur during the minimum time increment used for operation of the signals. A time resolution of one-tenth of a minute (six seconds) is used in the model and appears to be adequate for railway operations.[1] The following paragraphs describe the general program operation.

Initial Dispatch

The model enters a train into the system at the time specified by the user and begins to accumulate statistics at its time of dispatch. If the train cannot be dispatched at that time because there are other trains competing for main line facilities, the train is held in the siding or yard at its dispatching point.

Single-Track Operation

Two levels of control are used in the simulation:

. micro-resource or signal system control; and

. macro-resource or dispatcher control.

The micro-resource allocation responds like an automatic block signal system and controls train separation. Both the block spacing and the number of signal aspects used to space following trains can be specified by the user.

1/ The program takes advantage of PL/I's fixed-decimal notation by defining time to one digit to the right of the decimal, thus conserving data storage requirements.

The macro-resource allocation functions as a dispatcher, releasing or holding trains at stations (or control points) depending upon the logical and physical constraints of the rail line. The macro-resource control resolves conflicts between trains requiring the use of the sam line segment. The conflicts are resolved considering train priorities and physical characteristics, and line facility characteristics and ava ability. The dispatching procedure projects each train's movements a minimum of two sidings ahead of its current position in an effort to resolve potential conflicts. To replicate real-life operation, however, operations are not generally projected beyond this point, and actual train operation is not necessarily optimized overall. Within the confines of its foresight and the physical contraints of the track, a non-priority dispatching procedure releases trains on a first-come basis. A simulation using priorities gives preference to the higher priority train unless this would lead to an irresolvable conflict. These approaches tend to favor line capacity in the first case, and service to high priority trains in the second; other operating and priority rules could be used.

In an effort to increase train flow, the macro-resource allocation procedures fleet trains (i. e., allow them to follow closely on a single track). Trains are released under minimum following rules as controlled by the signal logic unless their release would cause an overflo situation at the downstream sidings. Depending upon the nature of the dispatching schedule, the necessity to resort to substantial fleeting in both directions on a single-track line can lead to eventual dispatching control failure or siding overflow where two fleets meet.

Multiple-Track Operation

Multiple-track operation provides for either single- or double-direction running of trains on each track. Automatic block signal control is used on double-track operations. The allowable directions of operation on a track may be changed for different segments provided the transitions are logically consistent. Trains are held at the end of double track and released to the single track when it is available in th same manner as trains are released from sidings or yards to single track. Trains reaching the beginning of double track are removed fro the single track provided the next block is available, with a delay whe appropriate to account for the slowdown in speed for the turnout.

Double-running tracks are used to allow faster trains to overtake slower trains. The model first searches for a route that will allow the overtake with no delay to either the overtaken train or to opposing trains. If such a route is unavailable, it will find a route which will delay only the overtaken train, provided it is lower priority. Failing that, it will delay opposing traffic if that traffic is of lower priority. Only as a last resort will it delay both the overtaken and opposing trains, and only if all are of lower priority.

Trains may be specified to originate or terminate on specific tracks. Permissible paths through interlockings can be restricted. When a train changes tracks, it incurs delays by slowing down for the crossover; it also blocks conflicting movements as it moves through the interlocking.. An interlocking may be taken out of service temporarily to represent at-grade crossings of other rail lines, drawbridges, maintenance time requirements or plant failure.

Trains may also be turned enroute to replicate local freight trains which run out and back or work trains which shuttle back and forth.

Termination

Certain conditions define the termination of a train and the completion of a simulation. Trains reaching their final destination are immediately removed from the system, and system statistics are adjusted accordingly. A train's arrival at its final destination is not subject to available yard or siding space at the terminal point. A train arriving at its final station and not, however, on the correct destination track will be delayed for conflicting movements at the interlocking and for the time to cross to the correct track.

The simulation is normally terminated when the desired period of simulation has elapsed. Other reasons for terminating the simulation include completion of all scheduled train operations, lockup of the system, and detection of error conditions. Any abnormal termination causes the status of the system to be displayed. Thereby, the situation causing the failure can be identified.

DATA REQUIREMENTS

The train dispatching simulation allows for virtually complete freedom in defining operating conditions and parameter values. The data requirements for the simulation are grouped into the following six categories:

- Basic Parameters -- The basic parameters include an alternative label for run identification purposes, an estimate of time lost when starting and stopping trains, the duration of the simulation period, and the numbers of parameters associated with the other categories of input.

- Track Configuration -- The line is described by mileposts and station names, which are located at sidings, interlockings, and other key points, and by the numbers of tracks and direction of movement allowed on each track in the segments between these mileposts. Siding and yard capacities (in terms of length and number of trains) and track connections at interlockings can be specified.

- Train Characteristics -- Train characteristics are described by class of train, number of locomotive units, nominal running times (without delays) between stations, and typical starting, stopping and crossover delays.

- Signal System -- The signal system is described in terms of the number of blocks in each segment, the number of signal aspects used for control, and the normal minimum desired following distance (in blocks) between trains.

- Dispatching Schedule -- Each train to be dispatched is described with a train identification label, length, origin, destination, initial dispatch time, train class, and priority. Wherever appropriate, values not specified by the user are assumed for parameters.

- Report Information -- Additional information can be provided to make the reports generated by the program more readable. Such information includes station names, activity descriptions, trains to be summarized, and categories of summaries to be used.

PROGRAM OUTPUTS

The simulation program produces five basic types of output:

- an "echo" print of input values and conditions for identification of the particular configuration being tested;

- a detailed movement record of all trains including cumulative running time and delay statistics;
- summary train statistics by train, train type, origin and destination showing total elapsed time, scheduled and unscheduled delays, and net running time;
- operating statistics and problem condition messages; and
- an optional time-distance (stringline) plot.

Error conditions detected on input or during program execution are also appropriately identified.

Train movements are reported by station, milepost, activity, day, and time of simulation. They can, therefore, be used to prepare stringline plots of the operations of the rail lines. The operating statistics presented indicate progress through the simulation and include the number of locomotives in the system for each hour, the maximum number required and the time at which the maximum occurs, plus other program informational messages. The optional stringline plot, which is useful for displaying results, is produced from the train movement record output by a separate program which generates a tape for a Calcomp plotter. In the plot, different tracks are represented by different line characteristics.

III. THE PARAMETRIC ANALYSIS

STRUCTURE OF THE ANALYSIS

The parametric analysis was structured in three stages. First, the parameters were defined and the ranges of practical variation we determined. Next, a structured approach to analyzing the possible variations and combinations of parameters was developed. Finally, a means of consolidating results and extending their applicability to the full range of interest was designed.

After data on the actual operations of several lines were examine the definitions and ranges of certain parameters (e.g., mix of trains were constrained because of the difficulty in developing a single measure of their variance. Others, such as timetable and train order operations, were not examined because of limitations of the model. Th individual parameters and the ranges of investigation are described i the next section.

The approach to the analysis was chosen from several alternative after considering the most likely applications of the results. Becaus of the large number of parameters and the virtually infinite number o possible combinations of values, even a comprehensive examination all reasonable combinations was too great a task. Thus, some type sampling approach was necessary. Since a thorough sampling was n possible within the scope of the project, a process starting from a "t ical" base case and then choosing a progressive set of variations was developed. First, a set of base cases was designed. Then a series cases with changes in only one parameter at a time was examined to develop the sensitivity of capacity to each parameter in relation to th base case. Finally, selected combinations of variations were tested to estimate the sensitivity to joint variations. Alternatives involving idealized base cases (either "best" or "worst") were rejected as star ing points because they are not reasonable representatives of existin rail lines. Results obtained using extreme base cases would probabl not be reliable for the mix of characteristics of typical rail lines. T base cases selected are hypothetical but representative of actual rail lines.

The basic approach to the parametric analysis was to define base cases for single and double track lines to represent normal utilization of capacity. Modifications to the base cases were made for those parameters judged to be potentially useful policy variables. Base cases and their modifications were translated into appropriate data input files, which described the line and trains to be run. The model was run for several different train volume levels for each base case and modified case.

The results of these runs show how average dispatching delays change with the increasing volume for each alternative set of conditions and how changes in the various policy variables impact upon average delay at a given volume level. The parametric analysis presents these results in a quantitative, unambiguous, usable form and generalizes them to any reasonable combination of the policy actions tested.

DEVELOPMENT OF BASE CASES

Three primary base cases, each for a 150-mile line, were identified:

- the single-track base case;
- the double-running double-track base case; and
- the single-running double-track base case.

Additional base cases were used for certain special analyses.

Single-Track Base Case

The single-track base case was intended to be a realistic case of moderate capacity utilization serving primarily as a foundation for improvement modifications but also serving as the base for realistic capacity decreasing changes. As shown in Table 1, there were 18 stations on the base case line, dividing the 150 miles of single track into 17 segments, with an average spacing of 8.82 miles and a standard deviation of spacing of 3.87 miles. This configuration was chosen after examination of station data for a number of typical existing single track lines in this country. At these stations were single sidings (except where noted), accessible to trains traveling in either direction. A siding can accommodate a train whose length is no greater than its own. Yards were included where substantial numbers of trains performed work.

TABLE 1

LINE DESCRIPTION

Station Number	Milepost	Siding Length (Feet)	Number o Siding Trac
1	0	10,500	Yard
2	1	7,500	1
3	6	5,500	1
4	17	10,500	1
5	30	7,500	1
6	39	10,500	Yard
7	50	5,500	1
8	64	7,500	1
9	70	10,500	1
10	85	7,500	1
11	89	10,500	Yard
12	99	5,500	1
13	111	10,500	1
14	118	5,500	1
15	128	7,500	1
16	135	10,500	1
17	144	5,500	1
18	150	10,500	Yard

A signal system with three indications was employed. The number of signal blocks assigned to a segment was based on an average block length of 1.6 miles. All fractional results were truncated, and every segment was assigned at least one block, regardless of length.

Four classes of trains were dispatched in the simulation runs. See Table 2 for train characteristics. Priorities were assigned based on the class of train with Class 1 as the highest and Class 4 as the lowest. One direction of travel was given priority over the other for trains of the same class. Class 1 represented passenger and other high priority short trains. Class 2 represented express freight trains. Class 3 represented ordinary freight trains. Local freight trains, with frequent stops, were represented by Class 4.

Time losses due to acceleration and deceleration are shown in Table 2. The stopping penalty includes time lost when passing through a siding. The crossover penalty includes both deceleration and acceleration when changing tracks in double-track runs.

In the base case, the same number of trains traveled in each direction. The dispatching schedule was repeated for each day of the simulation and every train was assigned to cover the entire length of the line. For simulations of two days (the predominant type of run), the base case utilized a train file containing 144 separate trains (72 trains per day). When all 72 trains per day were used, trains were dispatched every 20 minutes (every 40 minutes for trains in the same direction). For lower volumes of traffic, a less uniform distribution resulted. The train file was ordered so that it could be truncated at the desired number of trains per day (trains at the same time on different days being contiguous in the file) and still produce a fairly uniform distribution of train classes and dispatch times throughout the day. The class distribution of trains by time of day for the entire base case train file is shown in Table 3. In addition, for the stations listed, scheduled stops were specified for trains of the designated classes as shown in Table 4.

For the base case, a set of "rare events" (unscheduled incidents) was designed to simulate train and track failures and similar unplanned occurrences that might affect operations on a typical line. The rare events for the base case consisted of three types:

- a track segment out of service for a specified time (e.g., track maintenance time, which was chosen for minimum interference with train movement);

TABLE 2

TRAIN CHARACTERISTICS

Class	Length (Feet)	Average Running Speed (Miles per Hour)	Time Penalties (Minutes)		
			Starting	Stopping	Crossovers
1	1,500	50	1	2	1
2	3,000	40	2	6	3
3	5,000	25	4	8	5
4	3,000	25	2	5	3

TABLE 3

TRAINS BY CLASS AND DEPARTURE TIME FOR FULL BASE CASE TRAIN FILE

DEPARTURE TIME

HOUR	12 to 1 am	1 to 2 am	2 to 3 am	3 to 4 am	4 to 5 am	5 to 6 am	6 to 7 am	7 to 8 am	8 to 9 am	9 to 10 am	10 to 11 am	11 to 12	12 to 1 pm	1 to 2 pm	2 to 3 pm	3 to 4 pm	4 to 5 pm	5 to 6 pm	6 to 7 pm	7 to 8 pm	8 to 9 pm	9 to 10 pm	10 to 11 pm	11 to 12	TOTAL	Percent
Class 1						1	2	1						2				2	2						10	13.9
Class 2	1		1		2			2	1		1		2	1	1	1	2			1	1	1	2		20	27.8
Class 3	1	3	2	2	1	2	1		2	3	2	3	1		2	2	1	1	1	2	2	2	1	1	38	52.8
Class 4	1			1																				2	4	5.5
TOTAL	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	72	100.0

TABLE 4

SCHEDULED DELAYS
(MINUTES)

Station No.	Class 1	Class 2	Class 3	Cla
2	--	--	--	2
5	--	--	--	4
6	5	--	20	1
7	--	--	--	3
10	--	--	--	1
11	--	20	30	5
12	5	--	--	-
14	--	--	--	1
15	--	--	--	2
Total	10	20	50	20

- a train failure of a specified duration at a specified time; and

- a train, track, or signal failure of a specified duration for a train upon arrival at a specified location.

A different segment of track was taken out of service for three hours each day of the simulation. There were 18 instances of a failure of a train at a time for the entire train file (when all 72 trains per day were dispatched). There were 22 instances of a train or track failure at a location. Individual failures were either 10, 30, or 80 minutes in duration. The overall average incidence of such delays for the entire file was 8.96 minutes per dispatched train. No rare events were specified for Class 1 trains. The overall failure rate used was based on failure data presented in an MIT report on Railroad Car Movement Reliability.[1]

Run-specific data input to the simulation model included the duration of the simulation in simulated time and that portion of the period for which summary statistics would be calculated. In the base case and most other runs, two days were simulated and statistics were obtained for the trains dispatched during the middle 24 hours, thus minimizing network loading effects and ensuring that trains dispatched near the end of the summary period reach their destinations before the simulation was terminated.

The line characteristics and train schedules are representative of a moderate to low capacity single-track main line with a relatively uniform time distribution of trains. Although not the lowest possible capacity line, it represents one with a significant potential for improvement. Most of the cases analyzed were those with potential for improving capacity because they would provide the most useful basis for analysis. Generally, however, the base train speeds are probably lower than normally found on main lines, and failure rates higher than normal.

1/ A.S. Lang and R.M. Reid. <u>Railroad Car Movement Reliability: A Preliminary Study of Line-Haul Operations</u>. MIT Department of Civil Engineering. October 1970.

Double-Track Base Cases

The double-track base cases were equally as conservative on trai speeds and failure rates but were generally higher capacity facilities with double crossovers at every station. The double-track double-running base case included many of the same basic assumptions as th the single-track base case. Bidirectional signaling was used on both tracks. The sidings at stations 1, 3, 4, 8, 9, 12, 13, 16, 17, and 18 were removed as unnecessary on a double-track line. Because yards at the ends of the line were assumed to be on opposite sides of the lin trains had to switch tracks to go from one yard to the other.

The single-running double-track case is the same as the double-running version, except that all trains are generally confined to the track signaled for the direction in which they are running.

MODIFICATIONS TO THE BASE CASES

Each simulation fell into one of four categories:

- Base case.
- Variation from a base case for a continuously variable parameter. Such variations involved parameters or policy variables that could be easily and meaningfully varied along a numerical scale (e.g., changing the speed of trains, above and below the base values).
- Variation from a base case for a discrete parameter. Such deviations generally involved parameters that could not be easily or fruitfully varied along a continuous numerical scale (e.g., train class priorities).
- Joint or multiple variations from the base case for two or more parameters simultaneously. Such runs were made to test for interactions among parameters (e.g., increasing the speed and decreasing average block size simultaneously to note whether the impact on average dispatching delay is significantly greater or less than the sum of the individual impacts).

The cases actually simulated are presented in the next section. The descriptions of the modifications given below will permit the user to interpret the cases.

In addition, tests of running time versus track speed, profile, and train characteristics were made using the Train Performance Calculator. The results of these tests were effectively isolated from the TDS analysis by incorporating the effects of track speed, profile, and train weight and power in the speed parameter as explained later in this section.

Continuously Variable Parameters

Train Running Speed

Two types of speed parameters were tested:

- Proportional changes from the base speeds by train class. The changes were a 33-percent decrease and a 40-percent increase.
- Uniform speeds for all train classes. The speeds selected were 8.0, 25.0, 32.9 (the weighted average of the base case speeds over all classes), 50.0, and 70.0 mph.

Station/Siding Spacing

For the given 150-mile line, the number of stations (siding or crossover locations) were increased or decreased so as to alter the base case average of 8.8 miles between them. The values selected were:

- 31 stations with a mean station spacing of 5 miles;
- 11 stations with a mean station spacing of 15 miles; and
- 8 stations with a mean station spacing of 21.4 miles.

Average Signal Spacing

Two changes were made to the base case, which had a 1.6-mile average block length with three indication signals. They were:

- a 1-mile average block length, with a 4 indication signal system; and

- a 3-mile average block length.

Train Lengths

In the base case, all trains could fit into all sidings. Other cases were tested in which train lengths were altered in two ways:

- increased by a factor of 1.5 for each class of train (which produced some trains which could not fit in the shorter sidings); and

- doubled.

Double Length Trains in One Direction

The base case was changed by doubling the sizes of all trains in on direction and eliminating every second train from the schedule in that direction so that the car volume stayed the same.

Directional Imbalance of Train Dispatches

In the base case, the dispatching of trains over the line was approximately balanced by direction over the course of a day. The following two variations applied tested the impact on line capacity of dispatching more trains in one direction than the other over the course of the day.

- a 2:1 directional imbalance (two trains dispatched in one direction, one train in the other); and

- a 4:1 directional imbalance.

Partial Double Track

The base cases were modified by alternating stretches of single track and double track. The cases tested were labeled with the approximate ratio of single track segments to the total number of segments. (For analysis purposes these cases are quantified as the fraction of lin length which is double track.) In general, the distribution of double-running double track segments was made as uniform as possible in eac of the cases. The following cases were tested:

- 1 out of 3 segments single track (i.e., 70% of line length double track).
- 1 out of 2 segments single track (i.e., 53% double track).
- 2 out of 3 segments single track (i.e., 35% double track).

Discrete Deviations From Base Case

No Priorities

A set of runs was made in which priorities for all classes of trains and both directions of movement were the same.

Double the Number of Sidings at a Station

The number of trains of a given length that could be held by sidings was doubled for each station by adding additional sidings at each station.

Uniform Station Spacing

Although the base case had a given average station spacing over the line, individual stations were not uniformly separated by this average value. A case was constructed to test the impact of making the individual station spacings more uniform.

Train Dispatch Peaking

The impact of peaking train dispatches during the day was examined by using two peaking cases. Each had roughly 40 percent of the trains for a given day and in each direction dispatched during a four-hour period. The remaining trains were dispatched according to the base case throughout the rest of the day. The two cases were:

- separate peaks - four-hour peaks for the two directions, approximately 12 hours apart; and
- coincident peaks - peaks for both directions occurring during the same four-hour period.

No Rare Events

Runs were made for the base case, but without any of the base case rare events simulating train and track failures and track maintenance interruptions.

No Intermediate Block Signals Between Stations

This alternative tested the impact of having only one signal block between adjacent stations on single track.

Alternate Direction Crossovers

A double track double-running modification was tested with crossovers in only one direction (instead of two directions) at a station; the crossover direction alternated from one station to the next.

Additional Considerations

Two aspects of the modifications should be noted. First, some of the modifications involved more than the primary change(s) specified; and second, a special multiple modification of the single track base case was conducted.

Some changes require secondary alterations to be made for consistency or to reflect realistic operating conditions. When necessary, such changes were incorporated into the input data, but they were not specifically referred to in the modification labeling unless they represented substantial, non-obvious changes in their own right. Examples of such secondary alternations were changing rare event times when changing train dispatching times and increasing the incidence of rare events when using longer, more incident-prone trains. The second item of note involves the creation of the "ideal case." This case involves six major improvements of the base case in an attempt to simulate a very realistic, high-capacity single-track line. It is a useful foil to the modest capacity base case and suggests the upper limit of realistic single-track operations. The specific modifications used in this case were:

- a uniform 50 mph speed for all trains,
- no rare events,
- no priorities,
- short segments,
- uniform segments, and
- one-mile blocks with a four-aspect signal system.

EVALUATION OF THE RESULTS

Evaluation of the results suggested that the occurrence of system lockup (when the volume of trains exceeded the ability of the model's dispatching logic) was unstable, i. e., it was very sensitive to minor changes in schedule and line configuration. Thus, the occurrence of a system lockup itself could not be used to estimate actual line capacity. In actual practice, a dispatcher would probably be able to handle a larger volume of traffic but only by being much more cautious in his dispatching. This would result in a significant increase in delay as he held trains back until he could satisfactorily plan meeting points. The net result would be that the capacity in terms of the number of trains run, at least during peak periods, would be effectively reduced. It might be noted that line capacity is also a function of the amount of trackage assigned to a dispatcher. Current practice in CTC territory results in the assignment of about 6,000 to 8,000 train-miles of operations per day to one dispatcher. When this figure is exceeded or emergencies occur, the efficiency of train movement appears to deteriorate. Dispatchers who control territories with a large number of passenger trains operating on regular schedules may control more train-miles of operation. Fewer train-miles can be controlled by a dispatcher with timetable and train order operations.

Since true line capacity can be as much a function of the amount of time a dispatcher can devote to bottleneck areas as of the line and train characteristics, true line capacity cannot be defined as the ultimate logical ability to move trains. Therefore, capacity is, in this report, defined as a function of delay. To eliminate the element of available dispatcher time, it is implicit in this analysis that a reasonable number of dispatchers be assigned for the traffic to be handled. The relationship between the average dispatching delay per train and the train volume for a particular line was found to be far more stable than the presence or absence of a lockup. Thus efforts were focused on establishing a functional method for estimating the line capacity from this relationship.

Although the relationship between average delay per train and volume was not clearly linear for many cases, no other general functional form appeared to be consistently more appropriate. At least some of the nonlinearities appeared to be discrete jumps introduced by the specific prototype data used. For example, an increase in volume from 24 trains per day to 28 might involve the addition of one or more trains dispatched at particularly sensitive times in the schedule. Such discontinuities were reduced by performing regressions on a number of runs, all representing the same case but at different volume levels.

It was found that a square function also could be used and would provide a more conservative (higher) estimate of delay for higher traffic volumes. Where the availability of a dispatcher's time is a critical factor, this might be a more desirable curve form to use, since average delay to each train could be expected to increase as the square of the number of trains. Figure 1 shows curves which combine both the linear and square curves for several typical rail lines. While decreasing slope is generally indicative of increased capacity, other factors must be considered. The use of average delay for the single running double track line is particularly misleading. The other lines all represent situations where high priority trains receive relatively less delay than low priority trains. Since there are few opportunities for faster trains to overtake slower trains on single-running track, priorities cannot generally be honored and high priority trains receive excessive delays. While high priority trains can be expedited on double-running tracks, low priority trains incur even greater time loss than high priority trains save. Time lost for changing tracks further increases average delay for double-running. Nonetheless, the ability of double-running lines to honor priorities and the flexibility of such lines during periods of long on-track maintenance time or when train or track failures occur makes double-running much more desirable.

Since a linear form is easier to examine mathematically and since no consistently better functional relationship was found, it was decided to use a linear relationship between delay and volume which went through the origin. The squared relationship has been carried through much of the analysis, however, for those who might wish to use this form. A single parameter, the slope of the line relating delay per train to number of trains per day, or "delay slope," approximates the relationship for a given case. Delay slope is defined as:

$$\text{Delay Slope} = \frac{\text{Dispatching delay per train (hours)}}{\text{100 miles of line}} \div \text{trains per day}$$

On the average, the delay to a train increases linearly with the length of the line, and the number of trains dispatched. The delay slope values have been normalized to a 100-mile line and a 24-hour period of operation. Delay slope was estimated by a least-squares linear regression fit through the origin, since there would be no inherent dispatching delay with no trains.

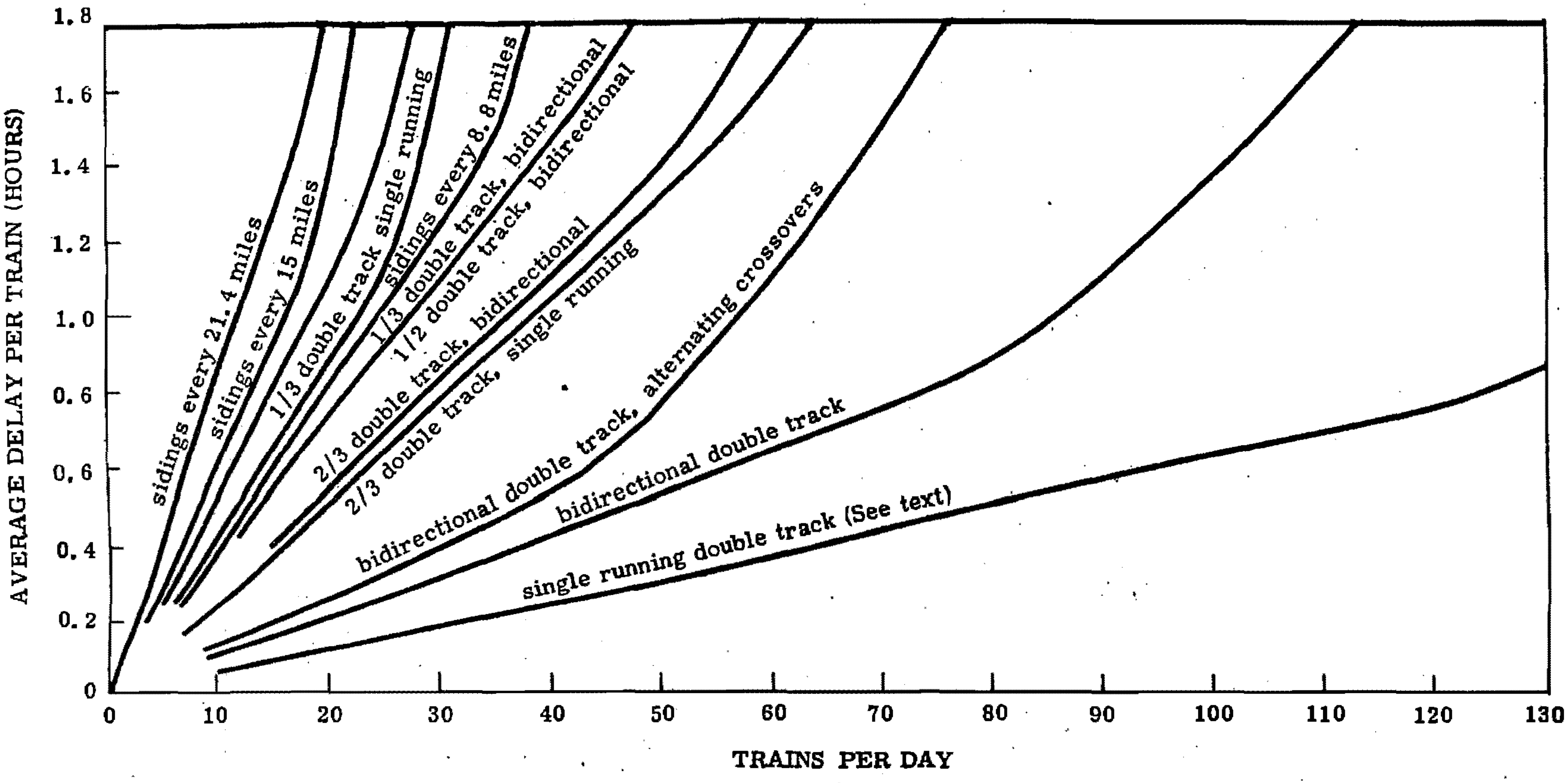


FIGURE 1: TRAIN VOLUME - AVERAGE DELAY RELATIONSHIPS FOR ALTERNATIVE CONFIGURATIONS OF 100-MILE RAIL LINE

For many of the cases (almost all of the single track cases), several runs were made at various volumes. However, after the form of the relationship had been established, a number of cases (especially those involving more expensive computer runs) were tested at only one volume level. The linear approximation was, therefore, the line from the origin through this point in such cases.

The results of the individual delay slope estimations for each of the modeled cases are included in the following section. A comparison of Figures 2 through 6 shows how the delay slopes change with several major policy variables which can be meaningfully represented along a numerical scale.

Figure 2 shows an inverse relationship between delay slope and train speed for uniform speeds. As speed increases, the average train delay asymptotically approaches zero. Conversely, decreases in train speeds cause a significant increase in dispatching delay. Figure 3 illustrates a similar relationship between delay slope and average train speed. In this case, however, the runs were with different train speeds for different classes, as in the base case. The average train speed is used to characterize the general speed level of the case. Data from both single track and double track runs are included in this figure. As expected, slope values are consistently lower for the double track runs than for the single track runs. As with the uniform speed cases, slope decreases with increasing speed. Unlike the previous graph, a curved shape is not evident. However, this lack of curvature may be due more to the smaller number of observations per case and smaller range of speed values in the second graph (note also that the vertical scales of the two graphs differ) than to any likely difference in the actual relationship involved.

Figure 4 illustrates the relationship between delay slope and average siding or crossover spacing on the line. For the single track line, average delay per train appears to increase fairly linearly with increasing siding separation, as might be expected. No clear relationship is evident from the limited number of observations for the double-running double-track case, over the tested range of spacing values (distance between crossovers). As always, the double-track delay slopes are much less than the corresponding single-track values.

The variation of slope with average signal block length for the three base cases is shown in Figure 5. In each case, because only three observations were made, the shapes of the curves are not necessarily

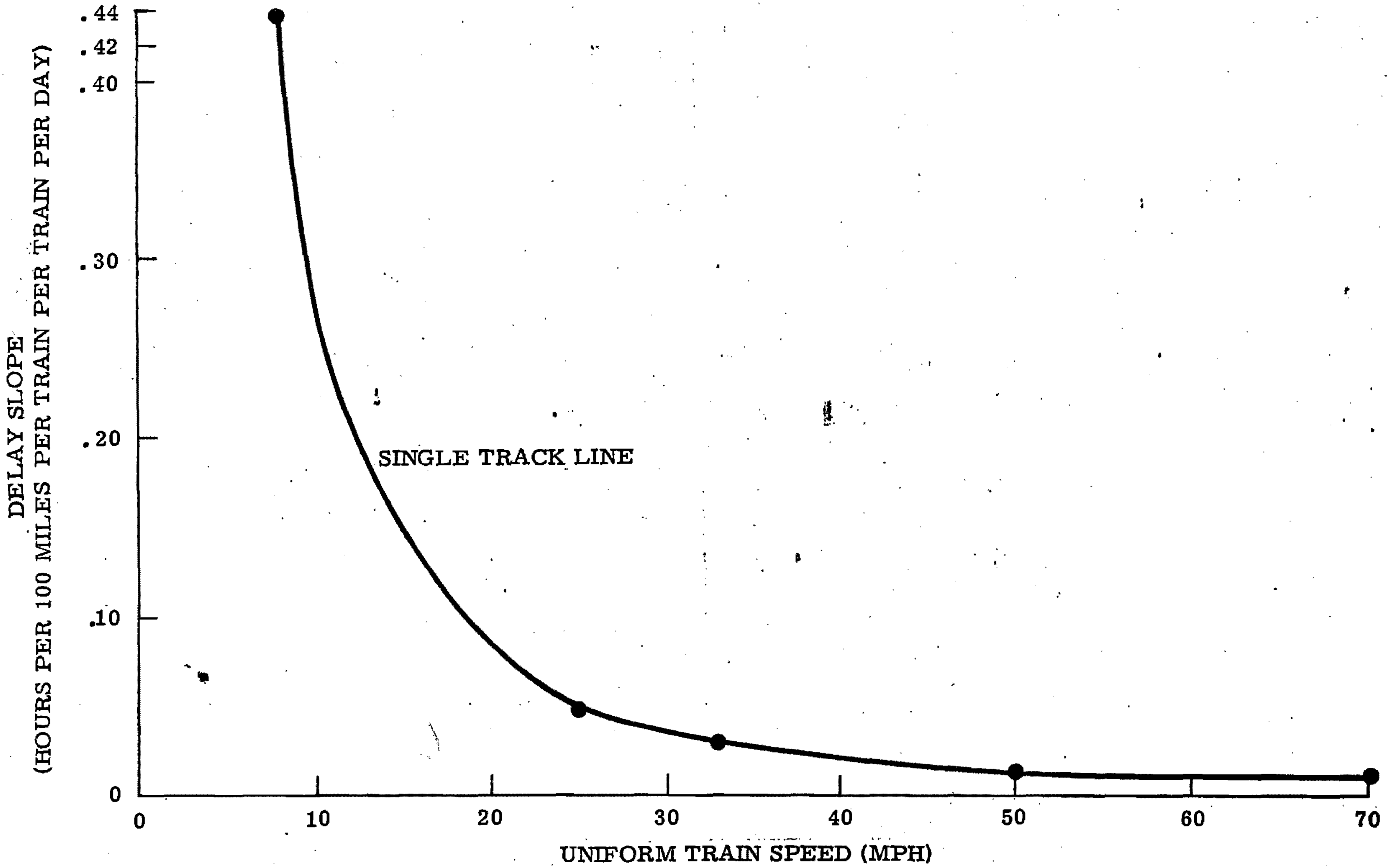
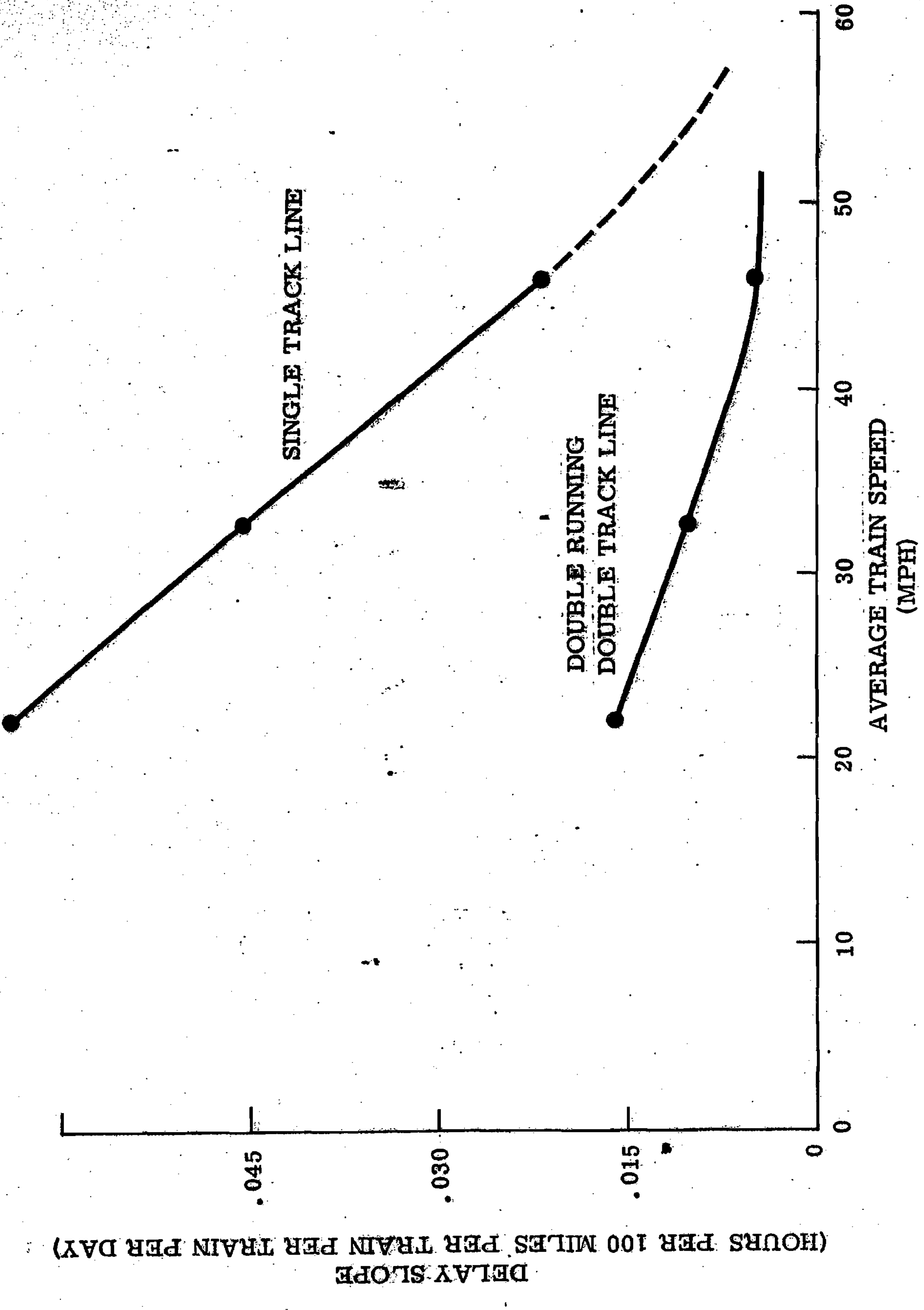


FIGURE 2: DELAY SLOPE VS. UNIFORM SPEED

SINGLE TRACK LINE
DOUBLE RUNNING
DOUBLE TRACK LINE
AVERAGE TRAIN SPEED
(MPH)
0
10
20
30
40
50
60
DELAY SLOPE
(HOURS PER 100 MILES PER TRAIN PER DAY)
0
.015
.030
.045

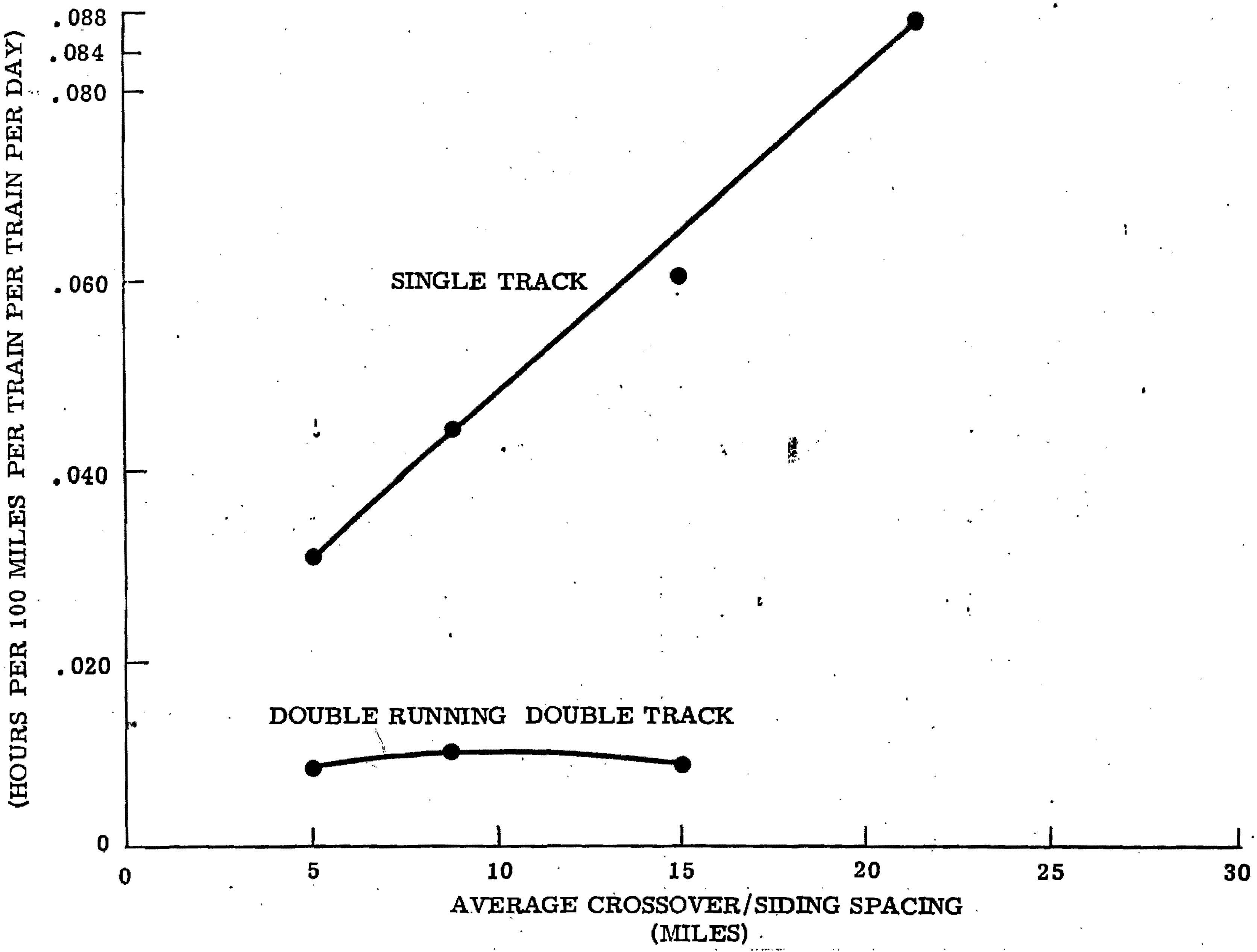
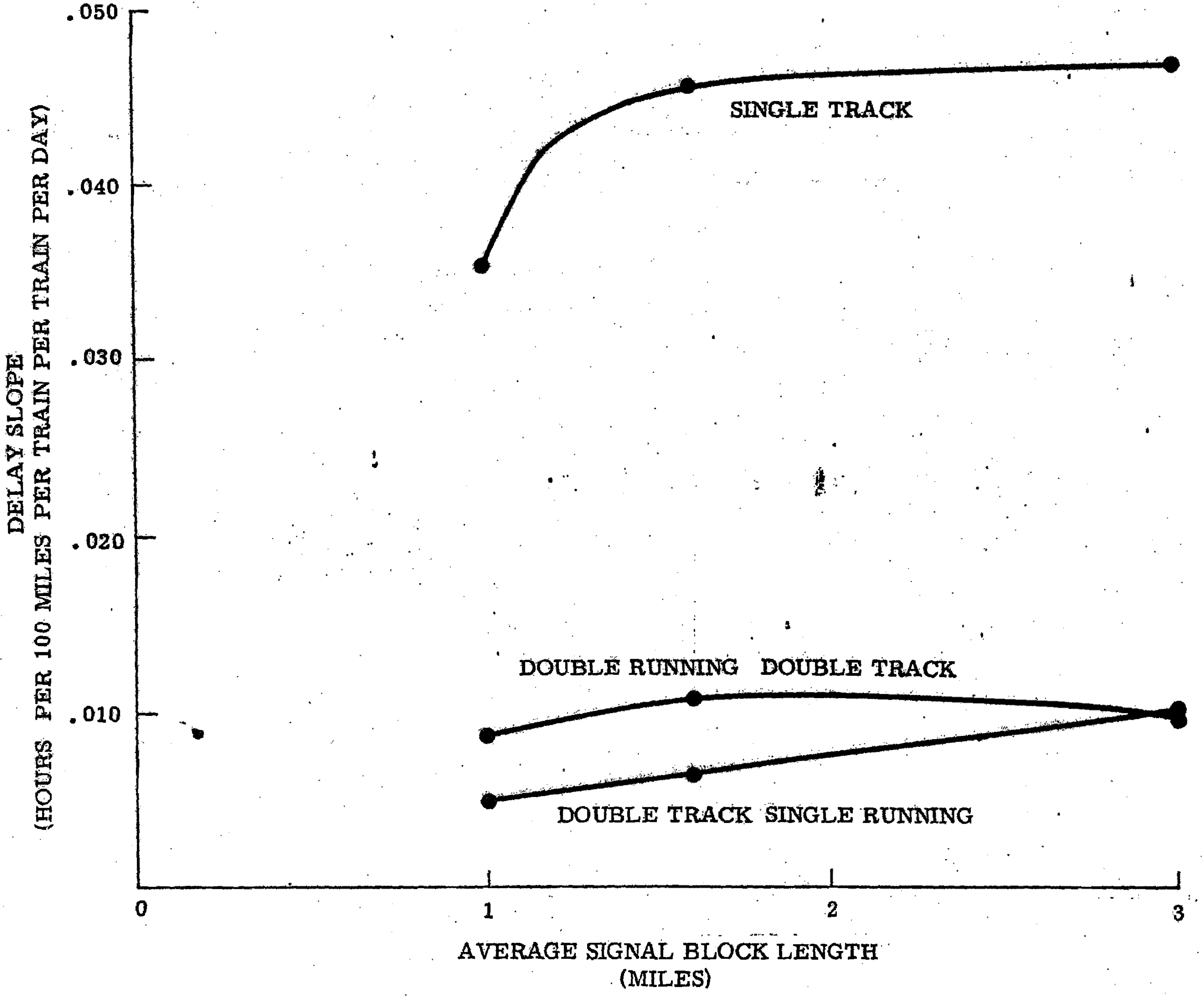


FIGURE 4: DELAY SLOPE VS. CROSSOVER/SIDING SPACING


.050
.040
.030
.020
.010
DELAY SLOPE
(HOURS PER 100 MILES PER TRAIN PER TRAIN PER DAY)
SINGLE TRACK
DOUBLE RUNNING DOUBLE TRACK
DOUBLE TRACK SINGLE RUNNING
0
1
2
3
AVERAGE SIGNAL BLOCK LENGTH
(MILES)

conclusive. However, both the single-track and double-track single-running base configurations show definite increases in delay slope with increasing block size. The slope for the single-track case appears to increase steeply between one and two miles and then more gradually, as signal block size approaches siding spacing. The increase appears to be more linear for the single-running double track case over the same range of values. For the double-track double-running case, the slope appears to reach a maximum around two miles, but the slightness of the peak and few points involved do not permit an assumption stronger than that of a relatively weak impact of block size on this configuration over the given range of values.

Figure 6 presents the relationship observed between delay slope and the fraction of the 100 miles of line which is double-track. The zero value is simply the single-track base case, whereas the upper limit of 1.0 represents the double-running double-track case. As expected, there is a strong inverse relationship between the delay slope and the fraction of line which is double track. The curve also suggests that rather small decreases in double track can have quite a large impact on congestion on a line that is totally or almost completely double-track. On the other hand, rather large increases in the double-track fraction must be made at the single-track end to achieve significant improvements (decreases) in slope values.

ANALYSIS OF PROFILE, SPEED LIMITS, AND TRAIN PERFORMANCE

A separate analysis was conducted of the impacts on line capacity of three factors--profile, speed restrictions, and train performance capabilities--that could all be reduced to an equivalent average speed parameter for inclusion in the parametric analysis. The relationship among speed, profile, and train performance capability (power-to-weight ratio, p/w) was determined. Then a procedure to consolidate this relationship with specific line profile and speed limit characteristics was developed.

The effects of profile and train performance capability on speed were determined using the train performance calculator. A series of tests were made to determine the limiting or balancing speed of typical trains with p/w's of 1.0 to 3.0 horsepower per ton (hp/ton), on grades ranging from -0.5 to +1.5 percent. The tests produced the series of

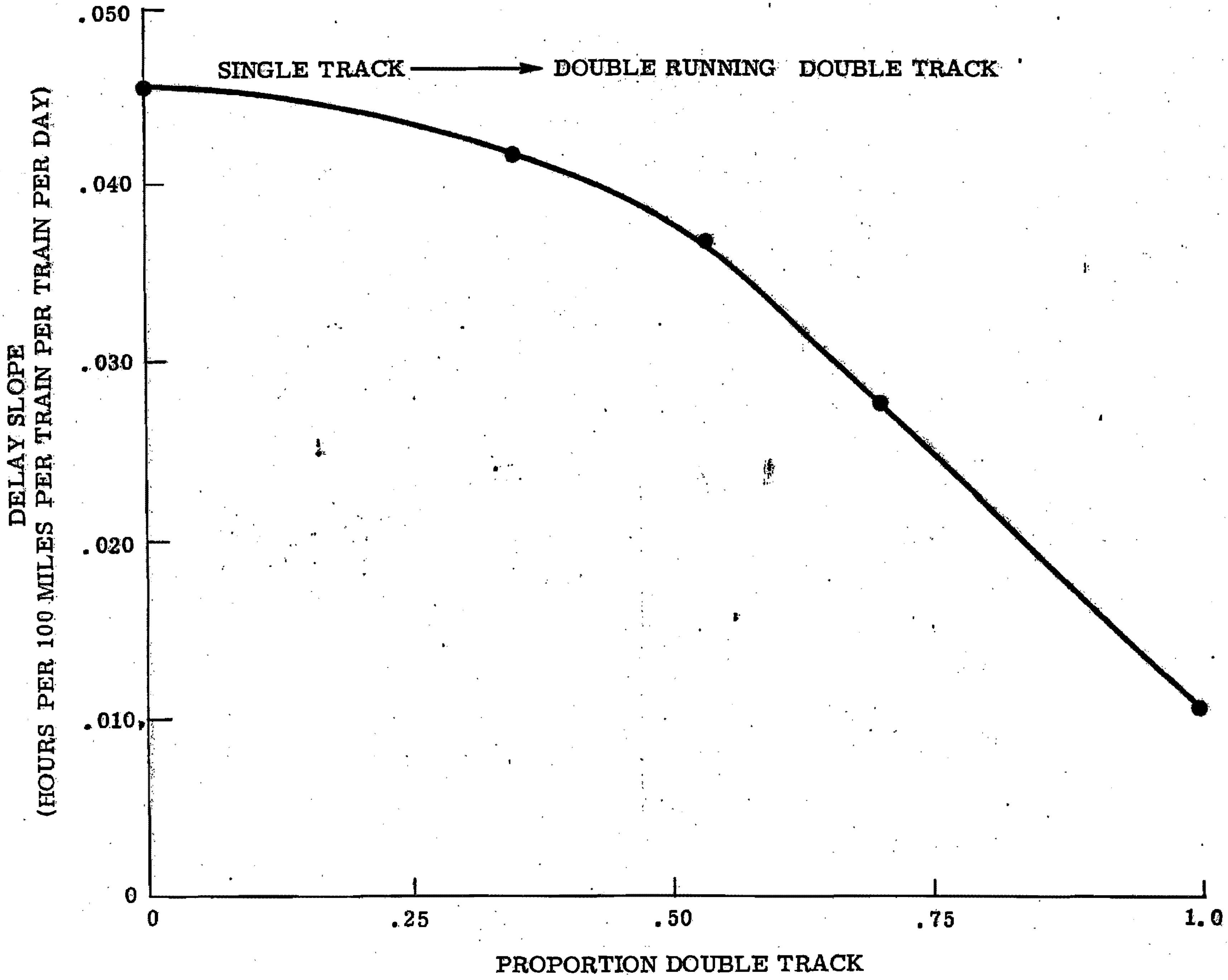


FIGURE 6: DELAY SLOPE VS. PROPORTION OF DOUBLE TRACK

curves in Figure 7. It should be noted that the train with the p/w of 3.0 hp/ton represents a TOFC train, while the others are typical freight trains. Consequently, the curves have slightly different characteristics.

A procedure which can be used to consolidated various line characteristics into an equivalent average speed is as follows:

- The line being analyzed is divided into segments with uniform speed limits for the class of train being analyzed.
- Each segment is subdivided again at each location where a significant change in grade occurs. The number of miles and fractional miles of each grade are added together. Generally, distances are to the nearest .05 miles and grades are grouped within a .05 percent range.
- All grades which would allow a train of a particular hp/ton to exceed the speed limit for the segment can be combined.
- For each grade which exceeds the grade at the speed limit, the running time is computed as follows:

  $$T_m = \frac{60}{S} \times L_g$$

  where:

  T is the running time in minutes,

  S is the balancing speed for the grade (from Figure 7),

  L is the length of line at this percent grade in miles.

  For segments where grades do not restrict speed to the speed limit, the speed limit is used for S.

- Running time over the entire segment is calculated by adding up the running times for each grade.
- Running times are added for all segments.

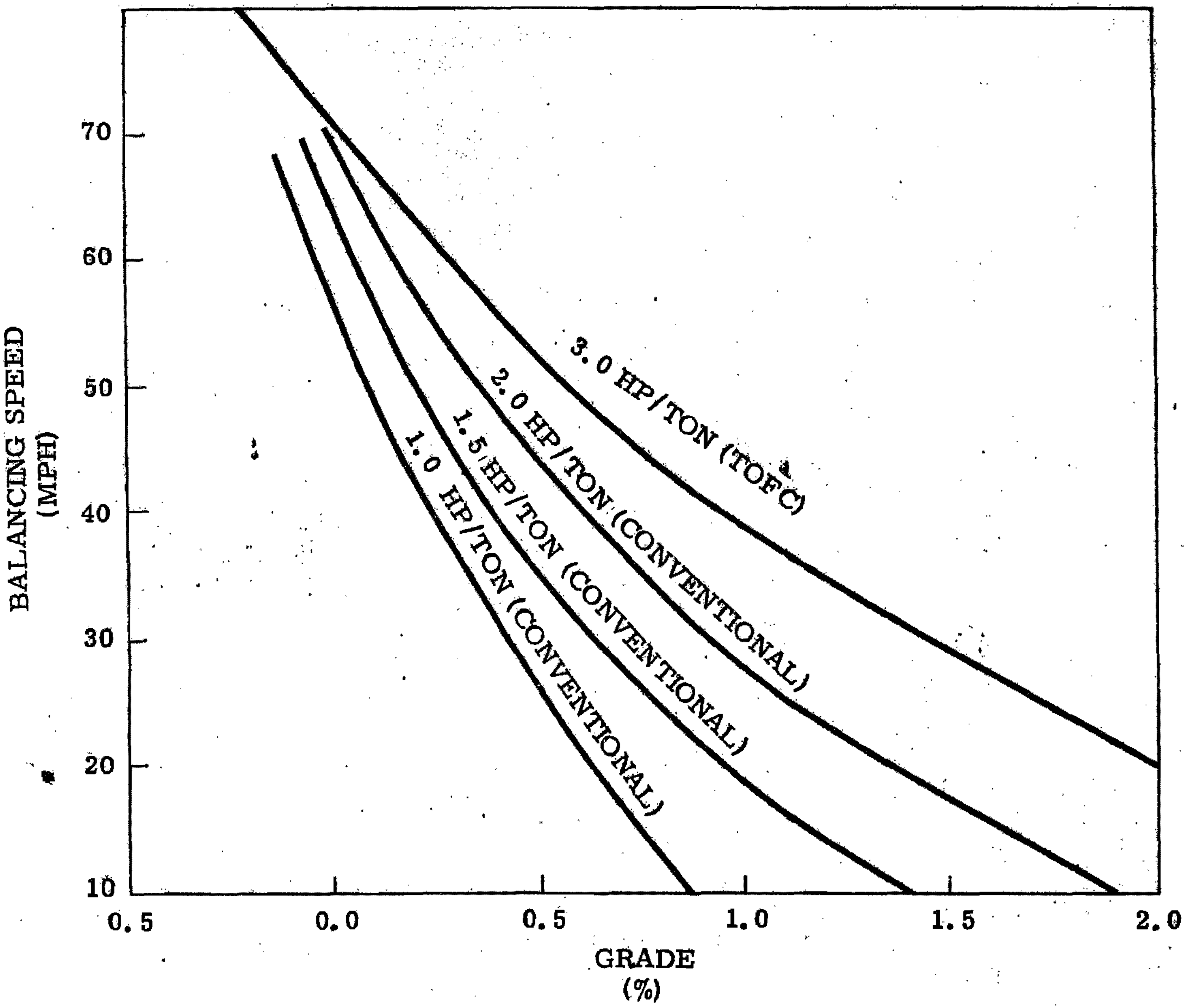


FIGURE 7: BALANCING SPEED OF FREIGHT TRAINS VS. GRADE

The inherent assumption in this process is that time lost while accelerating to a higher balancing speed will be offset by the time gained during deceleration to a lower balancing speed. On several lines of varying characteristics for which this procedure was tested, the results of run time estimates compared with TPC runs to within ± 1.5 percent.

Time lost due to stops should be added to any scheduled intermediate stops and to the run times calculated above. This consists of the stop duration plus acceleration and deceleration time losses. The latter time losses, when combined, appear to be essentially independent of grade, as shown in Figure 8. The primary factors appear to be running speed and hp/ton.

Nominal trip times developed by the above procedure can be converted to an equivalent average speed and the relationships between speed and delay can be used to estimate capacity according to the procedures in the next section. Care should be taken to analyze separately portions of lines which have significantly different average speeds.

The results of the parametric simulations discussed above were analyzed to develop a procedure which can be applied to typical American rail lines. The procedure and means of using it to analyze capacity and delay characteristics of rail lines are discussed in the next section.

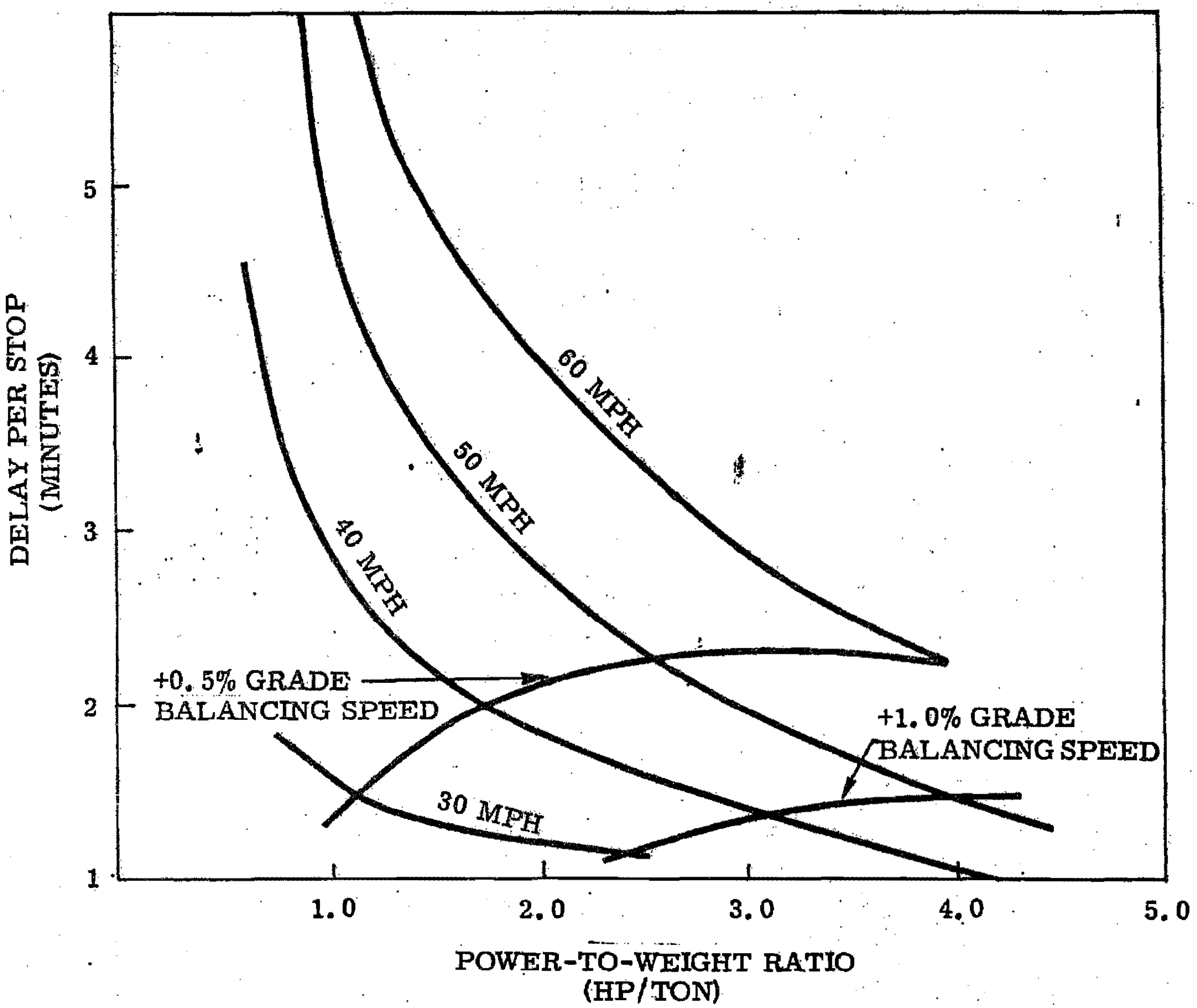


FIGURE 8: DELAY CAUSED BY STOPPING TRAINS

IV. APPLICATION OF PARAMETRIC ANALYSIS

This section presents the numerical results of the parametric analysis for the cases tested and the empirically derived procedures that will enable the user to generalize these specific case results to estimate line capacity for a wide range of single, double, and partially double track lines. Given an appropriately defined real or hypothetical rail line, the procedure enables a user to estimate the average dispatching delay per train at a given daily train volume, the maximum total running time for all trains of a given class at a given train volume, and the line capacity for conditions of constrained maximum total running time. The estimates obtained from applying these procedures to the parametric results of the simulation analysis are approximate; however, an examination of the overall results supplies a great deal of information on the relative impacts of various possible actions on line capacity and average delays. The procedures discussed in the following paragraphs generalize and extend the specific numerical expressions of these relationships obtained for specific cases via the simulation modeling.

ESTIMATION OF AVERAGE DISPATCHING DELAY

Delay Slope Coefficients

The primary relationship derived from the simulation runs was the one between the average dispatching delay per train and the daily train volume on the line. A simple linear relationship through the origin (net delay assumed to be zero at zero volume) appeared to be the most satisfactory solution to finding a simple functional form which would permit sample comparisons among various cases. The relationship could, therefore, be reduced to a single coefficient for a given case, representing the increase in average delay per train per additional daily train volume.

A least squares regression for a line through the origin was used to estimate the delay slope for each case tested. The average delay tends to increase more than linearly at high volumes because delay becomes infinite when capacity is reached. Therefore, a "square slope" coefficient, defining a linear relationship between average delay and the square of daily volume, was calculated for high volumes. The resulting relationship could be employed for very high volume and average delay levels. The procedures described below are valid for the application of linear slope values for moderate traffic volume cases. Once the delay slope for a case has been established, it can be used to estimate average delay for a given volume level or vice versa.

The basic relationship between average delay per train and the number of trains per day is:

$$\boxed{A = K_o n} \qquad (1)$$

where

A = average delay per train,

K_o = delay slope, and

n = number of trains per day.

Table 5 summarizes the values for K_o and other pertinent numerical results of the base cases and single modification model runs on a case basis; Table 6 summarizes multiple modification runs. On each line of the tables, following the case number and description, are the number of tracks (1 or 2) and a designation of double-running (D) or single-running (S) operation for double-track cases. The next column specifies the number of the case used as the base in the calculations described below. Most often, the base used is the appropriate primary base case (Nos. 1, 26, or 43); however, sometimes another case is chosen as a more appropriate base. In these cases, the fractional slope modification coefficients and related values (to be discussed below) pertain only to the net modification between the case at hand and the specified base. For example, when the 32.8 mph case (case 7) is used as the base for the 50 mph modification (case 8), only the overall speed level is changed. The change from mixed speeds by class to uniform speeds that would also have been involved if the primary single track base case were used as base has thus been excluded, and the change represents only a speed change.

The columns headed K_o and K_s present the linear and square slope coefficients for the various cases tested, adjusted for a 100-mile line. Where * appears in the K_s column, the number of runs made was insufficient for that case to justify the estimation of a least squares fit for a square coefficient. If the user wishes, he may approximate the value of K_s as 0.05 K, which has been demonstrated to be empirically valid for the values of K_o below 0.09.

TABLE 5

CASE SUMMARY OF SIMULATION RESULTS - SINGLE MODIFICATIONS

Case No.	Modifications From Primary Base	No. of Tracks D: Double Runs S: Single Runs	Base Case No.	K	K_s	P_i	f_{ci}
1	Single Track Base Case	1	--	0.04538	0.001867	-1	0.9450
2	5-mile Segments	1	1	0.03108	0.001324	-0.561	1.7752
3	15-mile Segments	1	1	0.06026	0.003625	+0.510	1.9486
4	21.4-mile Segments	1	3	0.08728	0.004935	+0.353	2.8556
5	Uniform Segments	1	1	0.03367	0.001360	+1	0.7897
6	33% Decrease in Speeds	1	1	0.06421	0.004277	-0.395	0.4154
7	40% Increase in Speeds	1	1	0.02226	0.000713	+0.333	0.1395
8	8 mph Uniform Speed	1	9	0.43867	*	-1.030	0.1124
9	25 mph Uniform Speed	1	10	0.04592	0.002761	-0.270	0.2140
10	32.8 mph Uniform Speed	1	1	0.03029	0.001119	+1	0.7062
11	50 mph Uniform Speed	1	10	0.01266	0.000435	+0.415	0.1221
12	70 mph Uniform Speed	1	11	0.00991	0.000261	+0.333	0.4799
13	1-mile Blocks, 4 Aspects	1	1	0.03515	0.001423	-0.462	1.5379
14	3-mile Blocks	1	1	0.04863	0.001919	+0.609	1.1475
15	1 Block Between Stations	1	1	0.12203	0.006365	+1	2.8890
16	Double Siding Lengths	1	1	0.03932	0.001317	+1	0.9170
17	1.5 Length Trains	1	1	0.04681	0.002020	+0.400	1.0806
18	Double Train Lengths	1	17	0.05609	0.004409	+0.286	1.8823
19	Double Length, One Way	1	1	0.05694	0.003516	+0.667	1.4053
20	Coincident Peaks	1	1	0.04179	0.002345	+0.824	0.9049
21	Separate Peaks	1	1	0.03329	0.002077	+0.824	0.8866
22	1:2 Directional Imbalance, No Rare Events	1	25	0.03189	0.001039	+0.667	0.7834
23	1:4 Directional Imbalance, No Rare Events	1	22	0.02563	0.000870	+0.667	0.7273
24	No Priorities	1	1	0.02981	0.001183	+1	0.6569
25	No Rare Events	1	1	0.03730	0.001540	+1	0.8219
26	Double Track, Double Run Base	2 D	43	0.01067	0.000137	-1	0.6029
27	2 in 3 Segments Single	2 D	28	0.04179	0.001235	-0.424	0.7436
28	1 in 2 Segments Single	2 D	29	0.03685	*	-0.271	0.3438
29	1 in 3 Segments Single	2 D	26	0.02759	*	-0.353	0.0677
30	5-mile Segments	2 D	26	0.00856	*	-0.561	1.4819
31	15-mile Segments	2 D	26	0.00840	0.000162	+0.510	0.6260
32	Uniform Station Spacing	2 D	26	0.00913	0.000164	+1	0.8563
33	33% Decrease in Speed & Uniformity	2 D	26	0.01645	*	-0.395	0.3343
34	40% Increase in Speeds & Uniformity	2 D	26	0.00547	0.000100	+0.333	0.1349
35	1-mile Blocks, 4 Aspects	2 D	26	0.00856	*	-0.462	1.8122
36	3-mile Blocks	2 D	26	0.00958	*	+0.609	0.8346
37	Coincident Peaks	2 D	26	0.01022	*	+0.824	0.9496
38	Separate Peaks	2 D	26	0.00767	*	+0.824	0.6698
39	1:4 Directional Imbalance, No Rare Events	2 D	41	0.00752	0.000092	+0.667	0.9716
40	No Priorities	2 D	26	0.00767	*	+1	0.7188
41	No Rare Events	2 D	26	0.00767	*	+1	0.7187
42	Alternate Direction Crossovers	2 D	26	0.01338	0.000308	+1	1.2520
43	Double Track, Single Run Base	2 S	26	0.00643	0.000052	+1	0.6029
44	2 in 3 Segments Single	2 S	45	0.05574	*	-0.875	0.3286
45	1 in 3 Segments Single	2 S	43	0.02630	*	-0.353	0.0185
46	40% Increase in Speeds and Uniformity	2 S	43	0.00276	*	+0.333	0.0804
47	1-mile Blocks, 4 Aspects	2 S	43	0.00522	*	-0.462	1.5895
48	3-mile Blocks	2 S	43	0.01011	*	+0.609	2.1023
49	Coincident Peaks	2 S	43	0.00856	*	+0.824	1.4139
50	No Rare Events	2 S	43	0.00467	*	+1	0.7257

* Not Calculated.

TABLE 6

CASE SUMMARY OF SIMULATION RESULTS - MULTIPLE MODIFICATIONS

Case No.	Modifications from Primary Base	No. of Tracks D: Double Run S: Single Run	Base Case No.	K	K_s	No. of Modifications by Direction	f_{om}	τ_f
1	40% Increased Speed, Very Long Segments	1	1	0.05053	0.002104	↓↑	1.1135	1.1156
2	No Rare Events, No Priorities, 32.8 mph	1	1	0.01993	0.000626	↓↓↓	0.4391	0.9462
3	50 mph, 1-mile Blocks	1	10	0.01416	0.000389	↓↓↑	0.4677	1.1549
4	50 mph, 1 Block Between Stations	1	10	0.04170	0.001461	↓↓↑	1.3768	1.1994
5	1:4 Directional Imbalance, No Priorities, N.R.E.	1	25	0.02302	0.000733	↓↓↓	0.6172	1.2206
6	No Priorities, 32.8 mph	1	1	0.02377	0.000891	↓↓	0.5237	1.0150
7	No Rare Events, 32.8 mph	1	1	0.03370	0.001106	↓↓	0.7427	1.2125
8	50 mph, No Rare Events, No Priorities	1	1	0.01239	0.000362	↓↓↓↓	0.2730	0.9688
9	Ideal Case	1	1	0.01037	0.000167	↓↓↓↓↓↓↓	0.2285	1.0087
10	1 in 2 Segments Single Track, 1:4 Directional Imbalance, N.R.E.	2 D	26	0.01852	*	↑↓↓	1.7361	1.0121
11	1 in 2 Segments Single Track, 40% Increased Speed	2 D	26	0.02130	*	↑↓	1.9964	1.1261
12	1 in 2 Segments Single Track, No Priorities	2 D	26	0.03204	*	↑↓	3.0034	1.2095
13	40% Increased Speed, 1-mile Blocks	2 D	26	0.00556	*	↓↓	0.5209	1.1436
14	33% Decreased Speed, Alternate Crossovers	2 D	26	0.03241	*	↑↑	3.0381	1.6938
15	33% Decreased Sreed, Uniform Station Spacing	2 D	26	0.01400	*	↑↓	1.3125	0.9942
16	Alternate Crossovers, 40% Increased Speeds	2 D	26	0.00861	*	↑↓	0.8073	1.2565
17	Long Segments, Alternate Crossovers	2 D	26	0.00870	*	↓↑	0.8160	0.8276

* Not Calculated

↑Modification which increases slope.

↓Modification which decreases slope.

The simulation runs have provided delay slopes for specific line cases. The next section shows how this information can be extended so that delay slopes can be estimated as a function of the base case K_o for other magnitudes of modifications from a base case and for combinations of modifications. This enables estimation of a delay slope for any case that can be expressed as a combination of tested modifications of a tested case (most validly, the appropriate "base" case).

Delay Slope Adjustment Factors

Approaches to Estimating Effects of Changes in Parameters

Two alternative methods were examined to estimate the effects of delay slopes of changes in parameter values. One method was based on the classical elasticity concept. Elasticity was defined as the proportional change in delay slope per proportional change in a parameter (e. g., the percent change in delay slope per one percent change in speed).

To ensure that the elasticity calculated would be the same whichever of the two cases (base or modified) were used as a starting point, an average of the "before" and "after" values of the parameter were used. With the elasticity approach, the effects of multiple modifications on delay slope were estimated by summing elasticities, factored by the magnitude of change involved for the several modifications. The resulting compound elasticity, when added to unity, was then multiplied by the appropriate base case delay slope to obtain the delay slope for the multiple modification case.

A second approach to estimating the effects of changes in parameter values treats delay slopes as fractions. Delay slopes for modified cases were developed as fractions of the base case delay slope. The fractions were normalized to a unit fractional modification by taking the p_i^{TH} root of the fraction, where p_i is the fractional change in the policy variable underlying the modification. Alternatively, this normalization could have been approached by dividing the signed difference between the slope ratio and 1 by p_i and adding the result to 1.

Suggested Method for Modifying Delay Slopes

The two approaches described above were evaluated by comparing (a) the delay slope adjustment factors calculated when each method was applied to individual modification coefficients to (b) the slope change actually observed in the multiple modification simulation runs. The ratio of these values for each case was denoted by τ_A for the first method and τ_F for the second method. A τ value of 1.0 indicated that the method exactly estimated the change in delay slope for a multiple modification, based on individual modification results. The second method, using fractional factors, was chosen over the elasticity method as generally producing τ values closer to unity.

The fractional approach defines the basic relationship between delay slopes for two different values of a parameter as:

$$\boxed{f_{oi} = \left(\frac{K_i}{K_o}\right)^{-P_i}} \tag{2}$$

where:

f_{oi} = the delay slope adjustment factor,

K_i = the delay slope for the change in parameter i,

K_o = the delay slope for the base case, and

P_i = the percent change in parameter i.

P_i is calculated as:

$$\boxed{P_i = \frac{(V_i - V_o)}{1/2(V_i + V_o)}} \tag{3}$$

where:

V_o = the value of the parameter in the base case, and

V_i = the changed values of the parameter.

Once values of f_{oi} are known, from simulations (see Table 5 and 6), equation (2) can be solved for K_i.

$$\boxed{K_i = K_o\left(f_{oi}\right)^{P_i}} \tag{4}$$

In Tables 5 and 6, the last columns summarize the numerical results of applying this method to the various cases. The column labeled p_i in Table 5 gives the fractional changes from their base cases for single modification cases. The column labeled f_{oi} gives the linear slope adjustment fractional factor. In some instances, several cases were investigated, representing modifications of different magnitudes of the same policy variable (such as speed). Slightly different f_{oi} values were obtained, reflecting the fact that the impact of a modification may not be constant on a percent change basis over the entire range of possible values. Thus, when more than one f_{oi} value exists for a given modification, it is best to apply the one corresponding to a policy variable level closest to that involved in the current calculations. For multiple modification cases, the value in the corresponding column of Table 6 is the appropriate f_{oi} value.

Table 7 presents the policy variables that correspond to the modifications tested and the units in which the V_o and V_i were expressed. The fact that p_i is a fractional quantity minimizes the possible influence of the choice of units on the results.

To combine the individual fractional factors, $\left(f_{oi}\right)^{P_i}$, in multiple modification cases, the method is more complicated. Two considerations control the method definition. First, slope values should never become negative because no matter how many beneficial modifications are made, it is not possible to operate with negative delay. Second, as noted from the tests of multiple modifications, for modifications that do not interact very strongly, simultaneous application of such improvements (or disimprovements) tended to achieve a combined impact which was not as strong as that predicted by a simple multiplication of the individual fractions (i.e., if two improvements each double capacity, both improvements together do not quadruple capacity).

The method which was defined yields a factor which is the net product of two components. One is calculated from all of the individual slope-increasing modifications and the other from all of the slope-decreasing modifications. These components were defined as unity if no modifications of the corresponding type are involved. The component for slope-increasing modifications is the sum of the corresponding individual fractional factors, raised to the power of the fractional modification involved, minus one less than the number of these factors.

TABLE 7

POLICY VARIABLES UNITS

Type	Modification	Policy Variable	Unit (V_i)	Base Value (V_0)
A	Change block size	Average block size	Miles	1.5 miles
B	Change train priority	Train priority	No priority: 3/2 Base priorities: 1/2	1/2
C	Change station spacing	Average segment size	Miles	6.82 miles
D	Select uniform or non-uniform speed	Train speed uniformity	Base speeds by class: 1/2 Uniform speeds: 3/2	1/2
E	Change uniform speed	Uniform train speed	mph	32.8 mph
F	Change proportional speed	Average train speed	mph	32.8 mph
G	Change siding capacity	Siding capacity	Base capacity: 1/2 Double capacity: 3/2	1/2
H	Select uniform or non-uniform segments	Segment uniformity	Non-uniform: 1/2 Uniform: 3/2	1/2
I	Select dispatch peaking or non-peaking	$\frac{\text{Fraction daily volume in peak}}{\text{Fraction of day in peak}}$	Peaking fraction	1
J	Select rare events or no rare events	Presence of rare events	Rare events: 1/2 No rare events: 3/2	1/2
K	Change train length	Train length as fraction of base length	Train length as fraction of base length	1
L	Change directional imbalance	$\frac{\text{No. of trains in heavy direction}}{\text{No. of trains in light direction}}$	Directional imbalance fraction	1
M	Select base blocks or 1 block between stations	Same as Modification	Base block configuration: 1/2 1 block between stations: 3/2	1/2
N	Select full crossovers or alternate directional crossovers	General double track crossover flexibility	Full: 1/2 Alternate: 3/2	1/2
P	Change fraction double track	Fraction of line mileage with double track	Double: 1 1-in-3 single: .7 1-in-2 single: .533 2-in-3 single: .3467 single: 0	0 or 1

The component for slope-decreasing modifications is exactly analogous, except the exponentially adjusted fractions are inverted before adding, so that the fractions added are always greater than or equal to one. The slope-increasing component is then divided by the slope-decreasing component to obtain the final multiple modification factor. This factor is multiplied by the base case slope to obtain the slope for the modified case. For an observed multiple modification, m, a factor, f_{om}, can be calculated:

$$\boxed{f_{om} = \frac{K_m}{K_o}} \tag{5}$$

where K_m is the delay slope for the multiple modification case. The estimated value of f_{om}, synthesized from the individual component modifications of m would be:

$$\boxed{\hat{f}_{om} = C_I \, C_D^{-1}} \tag{6}$$

where C_I is the component for factors which increase the slope and C_D is for the factors which decrease the slope. C_I can be defined as:

$$\boxed{C_I = \left(\sum_{\substack{f_{oi} \geq 1 \\ i \in m}} f_{oi}^{P_i} \right) - (N_I - 1)} \tag{7}$$

where N_I is the number of slope-increasing modifications. Thus

C_I equals one if N_I is zero.

Conversely, C_D is defined as:

$$\boxed{C_D = \left(\sum_{\substack{f_{oi} < 1 \\ i \in m}} f_{oi}^{-p_i} \right) - (N_D - 1)} \tag{8}$$

where N_D is the number of slope-decreasing modifications, and

C_D equals one if N_D is zero.

By calculating the ratio of the actual to the synthesized multiple modification factor:

$$\tau_F = \frac{f_{om}}{\hat{f}_{om}}$$

it is possible to determine the degree to which the synthesized factor can be used to estimate the actual factor.

When a multiple modification is considered for which f_{om} is not available, $\hat{f}_{om}$ can be used, by assuming that τ_F is acceptably close to 1. The new delay slope, K_m can be estimated by substituting (7) and (8) in equation (6), equating to (5), and solving for K_m.

$$\boxed{K_m = \left[\left(\sum_{\substack{f_{oi} \geq 1 \\ i \in f}} f_{oi}^{\;p_i}\right) - (N_I - 1)\right]\left[\left(\sum_{\substack{f_{oi} < 1 \\ i \in f}} f_{oi}^{\;-p_i}\right) - (N_D - 1)\right]^{-1} K_o} \quad (9)$$

In Table 6, a set of arrows in the last column indicates the number of slope-increasing and slope-decreasing modifications involved in the cases tested, each having their respective p_i's. The final column, labeled τ_F, gives the performance ratios for the multiple modification cases tested. A τ_F larger than 1 indicates that the fractional approach to synthesizing multiple modification slope values from individual modification values would underestimate the actual run results in this particular case. The opposite is true for τ_F values smaller than 1. The f_{oi}, f_{om} and τ_F values in Tables 5 and 6 were calculated for the linear slope coefficients only. However, the same procedure can be applied to calculate the corresponding values for the "square slope" coefficients.

Examination of the column headed τ_F shows the accuracy with which the fractional approach could reproduce the observed results. Values range from 0.82 to 1.69, although most are between 1.00 and 1.25 and the average is 1.12 for the 17 multiple modification observations. No particular pattern is apparent in these errors of estimation. The implication of these limited observations, if representative, is that this method on average underestimates the impact of multiple modifications by about 12 percent. The variation of these τ_F values about 1 was significantly less than that for the corresponding τ's calculated using the elasticity method. This was a major factor behind the choice of the fractional over the elasticity approach.

Additional multiple modification runs might yield a pattern in the fluctuation of τ_F values with the types of individual modifications involved (reflecting interactions among individual modifications). If this pattern were consistent, it might provide the basis for estimating τ_F's, according to the combination of modifications involved, to be used as an adjusting factor. Lacking such detailed information and given that the τ_F's are not too different from 1.0 in the cases tested, the second method is presented with the assumption that a τ_F equal to 1.0 can be used.

ESTIMATION OF LINE CAPACITY

A number of definitions of capacity were considered in attempting to develop the most useful definition. Ultimate capacity, where absolutely no more trains can be forced through the line, is too unstable and dependent upon precisely how trains are scheduled and what failures occur. An economic capacity, where an optimal balance between operating and capital costs would occur, is not within the scope of the project and would probably be too site specific for a general analysis such as this. Other possible definitions, such as an arbitrary percent delay of total running time or an operationally stable capacity where a line could recover from a disruption in service of moderate length (e.g., 4 hours) and return to normal service levels, were also rejected as too arbitrary or unstable. The most useful and stable definition appears to be one based on the maximum allowable time for the most delayed train to traverse the line. It was discovered that maximum time could be related to average delay and would allow the user to define capacity constraints based on either minimum level of service (maximum acceptable trip time) or minimizing the need to recrew trains because of the 12 hour on-duty time limitation imposed by the Hours of Service Law. It should be noted that since the parametric runs were designed to represent "typical day" operations, this approach would not eliminate all trains that exceed the time limit. Unusual delays or catastrophic failures could still result in some trains exceeding the time limit.

In order to incorporate this definition into the estimation procedures being discussed, it was necessary to establish a relationship between the average dispatching delay and the maximum total running time for a normal freight train (Class 3) in the simulation runs. The relationship of primary interest was between average delay and maximum total running time of a Class 3 train, the slowest group of line haul trains. It was also found that for single-track cases, train peaking and directional imbalance were factors that had to be taken into account separately in specifying the relationship.

After experimentation with several functional forms using stepwise multiple regression on the data from individual simulations, it was decided that the primary relationship could best be specified as that between maximum total running time and the square of average delay multiplied by speed.

AVERAGE DELAY - MAXIMUM TRIP TIME RELATIONSHIP

This section presents the final regression equations relating average delay to maximum trip time for single-track and double-track cases generalized for lines of any length. The formulas are then solved for average dispatching delay in terms of all other variables so that average delay at the time limit can be estimated. Once this average delay has been calculated, the delay slope can be used to estimate the capacity for the time limit. The basic equation for capacity is:

Linear Delay Relationship

$$C = \frac{A_c}{K}\left(\frac{100}{L}\right) \tag{10}$$

where:

C = capacity of the line in trains per day,

A_c = average delay per train (in hours, exclusive of scheduled delays),

K = delay slope (for a 100-mile line), and

L = the length of the line in miles.

To determine A_c the following equations were developed. Since the basic relationship between maximum trip time and A_c was square, a quadratic solution was necessary for the single-track case. Since only the positive value of A_c is a reasonable solution, the quadratic formula gives us:

Single Track (11a)

$$A_c = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

$$a = 0.04325(S)\left(\frac{150}{L}\right)^2 = \frac{973.125\ S}{L^2} \tag{11b}$$

$$b = \left(\frac{150}{L}\right)(0.44851\ P + 1.01139\ D) = \frac{1}{L}(67.2765\ P + 151.7085\ D) \tag{11c}$$

$$c = 1.41432 - M\left(\frac{150}{L}\right) + \frac{150}{S} + I \tag{11d}$$

where:

M = the maximum allowable total running time (12 hours less allowance for terminal time),

S = the speed of slowest class of through freight trains (mph),

P = the dispatch peaking factor: $\left(\frac{\text{trains peak hour during peak}}{\text{trains peak hour off peak}}\right) -1$

D = the directionality factor: $\left(\frac{\text{trains in dominant direction}}{\text{trains in opposite direction}}\right) -1$, and

I = the amount of imposed delays on regular freight trains (such as required stops, including the start and stop lost time).

Double Track

$$A_c = 0.031274\ L\sqrt{\frac{1}{S}\left[M\left(\frac{150}{L}\right) - \frac{150}{S} - I - 1.84636\right]} \tag{12}$$

Once A_c is calculated with the appropriate formula for a given line and maximum running time for a freight train, line capacity is estimated using equation (10).

A SAMPLE APPLICATION

The following hypothetical case demonstrates the use of the above procedures and parametric results to estimate line capacity and average delay. The case is a multiple modification of the single-track base case. The modifications are as follows:

1. uniform speed of 55 mph for all trains,
2. 4:3 directional dispatching imbalance,
3. average block length of 1.9 miles,
4. average siding spacing of 11 miles, and
5. line length of 200 miles.

Modification 1 is actually two separate modifications: (a) the base speeds are made uniform for all train classes and (b) the speeds are increased from a base average of 32.8 mph to 55 mph. These are treated separately for a more accurate estimate.

The resulting six individual modifications will be referred to as 1a, 1b, 2, 3, 4, and 5, respectively. The first step is to calculate the delay slope for this modified case, by applying the appropriate adjustment factor to the base case slope value of 0.04538 from Table 5. Note that line length only becomes a consideration in the later calculations.

Referring to Table 7 for the correct units:

	Modification 1a	Modification 1b	Modification 2	Modification 3	Modification 4
	Type D	Type E	Type L	Type A	Type C
V_o	1/2	32.8	1	1.6	8.82
V_m	3/2	55	1.33	1.9	11

Applying formula (3), we obtain the fraction changes in the parameters:

$$P_{1a} = \frac{3/2 - 1/2}{1/2\ (1/2+3/2)} = 1$$

$$P_{1b} = \frac{55 - 32.8}{1/2(32.8+55)} = 0.506$$

$$P_2 = \frac{1.33 - 1}{1/2(1+1.33)} = 0.283$$

$$P_3 = \frac{1.9 - 1.6}{1/2(1.6+1.9)} = 0.171$$

$$P_4 = \frac{11 - 8.82}{1/2(8.82+11)} = 0.220$$

Modifications 1a, 1b, and 2 are slope-decreasing modifications, whereas Modifications 3 and 4 are slope-increasing. Thus, N_I and N_D in formulas (7) and (8) are 3 and 2, respectively. Appropriate single modification adjustment fractions are obtained from Table 5.

If none of the f_{oi} values in the table corresponds to the modification desired, the f_{oi} for a modification value closest to the desired value should be used. Although f_{oi} values are not independent of the magnitude of the change, they are independent of the sign of the change, as long as the appropriate base is used for the respective f_{oi}. The f_{oi} for the five basic modifications are:

f_{o1a} = 0.7062

f_{o1b} = 0.1221

f_{o2} = 0.7834

f_{o3} = 1.1475

f_{o4} = 1.9486

Substituting into formulas (7) and (8), we obtain:

$$C_I = 1.1475^{0.171} + 1.9486^{0.220} - (2-1)$$

$$= 1.1819$$

$$C_D = 0.7062^{-1} + 0.1221^{-0.506} + 0.7834^{-0.283} - (3-1)$$

$$= 3.3855$$

Then, from formula (6):

$$\hat{f}_o = (1.1819)(3.3855)^{-1}$$

$$= 0.34911$$

Finally, from formula (9) (which incorporates some of the above intermediate steps), we get:

$K_F =$ (0.34911)(0.04538)

$K_F =$ 0.015843 hours of delay per train per 100 miles of line

This delay slope makes it possible to estimate average delay at a given daily volume for the line (after making the adjustment for the 200-mile line length). However, in order to estimate the line's capacity, specification of the relationship between average delay and maximum running time is needed.

Because this is a single track line, formula (11) is used. For this line:

S = 55 mph

P = 0

D = 3/2 - 1 = 0.5

L = 200 miles

I = 1.233 hours (delays imposed in the simulation cases, assumed not to change in this modification)

For the purposes of calculating line capacity, a maximum total running time, M, or 10 hours will be assumed. Substituting into formulas (11a) through (11d),we get:

$$a = \frac{973.125(55)}{(200)^2} = 1.338047$$

$$b = 1/200\ [67.2765(0) + 151.7085(0.5)] = 0.379371$$

$$c = 1.41432 - (10)\left(\frac{150}{200}\right) + \frac{150}{55} + 1.233 = -2.125407$$

$$A_c = \frac{-0.379371 + \sqrt{(0.379371^2 - 4(1.338047)(-2.125407)}}{2(1.338047)}$$

$$A_c = \frac{-0.379371 + 3.394028}{2(1.338047)}$$

$$A_c = 1.127$$

Thus, at capacity, the average dispatching delay on this line is about 1.1 hours. The estimated number of trains at capacity for this line, obtained from formula (10) is:

$$C = \frac{1.127}{0.015843}\left(\frac{100}{200}\right) = 36 \text{ trains per day}$$

In analogous fashion, the parametric results can be applied via the procedures described above for other cases of interest. To estimate average delay at any given volume, or capacity volume for the line of interest, it is necessary to approximate the line's characteristics by a combination of tested modifications from one of the tested base cases. The parametric results (summarized in Tables 5 and 6) can also be visually examined for the relative impacts on average delay and capacity of various policy alternatives.

VALIDATION

Three validations of both the simulation model and the parametric analysis were performed. In addition to providing a measure of the accuracy of the two procedures, the validations provided useful insight into some of the considerations necessary to apply the procedures. The three rail lines against which the procedures were validated are:

- the Mattoon-Lennox, Ill. portion of the Penn Central north line (formerly the New York Central main line) betwen Indianapolis and St. Louis;

- the Penn Central main line between Boston and Selkirk, N.Y.; and

- the Canadian National main line between Sioux Lookout and Transcona (Redditt subdivision, east of Winnipeg).

Each has distinctly different traffic and track characteristics. Simulation validations for the first two lines were performed early in the project (before many model logic refinements were incorporated).

It was not possible to rerun the validations after the parametric analysis was completed because of limited funds. In all cases, it was not always possible to reconstruct precisely all delays from train sheets since dispatchers did not usually record every minor delay, particularly those due to restricting signals. Busy days were purposely chosen in an attempt to observe the lines closer to capacity. Transcribing errors when copying from train sheets were also a possibility.

Mattoon-Lennox

This 110-mile portion of the Penn-Central north line, which is located between Indianapolis and St. Louis, is single track, CTC with predominantly westbound movements. The Chicago and Eastern Illinois Railroad (C&EI) has trackage rights into St. Louis for the 71 miles between Pana and Lennox. Trains average about 33 mph, and the slowest show little variation, averaging about 30 mph. The characteristics of the line which were deemed to be significantly different from the single-track base case are:

- eleven-mile station spacing,

- average block length of 2.4 miles,

- directional imbalance of 5:1, and

- mostly uniform speeds.

The first two parameters are slope-increasing factors yielding a net increase in delay slope of 1.21; the second two are slope-decreasing factors yielding a decrease in delay slope of 0.515. Thus, the delay slope, K, for this line is:

$$K = 0.623 \times K_{base} = 0.0283$$

Data collected from dispatchers' sheets for the three-day period of February 7 to February 9, 1974 showed 18 trains the first day, 25 on the second, and 23 on the third, for an average of 22 trains per day. Not all of the trains operated the entire length, however, with C&EI trains operating only on the western end and a daily local turn on the eastern end. Thus, the average train-mile/track-mile density was 19. Using this value, average delay was calculated as:

$$A = 0.0283 \times 19 \times \frac{110}{100} = 35.5 \text{ minutes per train}$$

This compares with simulated and observed dispatching delays of:

Observed: 43.3 minutes per train

Simulated: 31.3 minutes per train

Parametric Analysis: 35.5 minutes per train

Estimating capacity according to the parametric analysis was considerably more difficult. At first, it appeared that the limiting trains would be the through trains (Penn Central's Indianapolis-St. Louis run, 230 miles), which use one crew and, thus, must cover this 110-mile segment in less than 6 hours. When this is used as a constraint, a capacity of nine trains per day is calculated. This would imply a very large number of enroute "recrews" (changing crews enroute for hours-of-service reasons), which was not the observed condition. Only one through train was actually recrewed. This inconsistency was apparently because westbound trains were given priority over eastbound trains. Thus the eastbound trains, primarily C&EI trains, suffered most of the delay.

If the maximum trip time observed for an eastbound train (5.6 hours for 71 miles) is used to estimated capacity, a value of 24 trains per day is calculated. The 5.6 hours was observed on the second day when 25 trains were dispatched. Other limiting values for eastbound trains could be used -- 8 hours produces a capacity of almost 40 trains per day. It was concluded that not only should the slowest trains be considered when estimating capacity, but also the lowest priority trains should be examined, since they generally incur the most delay.

Boston-Selkirk

While portions of the primarily double-track Boston and Albany division are CTC, significant amounts of single direction signalling occur. Except for a few passenger trains, the train speeds were essentially uniform (30 mph east of Springfield, 35 mph west). Therefore, it was decided to use the single-running double-track base case for comparison. The two significantly different parameters are:

- mostly uniform train speeds, and
- incidence of rare events six times the base case.

Since the rare events parameter was treated like a discrete parameter, it was necessary to assume that the discrete function developed could be extrapolated as though it were continuous. Using this procedure, a slope-increasing factor of 2.23 was developed for rare events, and combined with a 0.6029 factor for uniform speeds. This calculation led to an increase in the delay slope of:

$$K = 0.00643 \times 2.23 \times 0.6029 = 0.00864$$

The average observed density of trains for the two-day period analyzed (June 11 and June 12, 1974) was 23 train-miles per track-mile. Thus the average dispatching delays were:

Observed:	27.6 minutes per train
Simulated:	40.7 minutes per train
Parametric Analysis:	23.0 minutes per train

The discrepancy between the simulated and the observed values is at least partially due to coding errors that were never corrected because of funding limitations. If some consideration for the partial CTC had been made, the delay estimate from the parametric analysis would have been slightly higher.

Estimating capacity again produced some useful observations. The first attempt to define capacity by the slowest trains to traverse the entire route resulted in an estimate of zero capacity (i.e., it couldn't be done). The slowest trains averaged:

30 mph for 193 miles = 6.4 hours
Enroute work time = 2.0
Average dispatching delay = 1.9
10.3 hours

This average was within the 12-hour limitation, even if an hour is allowed for originating and terminating the train. However, since capacity is defined for the average delay at which no train is likely to outlaw (i.e., its crew exceeds the 12 hr. limit of the Hours of Service Law), the implication is that some of the slow trains would need new crews. In fact, several recrews were observed during the five days for which data were collected. With the high incidence of rare events and the normal randomness of dispatching delay it, in fact, was likely that some trains would outlaw. For those trains with less work enroute (1.5 hours), capacity was estimated at 30 trains per day (vs. 23 actually observed).

Sioux Lookout-Transcona

The Canadian National Railroad (CN) independently validated the simulation model to determine its ability to replicate operations on its main line. The subdivision used was substantially different from the single-track base case:

- higher average speed (37 mph),
- moderately uniform speeds,
- 8-mile station spacing,
- moderately uniform stations,
- very low incidence of rare events,
- one signal block between stations, and
- a few shorter sidings.

Since uniform station spacing and uniform speeds were treated as discrete parameters in the parametric analysis, the values for V_i for this line were estimated as halfway between those for the base case and those for the simple modification case. Since sidings were closer and more uniform than in the base, an 8-mile block length rather than single-block-per-segment value was used. Delay slope was calculated as:

$$K = K_{\text{base}} \times 1.164 \times 1.932 = 0.0274$$

At 25 trains per day, average delay is:

Observed (manual analysis):	1.70
Simulated:	1.77
Parametric Analysis:	1.66

Capacity is calculated as 23.5 trains per day although no outlawed trains were either observed or generated in the simulation. The actual line, however, considered to be very close to capacity by the CN, is of somewhat lower quality than that modeled since it has modified CTC, in which most sidings have power switches on only one end (the other end is a spring switch). This may work to a capacity advantage, since spring switches primarily limit opportunities for overtakes, thus speeding lower priority trains along.

V. CONCLUSIONS AND RECOMMENDATIONS

This section summarizes some of the characteristics of rail line capacity and discusses the potential usefulness of parametric analysis. Recommendations are presented on both verifying the more tentative conclusions and extrapolating the analysis into other potentially fruitful areas of investigation.

CONCLUSIONS

Previous sections have presented the numerical results of the simulation and a parametric method for applying the results to estimate capacity for a wide variety of rail lines. Some concluding observations should be made on the implications of the results.

The f_{oi} values in Table 5 can be used to indicate the relative sensitivity of average delay per train, and hence capacity, to the various parameters. Care should be used in interpreting these, however, since the relative magnitudes in some cases depend upon the rather arbitrary measures of change in some of the discrete or less quantifiable parameters; other measures of change could affect the magnitude of the f_{oi} values. When values of f_{oi} are less than 1.0, they represent average delay decreases with increasing parameter values, and when greater than 1.0, average delay increases with increasing parameter values. Thus, if an f_{oi} less than 1.0 is inverted, it can be compared with other f_{oi} values to determine the relative sensitivity of delay or capacity to changes in values of its parameter. Each of the parameters, together with some observations about its importance to rail capacity, is discussed below. Special considerations for the application of the results are also discussed.

Train Speed

The most important single parameter over the entire range of values examined is train speed. The f_{oi} of train speed ranges from 2.41 (inverted) for a 33 percent decrease in all train speeds to 8.90 for increasing uniform train speed from 8 to 25 mph. These values represent an elasticity of almost -2 (i.e., a 1 percent decrease in speed results in a 2 percent increase in average delay). This relationship holds for both single and double track and for uniform and non-uniform train speeds. Uniformity of speed for some given average speed may be equally as important in the determination of capacity. For a change from the distribution of speeds assumed in the base case to a uniform speed of the same overall average, an f_{oi} of 1.42 was calculated. Arbitrary values of 1/2 and 3/2 were assumed for V_o and V_i respectively, because of the difficulty in developing a representative measure of uniformity of speeds. If it were argued that this case actually represents a relatively small change in uniformity of speed, a substantially higher effective f_{oi} would result. For example, if V_o were assumed to be 1.0 and V_i were 1.5, then the effective f_{oi} would be 5.69. This has the same degree of significance as changes in train speed.

Siding Spacing

Line capacity is generally less sensitive to siding spacing than it is to speed. Values for f_{oi} range from 1.78 to 2.86, representing an average elasticity of about 0.5 (i.e., a 1 percent increase in siding spacing results in about a 1/2 percent increase in delay). It is interesting to note that the greater the average siding spacing, the more sensitive capacity is to siding spacing. Conversely, for short spacing, other factors tend to dominate. Uniformity of siding spacing is not nearly as important as average spacing unless, again, it is argued that the arbitrary V values used are inappropriate. If V_o is 1.0 and V_i is 1.5, which implies that the change in uniformity is not as great as that assumed in the analysis for Table 5, the f_{oi} would be 3.25 or about the same degree of importance as siding spacing.

Siding Capacity and Length

Line capacity in terms of number of trains per day is relatively insensitive to the doubling of siding capacity, if the capacity is supplied as parallel tracks (side-by-side). Even if the V values are changed substantially, this result is not affected. Doubling siding length, however, would effectively double the length of train which could be easily accommodated and, in theory, double the capacity in terms of cars per day. Other considerations, however, such as grades, yard lengths, and the significantly higher failure rates of longer trains must be considered. These alternatives are discussed under the heading Sidings and Train Lengths later in this section.

Signal Block Length

Signal block length does not appear to be a strong determinant of capcity; however, in the case of segments with no intermediate blocks, considerably higher sensitivity can be seen. This is a result not only of the greater average block length but also of the variability of block lengths. The f_{oi} for the single block case is calculated only on the change in block length and would be lower if a factor for variability were included.

Proportion of Single Track

The greatest sensitivity of capacity to any parameter tested can be seen in the introduction of small proportions of single track into a double track line, but the sensitivity diminishes rapidly as the proportion of line that is single track increases. This high sensitivity is in part due to the way the line was defined for the analysis. Trains in one direction were assumed to reduce speed at the end of double track to change to the single track, and vice versa. If high speed switches had been assumed at the end of double track, delay, and thus sensitivity, would have been reduced.

Crossover Spacing

Although not as significant as siding spacing on single track lines, crossover spacing on double track lines was next in importance after speed and block length in determining capacity. Some of the results are conflicting and tend to indicate that this parameter was underestimated in the basic variation tests. The most extreme underestimations of joint impacts of parameters occurred in tests involving crossover spacing and configurations. This is probably because only one or two points were developed for most double track runs, due to the greater expense of such runs and the general familiarity at that time in the analysis with the shape of the delay curve. Further, the base case design tended to mitigate the need for effective crossover location, since tracks were removed from service for maintenance only during slack periods of traffic, as should be done for planned maintenance. Crossovers, however, are most important when failures occur during critical periods. With these factors in mind, it should be noted that the sensitivity to crossover spacing was only marginally less than that to block length. Uniformity of spacing and alternate direction crossovers were somewhat less important, although the two most extreme cases of underestimating sensitivity to joint variations occurred with alternating direction crossovers. Note that this factor is less important with uniform train speeds, which had fewer overtakes. No tests for crossover impact were prepared for single direction running, since the model does not accurately replicate the train order operation required when running in reverse of the assigned direction on a track.

Sidings and Train Lengths

The impact of trains too long for sidings was very difficult to analyze because of the necessity to develop a single measure of the various proportions of trains which would fit in various proportions of sidings. Three specific test cases were examined in which the proportions of trains which would fit in sidings were varied. That is, for the double train length relationship, all the trains would fit in 41 percent of the sidings, 47 percent of the trains would fit in 70 percent of the sidings, and 14 percent would fit in all the sidings. No single scalar measure could be developed to represent this kind of relationship, therefore arbitrary pseudo-measures which could be used in combinational situations, but not as measures of change in train and siding length matching, were developed. Train length was used directly as the measure of change for the 1.5 and double train length cases. Train length was also used directly for the doubling of train lengths in one direction only, but since only half of the trains were modified (and to balance flows, the number of trains was halved in the long train direction), the measure is not truly comparable, and thus the values of f_{oi} are not comparable. The implications of the results are still quite clear, however, since the doubling of train length, even for only one direction, had a much more severe impact on capacity than increasing train length by 1.5. Thus it can be concluded that it is far more preferable to have most sidings long enough for all the trains, even if a few sidings cannot accommodate most of the trains, than to have most of the trains short enough to fit all the sidings, if the remaining trains cannot fit most of the sidings. A few much-too-short sidings are better than a few much-too-long trains.

Train Power and Weight

It was concluded that train power and weight could best be incorporated in the capacity analysis through their relationship to train speed over a segment. The power-to-weight ratio may be a strong determinant of train speed on lines with speed limits in excess of train capability on grades. Speed, as noted previously, is the strongest determinant of capacity. No specific direct relationships among power, weight, grades, speed limits, and capacity were developed, however, because of the multi-dimensional nature of the relationship. The processes described in Section III can be used to relate power, weight, grade, and speed, and those in Section IV to relate speed and capacity.

Train Priorities

Removal of all considerations of priority when dispatching trains on the base case line had a considerable effect on delay, reducing it by a third, thus increasing capacity by 50 percent. Again, no easily quantifiable measure for the use of priority considerations in dispatching trains could be identified because of the mix of priorities among trains and the priority rules which could be observed. Thus, the P and f are arbitrary and applicable only for situations similar to those used in this study. It should be noted that the priority rules used were not very rigid since trains were given absolute priority only two segments ahead of their current movement. If complete absolute priorities had to be used, delays would have been greater, and savings due to removal of priority constraints even more significant.

Traffic Patterns

Two aspects of traffic patterns were examined--peaking and movement directional imbalances. Interestingly, both peaked traffic patterns and directional imbalances reduced average delay per train. The presence of coincident peaks (traffic peaks simultaneously in both directions) only marginally reduced delay, but separate peaks (first one direction, then the other) significantly reduced delay. It must be emphasized that, particularly in the case of coincident peaks, the reduced delay does not translate into increased capacity. The distribution of delays is much less uniform than in the base case, and trains operating during the peak period, especially those of low priority, have great difficulty covering the line within the prescribed time limit. With separate peaks, the entire line tends to operate first in one direction, and then the other, much like the directional imbalance cases. While this is very beneficial for the expeditious movement of trains in the heavy volume direction, the few remaining trains in the opposite direction have great difficulty progressing against traffic, particularly if they are of low priority. Again, necessity to operate all trains within the Hours of Service law time limit restricts capacity. Similar patterns were observed on both single track and double direction running double track. With coincident peaks, the average delay is reduced slightly, but capacity is reduced substantially when the two opposing fleets meet. The lower priority trains are severely affected and thus have great difficulty in making the trip within the maximum time constraint.

Incidence of Service Disruption

Removal of all unscheduled incidents, or rare events, reduced delay by about 20 percent, and increased capacity by about 25 percent. Total removal of all disruptions is probably impractical under today's operating conditions, and thus these values tend somewhat to overstate the impact of incidents. Further, the incident rates used (about one in seven trains delayed an average of 40 minutes during a 150-mile trip) are probably high for many lines, further overstating their impact. Three somewhat compensating factors should be noted, however. First, the delay due to the incidents themselves is not included in these delay estimates, only that encountered by other trains when they are delayed by trains which incur the incidents. The average delay to the trains encountering incidents is constant regardless of the volume of traffic, and thus was subtracted from the total. Second, only minor incidents were represented, with a maximum duration of 80 minutes. Catastrophic disruptions were not analyzed, since these should not be used to determine average delay or typical capacity. The ability of various line configurations to absorb catastrophic disruption and maintain a level of capacity depends strongly on the nature and location of the disruption. Third, no attempt was made to reflect the increased track maintenance requirements necessary to decrease incidents or to accommodate a higher volume of traffic. To the extent that removal of track from service is required, the increased maintenance necessary to reduce incidents may tend to offset the delay due to the incidents.

SUMMARY OF CONCLUSIONS

The following definite conclusions evolve from the tentative conclusions discussed above. These relate to:

- the ability of existing rail lines to absorb traffic increases;
- the impact of operating speed on capacity;
- the impact of Centralized Traffic Control (CTC) on capacity for multiple track; and
- the estimation of potential reserve capacity of a line.

The ability of rail lines to absorb considerable increases in traffic without major changes in line or operating characteristics must be questioned. Line capacity was found to be considerably less than widely believed. Capacity is not so much a function of the capability to move trains over a line at all, as it is of the ability to move trains over a line without undue delay. Delays generally exceed acceptable limits before a line will lock up.

The most important parameter in determining capacity, other than the number of tracks, is operating speed. Theoretical capacities for single and double track can only be approached as trains are run at moderately high uniform speeds. The greater the distribution of train speeds, the more the interaction among trains and the greater the delay.

It is interesting to note that CTC on double track is essentially unnecessary to reach theoretical capacity since overtakes would be unnecessary if all trains operated at the same average speed. In fact, CTC may actually increase average delay under normal operations. Its primary usefulness, other than to provide flexibility in the event of track blockage, is to increase the level of service to high priority trains. This improvement in service comes at the cost of increased delay to lower priority trains.

This work can be used as a guide for investigation of alternatives for improving capacity of a line. It can also be used to estimate the potential reserve capacity of a line. Because of its approximate nature, any marginal capacity results should be investigated in depth using detail procedures such as the train dispatching simulation.

RESEARCH RECOMMENDATIONS

Two areas of further research are suggested by the results of this project. First, more detailed analyses should be conducted on those parameters for which results obtained were not conclusive. Second, several additional areas of research should be considered.

More detailed analysis should be conducted to:

- expand the level of detail for the continuous variables, especially those for which anomalies were detected;
- develop continuously variable representations for those parameters treated as discrete in this analysis; and

- develop further the procedures for performing multiple modification analyses, including expanding the understanding of factors affecting the values if the fractional method is used.

New areas of research which should be considered include:

- extending the analysis to three or more main tracks;
- determining the impact of increased on-track maintenance with higher traffic volumes;
- examining the impact of major disruptions on long-term capacity, including the nature of recovery from such disruptions; and
- investigating train order operation parametrically to provide a reference point for comparison.

APPENDIX

BASE CASE TIME-DISTANCE DIAGRAMS

TIME-DISTANCE DIAGRAMS

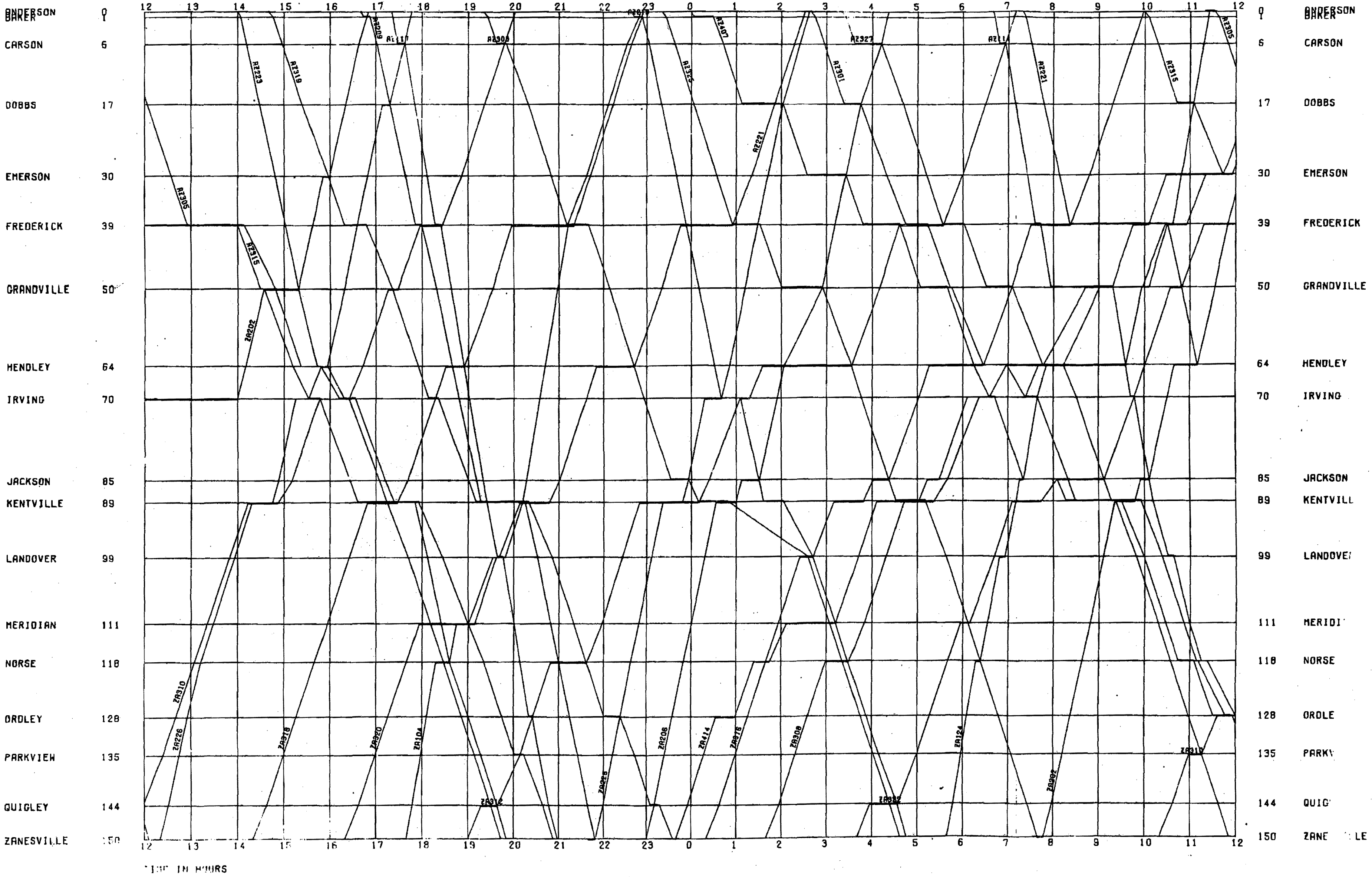
Single-Track Base Case

Single-Running Double-Track Base Case

Double-Running Double-Track Base Case

All best parameters (ideal) Single-Track Case

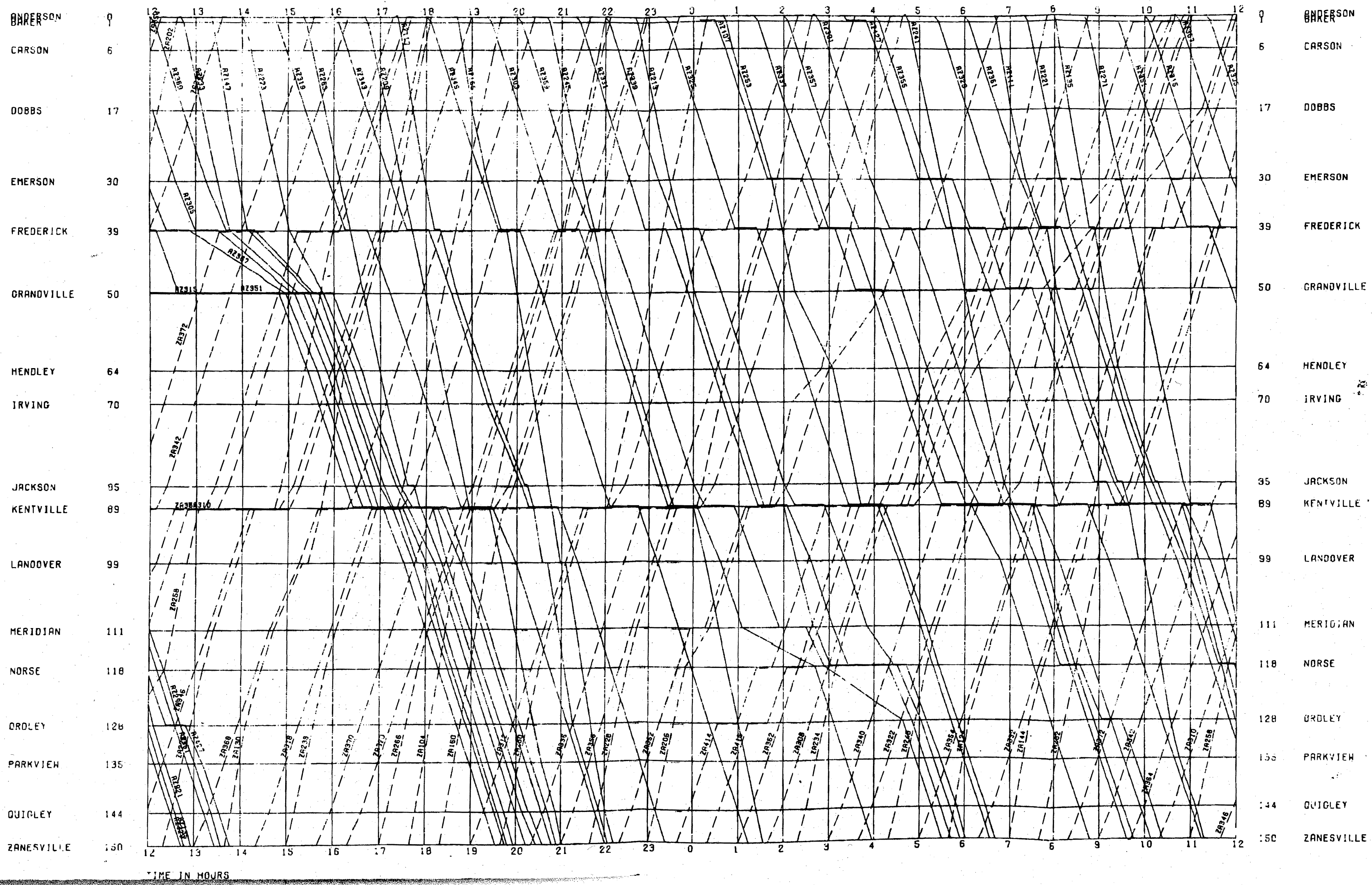
PMM & CO.
TRAIN DISPATCHING MODEL
BASE SINGLE TRACK CASE W/EVENTS;AVE. SEG.



PMM & CO.

TRAIN DISPATCHING MODEL

BASE DOUBLE TRACK CASE W/EVENTS;AVE. SEG.



PMM & CO.
TRAIN DISPATCHING MODEL
BASE DOUBLE TRACK CASE W/EVENTS;AVE. SEG.

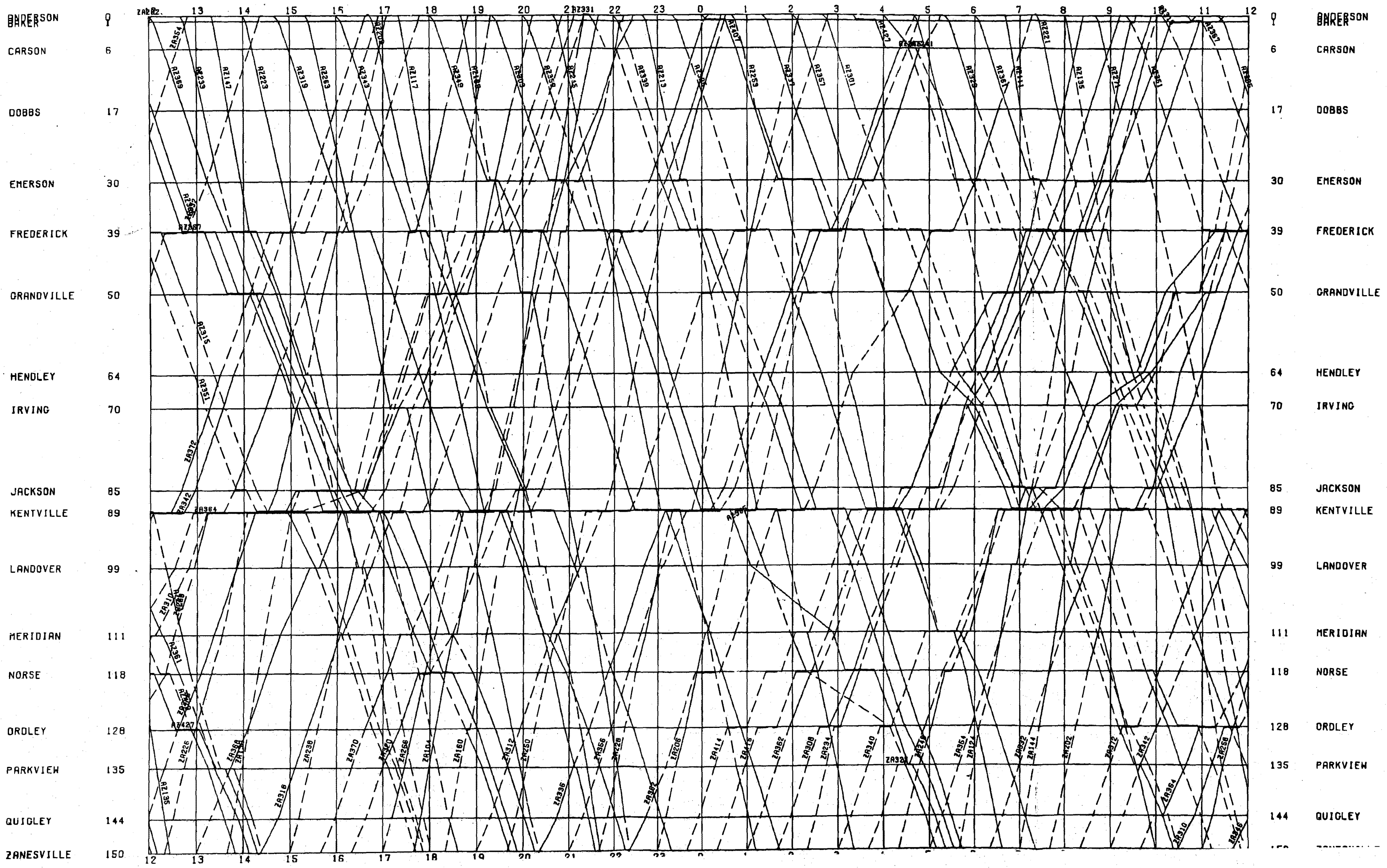
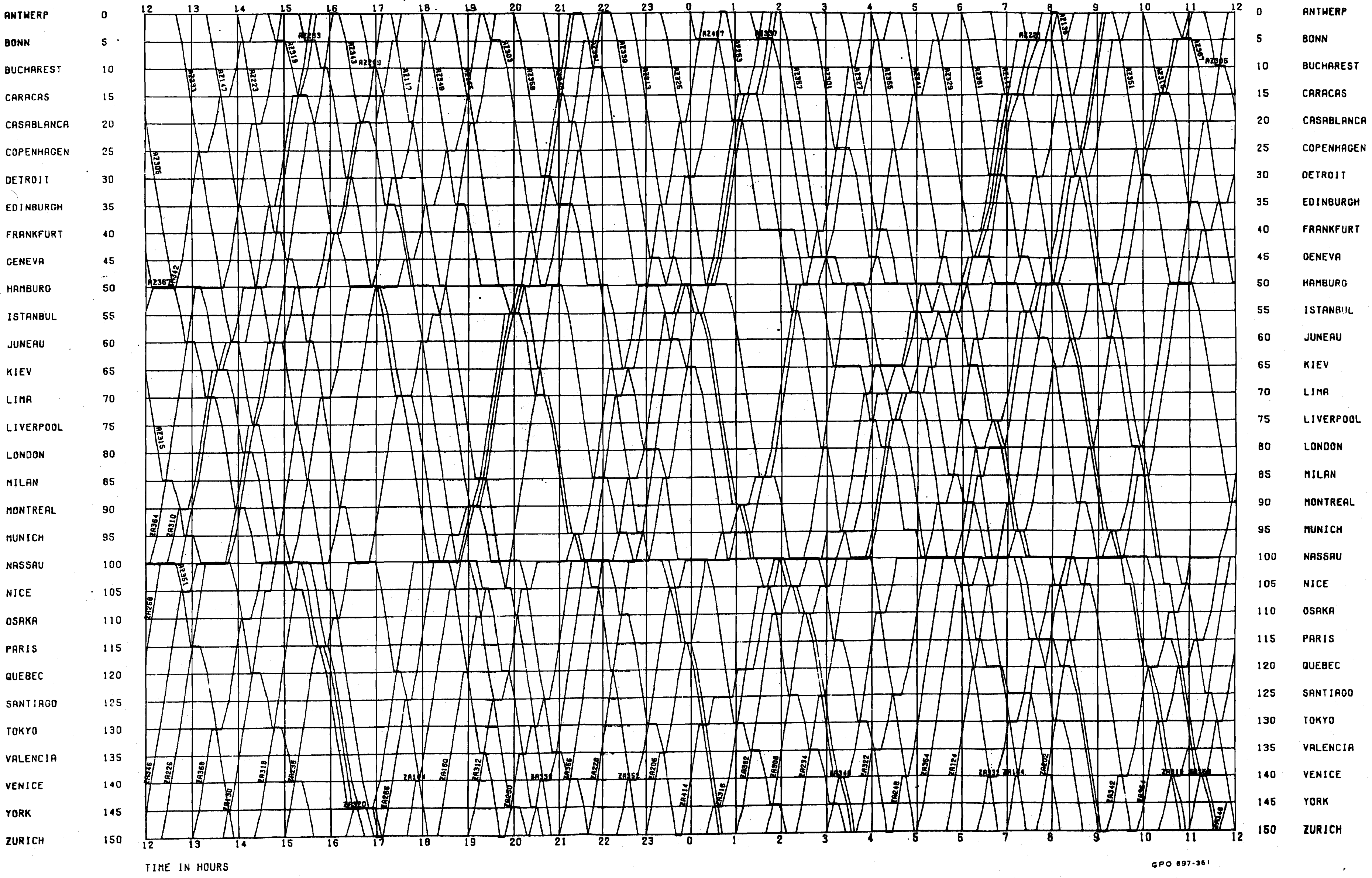


FIG. A.10.

TRAIN DISPATCHING MODEL

BASE SINGLE TRACK CASE ALL BEST;SHORT SEG.



GPO 897-361