PB85-149516

U.S. Department of Transportation

Federal Railroad Administration

Analysis of Rail Defect Data from the Burlington Northern Railroad and the Atchison, Topeka and Santa Fe Railroad

Office of Research and Development
Washington DC 20590

G. A. Mack
A. T. Hopper
H. C. Meacham

Battelle Columbus Laboratories
505 King Avenue
Columbus Ohio 43201

DOT/FRA/ORD-84/06
DOT-TSC-FRA-84-2

July 1984
Final Report

This document is available to the Public through the National Technical Information Service, Springfield, Virginia 22161.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FRA/ORD-84/06	PB85 149516/AS	

4. Title and Subtitle	
ANALYSIS OF RAIL DEFECT DATA FROM THE BURLINGTON NORTHERN RAILROAD AND THE ATCHISON, TOPEKA AND SANTA FE RAILROAD	5. Report Date July 1984 6. Performing Organization Code TSC/DTS-73
7. Author(s) G.A. Mack, A.T. Hopper, H.C. Meacham, M. Bush*	8. Performing Organization Report No. DOT-TSC-FRA-84-2
9. Performing Organization Name and Address Battelle Columbus Laboratories** 505 King Avenue Columbus OH 43201	10. Work Unit No. (TRAIS) RR419/R4304 11. Contract or Grant No. DOT-TSC-1708
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Railroad Administration Office of Research and Development Washington DC 20590	13. Type of Report and Period Covered Final Report February 1980 - June 1982 14. Sponsoring Agency Code RRD-10

15. Supplementary Notes

**Under contract to: U.S. Department of Transportation
Research and Special Programs Administration
Transportation Systems Center
Cambridge MA 02142

*Now Employed by Jet Propulsion Laboratory
California Institute of Technology
Pasadena CA 91103

16. Abstract

Rail descriptive and defect occurrence information from the Burlington Northern and Atchison, Topeka and Santa Fe railroads is given and analyzed. Track data includes information on track type, when laid, maintenance schedules, etc. The objectives and approaches of several comparative analyses are summarized. Results indicate that year laid is an important factor associated with defect rate, with effects of traffic loads also often apparent, although time effect of tonnage over track on older rail was not determined. Recommendations for future work include refinement of the regression analyses performed in the study, and expansion of the model, among others.

17. Key Words	18. Distribution Statement
Rail Defects Inspection Strategy Defect Clusters Defect Occurrence	DOCUMENT IS AVAILABLE TO THE PUBLIC THROUGH THE NATIONAL TECHNICAL INFORMATION SERVICE, SPRINGFIELD, VIRGINIA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
UNCLASSIFIED	UNCLASSIFIED	74	

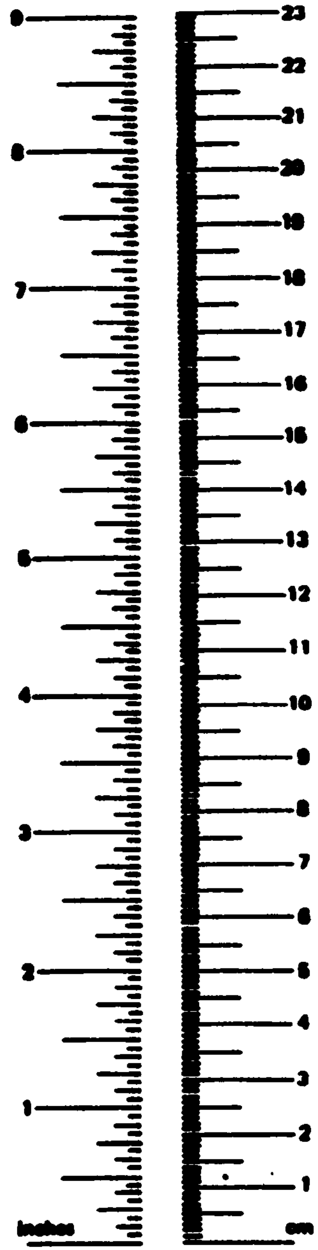
Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

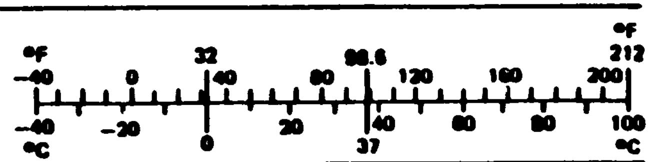
Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detail tables see NBS Misc. Publ. 286, Units of Weights and Measures. Price $2.25 SD Catalog No. C13.10:286.



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



PREFACE

This report covers work to support the development of better inspection program specifications for control of rail defects in railroad track. It has been made possible by the cooperation of the Burlington Northern and the Atchison, Topeka and Santa Fe railroads which supplied basic rail-related information.

The analyses were carried out at Battelle Columbus Laboratories. The work was sponsored by the U.S. Department of Transportation, Federal Railroad Admministration, Office of Research and Development, Track Safety Research Division, Washington DC. The report was prepared by the U.S. Department of Transportation, Transportation Systems Center, Cambridge MA.

Our appreciation is given to Marilyn Bush, formerly of the TSC, who provided coordination, organization and technical assistance to complete this report.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT.)

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

The Association of American Railroad's Ad Hoc Committee on Track Safety Standards has been developing an inspection program specification for control of rail defects in track. The specification is a cooperative effort on the part of the Office of Research and Development of the Federal Railroad Administration (FRA), the Association of American Railroads (AAR), and the Transportation Systems Center (TSC). The specification is based on an in-depth study of rail defect occurrence data from four railroads. This report presents analyses of data from two of those railroads - the Burlington Northern (BN) and the Atchison, Topeka, and Santa Fe (ATSF) - conducted by Battelle Columbus Laboratories, for the U.S. Department of Transportation, Transportation Systems Center.

This report is divided into four sections. The first section describes all available data from the BN Railroad and includes information such as type of track, when it was laid, maintenance schedules, etc. The second section, similar to the first, describes available data from the ATSF Railway. The third section describes the objective, approach, and results of several comparative analyses of the two railroads. The last section summarizes results and recommends future work.

2. BURLINGTON NORTHERN RAILROAD DATA

Battelle Columbus Laboratories, received a magnetic tape from the Burlington Northern Railroad (BN) which contained two files, a defect file and a traffic file. Both of these files had to be modified for use with the CDC computers at Battelle; the traffic file, especially, required an extensive amount of modification.

2.1 DEFECT FILE

Each record in the defect file contained information about a single defect found in the BN system. The file primarily described those defects found during the period from January 1979 through June 1981. Altogether, there were 45,258 defects in the file, representing approximately 200 different line segments. Many of these line segments, however, were quite small and contained only a few defects. The following information was given on each defect record:

Line Segment number.

Track number (single, main line 1, 2 or 3).

Rail statistics information (corresponded to the track segment which contained the defect).

- Ending milepost of track segment in which the defect occurred.
- Rail position (indicated how the defective rail was references in the traffic file: "T" indicated that both rails of the track segment had identical characteristics and thus were stored as one record in the traffic file; 1 ... 6 indicated that the rail that contained the defect was stored as a separate record in the traffic file).
- Year rail laid.
- Year rail relaid (if relaid, otherwise blank).

Track mile distance (length of the track segment that contained the defect).

Rail weight.

Rail section (e.g., RE, RS, RB, ASCE).

Continuous weld (bolted or welded).

Material condition (new or second-hand when laid).

Defect Information

Detector car number (or Service failure).

Date defect found (year, month, day).

Defect milepost

Position of rail (for rail that contained the defect: 1=S. Tangent; 2=S. Low; 3=S. High; 4=N. Tangent; 5=N. Low; 6=N. High).

Type of defect.

Year rail rolled.

Kind of steel (e.g., heat-treated, head-hardened).

Manufacturer (e.g., CF&I, Bethlehem).

Defects found by hand-held probe and defects found as a result of accidents were both recorded as service defects. The detector car codes in the defect file were as follows: 975103-975105, 975111-975120. All the cars were hi-rail cars, owned by Burlington Northern, and all, except cars 975103 through 975105, were ultrasonic.

2.2 TRAFFIC FILE

Each record in the traffic file corresponded to a segment of track with the same track and traffic characteristics. Because the traffic data base is constantly updated by Burlington Northern, the traffic file that was received represented the current track status at the time of data tape was prepared (summer 1981). There were 17,685 records in the traffic file. The following information was given in each traffic record:

Line segment number

Beginning station name (of the track segment).

Ending station name (of the track segment).

MGT-1978.*

MGT-1979.

MGT-1980.

Track number (single, mainline 1, 2 or 3).

Beginning milepost (of the track segment).

Ending milepost (of the track segment).

Rail position (indicated to which rail of the track that the traffic record corresponded: "T" indicated both rails; 1=S. Tangent; 2=S. Low; 3=S. High; 4=S. Tangent; 5=N. Low; 6=N. High).

Year rail laid.

Year rail relaid (if relaid).

Track mile distance (length of track segment).

Rail weight.

Rail section.

Continuous weld (bolted or welded).

Material condition (new or second-hand rail when laid).

Kind of steel.

Manufacturer.

Cumulative MGT.

Last detector car test data.

*MGT-million gross tons; indicates the amount of tonnage over the track for the specified year.

Next detector car test date.

Number of detector car inspection - 1st reporting year.

Number of detector car found defects - 1st reporting year.

Number of SERVICE found defects - 1st reporting year.

Defects per mile (detector car plus service failure) - 1st reporting year.

Same as the above 4 items - 2nd reporting year.

Same as the above 4 items - 3rd reporting year.

There were some other items given in each traffic record, but they were not used in this study. A complete listing of the contents and format of a traffic file record is given in Figure 1. No car movement data were available in this file.

2.3 SELECTION AND PREPARATION OF THE DATA FOR ANALYSIS

2.3.1 Data Selection

Our goal was to select for analysis several representative line segments from all of the Burlington Northern line segment data. To accomplish this, two computer runs were made, one on the defect file and one on the traffic file. The run on the defect file gave a breakdown of the total number of defects by line segment, while the run on the traffic file gave a breakdown of total track miles by line segment. This plot is given in Figure 2 for those line segments which had more than 100 track miles. From this plot, 16 line segments were initially selected (indicated by check marks) as a representative sample with respect to defect-per-mile rates. A Burlington Northern official confirmed that the sample did, indeed, represent a wide cross-section of the Burlington Northern track system. A short time later, 10 additional line segments were included in the study. Because these additional segments differed in the number of reported service failures from the original 16 segments, a comparison of these two sets of segments was desired. Table 1 gives a breakdown of the

17.11.30   SEP 13,1981

```
01  RAILSTAT-FOCUS-RECORD-WS.                                         RDS045
    05  BOUNDARY-FILE-FIELDS.                                         RDS045
        10  LINE-SEGMENT-NUMBER -        PIC X(4)   VALUE SPACES.  RDS045
        10  MAINTENANCE-DIVISION         PIC X(5)   VALUE SPACES.  RDS045
        10  ROADMASTER-DISTRICT          PIC X(5)   VALUE SPACES.  RDS045
        10  OPS240C-SEGMENT              PIC X(4)   VALUE SPACES.  RDS045

        10  STATION-TO-STATION-RAILSTAT.                            RDS045
            15  BEGINNING-STATION        PIC X(10)  VALUE SPACES.  RDS045
            15  ENDING-STATION           PIC X(10)  VALUE SPACES.  RDS045
        10  LIFE-INDICATOR               PIC X(01)  VALUE SPACES.  RDS045
        10  HISTORICAL-ROUTE-TYPE        PIC X(01)  VALUE SPACES.  RDS045
        10  CURRENT-ROUTE-TYPE           PIC X(01)  VALUE SPACES.  RDS045

        10  78-PHYSICAL-DENSITY -        PIC X(02)  VALUE SPACES.  RDS045
        10  78-MILLION-GROSS-TONS        PIC S999V9999   COMP-3    RDS045
                                                VALUE ZERO.        RDS045
        10  79-PHYSICAL-DENSITY -        PIC X(02)  VALUE SPACES.  RDS045
        10  79-MILLION-GROSS-TONS        PIC S999V9999   COMP-3    RDS045
                                                VALUE ZERO.        RDS045
        10  80-PHYSICAL-DENSITY-         PIC X(02)  VALUE SPACES.  RDS045
        10  80-MILLION-GROSS-TONS        PIC S999V9999   COMP-3    RDS045
                                                VALUE ZERO.        RDS045

        10  FILLER                       PIC X(6)   VALUE SPACES.  RDS045

        10  BASE-YEAR-80.                                          RDS045
            15  UNIT-TONNAGE             PIC S999V9999   COMP-3    RDS045
                                                VALUE ZERO.        RDS045
            15  OTHER-TONNAGE            PIC S999V9999   COMP-3    RDS045
                                                VALUE ZERO.        RDS045
        10  81-MILLION-GROSS-TONS        PIC S999V9999   COMP-3    RDS045
                                                VALUE ZERO.        RDS045
        10  82-MILLION-GROSS-TONS        PIC S999V9999   COMP-3    RDS045
                                                VALUE ZERO.        RDS045
        10  83-MILLION-GROSS-TONS        PIC S999V9999   COMP-3    RDS045
                                                VALUE ZERO.        RDS045
        10  84-MILLION-GROSS-TONS        PIC S999V9999   COMP-3    RDS045
                                                VALUE ZERO.        RDS045
        10  85-MILLION-GROSS-TONS        PIC S999V9999   COMP-3    RDS045
                                                VALUE ZERO.        RDS045

    05  RDS050-APPLIED-FIELDS.                                      RDS045

        10  TRDS116A-TRACK-RANGE.                                   RDS045
            15  TRACK-NUMBER -           PIC X(1)   VALUE SPACES.  RDS045
            15  PSEUDO-TR-X-NUMBER       PIC X(1)   VALUE SPACES.  RDS045
            15  TRACK-ABANDONMENT-CODE   PIC X(1)   VALUE SPACES.  RDS045
        10  DRDS121A-RAIL-STATISTIC.                                RDS045
            15  BEGINNING-MILEPOST-DISPLAY  PIC 9999.99.            RDS045
            15  ENDING-MILEPOST-DISPLAY  PIC 9999.99.-              RDS045
```

FIGURE 1. FORMAT OF EACH BN TRAFFIC FILE RECORD

```
                  15  RAIL-POSITION -          PIC X(1)   VALUE SPACES.  RDS045
                  15  YEAR-RAIL-LAID -         PIC X(3)   VALUE SPACES.  RDS045
                  15  YEAR-RAIL-RELAID         PIC X(3)   VALUE SPACES.  RDS045
                  15  TRACK-MILE-DISTANCE -    PIC S999V99       COMP-3  RDS045
                                                          VALUE ZEROS.   RDS045
                  15  RAIL-WEIGHT-             PIC X(3)   VALUE SPACES.  RDS045
                  15  RAIL-SECTION -           PIC X(2)   VALUE SPACES.  RDS045
                  15  CONTINUOUS-WELD-         PIC X(3)   VALUE SPACES.  RDS045
                  15  MATERIAL-CONDITION -     PIC X(1)   VALUE SPACES.  RDS045
                  15  LIFE-INDICATOR           PIC X(1)   VALUE SPACES.  RDS045
                  15  KIND-OF-STEEL-           PIC X(2)   VALUE SPACES.  RDS045
                  15  MANUFACTURER -           PIC X(2)   VALUE SPACES.  RDS045
                  15  MILLION-GROSS-TONS-      PIC S9999V999     COMP-3  RDS045
                                                          VALUE ZEROS.   RDS045
                  15  LAST-TESTING-DATE -      PIC X(6)   VALUE SPACES.  RDS045
                  15  NEXT-TESTING-DATE        PIC X(6)   VALUE SPACES.  RDS045

              10  DRDS126A-RAIL-RELAY.                                   RDS045
                  15  BUDGET-CLASS.                                      RDS045
                      20  BUDGET-CODE          PIC X(2)   VALUE SPACES.  RDS045
                      20  BUDGET-TYPE          PIC X(1)   VALUE SPACES.  RDS045
                  15  RELAY-PRIORITY           PIC X(2)   VALUE SPACES.  RDS045
                  15  EXPENDITURE-PROGRAM.                               RDS045
                      20  TYPE-OF-PROGRAM      PIC X(3)   VALUE SPACES.  RDS045
                      20  YEAR-OF-PROGRAM      PIC X(2)   VALUE SPACES.  RDS045
                      20  PROGRAM-NUMBER       PIC X(5)   VALUE SPACES.  RDS045
                  15  1ST-2ND-3RD-RELAY-PROGRAM.                         RDS045
                      20  1ST-PROGRAMMED-RELAY.                          RDS045
                          25  1ST-RELAY-YEAR   PIC X(2)   VALUE SPACES.  RDS045
                          25  1ST-RELAY-RAIL-WEIGHT                      RDS045
                                               PIC X(3)   VALUE SPACES.  RDS045
                          25  1ST-RELAY-KIND-OF-STEEL                    RDS045
                                               PIC X(2)   VALUE SPACES.  RDS045
                      20  2ND-PROGRAMMED-RELAY.                          RDS045
                          25  2ND-RELAY-YEAR   PIC X(2)   VALUE SPACES.  RDS045
                          25  2ND-RELAY-RAIL-WEIGHT                      RDS045
                                               PIC X(3)   VALUE SPACES.  RDS045
                          25  2ND-RELAY-KIND-OF-STEEL                    RDS045
                                               PIC X(2)   VALUE SPACES.  RDS045
                      20  3RD-PROGRAMMED-RELAY.                          RDS045
                          25  3RD-RELAY-YEAR   PIC X(2)   VALUE SPACES.  RDS045
                          25  3RD-RELAY-RAIL-WEIGHT                      RDS045
                                               PIC X(3)   VALUE SPACES.  RDS045
                          25  3RD-RELAY-KIND-OF-STEEL                    RDS045
                                               PIC X(2)   VALUE SPACES.  RDS045

          05  DEFECT-INFORMATION.                                        RDS045

              10  1ST-DEFECT-REPORTING-YEAR.                             RDS045
                  15  NUMBER-OF-INSPECTIONS-  PIC  99      VALUE 0.      RDS045
                  15  DETECTOR-CAR-DEFECTS -  PIC  99      VALUE 0.      RDS045
                  15  SERVICE-FAILURES-       PIC  99      VALUE 0.      RDS045
                  15  DEFECTS-PER-MILE-       PIC  99.99.                RDS045

              10  2ND-DEFECT-REPORTING-YEAR.                             RDS045
                  15  NUMBER-OF-INSPECTIONS   PIC  99      VALUE 0.      RDS045
                  15  DETECTOR-CAR-DEFECTS    PIC  99      VALUE 0.      RDS045
```

FIGURE 1. FORMAT OF EACH BN TRAFFIC FILE RECORD (CONTINUED)

```
           15  SERVICE-FAILURES          PIC  99       VALUE 0.       RDS045
           15  DEFECTS-PER-MILE          PIC  99.99.                  RDS045

       10  3RD-DEFECT-REPORTING-YEAR.                                 RDS045
           15  NUMBER-OF-INSPECTIONS     PIC  99       VALUE 0.       RDS045
           15  DETECTOR-CAR-DEFECTS      PIC  99       VALUE 0.       RDS045
           15  SERVICE-FAILURES          PIC  99       VALUE 0.       RDS045
           15  DEFECTS-PER-MILE          PIC  99.99.                  RDS045

       10  4TH-DEFECT-REPORTING-YEAR.                                 RDS045
           15  NUMBER-OF-INSPECTIONS     PIC  99       VALUE 0.       RDS045
           15  DETECTOR-CAR-DEFECTS      PIC  99       VALUE 0.       RDS045
           15  SERVICE-FAILURES          PIC  99       VALUE 0.       RDS045
           15  DEFECTS-PER-MILE          PIC  99.99.                  RDS045

       10  5TH-DEFECT-REPORTING-YEAR.                                 RDS045
           15  NUMBER-OF-INSPECTIONS     PIC  99       VALUE 0.       RDS045
           15  DETECTOR-CAR-DEFECTS      PIC  99       VALUE 0.       RDS045
           15  SERVICE-FAILURES          PIC  99       VALUE 0.       RDS045
           15  DEFECTS-PER-MILE          PIC  99.99.                  RDS045

       10  6TH-DEFECT-REPORTING-YEAR.                                 RDS045
           15  NUMBER-OF-INSPECTIONS     PIC  99       VALUE 0.       RDS045
           15  DETECTOR-CAR-DEFECTS      PIC  99       VALUE 0.       RDS045
           15  SERVICE-FAILURES          PIC  99       VALUE 0.       RDS045
           15  DEFECTS-PER-MILE          PIC  99.99.                  RDS045
```

FIGURE 1. FORMAT OF EACH BN TRAFFIC FILE RECORD (CONCLUDED)

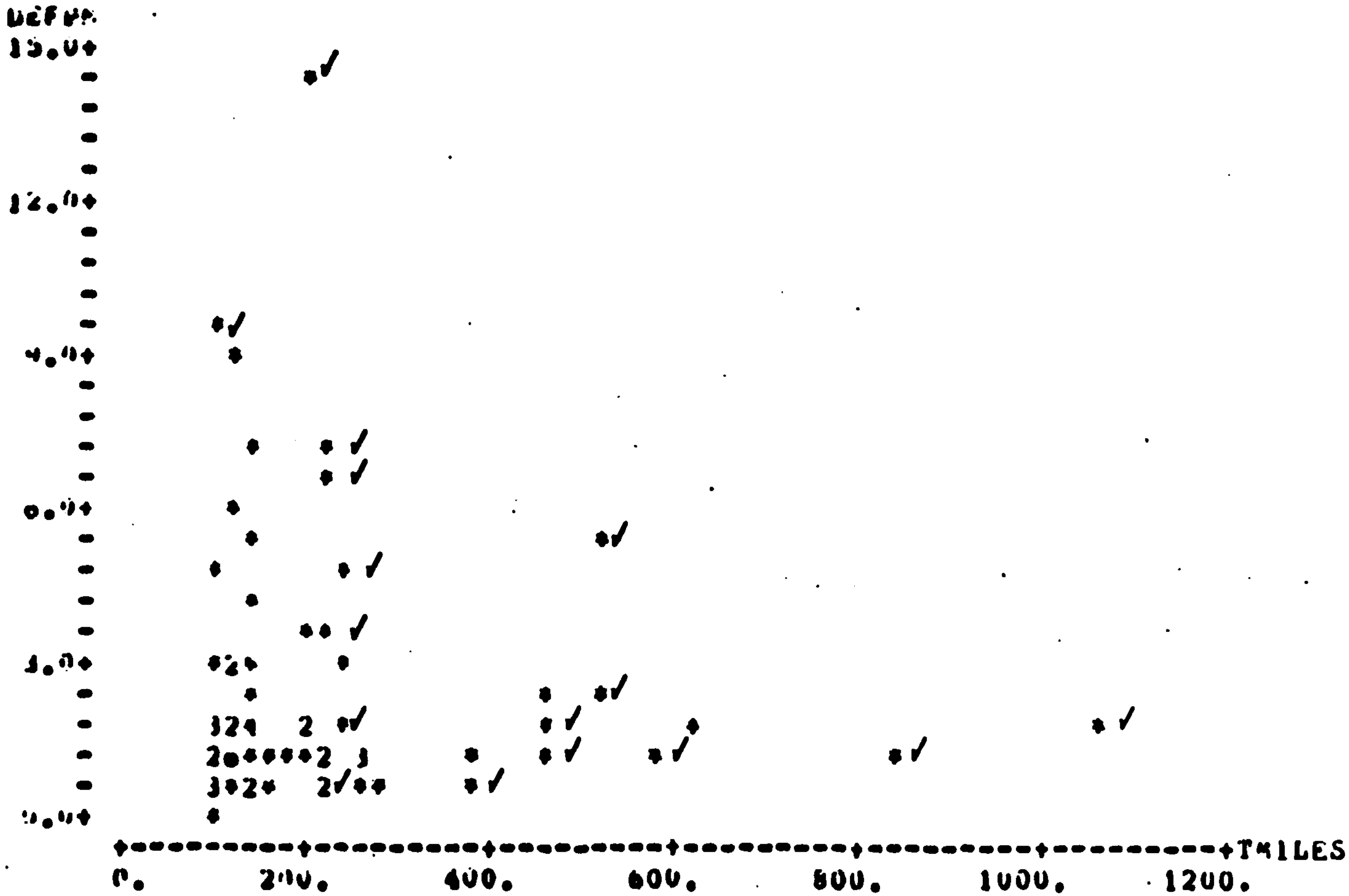


FIGURE 2. PLOT OF EACH BN LINE SEGMENT'S DEFECTS-PER-MILE RATE AGAINST TRACK MILES

TABLE 1. DESCRIPTION OF THE ORIGINAL 16 BURLINGTON NORTHERN LINE SEGMENTS SELECTED FOR ANALYSIS

Line Segment	Total Number of Defects	Total Number* of Track Miles	Defects Per Mile	Track Miles*			
				Single	ML1	ML2	ML3
1	1007	822.26	1.225	46.12	385.77	385.57	4.80
3	1135	621.56	1.826	105.29	260.22	256.05	
4	1813	1044.15	1.736	575.40	233.07	233.61	2.07
5	2690	514.01	5.233	505.46	4.27	4.28	
31	1031	104.13	9.901	104.13			
34	141	233.65	0.630	223.65			
35	979	459.02	2.133	397.43	30.79	30.80	
36	1277	505.32	2.527	327.10	88.86	89.36	
37	126	381.01	0.331	338.48	21.27	21.26	
197	1420	220.81	6.431	220.81			
240	2877	193.97	14.832	193.97			
362	786	221.64	3.546	221.64			
476	1165	237.35	4.908	231.45	2.96	2.94	
477	391	229.47	1.704	177.00	48.68	3.79	
485	440	451.53	0.974	446.61	2.46	2.46	
495	1484	208.59	7.114	208.59			
	18,762	6438.47		4323.13	1078.35	1030.12	6.87

* Does not include rail positions other than "T". Track segments which had a rail position other than "T" were omitted from the study.

details on the original 16 line segments, while Table 2 presents the breakdown for the additional 10 segments. Track segments in the traffic file that had a rail position other than "T" were omitted from the study (along with their associated defects) because of the numerous extra computations that would have been required to properly use them. These track segments represented less than 1 percent of the total track miles.

2.3.2 Data Preparation

Analysis of the defect-per-mile rates required a knowledge of the existing track mileposts for all the track. The defect file gave information only about the track segments which had defects. If there were no defects found between two given mileposts, it was not known whether the track was in good condition there or whether no track existed between these mileposts. The defect file also did not contain any information about the amount of tonnage over the track; whereas, the track file did contain this information. Therefore, it was necessary to merge the information from the defect and traffic files. The traffic file contained information on all the track segments and therefore served as a descriptor of the "population" of all track.

Two merged data bases were created: Burlington Northern Data Base 1 (BNDD1) and Burlington Northern Data Base 3 (BNDD3). The objective for BNDD1 was to create a data base containing defect, track, and traffic information for 1-mile track segments (as much as possible) having a fixed set of traffic and track characteristics. Thus, BDNN1 was created by dividing the Burlington Northern system into 1-mile continuous track segments, each having a fixed set of track and traffic characteristics. Partial mile segments were created whenever the characteristics changed within a 1-mile stretch. Each record in BNDD1 represented one of these track segments. The following information was given for each segment:

TABLE 2. DESCRIPTION OF THE 10 ADDITIONAL BURLINGTON NORTHERN LINE SEGMENTS SELECTED FOR ANALYSIS

Line Segment	Total Number of Defects	Total Number* of track Miles	Defects Per Mile	Track Miles*			
				Single	ML1	ML2	ML3
2	762	566.17	1.35	499.38	32.36	32.93	1.5
12	42	99.48	.42	99.48			
13	233	250.31	.93	168.36	40.99	40.96	
16	232	183.50	1.26	156.20	13.66	13.64	
21	390	126.08	3.01	123.06	1.51	1.51	
137	29	49.67	.60	44.20	2.23	2.24	
142	23	27.62	.83	27.62			
178	0	47.86	0	47.86			
264	56	21.26	2.63	21.26			
376	729	124.36	5.86	124.36			
	2,486	1,495.31	16.89	1,311.78	90.75	91.28	1.5

*Does not include rail positions other than "T". Track segments which had a rail position other than "T" were omitted from the study.

Line segment number.

Beginning milepost of the segment.

Ending milepost of the segment.

Track number.

Year rail laid.

Rail weight.

Rail section.

Continuous weld.

Material condition.

Manufacturer.

Cumulative MGT.*

MGT-1978.

MGT-1979.

MGT-1980.

Number of detector car inspections - 1st reporting year.

Number of detector car inspections - 2nd reporting year.

Number of detector car inspections - 3rd reporting year.

Number of all type defects found - each of the 30 months, January 1979-June 1981.

Number of bolt hold breaks found - each of the 30 months, January 1979-June 1981.

Number of detail fractures - each of the years 1979-1981.

Number of engine burn fractures - each of the years 1979-1981.

Number of transverse defects - each of the years 1979-1981.

*MGT-million gross tons; indicates the amount of tonnage over the track for the specified year.

Number of horizontal split heads - each of the years 1979-1981.

Number of vertical split heads - each of the years 1979-1981.

Number of head and web separations - each of the years 1979-1981.

BNDD1 was used to construct the histograms of defect frequencies by milepost.

BNDD3 was similar to BNDD1, except the track and traffic characteristics were not required to be constant within a 1-mile track segment. In addition, partial mile segments (e.g., at the end of the line segment or near a gap in the track mileposts) were omitted. The objective for BNDD3 was to create a data base of only 1-mile segments to determine: (a) if rail defects cluster on the same section of track from year to year; and (b) the lengths of the rail defect clusters for each year. The use of partial segments in these studies would have biased the results, and thus partial miles were omitted from BNDD3. BNDD1 and BNDD3 data bases were created for each set of Burlington Northern line segments. There was 18,762 total defects in BNDD3 for the original 16 segments and 2486 defects in BNDD3 for the additional 10 segments.

Table 3 presents a comparison of the two sets of line segments (Set A - Original 16, Set B - Additional 10) by various track and traffic characteristics. Year rail laid (YRLAID) was divided into 10-year intervals with the midpoints given in Table 3. Similarly, cumulative tonnage (TONC) over the track (in million gross tons, MGT) was divided into 50 MGT intervals; average tonnage (TONAVG) over the track for each year from 1978 through 1980 was divided into five MGT intervals; and the number of detector car inspections (INS) for each year from 1979 through 1980 were divided into four inspection intervals for each year. The only exception is the "0" category for INS, which represents exactly zero inspections.

TABLE 3. COMPARISON OF TRACK AND TRAFFIC CHARACTERISTICS FOR THE TWO SETS OF BURLINGTON NORTHERN LINE SEGMENTS. SET A - ORIGINAL 16 SEGMENTS; SET B - ADDITIONAL 10 SEGMENTS

PERCENTAGE OF TRACK MILES

Yrlaid	LS Set A	LS Set B
1905	2.9	0.8
1915	2.6	2.6
1925	2.7	4.7
1935	1.5	2.1
1945	11.6	14.4
1955	14.2	38.5
1965	22.5	11.4
1975	41.8	25.5

PERCENTAGE OF TRACK MILES

Weight, lbs/yd*	LS Set A	LS Set B
56	1.0	0.0
65	0.2	0.0
66	2.1	0.0
72	0.0	0.3
75	0.4	0.0
77	2.2	0.0
85	0.1	0.7
90	3.4	9.5
100	0.0	0.2
110	0.0	0.1
112	25.8	27.9
115	9.5	31.7
119	0.0	0.1
129	0.7	7.9
131	0.0	2.1
132	26.0	18.1
136	28.0	1.4

*Data from the BN track charts.

TABLE 3. COMPARISON OF TRACK AND TRAFFIC CHARACTERISTICS FOR THE TWO SETS OF BURLINGTON NORTHERN LINE SEGMENTS. SET A - ORIGINAL 16 SEGMENTS; SET B - ADDITIONAL 10 SEGMENTS (CONTINUED)

PERCENTAGE OF TRACK MILES

Weld	LS Set A	LS Set B
0	36.5	61.0
1	63.5	39.0

PERCENTAGE OF TRACK MILES

Rail Section	LS Set A	LS Set B
GN	2.2	1.6
NP	0.1	0.0
OT	2.0	0.0
RA	3.4	6.2
RB	0.0	1.7
RE	74.8	59.6
TR	17.6	30.9

TABLE 3. COMPARISON OF TRACK AND TRAFFIC CHARACTERISTICS FOR THE TWO SETS OF BURLINGTON NORTHERN LINE SEGMENTS. SET A - ORIGINAL 16 SEGMENTS; SET B - ADDITIONAL 10 SEGMENTS (CONTINUED)

PERCENTAGE OF TRACK MILES

Cumulative MGT	LS Set A	LS Set B
25.000	12.0	16.7
75.000	19.5	8.7
125.000	5.5	12.3
175.000	19.1	8.9
225.000	8.0	4.1
275.000	11.8	5.3
325.000	6.7	11.7
375.000	2.5	9.2
425.000	2.6	7.7
475.000	5.3	5.3
525.000	5.2	3.5
575.000	1.4	3.7
650.000	0.2	2.9

PERCENTAGE OF TRACK MILES

Avg. MGT Per Year	LS Set A	LS Set B
2.5	9.5	8.3
7.5	12.3	17.0
12.5	4.1	6.3
17.5	41.7	14.4
22.5	10.1	17.2
27.5	4.8	10.3
32.5	1.8	21.4
37.5	12.7	0.4
42.5	2.2	0.6
50.0	0.8	4.0

TABLE 3. COMPARISON OF TRACK AND TRAFFIC CHARACTERISTICS FOR THE TWO SETS OF BURLINGTON NORTHERN LINE SEGMENTS. SET A - ORIGINAL 16 SEGMENTS; SET B - ADDITIONAL 10 SEGMENTS (CONTINUED)

PERCENTAGE OF TRACK MILES

# of D. Car Inspections	LS Set A	LS Set B
0	10.3	1.3
2	4.8	9.4
6	30.9	30.8
10	32.3	38.0
14	20.4	14.8
18	1.3	5.8

3. ATCHISON, TOPEKA, AND SANTA FE RAILWAY DATA

3.1 AVAILABLE DATA

The Atchison, Topeka, and Santa Fe Railway data bases used in this study were constructed from four primary sources;

- Rail Failure Statements for the years 1974-1979.
- Traffic Density Statements for the years 1974-1979.
- Track charts covering the Los Angeles, Middle, New Mexico, and Southern divisions.
- 1980 Operating Timetables for each of the Santa Fe divisions.

This information was provided in printed form and the data bases were constructed by encoding, transcribing, and keypunching the data.

The Atchinson, Topeka, and Santa Fe Railroad is geographically divided into 14 operating divisions: Valley, Los Angeles, Los Angeles Terminal, Albuquerque, New Mexico, Plains, Southern, Northern, Colorado, Middle, Eastern, Kansas City, Illinois, and Chicago Terminal. Each of these divisions is further divided into districts. The current study concentrated on five districts from four divisions, as follows:

New Mexico Division, First District.

Southern Division, Second District.

Middle Division, Fourth District.

Los Angeles Division, Third District.

Los Angeles Division, Needles District.

These districts were chosen, with the aid of ATSF personnel, to represent a variety of traffic conditions and geographies.

3.1.1 Rail Failure Statements

The Rail Failure Statements are actually composed of several reports. Three of these were used: defects found by detector cars, defects found by hand held probe, and service-detected failures. Each line in one of these reports contained information pertinent to one rail defect. These rail defect reports were organized by division and district according to the year that the defect was found and its location.

Table 4 summarizes the rail detector vehicles used for the five districts studied from 1974 through 1979.

3.1.2 Traffic Density Statements

The Traffic Density records contained information on the tonnage carried by the railroad between each specified pair of stations within a district. These data were broken down by directions of travel, and by how much of the tonnage was carried by ATSF versus foreign carriers. Also reported was the number of cars in three tonnage ranges (over 132 tons, 111-132 tons, and less than 110 tons) responsible for carrying the total tonnage over each section of track. These traffic records were organized by year and location (division, district).

3.1.3 Track Charts

The Track Charts were detailed maps of the track on a scale of 1 mile to 1-3/4 inches. They showed grade and curve geometry, gave details about all roads and streams that cross the right-of-way, and provided track and maintenance information. Organized according to division, district, and milepost, the charts reflected the current track condition as of 1979.

TABLE 4. ATSF DETECTOR CAR HISTORY

Year	L.A. Needles	L.A. 3rd	Southern 2nd	New Mexico 1st	Middle 4th
1974	3	3	7, L1, L2	11	9, L1, L2, L3, L4
1975	3	3	7, L1, L2	11	9, L1, L2, L3, L4
1976	3	3	7, L1, L2	11	9, L1, L2, L3, L4
1977	3	3	12	11	9
1978	3, 18	3, 18	12	11	15
1979	3, 18	3, 18	12	11	15

L1, L2, L3, L4 - Sperry leased, ultrasonic and inductions

3 - all magnetic, on rail
7 - all magnetic, road/rail
9 - magnetic and ultrasonic hi rail
11 - ultrasonic, road/rail
12 - ultrasonic, road/rail
15 - ultrasonic, road/rail
18 - ultrasonic, road/rail

The information from these charts was extracted manually and then encoded. Each extracted record represented a contiguous section of track having common characteristics. For example, for one entry the rail weight, year laid, rail position (i.e., curve or tangent track), etc., would be constant. A new record was created whenever an important characteristic changed. The information recorded for each of the track sections is as follows:

- Location (division, district, beginning and ending mileposts).
- Rail weight.
- Bolted or continuously welded rail.
- Year rail laid.
- Rail position (curve, tangent).
- Rail location (single, double track--north or south).
- Grinding record (year, number of passes).
- Surfacing record (year).
- Under track plow record (year one, year two - if any).
- Number of tracks.
- Curve number (if a curve).

3.1.4 Current Operating Timetables

Current ATSF Operating Timetables were used to obtain milepost locations for each of the stations listed in the Traffic Density Statements. These station mileposts were needed to match traffic information with the defect and track chart information. (See Figure 3)

3.2 DATA PREPARATION

After the data were collected from the four sources, three data bases were created: a defect file of rail defect and track information, a traffic file of

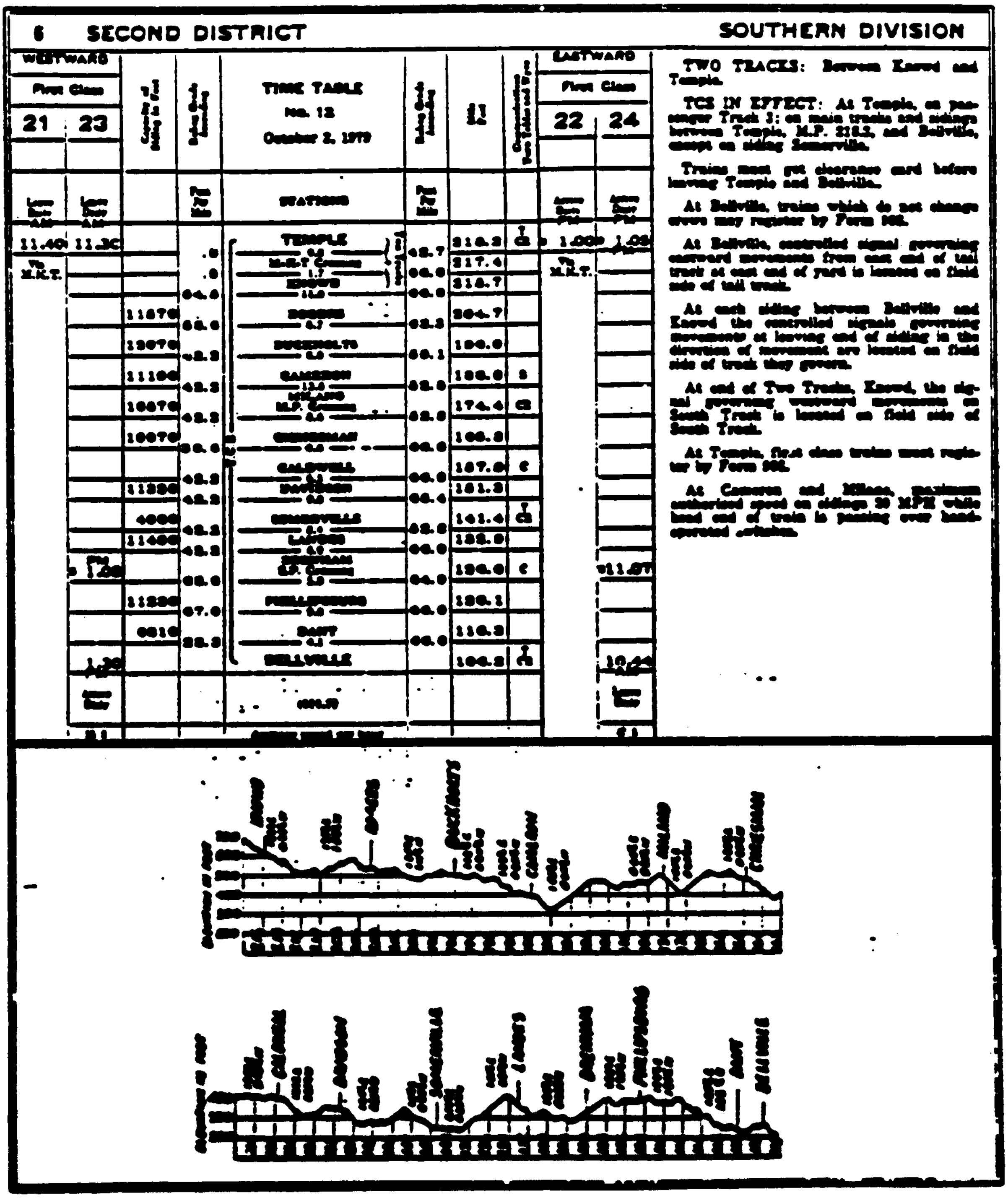


FIGURE 3. SAMPLE PAGE FROM THE CURRENT ATSF OPERATING TIME TABLES

tonnage and car movement data, and a maintenance file of track and maintenance information.

3.2.1 Defect File

Each record in the defect file contained information about a single rail defect found between 1975 and 1979. Altogether, there were 7520 defects in the file, representing 27 ATSF districts in 4 divisions: Los Angeles, Middle, New Mexico, and Southern. The following information was given on each defect record:

Division name.

District name.

Defect milepost.

Type of track (main, branch, siding, etc.).

Rail location (Single, double track -- north, or south).

Rail position (tangent, curve).

Type grade (level, ascending, descending).

Rail weight.

Mill (USS, CF&I, etc.).

Year rolled.

Type of failure.

Date defect found (year, month, day).

Defect found by (service, detector car number, audigage number).

Date defect fixed (year, month, day).

3.2.2 Traffic File

Each record in the traffic file corresponded to a segment of contiguous track having a constant set of tonnage and car movement history. A different set of tonnage and car movement data was given for each of the years between 1974 and 1979. The following information was given for each record and year:

Division name.

District name.

Beginning milepost of track segment.

Ending milepost of track segment.

Millions of gross tons (MGT).

Santa Fe only.

Foreign only.

Number of cars.

110-131 tons only.

over 131 tons only.

all cars.

Speed limit.

3.2.3 Maintenance File

Each record in the maintenance file represented a continuous track segment with a constant set of track and maintenance characteristics. The data represented the current status of the track as of 1979.

The following information was given on each segment:

Division name.

District name.

Beginning milepost (of the track segment).

Ending milepost (of the track segment).

Rail weight.

Bolted or continuously welded rail.

Year rail laid.

Rail position (curve, tangent).

Rail location (single, double track - north or south).

Grinding record (year, number of passes).

Surfacing record (year).

Under track plow record (year one, year two - if any).

Number of tracks.

Curve number (if a curve).

3.3 DATA SELECTION

The tonnage and car movement data for ATSF generally were not given separately for east and west traffic. Since it was not possible to obtain accurate traffic data for the separate tracks, all double track data were omitted for this study. Only single-track data were used and the details on the selected districts are given in Table 5.

3.4 CREATION OF WORKING DATA BASES

The four ATSF source data bases were merged to produce two data bases (Atchison, Topeka, Santa Fe Data Base 1 - SFDD1; and Atchison, Topeka, Santa Fe Data Base 3 - SFDD3) analogous to the two BN data bases. SFDD1 contained defect, track, and traffic information for 1-mile, single track segments having a fixed set of characteristics. SFDD3 consisted entirely of 1-mile track segments formed without regard to changes in the track and traffic characteristics. The following information was given for each track segment in SFDD1:

Division name.

District name.

Beginning milepost (of track segment).

Ending milepost (of track segment)

Rail weight.

Welded or bolted rail.

Year rail laid.

Rail position.

Grinding history.

Surfacing history.

Under track plow history

MGT for each of the years 1974-1979.*

Number of average weight cars (110-131 tons) for each year 1974-1979.

Number of heavy cars (>131 tons) for each year 1974-1979.

Number of all cars each year 1974-1979.

Speed limit.

Number of all type defects found each 6-month period from 1974-1979.

Number of all type defects (except damaged rail and worn rail) found each 6-month period from 1974-1979.

Number of bolt hole breaks found each year 1974-1979.

Number of horizonal split heads found each year 1974-1979.

Number of head web separations found each year 1974-1979.

Number of detail fractures found each year 1974-1979.

Curve data.

Bridge data.

Grade data.

Since the curve, bridge, and grade data were manually extracted from the track charts, these data were obtained for only approximately half the districts.

The analysis of ATSF data was complicated by the presence of relaid rail in several of the districts.** The actual number of years in service and

*MGT-million gross tons; indicated the amount of tonnage over the track for the specified years.

**A relaid rail is one which was previously used in a different location before being placed in service in its present position.

TABLE 5. DESCRIPTION OF THE ATSF DISTRICTS SELECTED FOR ANALYSIS

Division	District	Track Miles*	Number of Defects	Defects/Mile
Los Angeles	Cadiz	83.1	710	8.5
	Fourth	96.0	320	3.3
	Olive	6.0	4	0.7
	Second	60.3	79	1.3
	Third	29.2	280	9.6
Middle	Fifth	98.3	281	2.9
	Cushing	44.8	20	0.4
	Douglass	30.3	43	1.4
	Enid	116.5	0	0
	Oklahoma	148.4	396	2.7
	Strong City	152.9	952	6.2
	First	59.8	109	1.8
	Second	138.6	228	1.6
	Third	77.3	89	1.2
	Fourth	82.7	217	2.6
New Mexico	Carlsbad	184.2	656	3.6
	El Paso	254.2	932	3.7
	Rustler Springs	60.3	11	0.2
	First	205.4	694	3.4
Southern	Conroe	152.6	670	4.4
	Houston	20.0	2	0.1
	Lampasas	129.6	162	1.3
	Longview	188.0	108	0.6
	First	99.0	137	1.4
	Second	110.2	228	2.1
	Third	103.3	110	1.1

* Single track only.

cumulative tonnage over the track for such rail are not accurately known. For this reason, relaid rails were omitted from statistical analyses (Section 4.4 and 4.5) which examined the effects of track and traffic characteristics on rail flaw occurrences. Relaid rails were included in other analyses, such as the development of histograms and calculations to determine rail defect cluster recurrence and rail defect cluster lengths.

4. RESULTS

4.1 HISTOGRAMS OF DEFECTS BY MILEPOST

4.1.1 Objective

From the data collected, histograms were constructed; (1) to determine the rail defect occurrence rates of the various areas of Burlington Northern (BN) and the Atchison, Topeka, and Santa Fe (ATSF) railroads, and (2) to examine the track and traffic characteristics associated with these rates.

4.1.2 Approach

The BNDD1 data base was used in this part of the study so that all of the track could be included in the histograms. (Recall that partial miles were excluded from BNDD3.) Also, since the characteristics of a track segment were constant within each track segment in BNDD1, there was a unique set of characteristics associated with each segment.

The 1-mile segments in BNDD1 and SFDD1 were combined to form segments approximately 5 miles in length for the ATSF system and 10 miles for the BN. Exceptions occurred in the following situations:

1. Gaps in the mileposts.
2. Change of rail line (district or line segment).
3. Change of track number (e.g., single to double track).

In these instances, the interval width was less than 10 (or 5) miles, sometimes substantially less. For this reason, the actual length of each interval was given in addition to the track and traffic characteristics. For each interval the following information was given:

Interval length.

Line segment number or division-district names.

Track number (Burlington Northern only).

Beginning and ending mileposts of the interval.

Minimum, average, and maximum cumulative MGT for the interval.

Minimum, average, and maximum rail weight for the interval.

Minimum, average, and maximum year laid for the interval.

Average type of rail (bolted = 0, welded = 1).

Spike of asterisks which graphically displays the number of defects found (each asterisk represents two defects, except the last one, which may represent only one. If there were more than 60 defects, only 30 asterisks are displayed).

The average tonnage over the track (MGT), rail weight, year laid, and track type (i.e., bolted or welded) were actually weighted averages. For cumulative MGT, each track segment in BNDD1 had a value associated with it. The MGT value for each segment in a given interval was weighted by the percentage of the interval's total track miles represented by the segment. This weighted average was then reported on the histogram. For example, to obtain a weighted average, consider an interval of length 2.90 miles (which consists of two 1-mile segments and a partial mile segment of length 0.90). Suppose the MGT values for these segments are 557.1, 604.2 and 611.3, respectively. Thus, the weighted (i.e., weighted by the percentage of interval track mile represented by the segment) average MGT is computed to be.

$$(1(557.1) + 1(604.2) + 0.9\ (611.3))/2.9 = 590.16.$$

Similar computations were made to obtain the weighted averages for the other characteristics.

4.1.3 Conclusions

Examination of the distribution of rail defects by milepost revealed that defects do cluster and that many of the clusters appear to be associated with specific track and traffic characteristics. For example, clusters of rail defects are found in those lines having large values of cumulative tonnage over the track (TONC) and those with older rails. Rail laid in the 1950s, and earlier, generally has more defects than newer rail.

The Rail Defect distributions for the BN line segments and the ATSF districts are similar. The defects do cluster and often are associated with certain track or traffic characteristics, or both. That is, rail defects occur in certain sections of track, and not in other sections. However, some entire districts or line segments exhibited rail defects. BN line segment #376 and ATSF Los Angeles-Cadiz District are examples where rail defects exist on the entire track system. Even though the cumulative tonnage over the track (MGT) was not high on BN line segment #376, the rail was older. The ATSF Los Angeles-Cadiz District had 90-pound rail for its entire length, and in BN line segment #2 rail defects clustered only in certain stretches of track. Thus, rail defects cluster in converse track conditions - in track with high tonnage and old rail and also (contrary to expectation) in track with low tonnage and new rail. The latter case of rail defect clusters in track with low cumulative tonnage (100 MGT) was seen in the ATSF Middle Division - Strong City District and in BN line segment #4. These inconsistencies, along with other analyses done in this study, indicate that many factors interact to produce defects though certain ones (e.g., tonnage over the track and year laid) are more important that others.

4.2 CORRELATIONS OVER TIME

4.2.1 Objective

Historical correlations of BN and ATSF rail data were made to determine whether rail defects cluster on the same section of track from year to year.

4.2.2 Approach

The historical correlations of rail defect data were drawn from the BNDD3 data base, which covered the Burlington Northern Railroad for 1979, 1980, and half of 1981, and from the SFDD3 data base, which covered the Atchison, Topeka, and Santa Fe Railway from 1974 through 1979. The following categories of rail defects were examined:

All types combined (ALL).

All types except worn and damaged rail (ALLE).

Bolt hole breaks (BHB).

Horizontal split heads (HSH).

Head/web separation (HWS).

Detail fractures (DF).

Rail defect clusters were correlated on all the divisions and districts combined, each division separately, and for the ATSF New Mexico Division, First District alone.

4.2.3 Conclusions

The analysis indicates that rail defects cluster on the same section of track from year to year. In fact, track segments that had defects in one year were more likely to have defects in subsequent years. These conclusions hold for both the BN and the ATSF railroads.

In general, it seemed that the defect categories that contained the most defects also exhibited the strongest correlations. The all-types defect category had the highest correlations from one year to another, followed by bolt hole breaks and detail fractures. Horizontal split heads and head web separations showed the least correlation. These results held for all the divisions.

Rail defects were also found to cluster on the same section of track from one year to the next when only a single district was considered (the ATSF New Mexico First District was specifically examined).

However, defects clustering on the same section of track from year to year can be greatly affected by the variation in track quality across the system. Systems having segments of poor quality track as well as good quality track will tend to see more defects each year in the poor track and fewer defects in the good track. This will produce a correlation over time which is dependent on the amounts of poor quality and good quality track. Nevertheless, correlations exist. Track having defects in the past would have defects in subsequent years. However, it is the size of the correlations that is uncertain. From the ATSF data, therefore, significant correlations exist, and much of the correlation may have been caused by track and traffic characteristics which generally determine track quality.

4.3 SPATIAL AUTOCORRELATIONS

4.3.1 Objective

Spatial correlations of BN and ATSF rail data were made to determine the size of a rail defect cluster for one year.

4.3.2 Approach

Calculations of spatial autocorrelations of rail data in data bases BNDD3 and SFDD3 were applied to the following defect categories:

All type defects combined (ALL).

All types except worn and damaged rail (ALLE).

Bolt hole breaks (BHB).

Horizontal split heads (HSH).

Head/web separations (HWS).

Detail fractures (DF).

Calculations were made for all line segments combined, for all divisions combined, and for all years combined. In addition, correlations were calculated for each division separately for all types of rail defects. All of the above correlations were calculated using both the unadjusted and adjusted (for the number of components in the numerator and denominator) formulas.

Additional correlations were calculated (unadjusted formula only) for individual years. Only the category of all rail defect types combined was used. The following cases were considered:

- autocorrelations of each year with itself, 1974-1979, Lags 1 miles through 30 miles; all divisions combined; and each division separately.
- autocorrelations of each year with each other year (i.e., all possible pairs of the years 1974-1979), Lags 1 through 30 miles; all divisions combined; and each division separately.

4.3.3 Conclusions

Examination of the autocorrelation calculations for the various defect categories revealed that the categories of all types of rail defects and bolt hole breaks exhibited the highest values. For the Atchison, Topeka, and Santa Fe Railway bolt hole breaks were not the most common defects. There are many possible reasons for the high autocorrelation among bolt hole breaks - one of which may be that bolted rail occurs in stretches of older track. Thus, there may have been stretches of adjacent track miles having bolt hole breaks, followed by stretches of continuous-welded rail having relatively few bolt hole breaks.

The number of miles (lags) for which rail defect clusters persisted varied considerably for the different categories and divisions. Generally, the length of the defect clusters remained fairly large up to 10 or 20 miles. The adjusted autocorrelations showed the same general trend, except that the autocorrelations remained large for a slightly greater number of miles.

For the individual years, the length of the defect clusters remained fairly large, approximately 8 to 13 miles, except for the years 1974 and 1975, for which the distance was about 18 miles.

Although rail defects cluster from one year to the next, the length of a cluster tends to be smaller for succeeding years. For example, one year, on one section of track, a rail defect cluster was 13 miles long, but only 6 miles long the following year.

4.4 AID ANALYSES

4.4.1 Objectives

Automatic Interaction Detector (AID) analysis was applied to the BN and ATSF rail data (1) to determine empirically which track and traffic variables were most associated with differences in bolt hole break and detail fracture occurrence rates, and (2) to identify which combinations of values of these variables produced high defect rates and which ones produced low defect rates.

4.4.2 Approach

AID analysis uses a computer algorithm to systematically search a data set for associations among the variables. Given some response variable and a set of predictor variables thought to have an effect on it, the data set is divided into groups according to combinations of the values of the predictor variables. These groups are formed to highlight the differences in the response variables among the groups.

In this analysis each observation in the data set consisted of a 1-mile track segment having a constant set of track and traffic characteristics. The data bases BNDD1 and SFDD1 were used in this study, with all partial-mile segments omitted. Each observation contained a count of the number of bolt hole breaks (BHBs) and detail fractures (DFs) found during the time period under consideration: January 1979 through June 1980, for the Burlington Northern Railroad (BN) data, and the years 1974 through 1979 for the Atchison, Topeka, and Santa Fe Railway (ATSF) data. In addition, the track and traffic characteristics for the segment were also included.

Separate AID analyses were performed for bolt hole breaks (BHBs) and detail fractures (DFs). The response variables used were of two classes which indicated whether or not a defect was found in the track mile during the time period under consideration:

BHBIND = 1, if at least one BHB was found
         0, otherwise.

DFIND = 1, if at least one DF was found
        0, otherwise.

The number of DFs and BHBs were divided into two classes for several reasons:

1. The available data indicated that a substantial number of track miles did not have any bolt hole breaks or detail fractures during the period. To assist in understanding the differences, a comparative analysis was made on those segments with defects against those without.
2. The large number of track miles with no defects, along with the extremely large numbers of defects found in some track miles, indicated that an analysis based on the average number of defects per mile would not provide an adequate description of the typical track segment. Further, the relatively few segments with large numbers of defects would distort the analysis results.
3. Finally, the average value of the dichotomous variable for the track segments in each selected AID group provided an estimate of the probability of a defect occurrence for an arbitrary track mile having the associated set of track and/or traffic characteristics.

The predictor variables used in the AID analyses of the BN data are given in Table 6; the predictor variables used in the ATSF analyses are given in Table 7. AID analysis required that the values of each predictor variable be categorized into groups. The last columns of Tables 6 and 7 indicate the categories used. The midpoint of each category was used to label the category.

Several analyses were done for each rail defect type. Summaries of the AID groups selected for the final analyses are presented in Figures 4 through 7.

4.4.3 Conclusions

Bolt Hole Breaks. Examination of the AID trees given in Figures 4, 5, and 6 showed that year rail laid (YRLAID) was selected by all three of the

TABLE 6. PREDICTOR VARIABLES USED IN THE AID ANALYSES FOR BN DATA

Variable	Mnemonic	Categories
Line segment	S	Each segment individually
Year rail laid	YRLAID	10-year intervals
Rail weight	WT	WT ≤ 66; 72-77; 85-90; 100-112; 115-119; 124-132; 136
Welded or bolted	WELD	Each individually
Rail section	SECT	Each individually
Cumulative MGT	TONC	50-MGT intervals
Average MGT, 1978-1980	TONAVG	5-MGT intervals
Number of detector car inspections, 1978-1980	INS	1-4; 5-8; 9-12; 13-16; 17-20*

*There were some track miles which did not have any inspections from 1978 through 1980. These track miles were eliminated since they did not have the same chance of a rail defect occurrence being found.

TABLE 7. PREDICTOR VARIABLES USED IN THE AID ANALYSES FOR ATSF DATA

Variable	Mnemonic	Categories
Division	DIV	Each division individually
Year rail laid	YRLAID	10-year intervals
Rail weight	WT	90; 110; 112; 115; 119; 131; 132; 136
Cumulative MGT*	TONC	50-MGT intervals
Average MGT, 1974-1979	TONAVG	5-MGT intervals
Average number of heavy weight (> 131 tons), 1974-1979	HAVG	1000-car intervals
Average number of average weight (110-131 tons) cars, 1974-1979	AVGAVG	10,000-car intervals
Average number of total cars, 1974-1979	ALLAVG	100,000-car intervals
Rail position (curve, tangent)	POS	Each individually
Speed limit	SP74	10 MPH intervals

*The variable TONC was approximated by multiplying the average tonnage for 1974 through 1979 by the number of years since the track was laid.

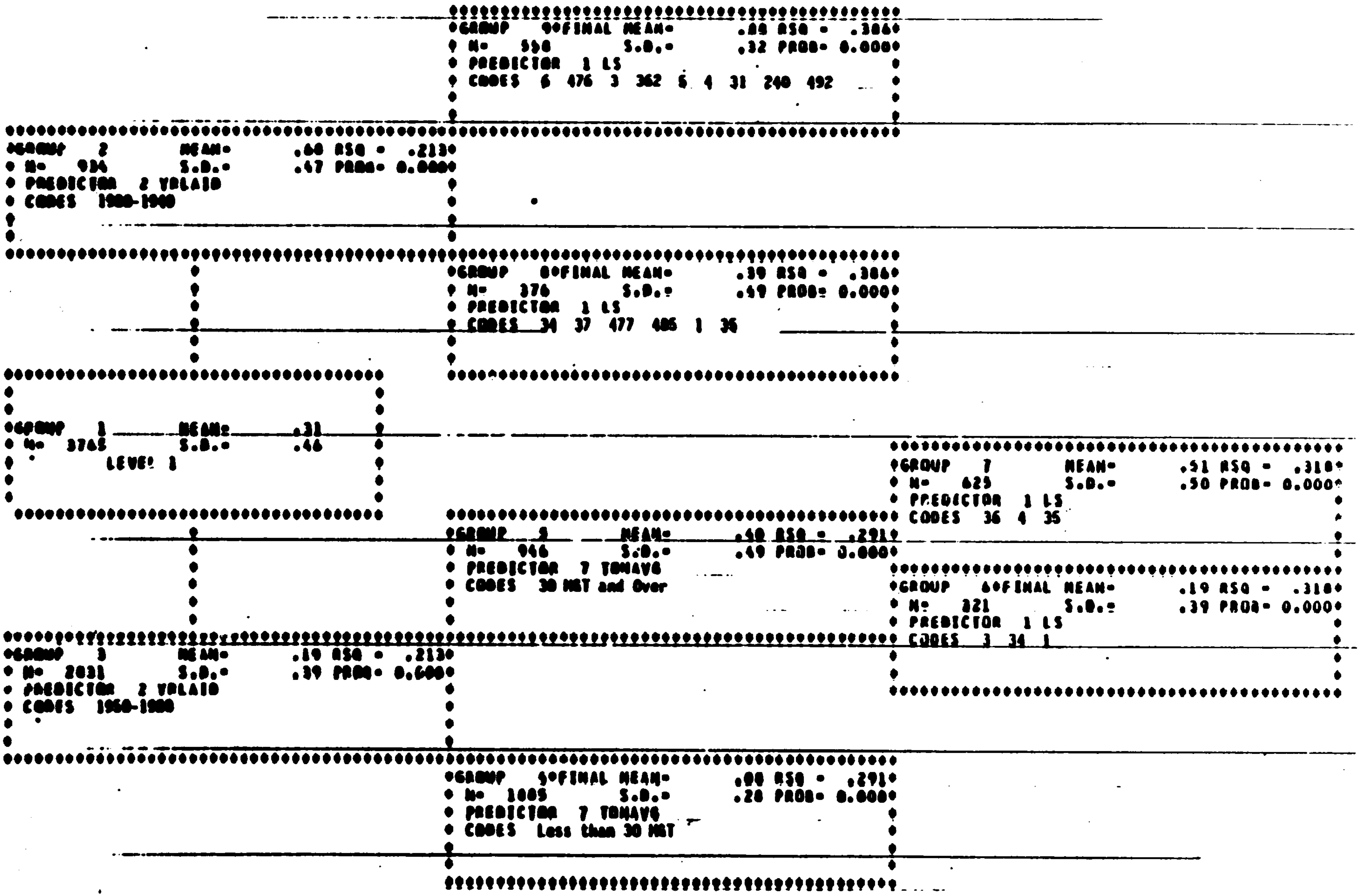


FIGURE 4. AID TREE FOR BOLT HOLE BREAKS - ORIGINAL 16 BN LINE SEGMENTS

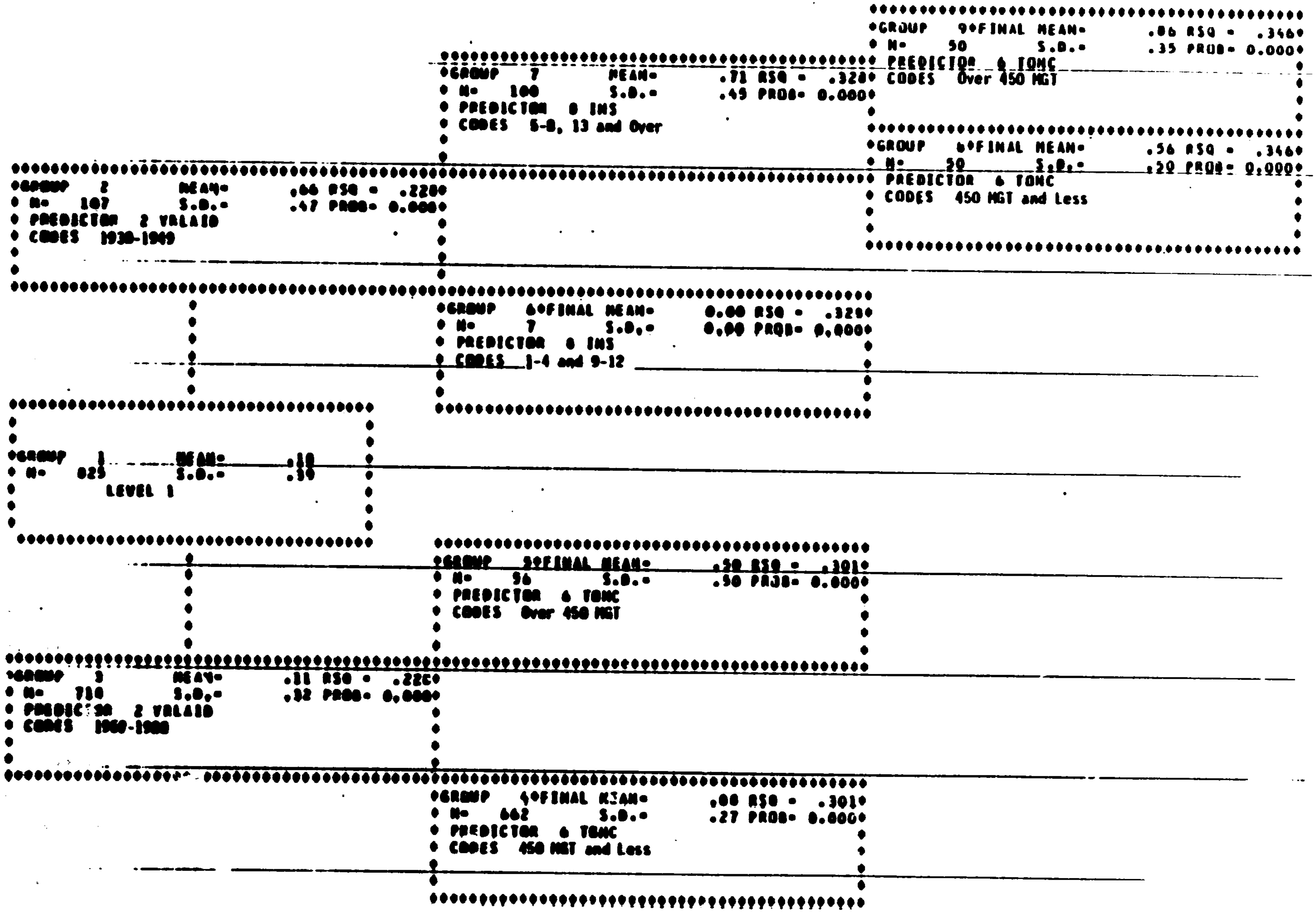


FIGURE 5. AID TREE FOR BOLT HOLE BREAKS - ADDITIONAL 10 BN LINE SEGMENTS

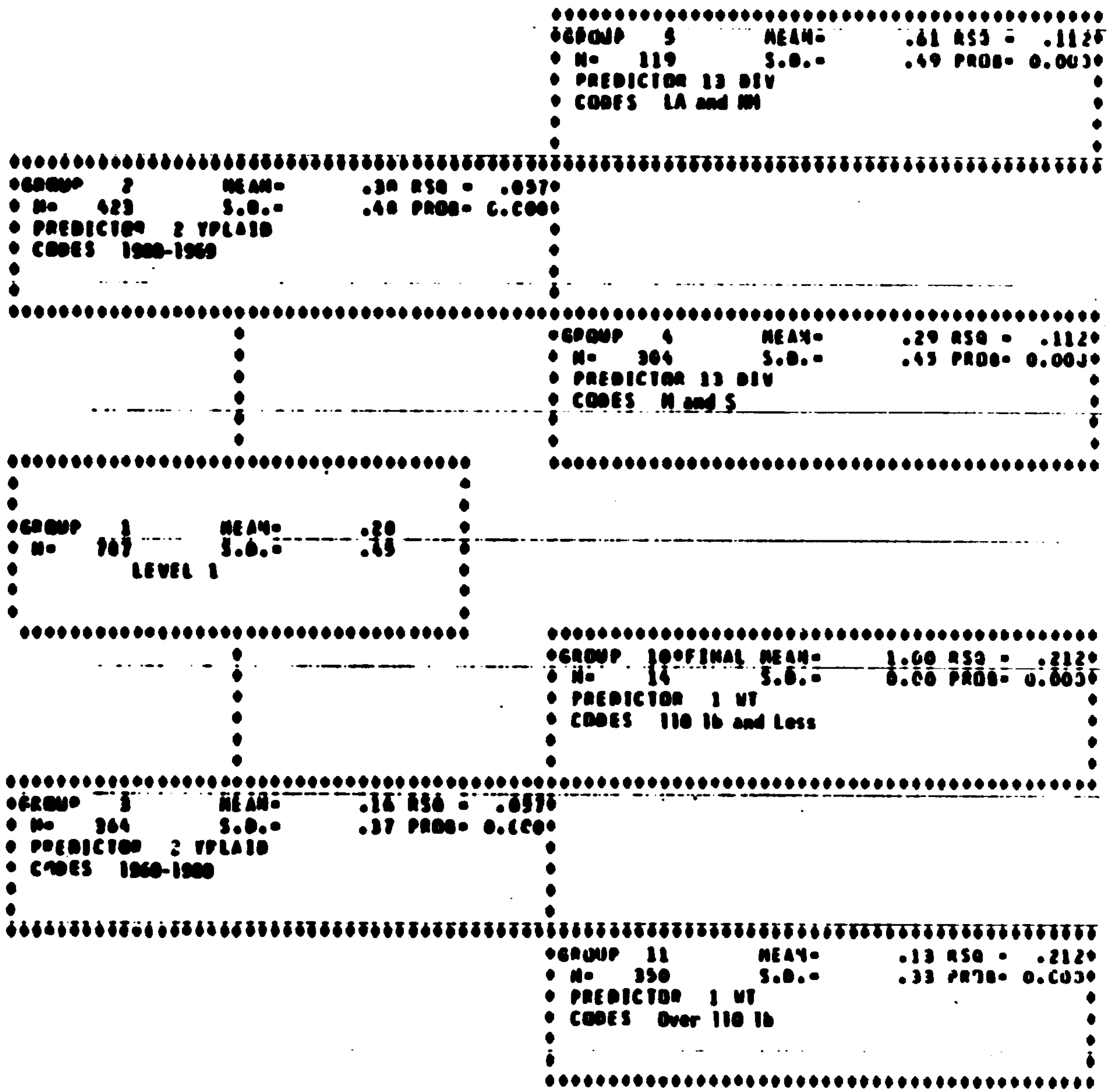


FIGURE 6. AID TREE FOR BOLT HOLE BREAKS - ATSF DATA

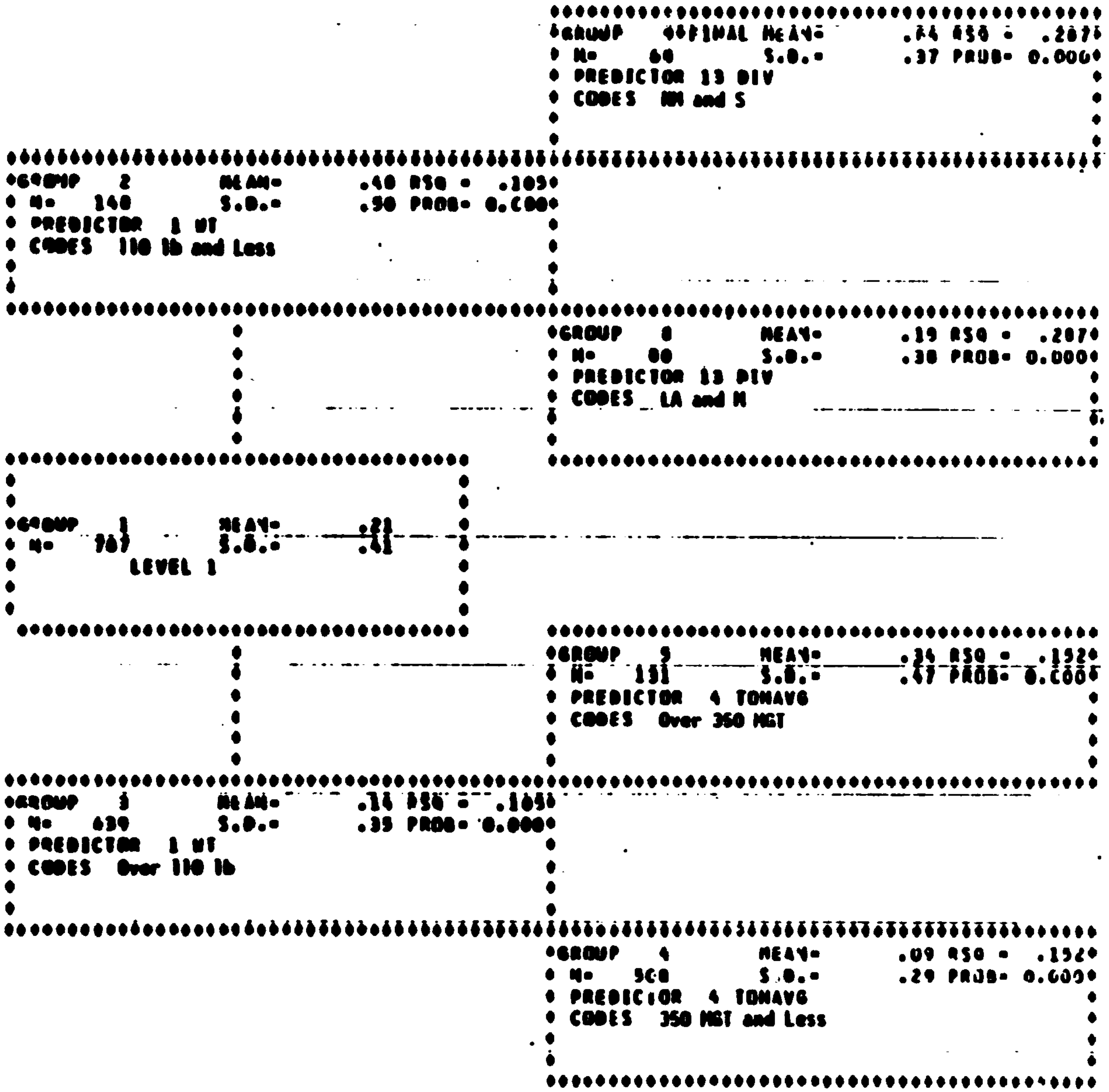


FIGURE 7. AID TREE FOR DETAIL FRACTURES - ATSF DATA

independent data sets as the single variable which provided the most discrimination between track with bolt hole breaks and track without bolt hole breaks. Both BN analyses selected 1950 as the separation point. Track laid prior to 1950 was placed in the high-defect rate group, while all newer rail was placed in the low-defect rate group. The ATSF analysis selected 1960 as the separation point. Table 8 presents the three data sets for the percentages of track miles that had bolt hole breaks for each of the selected AID groups. Note the extreme difference in percentages between the old and new rail, particularly for the two sets of BN line segments. The estimated probabilities of a bolt hole break (BHB) occurrence on the older rail ranged from 3 to 6 times higher than for newer rail.

It is interesting to note that in only one of the three data sets (the BN additional 10 line segments) did either of the tonnage variables, TONC and TONAVG, compare to the discrimination power of YRLAID. For the BN additional 10 line segments, the variable TONC ranked second to the variable YRLAID. They were separated at 450 MGT. The general lack of discriminatory power for tonnage at this first set of AID splits was most likely caused by the existence of some very old track in quite poor condition, yet with very little cumulative MGT. (The existence of track that produced a rather high defect rate for the lowest levels of cumulative MGT had been noted earlier in the study.)

The newer rail groups for both sets of BN line segments were subsequently split on tonnage variables. For the original 16 BN line segments, the split was made on TONAVG, with 30 MGT per year as the separation point. The high rate group had at least one BHB in 40 percent of the track miles, while the low group had only an 8 percent BHB occurrence rate. The newer rail for the additional 10 BN line segments was split on TONC, with 450 MGT as the separation point.

TABLE 8. PERCENTAGES OF TRACK MILES WITH BOLT HOLE BREAKS FOR THE SELECTED AID GROUP

AID Group	BN Original 16 Line Segments		BN Additional 10 Line Segments		ATSF Districts	
	%	Total Miles in Group	%	Total Miles in Group	%	Total Miles in Group
All data	31	3765	68	825	28	787
Older rail	68	934	66	107	38	423
Newer rail	19	2831	11	718	16	364

Figure 5 indicates that the occurrence rate was 50 percent for the high MGT group and only 8 percent for the low MGT group.

The newer rail group for the ATSF data was split on rail weight, with rail weights up to 110 pounds placed in the high-defect rate group (100 percent defect occurrence rate) and all heavier rail placed in the low-defect rate group (only a 13 percent defect occurrence rate). None of the tonnage or traffic variables had much discriminatory power for the newer rail.

The older rail groups for the BN 16 line segments and the ATSF were split on line segment and division, respectively. This indicates that general conditions of the various systems, rather than any particular single track or traffic variable, were responsible for differences in BHB occurrence rates. For example, the entire line segment #492 was in very poor condition. The line segments and divisions, placed in high and low groups, are presented in Figures 4 and 6 for the two data sets.

The additional 10 BN line segments had the older rail group split on the number of rail detector car inspections (IMS). However, only seven track miles were split off; thus, the split was not very meaningful. The subsequent split for the remaining track was a TONC, again at 450 MGT.

Summarizing the AID analyses of bolt hole breaks, it appeared that year rail laid, tonnage, and general differences among the track systems (i.e., ATSF divisions versus BN line segments) were the variables most associated with differences in rail defect occurrence rates.

Detail Fractures. No analysis of detail fractures (DFs) was done for either of the two sets of BN line segments because of the small numbers of such defects. Only 4 percent of the track miles in the original 16 BN line segments and 2 percent in the additional 10 BN line segments had any DFs. In the ATSF data, 26 percent of the track miles had at least one DF and, therefore, an analysis was performed. The summary tree of AID splits for ATSF detail fractures is given in Figure 7.

Rail weight was selected as the single variable most associated with DF occurrence rate for the ATSF. Weights of 110 pounds and less were placed in the high-rate group, which had a 49 percent DF occurrence rate, versus a 14 percent rate for the heavier rail. One possible factor that could have caused weight to be selected was its general association with year rail load (YRLAID). However, this did not appear to be the case for two reasons:

1. Examination of the AID output indicated that YRLAID itself had very little value in discriminating between track which had DFs and track which did not.
2. The ATSF Southern Division, Conroe District, consisted of two distinct sections of track, one of which had very few DFs and the other, many. The two sections of track were very similar in all the track and traffic characteristics except rail weight. The track sections which had many DFs were 110-pound rail, while the other track miles were 131-pound rail.

From the ATSF raw data, it appeared that detail fractures appeared to cluster within certain districts. The New Mexico Division, Carlsbad District, and Southern Division, Conroe District, are examples of districts which had relatively large numbers of DFs. In fact, the two divisions associated with these districts were placed in the highest DF occurrence rate group. The lighter rail group was split on division with these two divisions placed in the high group (an 84 percent DF occurrence rate). The remaining divisions, Middle and Los Angeles, had only 18 pereent of occurrence rates among their lighter rail.

The heavier rail group was split on TONAVG, with rail having ≥ 350 MGT placed in the higher rate group.

Summarizing the AID analysis of ATSF detail fractures, rail weight appeared to be the most important factor; however, curves were not considered in this analysis and may have played a hidden role. General differences among districts and tonnage appeared to be secondary factors. A second independent data set would have been useful in verifying the results of the AID analysis of rail detail fractures in ATSF rail.

4.5 REGRESSION ANALYSES

4.5.1 Objective

Regression analyses of the BN and ATSF rail defect data were performed to determine which of the track, traffic, or other variables are the best predictors of whether or not a track mile will experience a rail defect in the future. In particular, these analyses provided a means to assess the value of using a track mile's past defect record as a predictor.

4.5.2 Approach

Although the objectives of the AID analyses and regression analyses are similar, there are several important advantages to the regression analysis approach and one important disadvantage. The advantages are:

1. The development of a single prediction equation for the probability of a future rail defect which assesses the simultaneous influence of all the predictors.
2. The availability of an assessment of how well the equation predicts.

3. No requirement that the values of the predictor variables be grouped into categories.

The disadvantage is that a specific form must be assumed for the relationship between the response variable and the predictors. For simplicity, a linear model was assumed in all cases. That is, it was assumed that the probability of a future rail defect varies linearly with any changes in the predictor variables and that these changes were additive. For example, an increase in cumulative tonnage over the track (MGT) would increase the probability of a rail defect occurrence at the same rate, regardless of whether the rail was old or new. Although this assumption may not have been totally realistic, it was felt that it would provide an adequate approximation for the type of global analyses conducted.

The data bases BNDD1 and SFDD1 were used for these regression analyses, with all partial miles excluded. Analyses were conducted for bolt hole breaks individually for each of the two BN line segment sets and the ATSF data. Because of the low frequency of detail fracture occurrence among the BN data, detail fractures were analyzed only for the ATSF data.

The response variable used for the BN analyses was BHBIND80, where

BHBIND80 = 1, if the track mile had at least one BHB during 1980
0, if otherwise.

The predictor variables used were YRLAID, WT, WELD, TONC, TONAVG, and BHBIND79, where

BHBIND79 = 1, if the track mile had at least one BHB during 1979
0, if otherwise.

See Table 6 in Section 4.4 (AID Analysis) for a description of the other variables.

The response variables used for the ATSF analyses were BHBIND2 and DFIND2, where

BHBIND2 = 1, if the track mile had least one BHB from 1977 through 1979
0, if otherwise.

DFIND2, for detail fractures, was defined the same as BHBIND2. The predictor variables used were YRLAID, WT, TONC, TONAVG, HAVG, ALLAVG, AVGAVG, CVIND, BRGIND, GRDIND, and BHBIND1 (or DFIND1, for detail fractures), where

CVIND = 1, if the track mile had at least one curve
0, if otherwise,

and

BHBIND1 = 1, if the track mile had at least one BHB from 1974 through 1979
0, if otherwise.

BRGIND and GRDIND were defined like CVIND for bridges and grades, and DFIND1 was defined like BHBIND1 for detail fractures. See Table 7 in Section 4.4 (AID Analysis) for a description of the other variables. Note that none of the predictor variable values for the regression analyses were grouped into categories as was required for the AID analyses. In the regression analyses, for example, each individual YRLAID value was actually used. Since it was desirable to include the curve, grade, and bridge variables in the analysis, only those districts for which this information was available were used in the ATSF analyses.

The regression method used was "stepwise regression," where the variables are entered into the prediction equation in the order of each one's ability to improve the equation's predictive power.

4.5.3 Conclusions

Bolt Hole Breaks. For bolt hole breaks, the prediction equations obtained were:

Original 16 BN Line segments

BHBIND80 = 9.098 + .346 BHBIND79 - .00897 YRLAID + .00360 TONAVG + .101 WELD - .0043 WT + .0000589 TONC (1)

Additional 10 BN Line Segment

BHBIND80 = -.629 + .223 BHBIND79 + .000953 YRLAID + .000742 TONAVG - .0847 WELD - .00202 WT + .00470 TONC (2)

ATSF Data

BHBIND2 = 7.519 + .199 BHBIND1 - .00392 YRLAID - .00872 TONAVG + .000252 WT + .000245 TONC - .00000244 HAVG - .0328 BRGIND + .2626 GRDIND + .000000734 ALLAVG - .00000198 AVGAVG - .0230 CVIND (3)

Too much significance should not be placed on the sizes of the estimated coefficients since the appropriateness of the linear model and the possibility of significant interactions existing were not carefully examined, nor were the data carefully screened for the possibility that relatively few track miles were too greatly influencing the result. Rather, the equations should be examined in terms of the signs of the coefficients and their significance in the equation with respect to improving its predictive ability.

Equation (1) was considered to be the most reliable since it was based on the largest number of track miles (3814 miles). The positive coefficients for BHBIND79, TONAVG, WELD (1 = WELDED, 0 = BOLTED), and TONC indicated that the probability of a BHB in 1980 (P(BHB)) was increased for larger values of these variables. Each of these variables except WELD would be expected to be positively related to the probability of a BHB. The incorrect sign on WELD was probably due to the form of the assumed model on the relationship of WELD with some other variable, perhaps not in the equation, which was positively related with P(BHB). The negative coefficients for YRLAID and WT would be expected. As the YRLAID increases (i.e., the amount of newer rail increases) P(BHB) decreases. All of the variables except TONC were statistically significant as well as practically significant. (Practical significance means that the probability of a defect changes a fair amount when the predictor variable is changed.) For example, with a YRLAID coefficient of -.00897, P(BHB) is estimated to be decreased by .0897 for every 10 years newer that a rail is (i.e., -.0897 = -.00897 x 10), given that all other variables remain fixed. For TONC, the effect has no practical significance. An increase of 100 MGT is estimated to produce only a .0000589 x 100 = .00589 change in P(BHB).

An assessment of the adequacy of the equation's predictive ability revealed that the variables contained a very significant amount of predictive ability but that they were certainly not the only factors which affected the probability of bolt hole breaks, P(BHB). Other unknown variables not included in the model played a significant role in affecting P(BHB). Additional predictor variables and refinement of the model are needed to obtain a more accurate prediction equation.

It is noteworthy that the single most important predictor variable was BHBIND79. That is, a track mile's rail defect history was a better predictor of the likelihood of a future rail defect than any track or traffic variable. In fact, in every analysis but one, for both bolt hole breaks and detail fractures, for both BN data and ATSF data, the best single predictor was the track mile's past rail defect record. The one exception was the 10 additional BN line segments, in which BHBIND79 was the second best predictor. The coefficient of .346 for BHBIND79 in Equation (1) indicated that P(BHB) in 1980 was estimated to be .346 (34.6 percent) higher for a track mile which had a BHB in 1979 than for a mile which did not.

In Equation (2) all of the variables except YRLAID had the sign one would expect. However, only the variables TONC, BHB79IND, and WELD were found to be significant. Again, the predictive ability of the model was significant, but additional variables would be needed to obtain a highly accurate prediction equation.

In Equation (3) many of the variables had different signs from what one would expect. However, among those which had the wrong sign, none were statistically significant. The regression analysis results indicated that HAVG, BRGIND, TONAVG, WT, and CVIND did not have clear relationships with P(BHB). All the variables which had correct signs (TONC, YRLAID, BHBIND1, GRDIND and ALLAVG) were statistically significant except ALLAVG. In addition, all were of practical significance in the sense that a change in the value of the variable produced a non-negligible change in P(BHB). For example, a 30-year increase in YRLAID (i.e., newer rail) was estimated to produce a .00392 x 30 = .1176 drop in P(BHB).

<u>Detail Fractures</u>. The prediction equation obtained for DFs with the ATSF data was

$$\begin{aligned} \text{DFIND2} = {} & 5.901 + .2302\ \text{DFIND1} - 00275\ \text{YRLAID} - \\ & .0109\ \text{TONAG} - .00684\ \text{WT} + .000280\ \text{TONC} - \\ & .00593\ \text{BRGIND} + .2884\ \text{GRDIND} - .00000377\ \text{AVGAVG} \\ & - .1370\ \text{CVIND} + .000000926\ \text{HAVG}. \end{aligned} \qquad (4)$$

In Equation (4), the variables TONAVG, CVIND, and BRGIND have the wrong signs for their coefficients. Of these, CVIND and TONAVG were statistically significant. The reasons for the wrong signs were likely due to using the wrong form for the model, or unknown relationships among these variables and others not in the model. DFIND1, TONC, WT, ALLAVG, AVGAVG, and GRDIND were all highly significant variables, with DFIND1 being the single most important one.

The predictive ability of the overall equation was significant but, again, additional variables and refinement of the equation would be needed to obtain a highly accurate prediction model.

4.6 INDIVIDUAL RAIL LINE PROFILES

4.6.1 Objectives

Individual rail line profiles were constructed (1) to obtain a profile of each rail line with respect to its track and traffic characteristics and (2) to determine how the rail defect occurrence rates varied with the levels of each track and traffic variable.

4.6.2 Approach

Profiles were constructed for each line segment, or division and district (rail line). For each track and traffic variable, the entries in the profile presented the number of track miles (N) in the line which were classified into the associated category of the variable. In addition, the proportion of track miles in the category which had at least one bolt hole break (P(BHB)) or detail fracture (P(DF)) was given. Separate sets of profiles were prepared for bolt hole breaks and detail fractures.

4.6.3 Conclusions

Examination of the ATSF profiles for bolt hole breaks revealed the effect of year rail laid (YRLAID) on bolt hole break occurrence rates. This effect appeared to hold uniformly over the different rail lines. The BN profiles also showed this to be true. These results indicate that the effect of YRLAID, apparent from other analyses, was not due to data from just a few rail lines, but, rather, was universal over many lines. The BN profiles also revealed general trends in the probability of a bolt hole break P(BHB) over the levels of cumulative tonnage (TONC), and rail weight (WT). These effects were not as apparent from the ATSF profiles.

For detail fractures, the effect of WT was apparent in both the ATSF and BN profiles, although the BN data did not have many DFs. No trend seen for YRLAID was consistent with earlier analyses done on detail fractures. Although the number of DFs was too small for the BN data to provide a definite pattern, the rail section "TR" seemed to be associated with higher DF occurrence rates than rail section "RE". Those line segments which had a fair number of track miles of both rail sections had generally higher rail defect occurrence rates for "TR".

Since the profiles are all one-dimensional, care must be taken in drawing conclusions concerning relationships between rail defect rates and the levels of a given traffic or track variable. The trend which is seen may actually be due to the variable's relationship with another variable. For example, rail weight (WT) may appear to have an effect on the probability of a bolt hole break, P(BHB), but WT's relationship with year rail laid (YRLAID) may actually be the cause of the trend, or vice versa.

Finally, these profiles are useful for studying the composition of a specific rail line and for contrasting different lines. For example, the profile for BN LS 492 revealed that all the track miles had fewer than 50 MGT over the track and received an average of 6-10 MGT over the track per year from 1978 through 1980. Also, this track was primarily 90- and 115-pound rail.

4.7 FREQUENCY COUNTS OF RAIL DEFECTS BY TYPE AND DETECTION METHOD

4.7.1 Objective

Frequency counts of rail defects, by type and detection method, were used: (1) to determine the percentage of defects found by detector car versus those found by SERVICE; and (2) to determine how these percentages vary with rail defect type.

4.7.2 Approach

All rail defects in the rail defect file for the 16 BN line segments and the five ATSF districts were included in this part of the study, which covered all the months and all the rail positions.

Frequency counts of rail defects, broken down by defect type and detection method, were generated. The detection method corresponded to either service or the number of the rail detector car which found the defect. Service defects also included defects found by hand-held probe as well as those defects found as a result of a derailment. Rail detector cars 975103 through 975105 were magnetic, while all the others were ultrasonic. All the cars were hi-rail and owned by the Burlington Northern Railroad.

Figure 8 is a bar graph which compares rail defects detected by ATSF rail detector cars and rail defects detected by service, for the ATSF Middle Division, Fourth District. The x-axis represents months, starting with January 1974, and extending to December 1979. For each month, the number of service-detected defects was counted; the bar height (shaded) corresponds to that count. Thus, six rail defects were found in January 1974; four rail defects were found

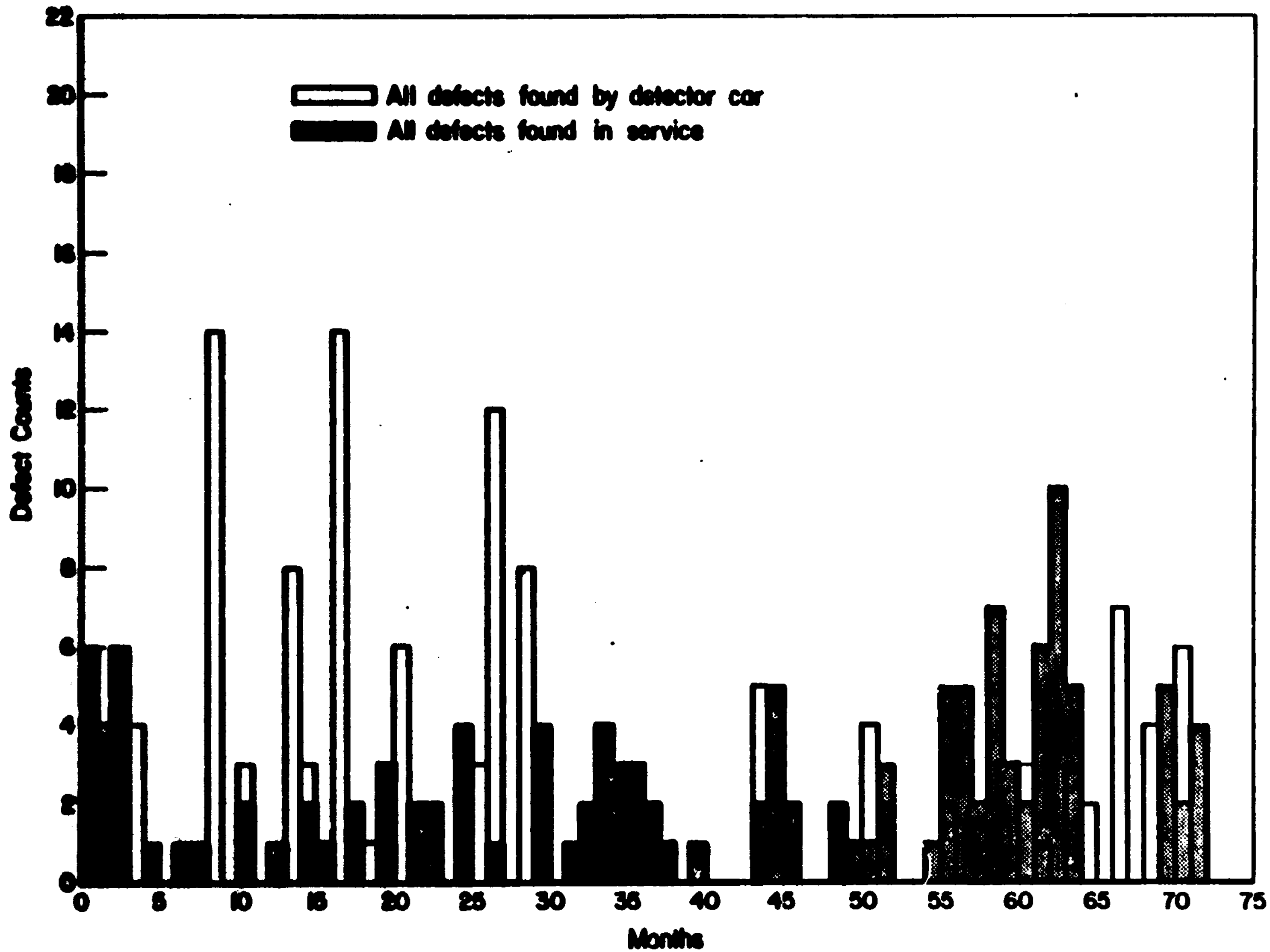


FIGURE 8. RAIL DEFECT DETECTION HISTORY FOR ATSF MIDDLE DIVISION, FOURTH DISTRICT

in February 1974; six rail defects were found in March 1974; and so forth. The second set of bars (open bars) represents defects found by detector cars. Inspections occurred in March 1974; April 1974; September 1974; November 1974; and February 1975; etc.

4.7.3 Conclusions

The percentage of all-type rail defects which were detected by service varied among the BN line segments. Burlington Northern Line Segments 485 and 492 had very few service-found defects. Bolt hole breaks, transverse defects, and vertical split heads represented the majority of rail defects in these lines. The number of detection car inspections for these line segments was average. The percentages of service-found defects for line segments 240 and 477 were quite high. Line segment #240 had an extraordinary number of service-found bolt hole breaks and head web separations, while line segment #477 had a large number of detail fractures found in service. Line segment #240 had fewer than normal detector car passes, which may partially explain the large number of service-found defects; however, the number of passes in line segment #477 was about normal.

Typically, 10 to 20 percent of all-type rail defects in a line segment were found in service. However, within each defect type these percentages varied. In general, detector cars found head web separations and detail fractures for the vast majority of the time, while only half the bolt hole breaks were found by detector cars.

5. SUMMARY AND RECOMMENDATIONS FOR FUTURE WORK

5.1 SUMMARY

The statistical analyses performed in this study were of a global rather than local nature. That is, the analyses were based on large aggregates of data, and emphasis was placed on identifying the general overall track and traffic characteristics that affect rail defect occurrence rates. The smallest units of data used were 1-mile track segments; and, though a single track mile may have several bridges, grades, and curves, global statistical analyses will not reveal the effects of curves, bridges, and other local characteristics. Thus, the discrimination between a rail defect that occurred at a bridge or curve and one that did not was not possible when a mile basis was used. A single count was given for all the rail defects for each mile of a given type.

The availability of such large amounts of data for the Burlington Northern and the Atchison, Topeka, and Santa Fe railroads provided an excellent opportunity to conduct global analyses for which the local characteristics tended to be "averaged out", thus allowing the effects of such factors as cumulative MGT (TONC) and year rail laid (YRLAID) to be seen. The averaging process was also enhanced by the use of an entire track mile as the observational unit. A specific local characteristic only exists for a small portion of a mile; therefore, when data is aggregated over the entire mile, the effect of the local characteristic is averaged out by the portion not possessing the characteristic. On the other hand, year rail laid (YRLAID) and other track/traffic variables tended to remain constant for the entire mile and their effects were therefore enhanced.

Some of the analyses conducted in this study were based on all types of rail defects combined, while others were based on specific types with particular emphasis on bolt hole breaks and detail fractures.

As one would expect, the results of the study indicated that year rail laid (YRLAID) was the single most important track or traffic characteristic associated with overall rail defect rate, particularly for bolt hole breaks. The effects of traffic loads (TONC and TONAVG) were also often apparent. However, traffic load generally was a secondary factor within a rail age category. The newer rail (i.e., rail laid between 1950 and 1970) was affected more by traffic loads than was the older rail. Older rail often received smaller loads, because of former, prevalent car capacity. Because of this, the true effect of high tonnage over the track (MGT) on older rail could not be determined.

For the analysis of detail fractures (DFs), the major portion of data came from the Atchison, Topeka, and Santa Fe Railway because the occurrences of detail fractures in the Burlington Northern Railroad data were not numerous enough for analysis. The ATSF data indicated that rail weight (WT) was the predominant global factor affecting detail fracture occurrences. However, the use of an entire mile as the observational unit did not permit a precise analysis of the effects of curves, bridges, and other local characateristics. The AID choice of rail weight (WT) as the factor most associated with differences in detail fracture occurrences was very interesting. Since AID selected year rail laid (YRLAID) as the most important factor for bolt hole breaks, the choice of rail weight (WT) for the detail fractures was due, not to track segments, but to true differences in the locations of the detail fractures and the bolt hole breaks. This tends to add significance to the selection of rail weight (WT) as a discriminatory variable for detail fractures.

In the regression analyses, the most important finding was the consistent selection of "past defect record" as the best predictor of future rail defects. This occurred for all three data sets and for both bolt hole breaks and detail fractures, with only one case in which "past defect record" was selected second. This indicates that no single track or traffic variable was as good a predictor of future rail defects as a track's past defect occurrence record. The correlation analysis confirmed this result. High degrees of correlation were found among rail defect rates over different periods of time, particularly for adjacent years. The regression and correlation results indicate that "past defect record" would be a useful variable in a track inspection formula.

The autocorrelation results also indicated that a strong positive relation exists among rail defect rates for neighboring miles of track. It is not clear, however, for how many neighboring miles the relationships exist. It is likely that the relationships result, somewhat, from the similarity of track and traffic characteristics for adjacent track. Future analyses, similar to the regression analyses, would be useful in determining the extent to which the similarity of major track and traffic characteristics (versus the prevalence of other unknown factors) is responsible for the autocorrelation.

The autocorrelations and the histograms of rail defects by milepost clearly indicate that rail defects cluster. A third piece of evidence supporting this conclusion is a comparison of the actual distribution of defects per mile against the Poisson distribution which assumes that defects occur at random locations. This comparison was done for the original 16 BNRR line segments combined. The results are given in Table 9. Note the larger-than-expected frequencies (if Poisson) for the lower and higher defects-per-mile categories. This indicates clustering. The large frequencies for the high categorties are the clusters,

TABLE 9. COMPARISON OF THE ACTUAL DISTRIBUTION OF RAIL DEFECTS PER MILE FOR ALL TYPE DEFECTS—ALL 30 MONTHS AGAINST THE POISSON DISTRIBUTION ($\mu = \overline{X}$)

Defects per Mile	Actual Number of Miles	Expected Number (if Poisson)	Chi-Square Contribution	Actual Percentage of all Defects	Poisson Percentage of all Defects
0	2818	328.9	18837.4	.449	.052
1	1133	970.0	27.4	.180	.155
2	580	1430.2	505.4	.092	.228
3	332	1405.9	820.3	.053	.224
4	254	1036.5	590.7	.040	.165
5	153	611.3	343.5	.024	.097
6	151	300.5	74.4	.024	.048
7	114	126.6	1.3	.018	.020
8	101	46.7	63.1	.016	.007
9	76	15.3	240.8	.012	.002
≥ 10	566	6.1	51391.5	.090	.001
	6278	6278.0	72896.0		

while the large frequencies for the low categories represent the track miles between the clusters.

In conclusion, there are a number of factors which can be used to identify areas of track likely to have rail defects in the future. These factors could be used to develop a track inspection formula for the purpose of guiding railroad companies in the optimal use of their inspection vehicles.

5.2 RECOMMENDATIONS FOR FUTURE WORK

The BN and the ATSF data bases contain a wealth of information which has yet to be fully tapped. Additional work is needed in the refinement of the regression analyses performed in this study. Little work was done to identify the correct form of the model that should be used. For this study, a simple linear form was assumed for all the variables, with no interaction terms included. It appeared from other analyses that an interaction term for year rail laid (YRLAID) and cumulative tonnage over the track (TONC) would be useful. Expansion of the model to include terms representing the condition of neighboring track would permit an assessment to be made of the portion of autocorrelation caused by factors other than the major track and traffic charaacteristics. Work is also needed to determine the optimal number of previous periods whose rail defect records should be used to predict future defects.

Selected segments of the available data would be useful in conducting analyses for local factors such as bridges and curves. In these analyses, smaller track sections, perhaps 500 feet, would be used as the observational unit. The results of the local analyses would be combined with the results of the global analyses. A comprehensive, yet simple, model containing both global and local charaacteristics would be developed for the purpose of predicting future defects.

Finally, more work is needed to compare the original 16 BN line segments with the additional 10 segments. A preliminary discriminent analysis did not reveal any significant differences in major track and traffic characteristics between the two sets. It is expected that the differences lay in either maintenance and inspection procedures or in local characteristics not yet analyzed.

114 copies