12-1995

An Examination of Railroad Capacity and its Implications for Rail-Highway Intermodal Transportation

David Bruce Clarke
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by David Bruce Clarke entitled "An Examination of Railroad Capacity and its Implications for Rail-Highway Intermodal Transportation." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Civil Engineering.

Frederick J. Wegmann, Major Professor

We have read this dissertation and recommend its acceptance:

Stephen H. Richards, Edwin Patton, Arun Chatterjee

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by David Bruce Clarke entitled "An Examination of Railroad Capacity and its Implications for Rail-Highway Intermodal Transportation." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Civil Engineering.

Frederick J. Wegmann, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor and
Dean of The Graduate School

AN EXAMINATION OF RAILROAD CAPACITY

AND ITS IMPLICATIONS FOR

RAIL-HIGHWAY INTERMODAL TRANSPORTATION

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

David Bruce Clarke

December, 1995

DEDICATION

This dissertation is dedicated to the memory of Dr. David Sedgwick Joy (1937-1995). Thanks, Dave, for being a mentor and friend. Without your help and advice, I wouldn't have walked this path.

ACKNOWLEDGEMENTS

In reaching the culmination of this dream, I was most fortunate to have been supported and encouraged by many fine individuals to whom I an eternally indebted. Without their help, I could not have accomplished this task. I would like to thank Dr. Frederick J. Wegmann, my Committee Chair, who provided continuous encouragement and guidance during, not just this effort, but my entire academic career. Drs. Arun Chatterjee and Edwin Patton generously spent much time discussing with me and reviewing the results of this research. Dr. Stephen H. Richards, Director of the Transportation Center, provided encouragement and financial support for this research, along with serving on my committee. I owe a special debt of thanks to my wonderful wife Shima, who showed unerring understanding and patience during this effort. Without her support, I could not have done this. My parents, James and Ellen Clarke, served as role models and provided continuous encouragement in my academic pursuits. I also thank my mother for her review of this manuscript. All errors remain, of course, my own. Finally, the use of the facilities of the University of Tennessee Computing Center is acknowledged.

ABSTRACT

After many years of decline in market share, railroads are now experiencing an increasing demand for their services. Service intensive intermodal transportation seems to be an especially promising market area. Since the historic decline in traffic has been accompanied by a reduction in network infrastructure, however, the railroads' ability to handle sizable traffic increases, at least in the short term, is in question. Since rail transportation is critical to the domestic economy of the nation, and is increasingly important in international logistics channels, shortfalls in railroad capacity are not desirable.

The published literature on railroad capacity is relatively sparse, especially in comparison to the highway mode. Much of what is available pertains to individual network components such as lines or terminals. Evaluation of system capacity, considering the interactive effects of traffic flowing through a network of lines and terminals, has received less attention. A tool specifically designed for evaluating freight railroad system capacity issues could be a useful addition to the rail analyst's toolbox.

The research conducted in this study resulted in the formulation and application of RAILNET, a multicommodity, multicarrier network model for predicting equilibrium flows within a railroad network. Designed for strategic planning with a short term horizon, the model assumes fixed external demand. The predicted flows meet the conditions for Wardropian system equilibrium. At completion, the solution algorithm

predicts the expected delay per train on each link, allowing the analyst to identify areas of congestion.

Following completion of the model, it was applied to a case study examining the railroad network in the southeastern U.S. The public use version of the Interstate Commerce Commission's Commodity Waybill Sample (CWS) provided flow data. The dissertation describes the procedure used to develop the case study and presents some results. The case points to major deficiencies in the CWS data which resulted in substantially less traffic in the network than is actually present. In general, given this limitation, the model behaved well and results appear reasonable, although not necessarily reflective of actual network conditions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Railroad capacity is a subject which has received relatively little published research attention, especially in comparison to the highway mode. This is probably due, in part, to the essentially private sector ownership of the rail system. The results of corporate research into topics such as capacity have competitive implications and are generally considered proprietary. Much of the published research has been on system components such as lines or terminals. Evaluation of system capacity, considering the interactive effects of traffic flowing through a network of lines and terminals, has received less coverage in the railroad literature, although such studies are commonplace in highway transportation.

After many years of decline in both freight and passenger market share, the railroads are now experiencing an increasing demand for their services. Since the decline in traffic has been accompanied by a reduction in railroad infrastructure, however, the ability of the remaining network to handle sizable traffic increases, at least in the short term, is in question. Empirical evidence certainly seems to indicate that capacity is constraining traffic growth in certain corridors. Since rail transportation is critical to the domestic economy of the nation, and is increasingly important in international logistics channels, shortfalls in railroad capacity are not desirable.

The purpose of this research was to develop an analytical framework for realistically predicting traffic patterns within the rail network and for evaluating the effects of these flows on capacity. Demand patterns for traffic, e.g. the traditional trip generation and distribution steps, are generated externally to the model. Unlike traditional highway traffic models, the assignment model developed considers multiple commodities, with each commodity having a potentially different set of costs and priorities. The model must also deal with the subdivision of the overall railroad network into subnetworks for specific companies, with transfers allowed only at designated points. The solution algorithm then assigns flows to the network so as to minimize the overall system transportation cost. This equilibrium approach should replicate the behavior of railroad management and produce facility volumes and performance levels closely approximating actual conditions.

The model allows policy makers to study congestion effects in the railroad system and to formulate and test options for network improvement. Although the research focuses on the intermodal component of railroad freight, because of its high service requirements and high current level of interest, the approaches developed are generally applicable to all rail freight flows.

Background

The following background material is intended to help the reader gain a perspective on the issues which underlie railroad capacity concerns: traffic growth and network reduction.

Growth in Traffic

During the 1990s, rail market share increased slightly after almost 50 years of decline. In terms of market share, railroad traffic reached a zenith in the U.S. during the 1920s. From this point, with the exception of the World War II period, competition from pipeline, automobile, airline, and motor carrier transportation steadily reduced railroad traffic share. In absolute terms, railroad freight traffic has continued to grow, but not at the rates of competing modes.

Railroad executives are optimistic that the current resurgence in traffic ends the long-term loss in rail traffic to competing modes, notably motor carriers and barge lines. High value freight is returning to the railroads in the form of service sensitive intermodal business. The railroads are finding, ironically, that motor carriers, once their nemesis, are becoming important customers as competition gives way to partnership. Despite contractions in Amtrak`s long haul rail passenger network, interest in short haul commuter service and regional rail passenger service is also high, promising additional business for rail carriers.

Table 1-1 provides an overview of recent intermodal activity by 10 U.S. Class I railroads (those companies having at least $251.4 million in gross annual operating

Table 1-1. Railroad Intermodal Volumes, 1994

Railroad	Units Moved (000's) 1994	Units Moved (000's) 1993	Percent Change
Conrail	1,611,852	1,372,787	17.4
Union Pacific	1,544,954	1,346,450	14.7
Southern Pacific	1,451,522	1,204,966	20.5
Santa Fe	1,416,392	1,218,889	16.2
Norfolk Southern	1,127,385	992,850	13.6
Burlington Northern	1,126,978	1,064,331	5.9
CSX Transportation	889,169	807,698	10.1
Chicago & North Western	766,451	729,685	5.0
Florida East Coast	323,400	324,186	(0.2)
Soo Line	242,877	209,992	15.7
Illinois Central	133,396	87,264	52.9
Kansas City Southern	92,168	63,113	46.0
Grand Trunk Western	35,470	39,916	(11.1)

Source: Association of American Railroads

revenue) and three major Class II railroads (companies having at least $20 million in annual revenues, but not reaching the Class I threshold). Figure 1-1 uses data from the Association of American Railroads (AAR, 1994) to show the industrywide intermodal volume trend for the years 1975-1994. During this period, intermodal freight traffic grew at an annual rate of 6.6%. Note that the data overestimate total loadings by about 7 percent, as each rail segment in a intermodal shipment where interchanges take place by highway may be counted as a separate shipment.

Despite its flattening in 1995, the long-term outlook for intermodal traffic growth appears strong provided that the railroads can provide service competitive with motor carriers. Growth trends for the past four years average 9.1 percent annually, with traffic increasing 14.1 percent in 1994. In discussing projected growth, railroad officials interviewed during a recent University of Tennessee study (Chatterjee et al, 1995) were extremely bullish. They pointed to new partnerships with motor carriers and continuing growth in doublestack demand as major reasons for optimism. The effects of increasing highway congestion, rising costs, and a severe driver shortage have made intermodal service an attractive alternative to over-the-road service for motor carriers.

Norris (1994) reports that, industrywide, between 20 and 25 percent of all rail revenues came from intermodal operations. When intermodal revenue growth is compared to the relatively flat revenue growth in other rail areas, it appears clear that the railroads must continue to support intermodal operations. If intermodal service reliability declines significantly, however, this business will return to the highways. In its annual

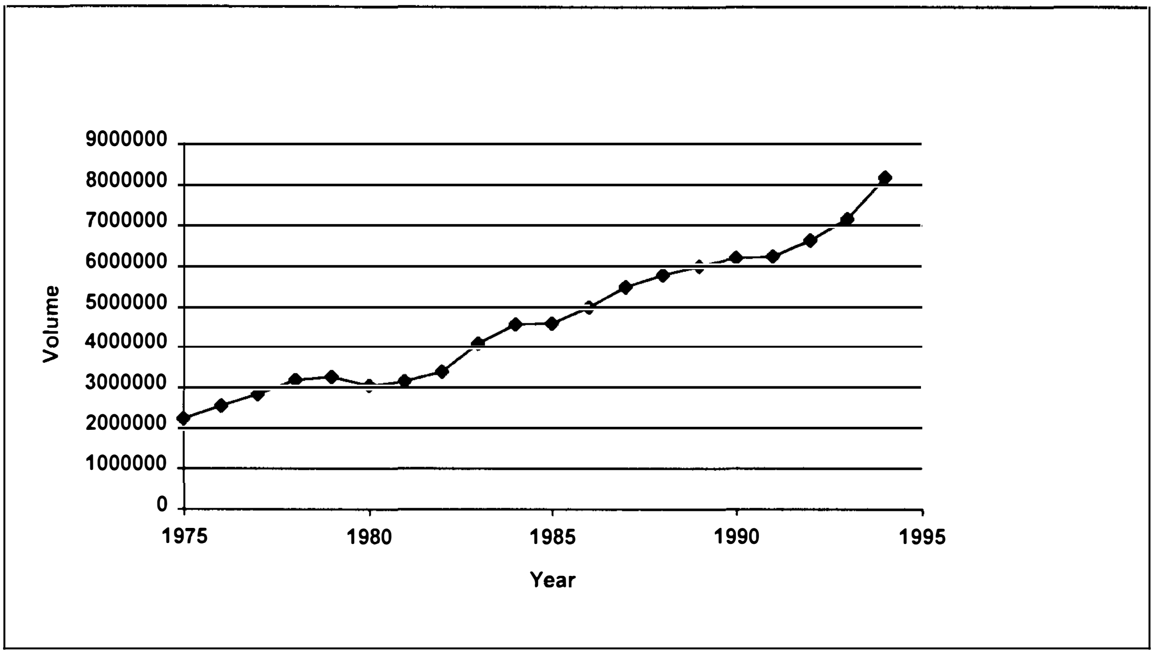

Figure 1-1. U.S. Railroad Intermodal Volumes, 1975-1994.

Source: Volume data from Railroad Facts, 1995 Edition, Association of American Railroads, Washington, DC.

survey of traffic and distribution managers, the Intermodal Association of North America (IANA, 1993) reports that 54 percent of its respondents felt that the biggest barrier to intermodal service was its slow and unreliable transit times. Decreasing transit time and improving reliability was felt to be the single most important factor in increasing intermodal use.

It should be noted that shipments of traditional rail commodities, especially coal, are also growing, although at a slower rate than intermodal. Welty (1995) writes that coal traffic accounts for almost 25 percent of rail carloads, 38 percent of rail tonnage, and 21 percent of freight revenues. While the revenue per ton is low, railroads enjoy coal traffic because it flows in large quantities, is easy to handle, and is not highly susceptible to modal diversion. *Progressive Railroading* (1995) provides AAR data showing that railroad coal carloadings approached 6.636 million in 1994, up 8.8 percent from 1993. National coal production is expected to rise about 2.5 percent in 1995, so rail carloadings should rise about the same amount. Since the largest production increases are in the low sulfur Powder River basin in Wyoming, western railroads will experience the bulk of the coal traffic growth.

Rail Infrastructure Changes

As an industry in decline during the period 1930-1980, railroad companies removed unneeded or underutilized infrastructure and deferred maintenance to reduce costs. Only since the passage of the Staggers Rail Act in 1980 have the railroads returned to a level of financial stability sufficient to maintain their infrastructure, although

downsizing continues. Despite the much improved financial picture, the railroad industry still cannot attract sufficient capital to make widespread capacity improvements to rights-of-way and terminals. During 1994, for example, only one major railroad, the Illinois Central, earned the cost of capital, estimated at 12.2 percent. The following sections explore these trends in more detail.

Reductions in route mileage

For many years, the U.S. railroad network has been considered overbuilt. During the period when railroads dominated surface transportation, promoters financed and constructed many miles of line which were not justified by existing or potential traffic. So long as railroads had little competition, these lines could survive on a thin traffic base. As competitive forces began to depress rates and revenues, however, railroad managers felt great pressure to reduce operating costs and increase the return on assets. Abandonment or downgrading of lines with marginal traffic became a key cost reduction strategy.

Until the 1980s, the Interstate Commerce Commission (ICC), which held regulatory authority over railroad abandonment, kept this process gradual. Line abandonment proceedings required volumes of supporting data and were frequently lengthy, with no guarantee to the carrier that the outcome would be favorable. As a result, railroad companies were unable to shed underutilized trackage at a rate necessary to offset the decline in revenues. During the 1970s, severe financial troubles among the northeastern (Penn Central, Erie Lackawanna, Reading) and granger railroads

(Milwaukee Road, Rock Island), which had large amounts of marginal or redundant track, threatened the health of the entire railroad industry. This forced Congress to reexamine and ultimately liberalize railroad abandonment regulations.

Since passage of the Staggers Rail Act in 1980, railroads have greatly accelerated the abandonment or sale of excess physical plant. Consider the reduction in size of the Class I railroad system. The 10 Class I freight carriers operate most of the major intercity and interregional trunk lines within the U.S. Although there are numerous smaller railroads, most of these serve local or regional markets. From a high water mark of 229,530 route miles (381,417 track-miles) in 1929, the Class I railroad network declined in size to 109,332 route-miles (183,685 track-miles) in 1994. Figure 1-2 shows these trends graphically. This reduction does not necessarily represent track abandonments. Changes in the financial threshold have removed numerous companies from Class I status, and many unnecessary Class I routes have been sold to smaller carriers.

Even with relative prosperity, line reduction trends continue today as the railroads attempt to increase overall traffic density and reduce costs. Under current trends, some observers predict an eventual Class I network size of 40,000 to 60,000 route-miles. By abandoning or selling intercity rail lines, as opposed to branches, the railroads are reducing the connectivity of the national rail network. This process of reduction assumes, to a large degree, that railroads have seen their day and that lost traffic will never return. The return of intermodal traffic and the renewed interest in regional passenger rail brings this assumption into question. From a public policy standpoint, therefore, the wisdom of

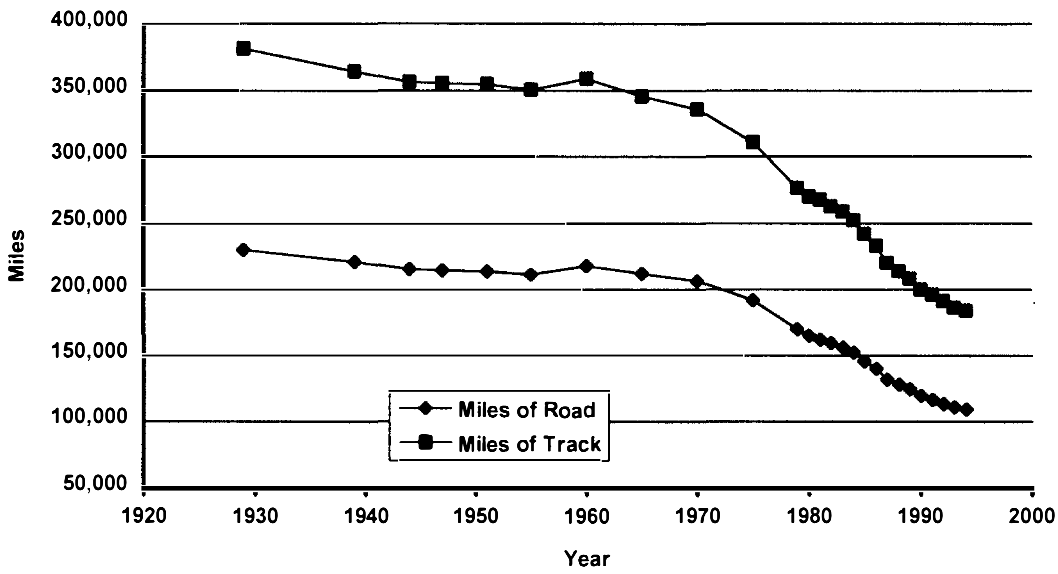

Figure 1-2. Class I Railroad Network Mileage, 1929-1994.

Source: Volume data from Railroad Facts, 1995 Edition, Association of American Railroads, Washington, DC.

wholesale reductions in the network merits careful study. Rail transportation has desirable environmental, energy, and economic characteristics. Given current environmental restrictions and citizen activism, relaying abandoned rail lines or constructing major new intercity railroad lines will be a difficult and expensive task.

Downgrading and elimination of facilities

In response to cost pressures, railroads have also downgraded or eliminated superfluous facilities along the remaining lines. Actions such as lowering track standards, reducing maintenance, tearing out second tracks and sidings, and removing signal systems occurred on a wholesale basis during the decades when the industry struggled financially. Under regulation, these actions were not as difficult for the carriers to accomplish as outright abandonment. Much of this infrastructure was thought unnecessary once passengers and high value merchandise freight left the railroads.

Although necessary for economic survival, infrastructure reductions decrease line capacity. To a degree, technological advances such as centralized traffic control allow single tracks to efficiently handle traffic volumes formerly handled on double track, especially given modern tendencies to run fewer, but longer and heavier trains. A single track cannot, however, maintain the same level of service as double track.

The remaining network, while possibly in its all-time peak physical condition, has been re-engineered to handle the railroads' characteristic bulk and low value cargoes, which are resistant to truck diversion. A notable characteristic of such cargoes is that they generally do not have demanding service standards. Railroads can therefore cut

track maintenance and lower operating speeds, run long but infrequent trains, and tolerate terminal line haul delays without suffering traffic losses.

Inadequate intermodal terminals

Intermodal terminal capacity limits may constrict near term intermodal traffic growth. Many railroad intermodal terminals were constructed during an initial intermodal boom in the late 1950s and early 1960s. In these early years, railroads based intermodal site selection largely on the availability of a surplus parcel of land. Access to the railroad track was considered far more important than access to the highway. Expenditures were kept modest since it was not known whether intermodalism would turn out to be a passing fad. Today, these terminals are often poorly located in relation to the intercity highway system, lack room for growth, and have configurations which are inefficient for access and internal circulation.

In an effort to reduce costs by concentrating volume, the railroad industry has closed many low volume intermodal terminals in favor of centralized facilities serving areas several hundred miles in diameter. This follows the principle used by the airlines in developing "hub and spoke" systems. In the intermodal system, railroads provide the service between hub terminals, with motor carriers handling pickup and delivery to the hubs. As the airlines have found, however, hub and spoke systems are prone to service reliability problems. Because each hub handles so many through connections, the effects of a terminal service failure will often disrupt traffic throughout the network.

The connection between terminals of different railroad companies is another potentially weak element of the intermodal system. Intermodal shipments involving several rail carriers must be interchanged, a process may exchange the entire loaded transport vehicle or just the container. In a large city, such as Chicago, direct rail interchange may take several days despite involving a relatively short distance. Many railroads therefore exchange intermodal containers over the highway, a practice which also allows them to retain scarce intermodal railcars. While faster than a direct rail interchange, this practice increases truck traffic within the intermodal terminal and on the urban street system.

The Capacity Debate

Although the traffic boom is presently having a positive effect on the railroad industry, as reflected in its record revenues and profits and strong stock prices, there is a potential cloud on the horizon. To retain and grow service sensitive traffic, railroads must offer a service level roughly comparable to that of competing motor carriers. This implies, among other things, achieving and consistently maintaining truck-like transit times. Reliability is also a key measure, since truckload motor carriers have an average on-time delivery rate of 95 to 97 percent.

Intermodal trains share many of the same service requirements as passenger trains, including relatively high operating speeds and consistent adherence to an operating schedule. To meet truck service levels, railroads need to operate short, frequent trains on a network with sufficient track capacity to minimize delays. To handle large volumes of

intermodal business, the railroad industry will certainly need to replace elements of the passenger infrastructure, such as passing track, signal systems, and additional main track. With capital in short supply, this will take time.

A key question raised in the formulation of this dissertation is the definition of "large volumes" of intermodal traffic. By mid-1994, many intermodal shippers were experiencing railroad service failures which seemed the result of capacity limitations. Service reliability in most intermodal lanes was not approaching that of motor carriers, and transit times were generally longer. With service in disarray, major railroads such as Conrail and Burlington Northern reduced or eliminated intermodal service in a number of corridors. The actual reasons for these retrenchments have become the subject of some debate within the freight transportation community. Industry analyst Hoffman (1995) flatly states the widely held shipper view, however, that terminal and line capacity constraints are to blame for poor intermodal performance.

The rail industry, on the other hand, seems to believe that its existing physical plant can accommodate any foreseeable increase in intermodal business. Interestingly, railroads frequently do cite capacity restrictions during negotiations with public agencies over additional passenger service. The industry has also admitted that increases in coal production, notably in the Powder River basin of Wyoming, are straining track capacity in certain corridors. The Union Pacific and Burlington Northern railroads, which serve Powder River Basin mines, presently have sizable programs underway to add additional track and so decrease coal train cycle times.

Despite the rhetoric, several facts are clear. First, as was shown previously, rail route and track-mileage is declining. Second, despite their overall decline in freight market share, by almost any measure U.S. railroads produce more transportation today than at any other time in history. Traffic densities on the network are therefore certain to be high, since the size of the network is at an all time low. Figure 1-3 shows a plot of daily freight train-miles/route mile for various years; current densities are at a peak, and the trend is sharply upward. Accepted railroad economic principles favor high traffic densities, so from an industry standpoint this is desirable. The implication, however, is that track capacity margins must be reduced. Options for absorbing additional traffic in the downsized network may be limited, especially in the short term.

The debate over railroad capacity has caught many transportation policymakers and analysts by surprise, since the conventional wisdom has been that railroads are a declining industry with excess infrastructure. At the same time, traditional sources of railroad expertise in the public sector are endangered. At this writing, it appears that Congress will sunset the Interstate Commerce Commission, the federal entity traditionally having regulatory authority over railroad service and infrastructure. The Commission also serves as an important reservoir of railroad policy expertise, which may or may not be retained within other agencies such as the U.S. Department of Transportation (USDOT). The Department is itself a candidate for downsizing, and rail analysis capabilities are especially vulnerable due to the essentially private sector

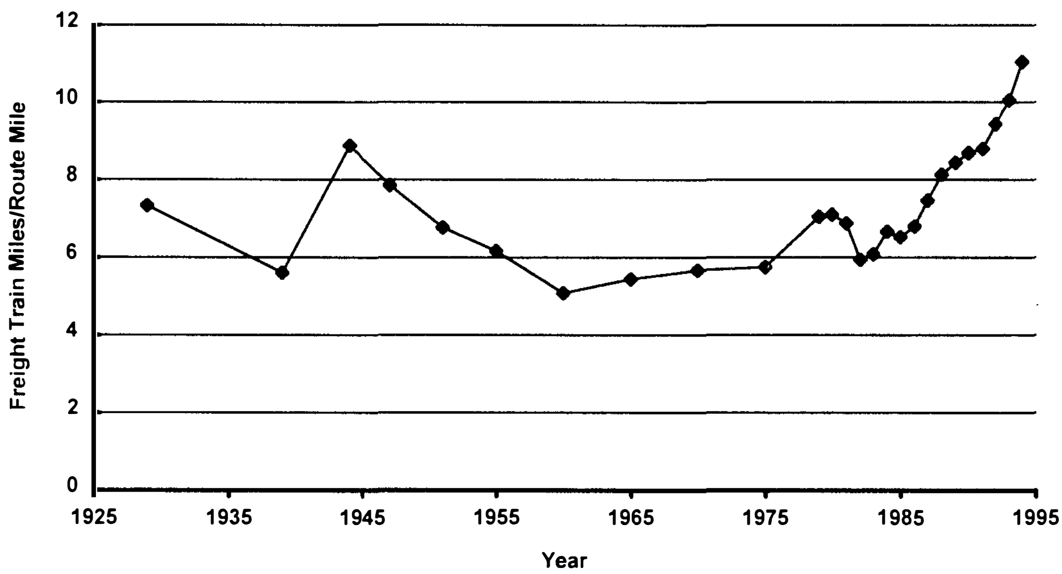

Figure 1-3. Daily Freight Train Miles per Route Mile, 1929-1994.

Source: Data from Railroad Facts, 1995 Edition, Association of American Railroads, Washington, DC.

ownership of railroad assets. The disappearance of dedicated railroad policy analysis expertise highlights the need for the network model developed for this dissertation.

Research Objectives and Scope

The overall goal of this research was to characterize the ability of the railroad system to handle additional intermodal traffic. The project consisted of two phases: model formulation and case study. These are described below.

Phase 1—Model Formulation

The initial phase of the research consisted of the formulation of a model for analyzing railroad system traffic flows and assessing the congestion effects of these flows. The research proposal envisioned adopting a network equilibrium assignment approach, similar to that used in the traditional four-step urban transportation planning process, for rail freight flows. Given equilibrium flow volumes, the link and terminal travel times at these flows can easily be determined. Travel times above some selected threshold would represent congestion.

The model formulation incorporates several key design assumptions. First, it is intended to be applied at a strategic level. The model predicts the average overall performance of the system over time without regard to temporary effects caused by peaking, accidents, etc. Second, freight flow volumes between origin-destination pairs are assigned externally. The model will not distribute the freight between arbitrary sources of supply and demand. It should, however, be capable of selecting the proportion

of freight leaving an origin by each serving carrier. Finally, the model will take a short term horizon regarding costs. In other words, the network is not assumed to have capital improvements, nor will costs change during the time horizon.

Issues to be addressed in formulating the model include the multicommodity nature of freight flows, development of a suitable objective function, development of congestion functions for rail network elements, and identification of pertinent railroad costs upon which to base the objective function. These findings are discussed in subsequent chapters.

Phase 2—Case Study

The second portion of the research focused on the ability of the railroad system to handle intermodal traffic, with the goal of identifying current and potential capacity limitations. This was accomplished using the analytical framework developed in Phase 1. Because of time and resource limitations, the research scope was restricted to the railroad system in the southeastern U.S. The study area was bordered by the Mississippi River on the west, and by the Ohio River and Mason-Dixon line to the north. These boundaries form effective cordon lines which coincide with natural divisions in the national railroad network. Selected rail lines external to this region were included to bring traffic to the appropriate cordon crossings. The resulting network is large enough to provide a meaningful test of the analytical approach.

Organization of the Dissertation

The remainder of this dissertation is divided into five chapters. Chapter 2 discusses railroad capacity, both at the facility level and at the network level. Using the premise that networks are composed of facilities, the chapter presents the underlying concepts of railroad facility capacity. Classical models for evaluating capacities of railroad lines and terminals are critically reviewed. The chapter then surveys various techniques for modeling flows in congested networks and discusses some previous applications of these techniques in rail freight transportation.

Chapter 3 presents a theoretical formulation for the multicommodity, multicarrier railroad network traffic assignment model. The network modeling structure is presented, and the traffic flow problem formulated as a mathematical program. Necessary and sufficient conditions for the solution of this problem are stated.

In Chapter 4, the solution algorithm for the model is presented and discussed. The algorithm uses a linear approximation approach decomposed by commodity. The computer codes developed to implement the solution are also described.

A case study demonstrating the use of the model is presented in Chapter 5. This study examines aspects of traffic flows on the mainline rail system in the Southeastern U.S. The results of the analysis are presented and discussed. The performance of the model is also examined.

The final chapter summarizes the research findings and presents some overall conclusions. Suggestions for further refinement of the model are then provided.

CHAPTER 2

REVIEW OF LITERATURE

As background to the development of the system level capacity model, this chapter presents a state of the art review of key foundations of the modeling effort. The chapter is divided into three sections. The first section discusses basic capacity concepts applicable to railroad systems. Section two then discusses congestion functions for railroad lines and terminals. The final section describes modeling approaches to network flow evaluation in capacitated networks. The topics addressed in each section will be used in formulating the assignment model.

Railroad Capacity Concepts

Capacity is, in general, a measure of the ability of a transportation facility or network to handle traffic. Methods for evaluating overall traffic performance under various facility design and traffic flow conditions are essential for the economical and efficient operation of transportation systems. As a discipline, capacity evaluation is extremely well developed in highway transportation. Rail capacity, by contrast, has received relatively little attention, although elements of highway capacity theory may be extended to railroads.

Definitions

Capacity is the maximum number of traffic units which can pass over or through a facility during a given time period under prevailing facility and traffic conditions. The maximum possible traffic flow on a facility is termed the *ultimate capacity*. Capacity analysis examines the relationship between traffic volume and vehicle performance (speed, travel time, emissions, etc.) on a facility. Congestion or capacity functions describe the relationship between total flow and vehicle performance.

In the railroad industry, a facility may be either a line haul track segment or a terminal. The traffic unit for a line haul track segment is usually the train, which is a set of vehicles operating as a unit. Terminal performance is more typically measured in terms of vehicle throughput, since the function of the terminal is to process single vehicles or vehicles in groups much smaller than train size. Given a measure of the mean number of vehicles in a train, line haul throughput can, if necessary, be expressed in terms of equivalent vehicles.

Railroad capacity is traditionally defined as the traffic volume above which the performance of a facility becomes unacceptable. The railroad definition of capacity is, therefore, somewhat analogous to the "practical" capacity definition formerly employed in highway engineering. The facility is capable of higher throughput, but traffic performance measures are not tolerable at these volumes. From this point in the dissertation, the term capacity, unless otherwise qualified, will employ the acceptable performance definition.

The rail industry normally uses a 24-hour day as the base time unit for capacity evaluation, although terminal capacity is sometimes expressed in terms of an 8-hour shift to allow correlations with staffing levels. For consistency, this dissertation uses the day as the time unit for terminals, also. Capacity is expressed, therefore, as trains per day (TPD) for a track segment or cars per day (CPD) for a terminal.

Capacity is normally measured as the total traffic in both directions on a rail line. This differs from highway practice, where capacity on certain facility types may be specified by direction. A railroad track is somewhat analogous to a highway lane. Unlike most highway lanes, however, single track railroads almost always handle bi-directional flows, so it is logical that railroad capacity reflects the total traffic flow. Where a railroad line has multiple main tracks, different tracks may be assigned to each traffic direction as in a highway. More often, however, the railroad company will, with signaling or operational controls, strive for bi-directional operation on each track. Such an arrangement increases operational flexibility and, in turn, capacity.

Performance Measures

Performance in railroad capacity evaluation is generally measured in travel time for line segments and car processing time for terminals. Delay, the difference between travel (processing) time actually experienced and the travel (processing) time under ideal conditions, is also commonly used.

Line Segments

The travel time for trains over a given line segment is a function of fixed conditions (line geometric characteristics, signal and control system characteristics, speed restrictions, train weight and power, etc.) and operational conditions (interference from opposing rail traffic, waits for rail traffic to clear at-grade crossings, dispatching delays, breakdowns, etc.).

Fixed Conditions. The best possible travel time which a train can achieve over a line segment occurs when only fixed conditions affect the time. The contribution of fixed conditions to travel time is quantitatively predictable using basic kinematic relationships, and, neglecting equipment reliability, is essentially deterministic. Such a travel time, in which the train is assumed to remain continuously in motion (unless forced to stop by normal operating practice), is called the *free running time*. The equivalent speed, called the *free running speed*, is determined by dividing the segment length by the free running time. *Track speed* is the maximum operating speed allowed within a subsection of a track segment, with the *average track speed* being the weighted average of the track speeds in the overall segment. Because of the time required to accelerate and decelerate, the free running speed cannot equal the average track speed. Free running speed only approximates average track speed when the train has adequate power.

Operational Conditions. Operational conditions impart a travel time component which varies as a function of traffic conditions. This component is probabilistic, since each train will encounter a different and random set of events which affect its travel time.

As traffic levels increase, the operational component of travel time increases. Models for this relationship will be presented in a subsequent section. The train speed computed by dividing segment length by the overall travel time, including operational effects, is called the *overall travel time.*

The effects of conflicting traffic depend upon the track configuration. On single track lines, trains traveling in the opposite directions must take turns using the track between passing points. One train must, therefore, wait for opposing traffic to clear. This waiting period, called *meet delay*, is not running time, since the train is not in motion. Obviously, multiple track line segments reduce or eliminate meet delay, since trains may pass on adjacent tracks without stopping.

Trains may also be affected by traffic in the same direction. Railroad operations require minimum headways for safety purposes. A train's speed must be reduced when it encroaches upon the headway of a preceding train. Signal systems increase line capacity by allowing the headways between trains to be reduced. Still, fast trains can be delayed behind slow trains moving in the same direction. Double track does not necessarily reduce such delays unless dispatcher controlled crossovers or sidings are provided to facilitate passing operations.

Terminals

Terminal processing time is a function of numerous factors, including terminal configuration, method of classification (flat switching, gravity, etc.), train arrival and departure rate, and the number of switch engines employed. Because terminal

configuration is extremely site specific, general relationships are difficult to predict. Like line segments, terminal processing time will have a fixed component and an operational component. Therefore, the concepts of an ideal free flow processing time and an average processing time reflecting traffic congestion effects are still valid.

Characteristics of the Performance Statistic

The flow of vehicles has random characteristics, so that performance measures for individual vehicles using a facility, given similar overall conditions, will differ. Operational conditions subject individual vehicles to a random number of delays. The number and duration of delays can each be hypothesized to be independent random variables following some statistical distribution. Any measure of vehicle performance at a given traffic flow level is, therefore, a stochastic value reflecting the expected value of the sum of a random number of random variables.

Service reliability is another important consideration in evaluating railroad capacity. Reliability reflects the measure of variance associated with the facility performance distribution. In many cases, a shipper will accept the generally higher transit time associated with rail provided that service is consistent. As traffic volumes increase, so do opportunities for incidents which will disrupt traffic flows and cause service failures. These considerations imply several things. First, a railroad must consider variability in travel time in addition to average vehicle performance when establishing capacity thresholds. Second, the variance of the performance distribution may differ at different discrete levels of flow.

Capacity Models for Railroad Facilities

In comparison to air and highway capacity, railroad capacity has received relatively little attention. In the design literature, neither Hay, (1982) in his text on railway engineering, nor the American Railway Engineering Association, in their Manual for Railway Engineering (1993) mention the subject. Railroad capacity did enjoy some research interest during the 1970s and 1980s when changes in railroad network structure and demand patterns raised concerns about capacity shortfalls. Much of the available literature dates from this period.

Congestion Models for Railway Track

The previous section introduced a number of definitions and concepts pertaining to railway facility capacity. In general, train performance measured in overall travel time on a railroad line degrades with increasing throughput. At very low traffic levels, fixed conditions predominate and travel time is very near free running time. The average travel time per train then increases non-linearly as the effects of operational conditions related to traffic flow become significant. At ultimate capacity, travel time increases asymptotically to infinity and flow ceases. Figure 2-1 shows the general form of this relationship, which applies to many types of transportation facilities.

Modeling Approaches

The literature contains a number of approaches for predicting railroad link travel time or delay as a function of volume. These fall broadly into two categories: empirical

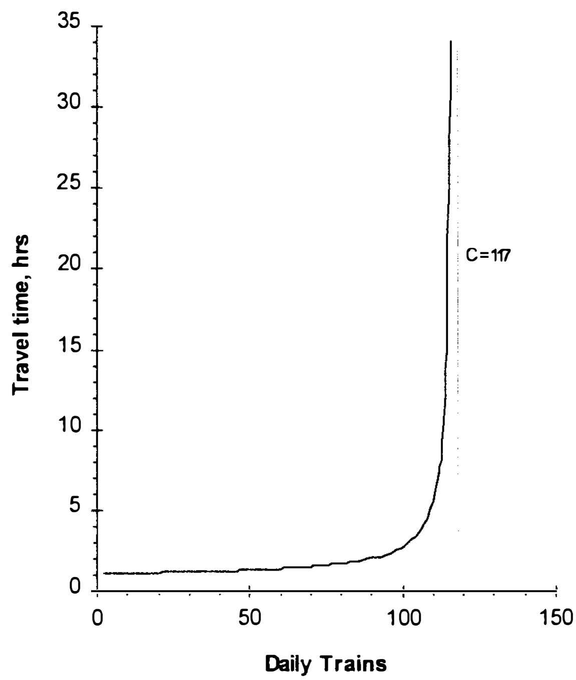

Figure 2-1. Travel Time—Volume Relationship.

and analytical. Realistic rail link travel time functions account for both the fixed and operational conditions previously discussed.

Empirical Functions. Empirical capacity functions predict the travel time/volume relationship for a particular railroad line based upon a fit to observed data. Empirical functions typically contain relatively few explanatory variables. The models must be calibrated to represent a given set of fixed and operational conditions. Applying a calibrated model to a line having different conditions is not appropriate.

Empirical function parameters are derived using observed data. Due to the difficulty of obtaining field observations, especially over a range of volumes, simulation is the preferred method for producing data points for model calibration. Simulation models are relatively inexpensive to run and can replicate conditions not easily observable in the field. Prokopy and Rubin (1975), Bronzini and Miller (1978), Bronzini and Sherman (1986), and Clarke (1982) describe the use of simulation for generating points on the delay function for specific railroad line configurations.

While there are a number of mathematical functions which approximate the theoretical travel time/delay curve, the most widely used are the hyperbolic function and the polynomial function.

Hyperbolic Travel Time Function. The hyperbolic travel time function is one function form which closely replicates the curve of Figure 2-1. The general function, presented by Mosher (1963), is expressed mathematically as follows:

$$T = T_A - \frac{C(T_0 - T_A)}{V - C}; \; T_A \leq T_0, \; V < C \qquad (2.1a)$$

where: T = travel time;

T_A = horizontal travel time asymptote;

T_0 = travel time at zero flow;

C = maximum capacity;

V = volume.

This expression is often reformulated in terms of delay, $D = T - T_0$. This yields:

$$D = \frac{V \, D_{50}}{C - V} \qquad (2.1b)$$

where: D_{50} = delay at 50 percent of capacity.

Using queuing theory, the equivalent function can be derived for an M/M/1 server. This provides an attractive theoretical foundation to the formula.

Since it contains essentially no terms describing the characteristics of the railroad line or its operations, the hyperbolic function must be calibrated for a particular configuration. In practice, families of curves are developed to account for ranges of conditions. This is not difficult, since the curve can be defined accurately using a few points.

One disadvantage of using the hyperbolic equation in practice is the behavior of the function at volumes at and above capacity. The function is undefined at $V = C$, and provides negative results when $V > C$. The common workaround for this problem

behavior is to linearize the function at $V \geq 0.90C$. In applications, this adds the overhead of testing V each time the travel time is calculated.

Polynomial Delay Function. Another common approach to delay or travel time prediction is the polynomial function, which has the following form:

$$T = T_0\left[1 + k_1 V + k_2\left(\frac{V}{C}\right)^{\gamma}\right] \tag{2.2}$$

where:

T	=	travel time at V;
T_0	=	travel time at zero flow;
C	=	maximum capacity;
V	=	volume;
k_1, k_2, γ	=	empirical constants.

The well-known BPR function for predicting highway link travel times at specified flow rates follows the polynomial form. In the BPR model, the K_1V term is omitted, presumably since its contribution to the total travel time is very small.

The polynomial function is intuitively attractive for several reasons. First, it explicitly considers the volume/capacity ratio, a flow parameter well known to transportation analysts. Second, the function has a consistent mathematical behavior. Travel time monotonically increases with volume, and the model has no abhorrent behavior at $V \geq C$.

Calibration of the polynomial function requires the determination of the empirical constants so that predicted travel times or delays closely approximate observed values.

Unlike the hyperbolic function, the shape of the polynomial function is greatly affected by the values of the calibration constants. The constants have no intuitive values and are strictly used to shape the curve. A trial and error process is used to select values for the constants which best fit the polynomial curve to an observed set of data. In general, this calibration process requires more observed values than does calibration of the hyperbolic function. Furthermore, polynomial function calibration requires observations at high volume/capacity ratios to accurately fit the asymptotic portion of the curve. These may be difficult to obtain in the field.

Analytical Models. Several researchers have proposed analytical models for predicting railroad link travel time as a function of volume, siding spacing, train speeds, and other pertinent characteristics. The attractiveness of the analytical approach is that, by explicitly considering factors affecting travel time, the models can be generically applied without the need for extensive and possibly expensive recalibration. In reality, no single model has yet been developed which can be applied to all possible line configurations, so families of models are developed (e.g. single versus double track). The models presented herein represent typical approaches to the evaluation of capacity and travel time using analytic formulations.

Poole Model. An early analytical model is described by Poole (1962) in his treatise on railroad cost evaluation. Poole states that line capacity is a function of the running time between sidings, the number and capacity of sidings, the time required to run through switches and siding tracks, the type of train control system, and the regularity

of train arrivals into the system. Poole's model considers a track section between passing locations which hold a single train. Extrapolation of the results to a given line requires the assumption that passing tracks are evenly spaced and that trains arrive uniformly. This being stated, the capacity model is as follows:

$$C = \frac{1440}{t + \frac{t}{2} + \frac{m}{2}} \tag{2.3a}$$

where: C = capacity, trains per day;

1440 = minutes in one day;

t = minutes to clear mainline between siding switches at full speed;

m = delay for each meet, excluding t.

The term $(\frac{t}{2} + \frac{m}{2})$ in equation (2.3a) represents the average delays per meet for the two trains involved. The model assumes that an equal number of trains in each direction are detained to enter the sidings, a typical result when trains have superiority by direction. To represent a control system such as Centralized Traffic Control (CTC) which permits the most advanced train to continue, Poole recommends that the $\frac{t}{2}$ term in the denominator of (3) be changed to $\frac{t}{4}$.

Poole also provides procedures to compute average travel time as a function of volume. For the case where all trains are of the same type, and assuming that the district

being evaluated takes longer than one day for a single train to traverse, the number of expected daily meets for a single train is equal to V, the daily number of trains entering the territory. At the time the single train enters the district, there will be $\frac{V}{2}$ opposing trains on the line. Another $\frac{V}{2}$ opposing trains will enter the district during the single train's journey. The travel time for a given volume, V, will then be:

$$T = T_0\left(\frac{24}{24 - V\left(\frac{t}{2} + \frac{m}{2}\right)}\right). \tag{2.3b}$$

A plot of this function for various volume levels follows the expected form of the travel time function. Poole offers function versions which consider factors such as mixed train classes and double track operations.

A limitation of Poole's approach is that meets and overtakes involve only two trains and that any potential interaction between multiple trains takes place two trains at a time. Thus, the model appears best suited to low train volumes. Despite its limitations, the model has seen application. Poole himself used it for numerous studies. Kresge and Roberts (1971) use a very similar function in their Colombian study, while Janic (1984) reports a virtually identical function used by the International Union of Railways for single track line capacity modeling.

Petersen Model for Single Track Lines. Petersen (1974) proposed an analytical model for estimating average travel times over single track railway lines. His model

considers delay because of train priority systems, meets, and overtakes. Train classes may have different operating characteristics. The model assumes that trains within a given class arrive uniformly during the study period. Sidings are assumed to be equally spaced and long enough to accommodate meets and overtakes. Petersen's model has similarities to Poole's in that it attempts to estimate the number of meets for a given traffic volume and the expected delay for a meet.

Given a set of free flow speeds, s, for I inbound trains and J outbound trains, create an index set $k = (-I, -I+1, \ldots -1, 1, 2, \ldots, J-1, J)$ such that for $i \in k$, $i < 0$ is an inbound train and $j > 0$ is an outbound train. The average transit time, T_i, for a train of class i, given line length d, is:

$$T_i = T_{0,i} + \sum_{j \in k} D_{ij} M_{ij} \qquad (2.4a)$$

where: $T_{0,i}$ = free flow travel time for class i;

= free flow travel time for class i, $= \begin{cases} d/s_i & i > 0 \\ -d/s_i & i < 0 \end{cases}$;

D_{ij} = constant delay incurred by train i when meeting a train of class j;

M_{ij} = the number of encounters (meets and overtakes) by a train of class i with trains of class j during a run.

Petersen considers a number of different cases for train interferences and derives an equation based upon the expected number of interferences for a given train class:

$$\frac{d}{v_i} = T_{0,i} + \sum_{j \in k} E_{ij} \, N_j \left(\frac{d}{v_i} - \frac{d}{v_j} \right) \tag{2.4b}$$

where: v_i = average speed for train class i;

$\frac{d}{v_i}$ = average travel time for train class i;

N_j = arrival rate for class i, trains per unit time;

$$E_{ij} = \begin{cases} -D_{ij} & j < i < 0,\ 0 < i < j \\ D_{ij} & \text{otherwise.} \end{cases}$$

Equation (2.4b) actually defines a set of $I + J$ simultaneous linear equations which can be solved for the $I + J$ variables $\frac{d}{v_i}$, the expected travel times for the train classes. Petersen also derives an equation for determining the expected delay, D_{ij}::

$$D_{ij} = S_i + \frac{p_{ij}^2}{2(l+1)} \left| \frac{d}{s_i} - \frac{d}{s_j} \right| \tag{2.4c}$$

where: S_i = delay to enter siding for train class i;

l = number of uniformly spaced sidings;

p_{ij} = probability that train of class i waits for a train of class j.

This formulation assumes that train arrivals for each class are independently and uniformly distributed during the time period.

Like Poole's model, Petersen's approach has the limitation that meets and overtakes involve only two trains and that any potential interaction between multiple trains takes place two trains at a time. Thus, the model also appears suited to low train volumes. In addition, the model considers only single track lines. Modifications of the

basic model form to remove these limitations have been presented by Daughety and Turnquist (1979), but the concept remains the same.

Janic Model. The approach employed by Janic (1988) to modeling travel time on single track railway lines relies on queuing theory. Janic's model assumes the following:

a) The single track line contains a critical bottleneck segment where most delay is incurred. This bottleneck section represents a server.

b) Trains are considered as customers, which may be members of various priority classes;

c) Arrivals of train categories are independent, Poisson processes. Service times of train classes are random variates with known distributions.

d) If a lower class train is being served, and a higher class train arrives, service of the lower class train is not interrupted.

The model produces the average delays for each train class according to the following system of linear equations:

$$\overline{W}_p = \frac{\overline{W}_0 + \sum_{i=p+1}^{P} \rho_i \overline{W}_i}{\left[1 - \sum_{i=p}^{P} \rho_i\right]}, \quad p = 1,2,\ldots P \tag{2.5a}$$

where: $\overline{W}_i$ = average delay of train class i, i = 1, 2,..., p, ..., P;

$\overline{W}_0$ = the expected value of the time spent by a newly arrived train waiting for another train to be served in the bottleneck segment;

ρ_i = $\lambda_i \bar{t}_i$;

λ_i = the arrival rate of trains of class i;

$\bar{t}_i$ = the mean service time for trains of class i.

Janic presents a methodology for estimating the values of $\bar{t}_i$. The value of $\overline{W}_0$ can be then be analytically determined as:

$$\overline{W}_o = \sum_{i=1}^{P} \lambda_i \frac{\bar{t}_i^2}{2}. \tag{2.5b}$$

The model is only valid for a stationary process. In other words, the time period of the analysis must be of sufficient length that the system is assumed to be in a non-transient state.

Greenberg Model. Another queuing based analytical approach for single track line delay is proposed by Greenberg, Leachman, and Wolff (1988). Their paper examines the case of slow speed operation on lines with widely spaced passing locations. Unlike the models of Petersen and Poole, the Greenberg model does not require equally spaced siding, uniform traffic patterns, or meets involving only two trains.

The busy period of the line in a given direction begins when a train arrives and finds the track idle or when the busy period in the opposite direction ends and at least one train is waiting to use the track. The model treats the busy period as an $M/D/\infty$ queue with directional Poisson arrival rate λ and service time T, where T is the running time for a train on the segment. Given the above definitions, Greenberg develops the following formulation for the expected delay D by direction on the line segment:

$$E[D_S] = P_N\left[\frac{1}{\lambda_N}e^{\lambda_S T} - \frac{T e^{\lambda_S T}}{e^{\lambda_S T}-1} + l\right] \quad (2.6a)$$

where: P_N = proportion of time that the track is occupied by northbound trains;

λ_N = Poisson arrival rate, northbound trains;

T = service time;

l = running time losses associated with taking and leaving siding.

In this case, the delay is for the southbound direction. Exchanging the S and N subscripts on the variables gives the equivalent formula for northbound delay. Assuming $\lambda_N = \lambda_S$, then:

$$P_N = P_S = \frac{1}{2}\frac{(e^{\lambda T}-1)(e^{2\lambda T}+1)}{1+(e^{\lambda T}-1)(e^{2\lambda T}+1)}. \quad (2.6b)$$

Congestion Models for Railway Terminals

Railway freight terminals consist of classification yards and loading/unloading facilities, such as intermodal terminals and freight transfer facilities. The relationship between volume and processing time for terminals is similar in form to that of line segments. Volume for freight terminals is typically measured in cars processed per unit of time, although tons loaded/unloaded could be used in the case of transloading terminals.

Factors influencing terminal capacity include facility configuration, operating mode (flat switched or gravity switched), traffic characteristics, storage capacity, and number of switch engines. For transfer terminals, the rate at which railcars are loaded and unloaded and the landside configuration are important additional determinants. At intermodal terminals, the highway side may influence throughput more than rail side. Truck traffic at rail-highway terminals naturally tends to peak around the cut-off times for pickup and delivery of trailers. In addition, many intermodal terminals have specific periods during which trucks can access the facility.

Like track segment models, railroad terminal performance models can take either an empirical or an analytical approach. Empirical models generally seem to be more practical for planning models.

Empirical Models. Empirical models for rail terminal capacity take virtually the same form as those discussed for rail lines. In fact, with appropriate substitution of parameter values and change of units, the line models can be applied to yards without change. Line segments are normally combined into groups having similar characteristics and the models calibrated to each group. Terminals, on the other hand, are highly individualized, so it will usually be necessary to calibrate the delay function for each terminal being modeled.

The difficulty with terminal model calibration lies in obtaining sufficient observations to perform the calibration. With track, this is not a great problem since the calibration process can usually draw observations from a population of similar lines. In

calibrating a model for an individual terminal, all data must come from that terminal. Unless the facility sees wide variations in traffic, the number of observations available for calibration is limited. One option is to use the hyperbolic model, which can be calibrated with a small number of observations. Simulation has also proven to be a useful workaround for generating data points, but yard simulation models must generally be custom developed for each application.

Analytical Models. The individuality of railroad terminals makes the development of generalized analytical models difficult. Nevertheless, models for various typical yard configurations have been proposed by several researchers. Queuing forms the basis of these approaches.

Petersen Model. Petersen (1977a,b) describes a queuing approach to rail yard modeling. In the first paper, Petersen defines some common yard configurations and describes the following yard operations: receiving and inbound inspection, classification or sorting, wait for connections, train marshaling and assembly, and outbound inspection and departure. He assumes that receiving and departure take constant time, but that the others are bottleneck operations which can be modeled using queuing theory. Petersen's suggested models include the M/M/S, M/D/S, and M/G/1 queues, but he does not give explicit formulations for them. He concludes by demonstrating that the models satisfactorily predict yard throughput times given existing conditions. In the second paper, Petersen then determines analytically how changes in various yard components

and operating practices change the processing times. It should be noted that in Petersen's work, the train is the measured processing unit.

Turnquist and Daskin Model. Turnquist and Daskin (1982) also employ a queuing approach to model classification yards, with their work representing an extension of Petersen's. Their paper presents queuing models for the classification and connection wait operations. The classification operation is based upon the $M^X/G/1$ batch arrival model, for which the average wait time for classification is shown to be:

$$E[T_1] = \left(\frac{1}{2\mu}\right)\left[\left(\frac{1}{1-\rho}\right)\left(\left(\frac{L_2}{L_1}\right) + \rho\mu^2\sigma^2\right) - 1\right] \tag{2.7a}$$

where: L_1 = mean train length, cars;

L_2 = second moment about the origin of train length;

λ = mean arrival rate of trains;

μ = mean classification service rate (cars/unit time);

ρ = $\lambda L_1/\mu$ = traffic intensity;

σ^2 = variance of service time distribution.

The connection delay time is determined using a simple batch server queue. Assume H is a random variable representing the time interval between successive outbound trains for a given block of cars, with $g(h)$ the probability density function for H. The expected connection delay is then:

$$E[T_2] = E(H)/2 + V(H)/2E(H). \tag{2.7b}$$

The total yard delay is the sum of equations (2.7a) and (2.7b).

Freight Traffic Assignment Models

The delay definitions presented in the previous section apply to individual facilities. Railroad capacity must be considered, however, in the context of a system of interrelated components. The total capacity of the system is determined by the interaction of these components. Table 2-1 summarizes some of the influences of various elements on capacity. Defining the relationships between these elements is one of the complicating factors in analyzing capacity.

One approach to general capacity evaluation is to model the railroad system in a network format, with places and junctions as nodes and terminals and connecting lines as links. Both track segments and terminals would have individual capacity functions which predict delay as a function of traffic volume. Given known flows of traffic within the system, an assignment approach can be employed to determine volumes on specific facilities. Transit times for various movements can then be determined by summing the delays incurred at terminals and during the line-haul. Practical capacity will be used to evaluate system performance.

The assignment problem can be formulated in several ways, depending upon the user's desire to replicate system behavior. Regardless of the technique, the basic idea is to place traffic between an origin-destination pair on a likely path or paths connecting the two points. The cumulative system flow pattern consists of the superimposed flows between all origin-destination pairs.

Table 2-1. Factors Influencing Railroad System Capacity

Factor	Influence
Track (line haul) characteristics	
Horizontal and vertical alignment	Train speeds and acceleration/deceleration characteristics
Number of tracks	Provides additional theoretical capacity; reduces traffic conflicts
Spacing of sidings/crossovers	Meet/pass delays
Junction configuration	Train operating speeds; traffic conflicts
Type of control system	Train headway and operating speed; flow stability
Siding/crossover geometry	Operating speeds; meet/pass delay
Equipment	
Supply	Constrains system throughput
Characteristics	Train productivity; operating characteristics
Terminal	
Classification procedure	Car throughput
Track length	Amount of switching needed to handle trains
Access/egress routes	Highway and rail vehicle rates to/from intermodal terminal
Gate processing	Highway vehicle entry and exit rate at intermodal terminal
Storage/handling capacity	Number of vehicles (rail/highway) in terminal
Lift capability	Physical transfer rate of packages: rail-highway (intermodal)
Configuration	Overall efficiency of terminal
Operations	
Operating plans and policies	Ability to achieve physical capacity
Labor rules	Labor requirements; ability to achieve physical capacity
Labor supply	Constrains train operations/unit of time
Traffic Characteristics	
Composition	Train characteristics; service requirements
Temporal demand	Service requirements; intensity of system use; peaking
Directionality	Traffic conflicts
Maintenance Requirements	Availability of system to provide capacity
System Reliability	Availability of system to provide capacity; flow stability

Early assignment problems in both freight and passenger transportation used an "all-or-nothing" approach in which traffic is assigned to paths without regard to congestion effects. The famous Colombian freight network study of Kresge and Roberts (1971), for example, used such an approach. Lansdowne (1981) also describes the same basic approach in his paper on rail traffic assignment. Refinements to this technique, much used in highway transportation, iteratively assign traffic to the network, with link loadings in each iteration being explicitly considered in computing travel time. Chang et al. (1981) used a similar procedure in their study of U.S. coal transportation by rail.

The "all-or-nothing" models have the weakness of being unrealistic in terms of the theory of user behavior in network flows. Wardrop (1952) described two types of flow behavior in networks. Wardrop's user equilibrium (UE) is a flow condition in which no individual user can unilaterally change paths and reduce his travel impedance. All users are assumed to have the same perception of impedance and to have perfect knowledge of the system. The UE problem is expressed by the following mathematical program per Sheffi (1985):

$$\min z = \sum_{a} \int_{0}^{v_a} T_a(v)dv \qquad (2.8a)$$

subject to

$$\sum_{k} f_{k}^{rs} = q_{rs} \quad \forall r,s; \qquad (2.8b)$$

$$f_{k}^{rs} \geq 0 \quad \forall k,r,s; \qquad (2.8c)$$

$$v_a = \sum_r \sum_s \sum_k f_k^{rs} \delta_{a,k}^{rs} \quad \forall a ; \qquad (2.8d)$$

Where: v_a = flow on link a;

T_a = travel cost on link a;

f_k^{rs} = flow on path k connecting O-D pair r-s;

c_k^{rs} = total cost of path k connecting O-D pair r-s;

q_{rs} = trip rate between origin r and destination s;

$\delta_{a,k}^{rs}$ = indicator variable = 1 if link a is on path k between O-D pair r-s; 0 otherwise.

Eash, Jansen, and Boyce (1981) describe an algorithm for iteratively solving this problem. Wardropian UE is frequently used in passenger traffic assignment algorithms, but is less applicable to freight networks since central authorities normally govern traffic flow in such systems. One place where UE might be applicable is in modeling the decision process by which individual shippers choose service providers.

A generalization of UE occurs when the perceived impedance is distinguished from the actual impedance. In this condition, called stochastic user equilibrium (SUE), the perceived impedance is treated as a random variable distributed across the population of users. This reflects reality rather more accurately than the assumptions of UE. Sheffi (1985) provides a thorough theoretical description of the approach. Loureiro (1994) describes the use of an SUE assignment approach in his multimodal freight network design model.

The system optimal (SO) assignment, also proposed by Wardrop, is the flow pattern which minimizes total impedance in the system. For each origin-destination pair in the network, the marginal cost of any path used does not exceed the marginal cost of any other path. Thus, transfers of flows to other paths cannot reduce total system cost, and the system is therefore in a state of minimum total cost. This is appropriate for situations where a single authority controls the network. Sheffi (1985) also defines the mathematical program to determine SO flows:

$$\min z = \sum_{a} v_a T_a(v_a) \tag{2.9a}$$

subject to

$$\sum_{k} f_k^{rs} = q_{rs} \quad \forall r, s; \tag{2.9b}$$

$$f_k^{rs} \geq 0 \quad \forall k, r, s. \tag{2.9c}$$

Normally, a railroad may be thought to operate in a system optimal fashion.

System equilibrium formulations for freight models have been proposed and or used by a number of researchers. Dafermos (1971) formulated an SE assignment model for examining multiclass flow problems, which include multicommodity freight flow assignments. Friesz and his colleagues (1981) describe the use of a multicommodity freight network equilibrium model which specifically attempts to reconcile the user-optimized (shipper) and system-optimized (carrier) aspects of the freight flow problem. This model performs a combined distribution, mode split, and assignment from the

shipper standpoint. The resulting origin-destination flows and generalized routes are used as inputs to a carrier submodel. This module computes system equilibrium flows for each mode/carrier. This model, while broader in scope than needed for this study, nevertheless contributes many useful ideas. Subsequent works by Harker (1986), Crainic (1990), and Guélat et al. (1990) further explore the theory of SE freight flow assignment.

The work described in the papers discussed above serves as a foundation for the research described in this dissertation. Chapter 3 presents a formulation for the multicommodity SE assignment problem. This model represents a synthesis of the concepts explored in the referenced works.

CHAPTER 3

MODEL FORMULATION

This chapter presents the formulation of the multicommodity freight network equilibrium model used in this dissertation.

Design Criteria and Objectives

Before the formulation of the model is presented, it is appropriate to set out a number of design criteria and to reiterate the design objectives.

The objective of the model is to predict, given a matrix of commodity flow demands between origin and destination pairs, the likely volume of flow on each link in a rail network. The flow patterns should accurately reflect the underlying decision logic used by shippers and railroad managers in routing traffic. Given a flow volume and a service function for each facility, the average travel time, and thus delay, can be calculated for that facility. Facilities having an excessive amount of delay can be targeted for additional study using more detailed modeling approaches such as simulation.

The model is intended to provide a strategic level view of network flows, rather than a tactical or operating viewpoint. To this degree, individual train operations are not replicated, nor are the flows considered in terms of traffic blocks which could be used for operations planning. The statistics provided represent average characteristics of the system. Peaking, traffic disruptions, and other transient phenomena are not addressed.

The time frame of the model is the short term. It is assumed that the network is fixed and that no improvements are made which would affect traffic flows. The analyst may, of course, use the model to test hypothetical improvements. These network changes must be specified exogenously, however.

The model formulation should be capable of reflecting:

- The flow of multiple separate commodity classes, each having a distinct pricing structure;
- The network topology of the modeled transportation system, including line haul arcs, terminals, and transfer points;
- Corporate ownership of network elements;
- Service characteristics of various network elements, such as line haul links and terminals; and
- Restrictions on the movement of commodities over specific carriers or network elements as needed to reflect operational practice.

Supply and Demand

This section describes the characteristics of the transportation supply and demand environment in the model.

Carriers

We assume that the transportation market consists of a set M of transportation providers or carriers $(m \in M)$. In this study, the carriers are railroads, although, in

general, this is not a requirement. The set M may include carriers representing different modes of transportation, although each carrier is assumed to be a single mode.

Carriers are assumed in the model to be cost minimizing entities. In economic terms, the firms are cost efficient. The carriers supply services, singly or in concert, between various origin-destination (O-D) pairs. An origin or destination may be a physical node in the network or an abstract node representing a demand centroid. This choice is left to the analyst. In general, however, because of the strategic planning orientation of the model, demand nodes represent centroids of mass for some shipper community in a region.

Demand

The problem contains a set W of O-D pairs. Some volume of a commodity or commodities flows between each O-D pair w in W. We denote the set of commodities as P, with p denoting an individual commodity. A commodity may represent a product, as in coal or grain, or a specific type of service, such as intermodal transportation. Empty cars returning to the point of loading may also be modeled as a commodity. It is assumed that each commodity has distinct cost characteristics.

The demand for transportation is fixed exogenously. Via measurement or some external procedure such as trip distribution or an input-output type model, the volume of flow for each commodity between each O-D pair is determined and provided as an input to the model. The model does not, therefore, replicate the decision making process of shippers in selecting markets for goods based upon economic principles.

The matrix of flow quantities between all O-D pairs is designated Q, with submatrix Q^p denoting the flow of commodity p. For consistency, units for all flows in Q are specified in a measure of weight, normally tons or metric tons. All flow values must be non-negative.

Network Structure

In scale, the modeled transportation network represents a region or nation. The topology of this network describes the physical transportation network with little aggregation or abstraction.

Links

Define L to be the set of all links in the network. For the most part, these links represent physical transportation facilities such as line haul track segments and classification yards or terminals. We may, in certain cases, add abstract links as in the case of a demand centroid connector. Associated with each link is a vector of attributes defining its physical and service characteristics.

In general, links in the real world network are undirected. For reasons which will become clear as the formulation proceeds, we represent the network as a set of N nodes and A directed arcs. Each undirected link is represented equivalently as a set of directed forward and reverse arcs.

There is no restriction against carriers of the same mode sharing a physical link $l = (i;j)$, $l \in L$, as in the case of joint track or trackage rights in the railroad industry. So

that we can model each carrier individually, we wish for the subnetworks to maintain separate representations for such shared physical facilities. The forward arc representing link l for carrier m is then specified as $a = (i,j,m)_l$. There may also be a corresponding reverse arc $a' = (j,i,m)_l$. The subscript accounts for the case where we have parallel physical arcs between i and j.

Each link l is represented, therefore, in the network by a set of forward arcs $\overline{A}_F = \bigcup_{m \in M}(i,j,m)_l$. If the link is undirected, then there is a corresponding set of reverse arcs $\overline{A}_R = \bigcup_{m \in M}(j,i,m)_l$.

Nodes

Nodes in the model physically represent junctions between line segments or locations where line characteristics change, as from single to multiple track. Nodes may also represent sources or sinks for traffic flow.

Connections between carrier subnetworks take place at a set T of designated transfer locations. The network is intermodal if transfers exist between carriers of different modes. Given a node $t \in \{N_m \cap N_n\}$, the transfer between carriers m and n at this node may be designated as $t_{m,n}$. Transfers are directed, and for transfer $t_{m,n}$, its counterpart $t_{n,m}$ may or may not be defined. Henceforth, we will use the designation t without subscripts to refer to an individual transfer.

In this model, transfers have a vector of cost attributes, but are assumed not to have capacity constraints or to experience congestion effects. If transfer congestion

effects are desired, the network structure can be modified by adding logical links through which flow to the transfer point must pass. We assume otherwise that carriers provide line haul service as necessary to handle transfer flows.

Complete Network

The complete network is therefore represented by $G = (N, A)$, where N is the set of nodes and A is the set of directed arcs which connect these nodes. The arcs represent the set of L physical and logical links. Each carrier m operates a subnetwork G_m which consists of N_m nodes and A_m directed arcs. The complete network therefore consists of the union of the carrier subnetworks, with $N = \bigcup_{m \in M} N_m$ and $A = \bigcup_{m \in M} A_m$. The set T of transfers defines connections where flows may pass between the subnetworks. We see that, in general, subnetworks may share nodes, as at transfers, but arcs are unique to a carrier. In other words, $A_m \cap A_n = \{\varnothing\}, \quad \forall\, m,n$.

Flows

The volume of commodity p on arc a is given by v_a^p. Likewise, the volume of commodity p through transfer t is v_t^p. Both v_a^p and v_t^p must be non-negative. The vector of network facility volumes for commodity p is:

$$v^p = \begin{pmatrix} (v_a^p),\, a \in A \\ (v_t^p),\, t \in T \end{pmatrix}.$$

The complete facility loading in the network, called the *load pattern*, is given by vector $v = (v^p, p \in P)$.

Next, we derive a relationship between path flows and arc/transfer flows. For a given O-D pair, w, the volume of commodity p flowing between w is q_w^p, $q_w^p \in Q^p$. Define K_w as the set of paths through the network connecting w. If, for w, i is the origin node and j is the destination node, a path k_w, $k_w \in K_w$, can be expressed as:

$$k_w = (i, n_1, n_2, \ldots, t_1, n_s, n_{s+1}, \ldots, t_2, n_u, n_{u+1}, \ldots, j).$$

Here, n_x represents an ordinary node in the chain and t_y represents a transfer. Alternately, the path may be expressed as a chain of arcs:

$$\begin{aligned} k_w = \big(&(i, n_1, m_1), (n_1, n_2, m_1), \ldots, (n_{s-1}, t_1, m_1), (t_1, n_s, m_2), (n_s, n_{s+1}, m_2), \ldots, \\ &(n_{u-1}, t_2, m_2), (t_2, n_u, m_3), (n_u, n_{u+1}, m_3), \ldots, (n_{u+z}, j, m_3)\big). \end{aligned}$$

Path k_w can be seen to consist of several subpaths, each of which belongs to a specific carrier:

$$\begin{aligned} k_w &= k_w^{m_1} + k_w^{m_2} + k_w^{m_3}, \\ k_w^{m_1} &= \big((i, n_1, m_1), (n_1, n_2, m_1), \ldots, (n_{s-1}, t_1, m_1)\big), \\ k_w^{m_2} &= \big((t_1, n_s, m_2), (n_s, n_{s+1}, m_2), \ldots, (n_{u-1}, t_2, m_2)\big), \\ k_w^{m_3} &= \big((t_2, n_u, m_3), (n_u, n_{u+1}, m_3), \ldots, (n_{u+z}, j, m_3)\big). \end{aligned}$$

Denote the flow of commodity p on path k_w as $\tau_{k_w}^p$, which must be non-negative. To assure flow conservation, the flows of p on all paths in K_w must sum to the total specified flow volume of p between O-D pair w:

$$\sum_{k_w \in K_w} \tau_{k_w}^p = q_w^p. \qquad (3.1)$$

The set of all paths between all O-D pairs over which commodity p might flow is $K = \bigcup_{w \in W} K_w$. The relationship between arc flows and path flows for p is expressed as:

$$v_a^p = \sum_{k \in K} \delta_a^k \tau_k^p \qquad (3.2)$$

where: $\delta_a^k = \begin{cases} 1 \text{ if arc } a \text{ is in path } k \\ 0 \text{ otherwise.} \end{cases}$

The equivalent relationship between transfer flows and path flows is:

$$v_t^p = \sum_{k \in K} \delta_t^k \tau_k^p \qquad (3.3)$$

where: $\delta_t^k = \begin{cases} 1 \text{ if transfer } t \text{ is in path } k \\ 0 \text{ otherwise.} \end{cases}$

Note that for a particular path k_w, the total flow is the vector $\tau_{k_w} = (\tau_{k_w}^1, \tau_{k_w}^2, \ldots, \tau_{k_w}^p)$ which contains a flow (possibly zero) for each commodity. The indexed set $\tau \equiv \{\tau_k, k \in K\}$ contains all path flows in the network. This set is called the *flow pattern*. The equivalent load pattern for arcs and transfers is constructed using the relationships in (3.2) and (3.3). The load vector for arc a is $v_a = (v_a^1, v_a^2, \ldots, v_a^p)$ and for transfer t is $v_t = (v_t^1, v_t^2, \ldots, v_t^p)$. The load pattern is then the indexed set $v \equiv \{v_a, a \in A\} \cup \{v_t, t \in T\}$, which is a restatement of the earlier definition.

Costs

Given a pattern of flows, we are now interested in determining the cost characteristics of those flows.

Flow/Cost Relationships

The cost of a flow pattern is equivalent to the cost of the corresponding load pattern. Thus, we may look at costs for loads on individual facilities.

Average Costs

The average cost of a flow unit of commodity p on arc a is given by s_a^p and on transfer t by s_t^p. Both s_a^p and s_t^p must be non-negative. The vector of network average facility unit costs for commodity p is:

$$s^p = \begin{pmatrix} (s_a^p), a \in A \\ (s_t^p), t \in T \end{pmatrix}.$$

Vector $s = (s^p, p \in P)$ provides the average unit costs for all facility/commodity combinations.

For a given commodity, the unit cost on a facility is normally considered to be a function of the load pattern. In general, we therefore can say that $s_a = s_a(v)$ and $s_t = s_t(v)$. Realistically, however, it can be questioned whether, for example, there are cost interactions between arcs or transfers representing different physical facilities. In our model, therefore, we assume:

a) The cost functions for a given transfer are not affected by the flows at other transfers or by arc flows. This infers that flows at $t_{m,n}$ do not interact with flows for $t_{n,m}$.

b) The cost function for an arc is not affected by transfer flows; and

c) The cost function for an arc is only affected by flows on arcs which represent the same physical link. There is no interaction between flows on separate physical links.

The real world railroad system behaves similarly.

Under assumption (c), the cost function for an arc can be affected by the flows on other arcs representing the same physical facility. The interaction between flows is apparent, for example, on a single track railroad line represented in the model by a forward arc and a reverse arc. The delay characteristics for such a line are a function of the total traffic in both directions. We then define $\overline{A}$ as a set of interacting arcs representing a physical link, $l = (i; j)$, $l \in L$, connecting nodes i and j. In general, for most railroad line classes where two-way traffic interacts, $\overline{A} = \overline{A}_F \cup \overline{A}_R$. In the case of non-interacting two-way traffic, as with directional double track, $\overline{A} = \overline{A}_F$ if $a \in \overline{A}_F$, otherwise $\overline{A} = \overline{A}_R$. It is apparent then, for arc flows, that we must evaluate a portion of the load pattern defined as $v_{\overline{A}} \equiv \{v_a, a \in \overline{A}\}$.

Based upon the above assumptions, and the definition of $\overline{A}$, the form of the average cost function can be made more specific for each facility type. The average cost

vector for arc a is now $s_a = s_a(v_{\bar{A}})$. Since each commodity can have a distinct cost structure, the vector equation may be expressed as a set of p-scalar equations:

$$\begin{aligned} s_a^1 &= s_a^1(v_{\bar{A}}^1, \ldots, v_{\bar{A}}^p), \\ &\vdots \\ s_a^p &= s_a^p(v_{\bar{A}}^1, \ldots, v_{\bar{A}}^p). \end{aligned}$$

Transfers have no interaction, and therefore, no equivalent to $\bar{A}$. The average cost vector for transfer t is $s_t = s_t(v_t)$, with the corresponding set of p-scalar equations:

$$\begin{aligned} s_t^1 &= s_t^1(v_t^1, \ldots, v_t^p), \\ &\vdots \\ s_t^p &= s_t^p(v_t^1, \ldots, v_t^p). \end{aligned}$$

Total Costs

The preceding section defined average cost relationships to the flow pattern. The total cost for the flow pattern is the practical measure of interest, however. As with the average unit cost, the total cost can be expressed in terms of the facility load pattern. The total cost for the flow of commodity p on arc a is $s_a^p(v_{\bar{A}})v_a^p$. The corresponding total cost for a transfer t is $s_t^p(v_t)v_t^p$. The total cost of the flow for product p is then:

$$\sum_{a \in A} s_a^p(v_{\bar{A}})v_a^p + \sum_{t \in T} s_t^p(v_t)v_t^p. \qquad (3.4)$$

The total system cost for the entire load pattern is:

$$\sum_{p \in P} \left(\sum_{a \in A} s_a^p(v_{\bar{A}})v_a^p + \sum_{t \in T} s_t^p(v_t)v_t^p \right). \qquad (3.5)$$

Facility Cost Functions

To compute costs, specific average cost functions which adhere to the requirements of the previous section are needed. These functions yield a generalized cost expressed as cost/unit of weight. First the case of arcs is examined and then that of transfers.

Arc Cost Functions

In this model, there are two distinct average cost functions for arcs. One function applies to arcs which model line-haul track segments, and the other applies to arcs representing terminals.

Line-haul cost function. The line haul average cost function is hypothesized to provide a generalized cost having a weight-distance based component and a time based component. The function has the form:

$$s_a^p(v_{\bar{A}}) = m_a^p l_{\bar{A}} + T_{\bar{A}}(v_{\bar{A}}) f_a^p h_a^p \tag{3.8}$$

where:

m_a^p	=	the cost per net ton-mile for commodity p on arc a;
$l_{\bar{A}}$	=	the length of the arc's physical link;
h_a^p	=	train cost per hour for commodity p on arc a;
$T_{\bar{A}}(v_{\bar{A}})$	=	travel time on arc a, given load pattern $v_{\bar{A}}$;
f_a^p	=	commodity conversion factor, weight to trains.

Subsequent sections discuss these terms and their explanatory variables.

Weight-distance cost term. The weight-distance component $m_a^p l_{\bar{A}}$ reflects cost elements such as track maintenance, equipment wear, allocated overhead costs, etc. Such items are normally measured as a cost per net or gross ton-mile of carriage. We use the $\bar{A}$ subscript on the length variable to denote a link specific attribute. Note that, given a gross-weight to payload ratio, m_a^p can be adjusted quite easily to reflect the gross ton-mile cost.

Time cost term. The second component of the cost function is the time cost of transporting the commodity over the arc. This term accounts for costs such as fuel, labor, time value of locomotives and equipment, and time value of the commodity being transported. These cost categories are measured in cost per unit time, typically dollars per hour. The discrete unit of many of these costs is the train, and travel time over a line segment is typically viewed on a per-train basis.

If the load pattern $v_{\bar{A}}$ can be converted to the equivalent number of trains, a congestion function of the type discussed in Chapter 2 can be used to compute the average travel time. To do this, we define for each commodity *p* and arc *a*, a factor f_a^p which converts the net weight of *p* to a number of equivalent trains:

$$f_a^p = \frac{\omega_m^p + \varepsilon_m^p}{\omega_m^p \chi_m^p \alpha_a} \tag{3.6}$$

where: $a \in A_m$

ω_m^p = weight of commodity *p* in a loaded car for mode *m*;

ε_m^p = tare weight of an empty car for commodity p on mode m;

χ_m^p = trailing gross weight of a train of commodity p on mode m;

α_a = calibration factor for arc a.

The number of trains V_a^p on arc a of commodity p is then $f_a^p v_a^p$. The total number of trains, $V_{\bar{A}}$, defined by load pattern $v_{\bar{A}}$, is

$$V_{\bar{A}} = \sum_{p \in P} \sum_{a \in \bar{A}} f_a^p v_a^p . \qquad (3.7)$$

This approach is similar to that employed by Crainic, Florian, and Léal (1990), who report good agreement with observed volumes on Canadian railroads.

There are several points related to this approach which should be noted. First, Equation (3.7) yields, in general, a non-integer number of trains. Since we are considering average flow, and not modeling detailed operations, this is acceptable. Second, the trailing gross weight of a particular train type does not include locomotive weights. Third, the arc calibration factor α_a is used to adjust train weights on arcs representing links with operating restrictions, such as grades or short sidings, which do not permit operation of the "average" train. It may also be used to increase weights. Finally, for a given product p, values of ω and ε are recommended to be constant for carriers which interchange traffic. Different values may be appropriate where transloading takes place at a transfer point. Otherwise, there will be a flow imbalance in

terms of cars at transfer points, although weight flow conservation constraints will not be violated.

Given a congestion function, the average travel time $T_{\bar{A}}$ for the arc can be determined as a function of the train volume $V_{\bar{A}}$. Since $V_{\bar{A}}$ is, in turn, a function of the load pattern $v_{\bar{A}}$, then $T_{\bar{A}} = T_{\bar{A}}(v_{\bar{A}})$. In formulating our assignment model formulation, we may use, in general, any congestion function. Chapter 4 will show, however, that the solution procedure requires the congestion function to meet certain criteria.

The time cost term needs to be expressed in terms of cost per unit weight. The product of $T_{\bar{A}}(v_{\bar{A}})h_a^p$ yields units of cost per train-hour. Multiplying this by f_a^p will yield units of cost per unit weight. The complete cost term is therefore $T_{\bar{A}}(v_{\bar{A}})f_a^p h_a^p$.

<u>Terminal cost function</u>. The form of the terminal arc average cost function is similar to that of the line haul arc. With terminals, however, the discrete unit of traffic is typically the car. In addition, terminal arcs are typically short in length, so weight-distance is not a major contributor to costing. The terms of the cost function therefore need to be modified to reflect this. It can be hypothesized that, in the case of terminals, the cost function contains one term expressing the processing costs per traffic unit and a second term related to the time costs of processing traffic. The cost function is:

$$s_a^p(v_{\bar{A}}) = \hat{m}_a^p \hat{f}_a^p + \hat{T}_{\bar{A}}(v_{\bar{A}}) \hat{f}_a^p \hat{h}_a^p \qquad (3.9)$$

where: $\hat{m}_a^p$ = the processing cost per car for commodity p on arc a;

$\hat{f}_a^p$ = commodity conversion factor, weight to cars;

$\hat{h}_a^p$ = car cost per hour for commodity p on arc a;

$\hat{T}_{\bar{A}}(v_{\bar{A}})$ = car processing time on arc a, given load pattern $v_{\bar{A}}$.

Equations (3.8) and (3.9) are structurally equivalent.

Processing cost term. The initial term in the terminal average cost function accounts for the cost of processing a carload of commodity p on arc a. Components of this cost include handling, loss and damage, and allocated terminal operating costs. Conversion of this cost to a generalized cost per unit of weight is accomplished by multiplying by $\hat{f}_a^p$, where

$$\hat{f}_a^p = 1 / \omega_m^p, \quad a \in A_m. \tag{3.10}$$

Time cost term. The second term accounts for the time value of the contents of a car of commodity p. A congestion function for terminal facilities predicts the car processing time as a function of car volume per unit of time. The number of cars $\hat{V}_a^p$ on arc a of commodity p is $\hat{f}_a^p v_a^p$. Derivation of the complete term, $\hat{T}_{\bar{A}}(v_{\bar{A}}) \hat{f}_a^p \hat{h}_a^p$, then follows the same logic as the equivalent term in the line haul cost function.

Transfer Cost Function

In the model, transfer locations have no congestion effects or capacity limits. The cost model for a transfer is designed simply to reflect a commodity specific cost per car for performing the transfer:

$$s_t^p = \tilde{m}_t^p \tilde{f}_t^p \tag{3.11}$$

where: $\tilde{m}_t^p$ = the cost per car of commodity p using transfer t;

$\tilde{f}_t^p$ = cars per ton of commodity p using transfer t.

The cost $\tilde{m}_t^p$ may reflect factors such as an average time cost for the transfer, administrative charges, or delivery costs.

Railroad routing practice usually minimizes the number of transfers, since a transfer normally represents delay to the shipment. Of the set of transfer points available to a large railroad, historic traffic patterns will favor a subset for the majority of interchange activity. Other interchanges will have relatively little traffic. If the predicted flow pattern is to replicate actual conditions, the transfer cost function should reflect this hierarchy.

Objective Function

The preceding sections provided a framework for defining the network, describing demand and load patterns, and defining costs for facility loadings. Of interest now is a mathematical expression which will produce the load pattern in the network.

In this model, the objective is to select the load pattern which minimizes total generalized costs. While not every shipper-carrier interaction results in minimum total costs in the real world, in a strategic planning model, this objective appears to have validity. In passenger transportation planning, it is often assumed that the flow pattern will reach an equilibrium state where no traveler can improve his travel cost by unilaterally changing routes. To achieve this user equilibrium flow pattern, travelers

must have perfect knowledge of the system. In a freight system, shippers might select a carrier on a user equilibrium basis, given a limited set of cost and service variables. Seldom, however, will the shipper have much knowledge of the carrier's network operations. Carriers control micro level routing decisions based upon the traffic tendered from the entire shipper population.

The use of generalized costs reflects total logistics costs, and, in an environment of competition, carriers and shippers will, it can be argued, work together to minimize total costs. Since the model is based upon fixed demands, the shippers are not explicitly included as agents. The generalized cost may, however, contain components, such as the time value of commodities, to implicitly represent shipper interests. These cost components decrease the utility of routes with poor service characteristics. From a carrier standpoint, since the time frame of the model is short term, rates are assumed to be fixed. By minimizing costs, a carrier will maximize the portion of revenue brought to the bottom line.

Mathematical Program

The load pattern at which total generalized costs are minimized is called the system optimum (SO). Mathematically, the SO load pattern can be determined using the following non-linear program:

$$\min Z = \sum_{p \in P} \left(\sum_{a \in A} s_a^p(v_{\bar{A}}) v_a^p + \sum_{t \in T} s_t^p(v_t^p) v_t^p \right) \tag{3.12}$$

subject to:

$$\sum_{k_w \in K_w} \tau_{k_w}^p = q_w^p, \quad \forall p, w \tag{3.1}$$

$$\tau_{k_w}^p \geq 0, \quad \forall p, w, k_w \in K_w \tag{3.13}$$

$$v_a^p = \sum_{k \in K} \delta_a^k \tau_k^p, \quad \forall a, p \tag{3.2}$$

$$v_t^p = \sum_{k \in K} \delta_t^k \tau_k^p, \quad \forall t, p. \tag{3.3}$$

The constraints (3.1) and (3.13) assure flow conservation on paths. Constraints (3.2) and (3.3) transform path flows into arc and transfer flows.

Necessary and Sufficient Conditions

The solution of the above problem will yield the desired SO flow pattern for the network provided that certain necessary and sufficient conditions are met. Convexity of the feasible region is guaranteed by the fact that constraints (3.1), (3.2), and (3.3) are linear equalities. A second requirement is that equation (3.12) be convex. This can be guaranteed since all of the arc and transfer performance functions are convex, positive, and monotone increasing, and, therefore, the product $s_a^p(v_{\bar{A}})v_a^p$ is convex over the range of flows v_a^p. The objective function will then be convex since the sum of a series of convex functions is itself convex. The mathematical conditions are defined in the following sections.

Necessary Conditions

The necessary conditions, which can be found in a number of texts, such as Sheffi (1985), are as follows:

$$\tau^p_{k_w}(c^p_{k_w} - \hat{c}^p_w) = 0, \quad \forall p, w, k_w \in K_w \tag{3.14}$$

and

$$c^p_{k_w} - \hat{c}^p_w \geq 0, \quad \forall p, w, k_w \in K_w . \tag{3.15}$$

Equations (3.1) and (3.13), the flow conservation constraints, must also be met.

Variable $c^p_{k_w}$ represents the marginal total cost for moving product p over path k_w:

$$c^p_{k_w} = \frac{\partial Z}{\partial \tau^p_{k_w}} . \tag{3.16}$$

The marginal cost, well known in economic theory, is the addition to total costs of adding an additional incremental unit of commodity p to the flow on path k_w. Variable $\hat{c}^p_w$ is the dual variable for the corresponding constraint in equation (3.1). According to the duality theory of linear programming, this dual variable is the cost of adding an increment of commodity p to the total flow between O-D pair w. Thus, $\hat{c}^p_w$ is also a marginal cost.

From equation (3.14), for O-D pair w flow of commodity p on path $k_w \in K_w$ is non-zero only when $c^p_{k_w} = \hat{c}^p_w$. Paths where $c^p_{k_w}$ is greater than the associated dual $\hat{c}^p_w$ receive no flow.

Although the marginal costs are expressed in terms of paths, equivalent arc and transfer formulations can be easily derived. Facility marginal costs are discussed in detail in Appendix A. From Appendix A, the incremental unit cost for arc $\bar{a}$ is

$$c_{\bar{a}}^{\bar{p}} = s_{\bar{a}}^{\bar{p}}(v_{\bar{A}}) + \sum_{p \in P} \sum_{a \in \bar{A}} \frac{\partial s_a^p(v_{\bar{A}})}{\partial v_{\bar{a}}^{\bar{p}}} v_a^p$$

and for transfer t

$$c_t^{\bar{p}} = s_t^{\bar{p}}(v_t^p).$$

The marginal cost for commodity p on path k_w is then

$$c_{k_w}^p = \sum_{a \in A} \delta_a^{k_w} c_a^p + \sum_{t \in T} \delta_t^{k_w} c_t^p$$

where $\delta_a^{k_w}$ and $\delta_t^{k_w}$ are indicator variables as in equations (3.2) and (3.3).

Sufficient Conditions

The condition for the existence of a unique minimum to the multicommodity SE problem is that the objective function be strictly convex. If the Hessian of Z (the matrix of second derivatives of Z) is positive definite, this is sufficient to demonstrate strict convexity, and, thus, the existence of a unique minimum. The Hessian, H, is positive definite if, for $v \neq 0$, $v^T Hv > 0$. In the formulation, elements of H relating to arcs are positive, since arc cost functions are strictly convex, positive, and monotone increasing. Transfers, however, have a linear cost function which yields a second partial derivative of zero. The reader can verify that, under these conditions, terms in $v^T Hv$ contain only arc

flows. By the criteria applied to arc cost functions, then, $v^T H v$ cannot be non-positive and H must be positive definite.

The properties of convex function addition can also prove the uniqueness of the result. We know that objective function is convex because the sum of convex functions is always convex. The objective function in this program is the sum of strictly convex functions (arc costs) and convex functions (transfer costs). If the result of the addition of convex and strictly convex functions is strictly convex, then the program will guarantee a unique minimum.

Strict convexity requires that, given any two distinct points x_1 and x_2,

$$z[\theta x_1 + (1-\theta) x_2] < \theta z(x_1) + (1-\theta) z(x_2)$$

for any value of θ, $0 < \theta < 1$. Let $f(x)$ be a strictly convex function of x, and $f(y)$ be a convex function of y. Two sets of points, (x_1, y_1) and (x_2, y_2), contain distinct values of x and y. If the sum of $f(x)$ and $f(y)$ is strictly convex, then

$$f[\theta x_1 + (1-\theta) x_2] + f[\theta y_1 + (1-\theta) y_2] < \theta[f(x_1) + f(y_1)] + (1-\theta)[f(x_2) + f(y_2)].$$

If $f(y)$ is convex, but not strictly so, then $f(y)$ must be linear on y, since $f''(y) = 0$. It is recognized, therefore, that

$$f[\theta y_1 + (1-\theta) y_2] = \theta f(y_1) + (1-\theta) f(y_2).$$

These terms cancel in the inequality, leaving

$$f[\theta x_1 + (1-\theta) x_2] < \theta f(x_1) + (1-\theta) f(x_2)$$

which we know to be true since $f(x)$ is strictly convex. Therefore, we have shown that the sum of convex and strictly convex functions is strictly convex.

Since transfer flow cannot occur in the objective function without arc flow, the objective function must always be strictly convex in the vicinity of the optimum, and, therefore, Z is a global minimum.

Given its linear term for either weight distance or car throughput and its polynomial term for time related costs, the arc average cost function is convex, monotone increasing, and everywhere positive. Transfer costs are linear, and, therefore, meet the same criteria. Under these conditions, the objective function is strictly convex and the algorithm will converge to a global minimum. If transportation firms exhibit economies of density, however, average unit costs decline with increasing volume to a point, and then increase as the firm devotes additional resources to handle traffic effects. This creates a U-shaped average cost curve which is convex, but not monotone increasing. Under this condition, the terms $s_a^p(v_{\bar{A}})v_a^p$ will not generally be convex, and, therefore, the objective function will be nonconvex. This means that the program solution will not have a unique value, and there will be multiple minima. The algorithm may converge to a minimum, but there is no guarantee that this is the global minimum.

Chapter 4 will discuss the implementation of the model.

CHAPTER 4

MODEL IMPLEMENTATION

This chapter discusses the implementation of the model. The first section presents the solution algorithm for the mathematical program stated in Chapter 3. Subsequent sections then describe the computer code developed to implement the solution algorithm.

Solution Algorithm

The mathematical program set forth in Chapter 3 can best be described as having a non-linear, multivariable, convex objective function with linear constraints. Solution approaches which provide insight into this particular programs are provided in a number of references. In his text on network flows, Hu (1969) discusses some of the unique issues associated with multicommodity flow formulations, namely that the constraint matrix is not unimodular and that the tremendous number of potential columns in the solution algorithm hint at a column generation based solution procedure. Dafermos (1971) examines the multiclass assignment problem and proposes a two-stage solution procedure which has as its heart a decomposition of the problem by class. Sheffi (1985) describes efficient two-stage algorithms for solving the single commodity, non-linear SO problem which might be extended for the multicommodity problem. These include linear approximation procedures such as the Frank-Wolfe algorithm. Guélat, Florian, and

Crainic (1990) describe a solution procedure similar to Dafermos' which is used in their network model.

Overview

The constraint set defines a convex polytope encompassing the feasible region. The heart of any solution procedure is as follows. First, obtain an initial feasible flow pattern, v. This will represent a point on the surface of the polytope. Then, with each step of the algorithm, find a new feasible extreme vector, w, which improves the objective function. The two vectors v and w define a line in n-space. Using a linear search procedure, find the value of θ which minimizes the convex combination of v and w,

$$v_{new} = (1-\theta)v + \theta w . \tag{4.1}$$

The algorithm continues until $v_{new} \approx v$.

The above procedure is generally referred to as a convex combinations algorithm. The important step of determining the new feasible extremal vector w is the critical step. The procedure is to use the gradient of the objective function to formulate a linear approximation to the objective function. Minimizing this linear approximation to the value of the objective function subject to a system of linear constraints has as its solution a corner of the feasible space. The objective function of this program is

$$\min Z(w) = Z(v) + \nabla Z(v) \cdot (w-v)^T . \tag{4.2a}$$

Terms $Z(v)$ and $\nabla Z(v)(v)^T$ are constants which may be omitted. This results in the revised objective function

$$\min Z(w) = \nabla Z(v) \cdot (w)^T = \sum_i \left(\frac{\partial Z(v)}{\partial v_i} \right) w_i . \qquad (4.2b)$$

The term $\dfrac{\partial Z(v)}{\partial v_i}$ is simply the marginal cost with respect to v_i. When the problem has the structure of a network, a feasible optimal solution for equation (4.2b) may be found using a straightforward shortest path algorithm.

In the multicommodity flow problem, the vectors v and w are of dimension $P(A+T)$. By decomposing the problem by commodity, the vector size may be reduced to $(A+T)$, which represents a substantial savings in computer storage. This approach was advocated in both the aforementioned papers by Dafermos and Guélat et al. During each iteration of the algorithm, a linear approximation subproblem is solved for each commodity, using marginal costs with respect to the flow of that commodity. Flows of the other commodities are held fixed.

We have mentioned that, for the multicommodity problem, the constraint coefficient matrix is not unimodular. This means that, given integer flows for each commodity, optimal arc and path flows will generally not be integer. In a strategic planning model such as this one, non-integrality of the solution is not a problem, since quantities are generally large and the solution represents, at best, average conditions.

Algorithm Description

The following paragraphs summarize the steps in the solution algorithm.

Step 0. Initialization

Determine an initial feasible flow vector, v. This can be done using an iteration of Step 1 with initial marginal costs corresponding to a zero flow state and $\theta = 1$ for each commodity subproblem.

Step 1. Flow Vector Update

For each commodity $p \in P$, perform the following sequence of steps:

a) Given v, compute marginal costs, c_a^p and c_t^p, for all arcs $a \in A$ and transfers $t \in T$.

b) For each O-D pair $w \in W$ having a corresponding flow $q_w^p \in Q^p$, solve the shortest path problem using c_a^p and c_t^p as facility costs. Assign q_w^p to this path.

c) Let y^p be the load vector resulting from Step 1b, with y being the corresponding overall load pattern. Using a one-dimensional search algorithm, solve the problem

$$\min\,(1-\theta)Z(v)+\theta Z(y)$$

$$\text{subject to: } 0 \le \theta \le 1 .$$

d) Let $v^p = (1-\theta)v^p + \theta\, y^p$.

Step 3. Stopping Criterion

The algorithm terminates if the iteration count exceeds a predetermined number or if the current value of the objective function is within a predefined tolerance of the previous value. Otherwise, return to Step 1.

Guélat et al. (1990) prove that convex combinations algorithms which decompose the problem by commodity will converge when the objective function and constraints are convex.

Cost Functions

The solution algorithm uses functions to compute two types of costs: marginal total costs and average total costs. Appendix A provides derivations for the marginal cost functions.

The total cost function forms were described in Chapter 3, without specific reference to the form of the congestion function used to compute arc travel times. The model uses the polynomial travel time function introduced in Chapter 2 and presented in more detail in Appendix A. This function is convex, monotone increasing, and everywhere positive. The polynomial function is also twice continuously differentiable. This makes it ideal for use in calculating both average unit costs and marginal unit costs.

In form, the arc cost functions always provide positive results, are continuous, and monotone increasing. Substitution of a U-shaped functional form which reflects economies of density, while possibly more reflective of transportation firms, would lead to a nonconvex objective function which has multiple minima. If we consider that our

network consists of major routes, all of which have a reasonable volume of traffic, then all routes may be considered to operate on the increasing side of the cost function if economies of density apply. Thus, the cost functions used in the model should approximate these conditions.

Shortest Path Algorithm

Step 1b of the solution algorithm uses a shortest path algorithm (SPA) to solve the minimum marginal cost path problem for each O-D pair $w \in W$ with $q_w^p > 0$. RAILNET uses a modified form of the standard Moore algorithm to generate these paths. Modifications to the SPA were made to account for some unique requirements of the model structure. First, the algorithm must produce paths which account for the decomposition of the overall network into a series of carrier subnetworks connected at transfer points. This is easily done using an arc-chain path rather than a node chain path. The arc-chain formulation also simplifies path tracing during the arc loading process. Second, if flow q_w^p has a designated originating carrier, the SPA must ensure that the path starts with this carrier. Appendix B describes the algorithm in detail.

Implementation

This section generally describes the data files and software used to setup and solve an assignment problem. Chapter 5 also discusses the use of the files and software in a case study.

Input Data Files

Problems are defined using a series of input data files, some required and some optional. These are:

- Problem parameter file;
- Link file;
- Transfer definition file (optional);
- Commodity flow file;
- Commodity/carrier cost file;
- Link cost exception file (optional); and
- Transfer cost exception file (optional).

The problem network is provided in the form of a link file and an optional transfer definition file. Both of these files are formatted text files which may be created using any text editor.

The link file defines the network topology and link attributes. Table 4-1 lists the contents of the link file. Links in the model are of five types:

(5) yard, directional;

(4) multiple track, bidirectional;

(3) multiple track, directional;

(2) single track, bidirectional;

(1) single track, directional; and

(0) dummy link, directional.

Table 4-1. Link File Structure

Field	Type	Description
ID	I5	Unique integer identifier for link
A_NODE	I5	Tail node for link
B_NODE	I5	Head node for link
LENGTH	F6.1	Link length in miles, ignored for terminal
TTIME	F6.2	Free flow travel time, mins. (car processing time, mins.)
CAPACITY	F6.1	Capacity in trains per day or cars per day (terminal)
K1	F10.6	Polynomial capacity function coefficient
K2	F10.6	Polynomial capacity function coefficient
GAMMA	F6.2	Polynomial capacity function exponent
TADJ	F6.2	Link train size adjustment coefficient
ML_CLASS	A1	Link FRA mainline class
LINK_TYPE	I1	Type of facility represented by link
NO_RRS	I1	Number of carriers using link
RR1	I3	Carrier 1, expressed as an integer code
RR2	I3	Carrier 2
RR3	I3	Carrier 3
RR4	I3	Carrier 4
RR5	I3	Carrier 5
RR6	I3	Carrier 6
RR7	I3	Carrier 7
RR8	I3	Carrier 8
RR9	I3	Carrier 9

A dummy link has no capacity attributes and is used simply to maintain network topology. A typical use might be to reach a demand centroid.

The transfer file defines nodes where flows may cross from one carrier subnetwork to another. This file is not needed if the network contains no transfers, as might be the case if a single carrier is being studied.

The demand characteristics of the problem are defined in a data file containing a series of records defining flows of commodities between O-D pairs. Each record contains an origin, destination, commodity, flow quantity, and optional origin carrier. Origins and destinations may be any nodes in the network.

The input cost file provided to RAILNET contains, for each carrier, cost terms by commodity for line haul arcs, terminals, and transfers. These cost terms are applied globally to all of the carrier's facilities. If adjustments are needed for specific facilities, the program can accept these from the optional link and transfer exception files. Values in the exception files override the default values. Should a particular commodity be restricted from a specific carrier, transfer, or facility in the real network, the analyst can use the cost and exception files to set costs for the commodity to very high values over all affected facilities.

Users may use different cost files with a given set of commodity and network files. Commodities in the cost file which are not in the commodity file will be ignored. If a carrier in the network is not found in the cost file, that carrier's facility costs for all commodities will be set to very high values. Carrier facility costs for a commodity not

found in the cost file will also receive very high values. In general, if there is any option, paths will not use these high cost facilities.

Model Components

A series of computer programs were written to implement the multicommodity, multimode assignment algorithm described previously. These programs are designed to run on an IBM-compatible 80386 personal computer under the DOS operating system. All source code is written in 32-bit Microsoft FORTRAN Powerstation 1.0.

The program suite used to implement the model consists of three components: NETBLD, COMMODTY, and RAILNET. The first two programs process the input network and commodity data into a form suitable for the third component, which is the algorithm solver. The following sections describe the basic function of these programs.

NETBLD

The NETBLD program preprocesses the input network and transfer files, producing output files suitable for the solution program. Given network and transfer files, NETBLD performs a number of processes. First, the program checks all of the data fields on each record in both files for validity. Second, it creates an equivalent arc representation for each link. Links are decomposed into separate arcs for each carrier using the link. Next, NETBLD creates an indexed representation of the network using a forward star format suitable for use in shortest path algorithms. Finally, if a transfer file is present, the internal network representation is adjusted to include transfers.

After processing is complete, NETBLD produces a series of binary files containing the processed network and transfer vectors, a vector of network carriers, and a parameter file. Provided that the network is not changed, these files may be used repeatedly by the solver. NETBLD also produces output reports listing network elements.

COMMODTY

COMMODTY processes the input commodity flow file, checking the data elements on each record for validity. This requires access to the data files produced by NETBLD for the target network. COMMODTY builds lists of commodities and demand nodes as it encounters them in the input file.

After reading the data elements on each record and storing them in vectors, the program orders the vectors by commodity, origin, start mode, and destination. It then builds a commodity index so that the subvector for each commodity can be accessed. In essence, each commodity subvector represents a sparse O-D matrix since O-D pairs with zero flow are omitted. This vector storage scheme for the flow matrixes greatly reduces data storage requirements, especially since P matrixes are needed.

After processing the commodity file, COMMODTY creates a series of binary files containing the vectors and indexes. The processed files may only be used with the network used in checking the input data. They are valid as long as the base network is not altered. The user may develop any number of flow cases using the same base

network. COMMODTY will produce a separate set of output files for each case. There are no limits on the number of commodities or demand nodes.

For each flow file, COMMODTY creates a problem parameter file describing the problem. This file includes a factor used by the model to convert the time period of the flows to the time unit of the congestion function. If, for example, flows are annualized and the congestion function is expressed in trains per day, the conversion factor might be 365.

RAILNET

RAILNET is the program solver. It takes as input the network and commodity files produced by NETBLD and COMMODTY. One additional file is needed, however, for RAILNET to define completely the problem. This file contains values for terms in the facility cost functions.

Given the required input files, RAILNET then solves the problem. The program, as written, can solve problems of virtually any size, subject to machine memory and disk limits. Arrays are dynamically allocated at run time. The program uses disk virtual memory to augment the available random access memory in the computer. Program termination occurs when a specified number of iterations completes or upon convergence of the objective function on two successive iterations to a specified tolerance. During execution, the user may, if desired, view intermediate results.

Upon completion, the program produces, at user option, two types of output files. A report file contains summary information and final link loadings, by carrier,

commodity, and direction. The listing also contains link travel times and delays for the assigned volumes. RAILNET will also produce two formatted output files containing, respectively, link volumes and delays and transfer volumes. The data in these files may be loaded into geographic information systems (GIS) for display or into other applications for further processing.

CHAPTER 5

CAPACITY EVALUATION CASE STUDY

This chapter presents a case study of the railroad network in the southeastern U.S. The objective of the study is to use the model to evaluate rail traffic flows in the region and to determine whether significant congestion problems exist. A base scenario will be examined.

Study Scope

The study network consists of all through routes in the southeastern U.S. The study region is bounded on the west by the Mississippi River, and on the north by the Ohio River to its intersection with the Mason-Dixon line, which then completes the boundary. This encompasses all of the states of Maryland, Virginia, Kentucky, Tennessee, North Carolina, South Carolina, Georgia, Florida, Alabama, and Mississippi, along with portions of the states of West Virginia and Louisiana. The region includes all of the ICC Southern Territory and a small portion of Official Territory.

The study region has several attractive characteristics. First, the boundaries form a cordon line with relatively few crossings. Key cities located along the boundary—Washington, Cincinnati, Louisville, Memphis, Baton Rouge, and New Orleans—are gateways or interchange points between the major southeastern carriers and railroads serving other regions of the country. Few interregional mainlines cross the boundary at

other locations. This somewhat simplifies accounting for interregional flows. Second, the territory is dominated by two major rail carriers, CSX Transportation and Norfolk Southern Corporation. Other major railroads—Illinois Central, Burlington Northern Santa Fe, and Kansas City Southern—enter the region in certain corridors, but have a limited overall role. This simplifies data collection and the evaluation of results. Third, the region has a number of well-defined intraregional corridors served by both carriers. Finally, the author's familiarity with the current railroad characteristics and traffic patterns in the region helps in developing scenarios for evaluation and appraising the model's predicted flows.

Network Construction

The first step in conducting the case study was to develop a network model of the railroad system in the region.

Base Network Evaluation

Because network construction is time-consuming, the use of existing network models is highly desirable. An ideal network will have all or many of the attributes, such as link lengths, carriers, speeds, etc., needed by the assignment model. In addition, the network will contain geographic coordinates to facilitate mapping and display, perhaps with a GIS. The network needs, of course, to reflect the topology of the rail network with reasonable accuracy. While some aggregation of features is acceptable in a planning model, excessive aggregation may lead to prediction errors.

An initial survey identified the following candidate networks: the Oak Ridge National Laboratory (ORNL) network; an enhanced version of the Federal Railroad Administration (FRA) network; and the rail component of the U.S. Geological Survey (USGS) TIGER files. The ORNL network has a very good set of attributes which are kept updated. It is highly collapsed in urban areas, however, and has geographic coordinates for nodes only. The FRA network is geographically accurate at 1:2,000,000 scale and contains line ownership, control system, and traffic density attributes. It contains reasonably good detail in urban areas, and attribute and network structure were validated during a previous research project. The TIGER files contain rail lines geographically at 1:100,000 scale, albeit with virtually no attributes. Although the data contains exquisite detail, it requires extensive editing to build topology, reflect carrier ownership and line attributes. This work had been started during a previous project, but completing it for this research seemed impractical. Based upon these findings, the FRA network was selected for use in the case study.

Network Editing and Revision

The next step in the development of the study was to load the FRA network into a GIS for editing and conversion to a format suitable for RAILNET. The TransCAD 2.1 GIS package was used for all network editing and display function. TransCAD has a suite of tools for network manipulation which greatly simplified the tasks associated with cleaning up and modifying the FRA databases. After processing, network data can be

exported from TransCAD to standard DOS text files. In this form, the data is suitable for use with other programs, including the program suite used in this model.

Despite the overall attractiveness of the network, extensive editing was required. An initial edit removed links and nodes not needed in the study. Except for a small subset of major rail lines extending to major gateways and demand centers, all of the network outside of the study region was deleted. Unless they served a major demand center, branch lines not forming part of a through route were deleted within the region. All binary nodes, such as those at state borders, former junction points, etc., were removed unless they represented a demand center. Finally, the network topology in major urban areas was examined and corrected as required. The rail lines ultimately retained met one or more of the following criteria:

a) annual traffic density of at least five million gross tons per mile;

b) service to a demand centroid as part of a through route; and

c) potential to serve as part of a through route with upgrading.

Criterion (c) was arrived at somewhat objectively. The other two criteria were based upon network topology and flow data contained in the FRA network.

Major lines outside the region were retained between the study boundary and Houston, Dallas/Fort Worth, Kansas City, St. Louis, Chicago, Detroit, Cleveland, Philadelphia, and northern New Jersey. These lines were retained so that the model would channel flows from far outside the region into the "proper" border crossing.

At this point, attributes required by RAILNET, but not present in the database, were added. This was done on a line segment by line segment basis for each carrier's trackage. Table 5-1 lists the data elements coded for each link, indicating those which were added during the editing step. Links were first checked to verify the owning railroad(s) and any tenant railroad(s) operating via trackage rights. Link attributes such as average speed, siding spacing, number of tracks, and signal system were obtained from or computed using railroad employee timetables and track profiles in the author's collection. In addition, each link was assigned a RAILNET line classification code and a capacity code. Capacity codes, which are discussed in detail in a subsequent section, denote segment terrain and track configuration characteristics. During attribute updating, links were split at points where the line changed from predominantly single track to double track.

The next step in the network development process was to identify the locations of interchanges between rail carriers using the Open and Prepaid Station List (OPSL), an industry tariff. Interchange nodes were selected in the TransCAD database and exported to a text file. Carriers participating in interchanges at each node were added to the file manually based upon the OPSL information.

The finished network consists of 670 nodes and 954 links. All links represent track segments, with a total of 34,251.2 route miles. No terminals were included in this model, since a detailed source of information on terminals was not available. The transfer nodes account for 444 carrier-to-carrier interchanges. The network is used by 12

Table 5-1. Link Fields in Network Data Base

Field	Added
Link ID	No
Length, miles	Yes
Average speed, mph	Yes
FRA tonnage density class	No
Link capacity class	Yes
Link type	Yes
Control system type	No
Number of tracks	No
Owning carrier alpha codes (3 fields)	No
Trackage rights carrier alpha codes (9 fields)	No
Passenger route code	No
Military route code	No

separate railroad companies. Table 5-2 summarizes route mileage by track type for each carrier. Since a sizable portion of the track in the network is used by more than one carrier, the mileage totals in the table exceed the figure given above. Figure 5-1 shows the network.

Capacity Function Calibration

The initial editing step did not include adding the travel time function parameters for each link. The scope of the study did not, of course, allow the calibration of the function individually for each line segment. Such an effort, which might involve modeling train performance on each line using a train performance calculator and then simulating operations to obtain estimates of delay at various traffic volumes, would simply have taken too long given the size of the network. An alternate approach, described below, was therefore developed.

Line classification variables

The network links were grouped into categories representing similar sets of performance characteristics. Two variables were used in this process, link capacity class and control system type.

Link capacity class. Using track profile and timetable data, line segments were cross-classified by terrain and track configuration. Both of these characteristics are known to influence line capacity.

Table 5-2. Route Mileage by Carrier and Link Type

Carrier	AAR Code	Link Type: 1-Way Single	Single	1-Way Double	Double	Total
Norfolk Southern	555	128.1	8772.5	424.6	1042.3	10367.5
CSX Transportation	712	—	10682.0	1035.5	786.0	12503.5
Burlington Northern	76	—	3201.8	—	268.2	3470.0
Union Pacific	802	—	2825.3	7.6	420.9	3253.8
Southern Pacific	721	—	1943.7	8.6	274.2	2226.5
Kansas City Southern	400	—	2191.0	0.5	10.9	2202.4
Illinois Central	350	—	1631.5	35.6	67.0	1741.7
Conrail	190	—	421.9	132.5	221.2	775.6
Florida East Coast	263	—	331.6	14.2	—	345.8
Paducah and Louisville	907	—	229.7	29.5	—	259.2
Apalachicola Northern	12	—	96.3	—	—	96.3
Meridian and Bigbee	462	—	51.0	—	—	51.0

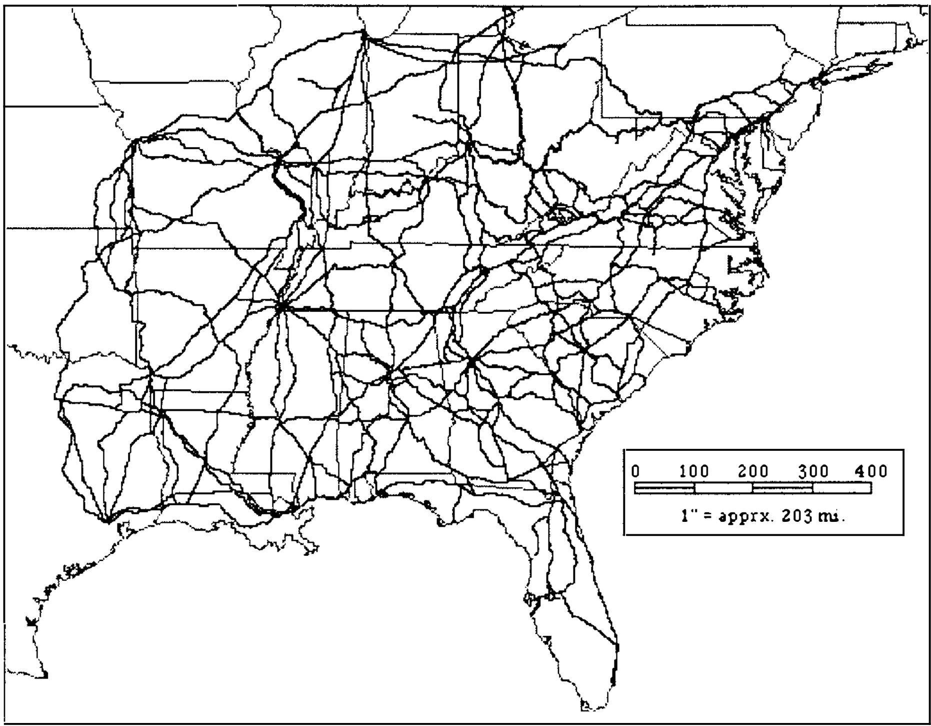

Figure 5-1. Railroad Network Used in the Case Study.

Terrain serves as a surrogate for line geometric conditions which affect both maximum train operating speeds and acceleration and deceleration characteristics. In flat terrain, track alignments have relatively little speed reducing horizontal and vertical curvature. Mountainous terrain reduces train speeds because of the generally substantial amount of horizontal and vertical curvature.

Track configuration reflects the ability of the line to accommodate train meets and passes. Double or triple track lines have a high capacity because meet delay is virtually eliminated. On single track lines, capacity increases as siding spacing decreases.

In this study, terrain assumed three discrete values: flat, hilly, and mountainous. Track configuration consisted of four categories:

a) Double or multiple track;

b) Single track, siding spacing ≤ 10 miles;

c) Single track, 10 miles > siding spacing ≤ 20 miles; and

d) Single track, siding spacing > 20 miles.

Classification of lines by terrain was done using railroad supplied speed and track alignment data. Hilly lines had repeated changes in vertical alignment, generally with grades in excess of one-half percent. Mountainous lines combined a high degree of curvature with sustained grades, normally of one percent or more. These definitions were subjectively arrived at based upon examination of the line data.

Line track configuration was derived analytically using siding information in operating timetables. In general, only sidings more than 4,000 feet long were considered

usable. For single track links, the distances between sidings or short sections of double track were summed for each link and the average computed. The link was then classified appropriately.

Control system type. The control system employed in handling train traffic is also a major influence on line capacity. Centralized traffic control (CTC) systems, which use remotely controlled signals and power operated turnouts to facilitate meets and passes, greatly reduce train delay. Automatic block signal (ABS) systems, generally incorporated into CTC but installed in a standalone fashion on many miles of railroad, allow headways to be reduced and train speeds to be increased. This increases the possible throughput and, thus, the line capacity. Under ABS, train crews normally operate siding turnouts manually, so meet and pass delay may not be alleviated substantially. Unsignalized lines have the lowest capacity because of the increased headway requirement, lower speed limits, and meet and pass delay.

The network database contained an attribute indicating whether the track segment was operated using CTC, ABS, or without signals. This was verified during the updating process.

Calibration of Capacity Functions

The classification process yielded 36 discrete control system/link capacity class combinations. Travel time function parameters needed to be developed for each one of these, still a substantial task if done using traditional techniques.

Fortunately, however, Bronzini and Sherman (1986) developed, using simulation, a family of hyperbolic delay curves in their study of railroad routing. Treating these curves as data and fitting the polynomial function to them seemed viable, since the curve parameters were provided. Bronzini's curve families are based upon combinations of control system (CTC, ABS, unsignaled), train power-to-weight ratio, region, and terrain (flat or hilly). His curves can be used provided that we assume a uniform power to weight ratio of appropriate value, select the southern region, and develop a method for estimating values for mountainous terrain.

As "datasets," the following of Bronzini's curves for CTC controlled, single track rail lines in the southern region were selected to represent each model terrain type:

a) Flat—Flat terrain, 2.5 horsepower per ton;

b) Hilly—Hilly terrain, 2.0 horsepower per ton; and

c) Mountainous—Hilly terrain, 1.7 horsepower per ton.

Since Bronzini had no mountainous type, a curve in his "hilly" family representing trains of lower power-to-weight ratio was used as a surrogate. The lower power has an effect on train performance similar to that of increased grades and curves. The curve selected had the lowest available ratio. Using CTC line curves as a start, Bronzini demonstrates how these curves may be adjusted, using coefficients, to reflect multiple track lines and other control systems.

The procedure used to fit the polynomial curve to the hyperbolic function generated data is now described. First, points on a Bronzini curve were reproduced for a

range of volumes up to maximum capacity using a Microsoft Excel spreadsheet. A heuristic search procedure was then used, based upon a recommendation by Margiotta (1995) to select values of k_1, k_2, and γ yielding the best fit. Over a range of exponent values, γ, a nonlinear regression procedure estimated values for the other coefficients. The regression procedure was forced to set the intercept to zero. The regression statistics (R2, t-statistic, etc.) were examined for each value of γ and the predicted values plotted against those generated by Bronzini's hyperbolic function. The exponent and estimated parameters which produced the best statistical fit to the observed data were selected.

The aforementioned procedure produced parameters for the polynomial function which fit the hyperbolic dataset reasonably well. Graphically, the hyperbolic and polynomial curves appeared very similar in form. Figure 5-2 shows an example of one of the fitted curves.

After developing polynomial curves for the three CTC single track configurations, the task of developing curves for the remaining configurations was addressed. Bronzini's study indicated that changes in the number of tracks and/or the signal system affects the physical capacity, *C*, as follows:

C(ABS) = 0.5* C(CTC),
C(unsignaled) = 0.32* C(CTC), and
C(double track) = 3.0* C(single track).

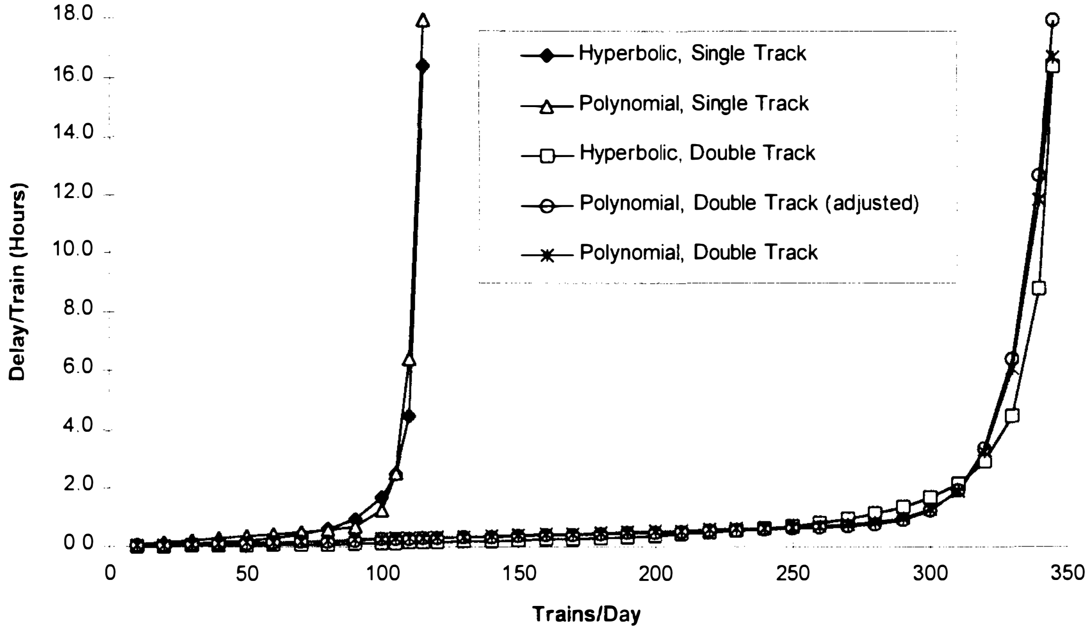

Figure 5-2. Example Delay Function Curves for CTC, Eastern Region, 1.3 HP/Ton.

Based upon the simulation work of Clarke (1982), it appears that, at least within the range of the case study values, doubling siding spacing on single track reduces physical capacity by 40 percent. In the Bronzini study, base single track siding was 10 miles.

Given the above factors, it is possible to modify the calibrated polynomial equations to account for the complete range of factors in the case study. Multiplying the physical capacity by the appropriate combination of factors is the first step. The second necessary step, developed through experimentation, is to divide the k_1 parameter by the same combination of factors. Figure 5-2 also shows a curve fit to the equivalent hyperbolic curve in this manner. The fit seems quite satisfactory, and is very close to a polynomial curve fitted using the regression approach.

Table 5-3 provides the calculated function parameter values for each terrain/track configuration combination for CTC lines. Tables 5-4 and 5-5 provide the function parameters for, respectively, ABS and unsignaled lines.

Final Network Production

Following the calibration of the link travel time functions and the completion of basic network editing, the network link and node data were exported from the GIS into text files. These files were read into databases created by Microsoft's FoxPro 2.5® relational database management system. FoxPro was deemed more flexible for subsequent network processing steps involving programming. In addition, the FoxPro database could, in turn, be exported into files having precisely the format expected by NETBLD. Based upon the signal system and line capacity class, appropriate travel time

Table 5-3. Line Capacity Class Codes, CTC

Code	Description	C	K_1	K_2	γ
1	Flat, double track	258.0	0.001723	29.52711	25.0
2	Flat, single track, sidings ≤10 mi.	86.0	0.005269	29.52711	25.0
3	Flat, single track, sidings >10 mi., ≤20 mi.	51.6	0.008615	29.52711	25.0
4	Flat, single track, sidings > 20mi.	31.0	0.014359	29.52711	25.0
5	Hilly, double track	216.0	0.003291	25.64741	25.0
6	Hilly, single track, sidings ≤10 mi.	72.0	0.009872	25.64741	25.0
7	Hilly, single track, sidings >10 mi., ≤20 mi.	43.2	0.016453	25.64741	25.0
8	Hilly, single track, sidings > 20mi.	25.9	0.027422	25.64741	25.0
9	Mountainous, double track	204.0	0.000750	48.53455	32.5
10	Mountainous, single track, sidings ≤10 mi.	68.0	0.002249	48.53455	32.5
11	Mountainous, single track, sidings >10 mi., ≤20 mi.	40.8	0.003748	48.53455	32.5
12	Mountainous, single track, sidings > 20mi.	24.5	0.006246	48.53455	32.5

Table 5-4. Line Capacity Class Codes, ABS

Code	Description	C	K_1	K_2	γ
1	Flat, double track	129.0	0.003446	29.52711	25.0
2	Flat, single track, sidings ≤10 mi.	43.0	0.010338	29.52711	25.0
3	Flat, single track, sidings >10 mi., ≤20 mi.	25.8	0.017230	29.52711	25.0
4	Flat, single track, sidings > 20mi.	15.5	0.028717	29.52711	25.0
5	Hilly, double track	108.0	0.006581	25.64741	25.0
6	Hilly, single track, sidings ≤10 mi.	36.0	0.019744	25.64741	25.0
7	Hilly, single track, sidings >10 mi., ≤20 mi.	21.6	0.032907	25.64741	25.0
8	Hilly, single track, sidings > 20mi.	13.0	0.054844	25.64741	25.0
9	Mountainous, double track	102.0	0.001499	48.53455	32.5
10	Mountainous, single track, sidings ≤10 mi.	34.0	0.004497	48.53455	32.5
11	Mountainous, single track, sidings >10 mi., ≤20 mi.	20.4	0.007495	48.53455	32.5
12	Mountainous, single track, sidings > 20mi.	12.2	0.012492	48.53455	32.5

Table 5-5. Line Capacity Class Codes, Unsignaled

Code	Description	C	K_1	K_2	γ
1	Flat, double track	82.6	0.005385	29.52711	25.0
2	Flat, single track, sidings ≤10 mi.	27.5	0.016154	29.52711	25.0
3	Flat, single track, sidings >10 mi., ≤20 mi.	16.5	0.026923	29.52711	25.0
4	Flat, single track, sidings > 20mi.	9.9	0.044871	29.52711	25.0
5	Hilly, double track	69.1	0.030850	25.64741	25.0
6	Hilly, single track, sidings ≤10 mi.	23.0	0.019744	25.64741	25.0
7	Hilly, single track, sidings >10 mi., ≤20 mi.	13.8	0.051416	25.64741	25.0
8	Hilly, single track, sidings > 20mi.	8.3	0.085694	25.64741	25.0
9	Mountainous, double track	65.3	0.002342	48.53455	32.5
10	Mountainous, single track, sidings ≤10 mi.	21.8	0.007027	48.53455	32.5
11	Mountainous, single track, sidings >10 mi., ≤20 mi.	13.1	0.011711	48.53455	32.5
12	Mountainous, single track, sidings > 20mi.	7.8	0.019519	48.53455	32.5

function parameters were added to each link record. In addition, certain other fields were calculated, such as the free flow travel time, and carrier alpha codes were translated to AAR numeric carrier codes required by NETBLD.

Traffic Flow Data

The next step in the case study was to develop the commodity flow matrixes defining the demand patterns. This process required two separate steps. The first step was to develop flows of actual commodities. The second step, equally important from a train volume standpoint, was to compute flow demands for empty cars. Railroad traffic consists of loaded cars and empty cars being returned for loading. Omitting empties will cause total train volumes to be dramatically understated in most cases.

Commodity Data

This section describes the selection of a traffic data source and the steps involved in producing the demand matrixes for carload traffic.

Source Evaluation

Analysts examining national railroad freight flows have relatively few options for obtaining demand data. Railroad companies, of course, maintain traffic flow databases, but are, in general, reluctant to release this data to the public. The author contacted several companies and was politely declined. The Census Bureau is presently completing a freight commodity flow survey which covers all modes. This was not yet available,

unfortunately, at the time this research was conducted. At present, the only public source of national railroad traffic data is the ICC Carload Waybill Sample (CWS).

Railroads terminating over 4,500 cars per year or five percent or more of a state's traffic provide a fraction (generally about one percent, but varying by the type of traffic) of their waybills to the ICC. From an analysis standpoint, waybills contain quite detailed data on the shipment. The waybill includes fields, for example, describing the shipment origin and destination, commodity, quantity, rate, revenue, freight car type and characteristics, intermodal characteristics, participating carriers, and interchange points.

After receiving a waybill, the ICC processes it to identify and correct errors and to derive some additional data elements. The agency then uses statistical procedures to expand the waybill's data to reflect the entire population of shipments it represents. Although there are some criticisms of this procedure from a statistical standpoint, it has been performed for many years and is well accepted. The CWS is an entire year's collection of processed waybills. The original fields for each waybill are preserved, so the analyst may always use an alternative expansion procedure. For additional information, the reader can contact the AAR, which publishes several documents describing the CWS and its contents.

The ICC makes the complete CWS available to Federal and state government agencies, where it is commonly used for policy analysis. The Commission cannot, by law, release this version to non-government entities. In its raw form, the file contains data considered proprietary to the carriers. To alleviate these confidentiality concerns,

the ICC creates a processed version of the waybill file in which fields containing sensitive data are masked. This dataset, called the public use version of the CWS, is available at no cost on several compact disks produced by the Bureau of Transportation Statistics of the U.S. Department of Transportation.

Several vendors offer commercial products based upon an enhanced version of the public use CWS. These vendors attempt to use other sources of data to restore the purged fields. While no doubt better than the public use CWS, these products were too expensive for this study's budget and could not be used.

For this case study, therefore, the CWS was the only feasible railroad traffic flow database. The most current release contains waybills for 1992. This establishes the time frame for the base level analysis in the case.

Processing

The raw waybill file contains individual shipment records, while the RAILNET model requires O-D flow matrixes for each commodity. A scheme for processing the waybill data to produce the desired matrixes was obviously necessary. Evaluation of the documentation supplied with the dataset and a cursory evaluation of some typical waybills in the file revealed a number of issues which the processing scheme needed to address. These were:

a) Reduction of the size of the waybill file to simplify processing. The complete file, which contained almost 98 megabytes of data comprising 396,670 records, was unwieldy to work with.;

b) Aggregation of individual commodities into categories having similar characteristics from a cost and service standpoint. The solution algorithm performs one iteration for each commodity during Step 1. Processing time increases, therefore, with the number of commodities.; and

c) Assessment of the impact of records containing purged values and, if practicable, correction of these records.

An examination of the CWS revealed a core set of data elements that appeared relevant to the case study. Table 5-6 describes these fields. Appendix C contains a listing of the complete CWS record structure.

Shipment origin and destination in the CWS are provided at the Business Economic Area (BEA) level. Business Economic Areas are multicounty reporting regions for economic activity. Each of the 181 BEA regions within the continental U.S. is centered around a major urbanized area. Although the actual shipment endpoints lie somewhere within BEA region boundaries, our flow analysis must use some arbitrary point within each BEA region as a surrogate endpoint.

File reduction. To reduce the CWS file to a workable size, records for shipments not originating or terminating within the region were deleted. Because of the geography of the study region, it seems a reasonable assumption that through traffic is minimal.

From the CWS documentation file, omission of the origin and/or destination BEA region appeared to be a major impact of data cleansing. The public use CWS omits a shipment's BEA region when fewer than three freight stations in the BEA region handle

Table 5-6. Characteristic Fields in the CWS File

Columns	Type	Description
16-19	A	AAR car type
20-23	A	AAR car mechanical designation
26-28	A	TOFC/COFC plan code
34	A	TOFC/COFC unit type
36-40	N	Standard Transportation Commodity Code
84	N	All rail/intermodel flag
89-92	N	Estimated short line miles
103	N	Number of interchanges
104-106	N	Origin BEA region
107	N	Origin ICC freight rate territory
126-128	N	Destination BEA region
129	N	Destination ICC freight rate territory
136-138	N	Nominal car capacity, thousands of pounds
139-142	N	Tare weight of car, in hundredweight
216-221	N	Expanded carloads
222-230	N	Expanded tons
242-247	N	Expanded number of TOFC/COFC units

the shipment commodity and there are not at least two more such freight stations than railroads in the BEA. To assess the magnitude of the impact, the origin and destination freight rate territory fields, which are not altered, were used to select all shipment records having one or both endpoints in Southern Territory. This subset was then augmented by selecting all records having an origin or destination in the BEA regions for Norfolk, Virginia, Washington, DC, and Baltimore, Maryland. These lie within the study boundary, but just outside of Southern Territory. This resulted in a file containing 127,627 shipment records. To simplify further processing, these records were loaded into a FoxPro database.

Commodity aggregation. The next step in the editing process was to segregate the records by commodity group. Each CWS shipment record contains a Standard Transportation Commodity Code (STCC) describing the commodity being transported. The STCC system has multiple levels, with each level being increasingly specific. Codes in the CWS contain five digits; the test case did not require this level of specificity. In most freight studies, commodities can be grouped into categories having similar cost, physical, and service characteristics. For the case, commodities were aggregated into six groups, with six additional groups added to represent empty cars. Table 5-7 lists these groups.

The rationale employed in selecting these groups is as follows. Intermodal traffic, encompassing a variety of commodity groups, was of specific interest in the case. Such traffic frequently consists of high value goods and is, accordingly, given expedited

Table 5-7. Commodity Grouping for Case Study

Commodity Description	ID	STCC
Coal	11	11xxx
Non-coal bulk materials	12	01xxx, 14xxx, 32xxx, 241xx, 287xx, 40xxx
Automobiles and auto parts	37	371xx
Other carload freight	39	(misc.)
Intermodal, TOFC/COFC	40	(any)
Intermodal, double stack	41	(any)
Empties, coal	99	n/a
Empties, non-coal bulk	98	n/a
Empties, auto traffic	96	n/a
Empties, other carload	97	n/a
Empties, TOFC/COFC	95	n/a
Empties, doublestack	94	n/a

service. The distinction between doublestack and conventional intermodal was made because doublestack can only be handled on that portion of the network having adequate clearances.

Automobiles and automobile parts are also high value commodities which the railroads handle in priority service. The equipment used cannot be handled on all portions of the network because of clearance restrictions.

Coal is of interest because of its high volume throughout much of the study region. Coal is a relatively low value commodity which is not highly service sensitive. Much coal moves in trainload lots between mines and electric generating stations.

The category of other bulk materials accounts for bulk commodities other than coal. These include grain, sand and gravel, cement, wood chips, and certain agricultural chemicals. These materials move in substantial volumes, although not at the magnitude of coal. Service requirements are generally similar.

The other carload freight category includes all remaining carload commodity groups. This traffic is generally of higher value than the bulk materials and, therefore, is more service sensitive.

Using query commands, the original dataset was divided into six subsets, each containing shipments for a particular commodity group.

Data editing. The commodity group databases were examined to determine the number of records affected by missing origin and/or destination BEA regions. Substantial numbers of records were affected by such omissions. The records with either

field missing were stripped out of the commodity files and set aside. Records with both fields set to zero were discarded, since without one known endpoint, the record was not correctable. Attempts were then made, for each commodity group, to determine the correct value for missing endpoints in the remaining records.

For some records, given the commodity code, shipment distance, and a known set of unused BEAs for the missing field, it was possible to estimate the appropriate value using additional data sources. Listings of automobile or truck plants, for example, helped fill in missing values for some automobile industry shipments. Similarly, utility industry data on steam generating stations helped fill in missing destinations on coal waybills.

If the specific BEA could not be determined, it was possible for some shipments to determine the likely direction. In this case, an appropriate gateway city could be substituted for the actual BEA region. In this manner, the shipment would take the correct route within the study region, even if the actual endpoint was unknown. Shipments from points in Florida, for example, to unknown locations in Southwestern Territory would most likely enter or leave the study region via the New Orleans gateway.

Intra-BEA flows were also identified during the correction process. Attempts were made to identify specific origin and destination points in the network for these shipments. This was successful in some cases, as in the movement of phosphates from mines in central Florida to the port of Tampa.

These correction procedures required much manual effort to implement. After completion, a large number of waybills were still uncorrected. These bills accounted for

1,608.255 carloads, or 48,048,713 tons of cargo. Nevertheless, the completed commodity datasets still accounted for a substantial amount of freight, as shown in Table 5-8. Whether the expanded carload data represents accurately the number of cars handled and the commodity tonnage is hard to determine. The AAR data reveal carloads by commodity, but not at the regional level. The flow datasets contain a mixture of originating and terminating carloads, further complicating an objective assessment. Finally, the waybill expanded tonnage is the billed tonnage, which may be higher than the actual tonnage.

O-D matrix generation. Following the correction process, the waybills for each commodity group were aggregated by OD-pair to develop a BEA to BEA flow matrix. Following this, translate tables mapping each BEA region to a network node were created. Using these tables, sparse O-D flow tables containing network nodes instead of BEA regions were created. These tables were in a form usable to RAILNET.

The procedure used to map BEA regions to network nodes is as follows. First. for each BEA region served by the modeled network—and this includes regions in the network area outside of the study boundaries—a network node representing the largest urban area named in the BEA region was selected. The node selection criteria were that it lie near the center of railroad activity in the area and that it be accessible to all of the carriers serving the urban area. If the urbanized area originates and terminates most shipments, this is probably an acceptable approximation. Whether this is uniformly the

Table 5-8. Commodity Group Tonnages

ID	Description	Carloads	Tons
11	Coal	1,891,092	180,378,074
12	Non-coal bulk materials	1,727,013	158,610,367
37	Automobiles and auto parts	39,280	904,760
39	Other carload freight	546,590	37,556,739
40	Intermodal, TOFC/COFC	605,946	9,147,968
41	Intermodal, double stack	198,112	3,095,244
99	Empties, coal	1,891,092	45,094,534
98	Empties, non-coal bulk	1,727,013	51,230,994
96	Empties, auto traffic	39,280	1,105,351
97	Empties, other carload	546,590	14,385,265
95	Empties, TOFC/COFC	605,946	5,280,328
94	Empties, doublestack	198,112	962,148

case is debatable, but, given the difficulty in matching the waybill data to other secondary data sources to identify specific shippers, seemed the only practical option.

BEA regions beyond the network were mapped to major rail gateways located along the periphery of the network. This was done subjectively based upon the proximity of the BEA region to the nearest network gateway. Database tables containing the BEA region to network node mappings were created. Separate mappings were required to account for commodity characteristics. Coal flows, for example, were directed to utility locations within the BEA region, rather than to an arbitrary centroid location. If more than one utility was found within the BEA region, the final flow table was adjusted to distribute the flow to nodes representing each plant. Intermodal flows were dispatched to locations of intermodal terminals within the region.

Empty Car Flows

A flow analysis of rail traffic is not complete without accounting for the movement of empty cars within the network. Empty cars, unless being shipped as products, do not move under waybills. The CWS, therefore, does not describe empty movements. In the railroad environment, it is not uncommon for one-half of the railcars in a general freight train to be empty. In unit trains, all cars are normally empty on the backhaul. Neglecting empties will seriously understate the total cars, and therefore, trains within the network.

Empty Return Rule

Because we have no data on empty movements, methods for synthesizing the movement of empty cars must be developed. One simple rule is for empty cars to move from shipment destination back to the origin. Many cars in dedicated service follow this pattern. In the case, flows for empty auto, general freight, coal, and bulk cars were generated using this rule. General freight cars and many non-unit bulk cars might be moved empty in the direction of the nearest available load. Since the flow volumes consist of different commodity and car types, the reverse movement rule seemed to be sensible.

Empty Distribution Rule

In the intermodal arena, it is assumed that cars can be sent to any demand center, since the cars and containers are not commodity specific. The rule developed in generating empty flows is based on minimizing the cost of distributing the empty cars from locations of surplus to locations of demand. For each centroid node, the inbound and outbound tonnages are calculated. If the outbound flow exceeds the inbound, the node demands empty cars. If the inbound flow exceeds the outbound flow, the node produces empty cars. For each such node, the surplus (deficit) tonnage is converted into an equivalent quantity (demand) of empty cars.

Given a series of supply and demand nodes, the problem is to allocate at least cost quantities from supply nodes to the various demand nodes. This is the well-known classic transportation problem, which is a bipartite assignment problem described in most

texts on linear programming. The flow calculation provides the quantities. Appropriate costs for each supply-demand pair are needed. A modification of RAILNET's source code yielded a program which would read the list of supply and demand nodes and compute costs based upon the shortest marginal cost paths between node pairs at zero flow volumes. Thus, marginal cost is equal to average unit cost. The path costs were then used in the formulation and solution of the transportation problem. The LINDO mathematical programming package on the University of Tennessee Computing Center VAXcluster actually solved the problem. Resulting flow volumes were entered into a RAILNET compatible flow table.

The flow tables for each of the commodity groups were copied to a single file for ultimate use in the case study.

Carrier Costing

The next stage of the case study was to estimate costs to use in the coefficients of the line and transfer cost functions. These coefficients should incorporate appropriate cost elements based upon a typical engineering type cost analysis. It is more desirable to use published costs from previous studies rather than to initiate a primary data collection effort, which was outside of the scope of the dissertation. These costs are general to the industry, although, where available, figures for the eastern region of the U.S. were selected since the case addresses this region. Given the sensitive nature of such data,

carrier specific cost data collection would probably have introduced an additional level of difficulty to the problem.

Generalized Commodity Costing

One of the features of the model formulation is its ability to accommodate different costs by carrier, commodity, and facility. Differences in costs between carriers are ignored in this model because of the difficulty in carrier specific data collection. Costs were varied by commodity, however, to reflect differences in service demand, equipment costs, and operating costs. The methodology employed in computing costs is now addressed.

Mileage-Based Costs

One term in the arc cost function determines the commodity cost component related to mileage traveled. Patton (1992) provides some synthesized cost data in his handbook on railroad costing. These data, which reflect actual industry costs for the CWS year, were used as the basis for allocated costs per gross ton-mile in the categories of administration, operations support, maintenance of way, maintenance of equipment, and overhead. Modifications to Patton's numbers were made to incorporate specific costs for certain commodities. Progressive Railroading (1995b), for example, provided detailed maintenance costs for utility coal hoppers.

Time-Based Costs

The second arc cost term computes costs based upon a function of train operating time. Four cost categories were included in the coefficient for this term: crew costs, fuel, equipment costs, and a commodity value.

Crew costs were computed using the wage rates given by Patton. The use of an hourly wage rate is slightly unrealistic, since crews are actually paid based upon the greater of a mileage or time basis. For this model, two-person crews were assigned to each train.

Fuel costs per train hour were determined using data given by AAR (1995). Calculations using various tables in the publication yielded an average nationwide fuel cost of $117.44 per train hour for 1992.

Hourly equipment cost figures were determined industry data sources. Locomotive hourly capital costs were based upon two 4,000 horsepower units per train, each costing $1.4 million in 1992. The total hourly locomotive capital cost of $100.20 assumes a 15-year lifetime, a 10 percent salvage value, a capital cost of 11.4 percent, and 3,500 operating hours per year. These assumptions follow Bronzini (1986) and AAR (1995).

The AAR tables (1995) state that the average freight car age in 1992 was 19.5 years. For 1975, the tables give representative prices for each car type. Representative car hourly costs are then taken from the car hire tables in The Official Railway Equipment Register (1994). Table 5-9 shows the rates used in the case study.

Table 5-9. Car Hourly Costs

Car category	Cost/hour ($)
Coal hopper	0.57
Bulk (open hopper, gondola, covered hopper)	0.32
Multi-level rack car	1.00
General freight (boxcar, tank car, etc.)	0.38
TOFC/COFC flat	0.58
Well car (doublestack)	1.51

Given the mathematical formulation of the problem as a cost minimization, the commodity value component ensures that high value freight receives high priority, as in actual railroad practice. Since the actual value of each commodity is unknown, revenue per train hour is used as a surrogate. High unit revenue commodities are assumed to have high value. Since the model is a minimization, high cost (revenue) commodities will receive priority in the assignment and will flow on the fastest, least congested routes. Lower cost commodities may incur higher levels of circuity and congestion and will then be diverted to alternate routes. Friesz et al. (1981) discuss the use of such an approach. The revenue per average train hour for each commodity class is calculated using the AAR statistic on average ton-miles per train hour and the data within the CWS on revenue, tonnage, and shipment mileage.

The cost figures obtained after performing the above calculations yielded lower than expected values for intermodal traffic. This could be due to the extreme pressure exerted on intermodal rates by truck competition. Railroads obviously, however, expedite intermodal shipments to retain the business. Table 5-10 summarizes the final cost coefficients.

Transfer Costs

The cost of a transfer is provided on a per-car basis. No actual data were available to allow for estimating actual transfer costs. To reflect actual railroad practice, the interchange costs should be high enough to prevent excessive numbers of interchanges for each specific movement. Accordingly, the average transfer cost needs to be high

Table 5-10. Cost Coefficient and Train Size Data

ID	Description	Cost/ Ton-Mile	Cost/ Train-Hr.	Transfer Cost/Car	Car Tare	Car Payload	Train Tonnage
11	Coal	0.00526	2,602.89	310.30	29.3	100.0	11,637
12	Non-coal bulk materials	0.00705	3,513.40	647.55	30.2	93.5	7,422
37	Automobiles and auto parts	0.00705	8,251.56	1914.22	52.9	43.3	4,810
39	Other carload freight	0.00705	3.618.00	607.70	33.4	87.2	7,960
40	Intermodal, TOFC/COFC	0.00705	3,280.06	721.06	38.9	33.6	3,625
41	Intermodal, double stack	0.00705	3,439.24	1898.16	82.6	138.6	4,424
99	Empties, coal	0.00526	291.71	40.00	0	29.3	2,637
98	Empties, non-coal bulk	0.00705	177.40	20.00	0	30.2	1,983
96	Empties, auto traffic	0.00705	208.20	20.00	0	52.9	2,645
97	Empties, other carload	0.00705	183.03	20.00	0	33.4	2,000
95	Empties, TOFC/COFC	0.00705	187.20	20.00	0	38.9	2,645
94	Empties, doublestack	0.00705	188.40	20.00	0	82.6	2,230

enough to penalize interchanges so that they will only occur when the penalty for circuity exceeds the interchange cost or when necessary for route continuity. The assumption was made that interchanges at general locations induce an average twelve hour delay for each car. The cost per car hour was estimated by dividing the revenue cost per train hour by the number of cars in the average train. Multiplying this by twelve hours yielded the cost per interchange for each commodity group. Costs at specific interchanges can be changed at run time to reflect run through agreements or expedited interchange.

Train Sizes

To equate flows in tons to equivalent trains and cars, statistics on average car tare weight, payload weight, and trailing train tonnage limits are needed for each commodity. The average tare and payload weight were computed for each commodity group using the waybill records. Train weights were estimated based on industry practices in the region, such as the 90-car unit coal train and the 50-car auto rack train.

Empty train weights equal the product of the number of cars in the equivalent loaded train and the tare weight of the car type. The tare is set to zero in the parameter table and the payload per car then becomes the empty car weight. This assures that the costing logic in the model will function properly.

Final tare, payload, and train weights for the commodity classes are provided in Table 5-10.

Facility Cost Adjustment

Costs on specific facilities were adjusted to reflect changes in costs caused by track profile, operating practices, commodity restrictions, or facility type.

All double track line segments had the maintenance of way component of the distance-based cost coefficient increased by 50 percent to account for higher maintenance costs associated with the additional track and structures. All other cost components were left unchanged.

All line segments having insufficient clearances for equipment used for a commodity group received extremely high cost penalties for these commodities. This effectively prevented the routing of these commodities over the segments. Restrictions were imposed where applicable on loaded doublestack well cars, loaded conventional TOFC cars, and multi-level rack cars. Empty doublestack well and TOFC cars were not penalized. Restricted clearance segments were identified using Railway Line Clearances (1994), an industry publication.

Line segments requiring helper locomotive operation received increased hourly fuel and crew costs. In general, train costs for helpers were increased to account for an additional set of locomotives, a train crew, and fuel. Only trains for commodity groups using helpers were penalized. As a rule, coal and bulk commodities always received helpers.

Model Operation

Given the assumptions and techniques described above, data tables were developed for the base scenario. Most tables were created in data base format for ease of manipulation. When all data had been entered, the data base tables were exported to formatted text files. NETBLD and COMMODTY were used to process the network and commodity files. Operation of each of these programs took only a few seconds. Correction of a few minor errors found in the case files required rerunning the programs. This did not prove to be an inconvenience.

After processing of the case files was complete, RAILNET then solved the base scenario. Initial execution of the program was performed on an IBM-compatible personal computer equipped with an 80386 microprocessor operating at 40 MHz, an 80387 math coprocessor, a 128 kilobyte memory cache, and 8 megabytes of random access memory (RAM). The computer operating system was MS-DOS 6.20.

RAILNET uses a 32-bit memory model which allows use of RAM above the DOS 640 kilobyte limit. The problem workspace occupied 745,610 bytes. Since this is well within the RAM capacity of the computer, fixed disk capacity and speed did not affect the processing time.

Complete processing of the test case on the 80386 equipped system took about two hours. An exact time was not obtained, since RAILNET contains no timer and the author performed other tasks while waiting for program completion. The run time for a problem of this size seems reasonable, and is certainly small enough to allow several

cases to be solved during a work day. Subsequent tests revealed that a personal computer equipped with an 80486 microprocessor running at 66 MHz could solve the problem in about eleven minutes. This time is astounding in comparison to the 80386, and should be encouraging in view of the general prevalence of the 80486 in new personal computers. The 80486 machine ran PC DOS version 6.3, revision 0.

Upon completion, RAILNET produced several output files containing problem results. A run listing file detailed flows by link, arc, and commodity. The listing also provided link travel times, delays, and train volumes. Another output file contained gross tonnages, train volumes, and delays by link. The file had a structure compatible with TransCAD 2.1 table import files. Thus, the file contents could be imported into the network GIS for manipulation and display.

Results

RAILNET solved the case in three iterations of the algorithm. The stopping criteria was a maximum of ten iterations or a difference in the objective function in successive iterations of 0.01 percent or less.

The initial objective function value was $13.081 billion, while the solution value was $12.911 billion. These numbers seem large, but the reader should keep in context that the flows represent one year and that the commodity values are a large portion of this total. The solution value represents a $169.6 million reduction from the original assignment. While a large number in absolute terms, the difference is relatively small as

a percentage. The reason for this appears to be the relatively uncongested nature of the flows in the network.

Overall Observations

The results of the case can be interpreted by studying the output link and transfer volume files. Intuitively, however, it is much easier to visualize the results graphically. The link volume output file was loaded into a TransCAD table and imported into the link database in the GIS. The program was instructed to plot link thickness as a function of tonnage. TransCAD selected six discrete ranges, based upon a specified scaling factor of 30 million gross tons:

- 0 to 6 million;
- 6 million to 12 million;
- 12 million to 18 million;
- 18 million to 24 million;
- 24 million to 30 million; and
- 30 million and above.

Figure 5-3 shows the overall solution flow pattern. Figure 5-4 provides a more detailed view of the study region, except for Florida. Based upon the visual evaluation, the solution flow pattern seems to reflect actual railroad traffic patterns, at least in terms of relative densities. On closer examination, however, the absolute magnitude of the density on virtually all links is dramatically understated.

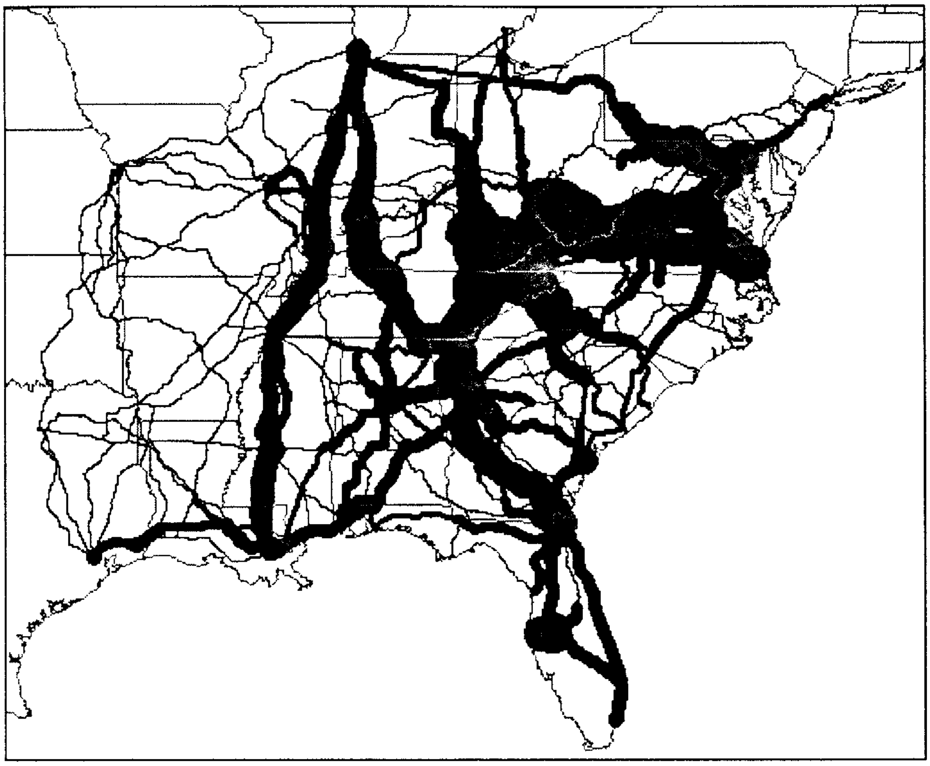

Figure 5-3. Flow Pattern for Case Study.

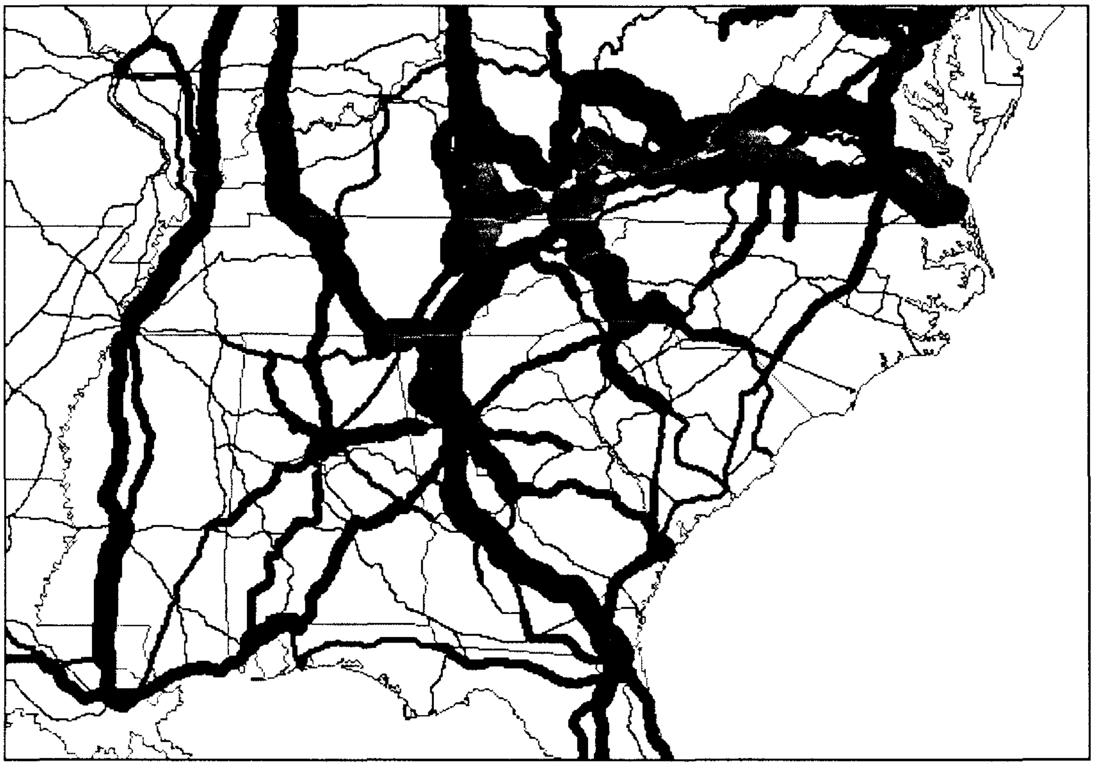

Figure 5-4. Enlarged View of Flows in Study Region.

Although actual gross tonnages for most lines were not known, the FRA mainline class can be used as a surrogate. FRA mainline classes are categories based upon annual gross tonnages. Consider, for example, lines with tonnages meeting the criteria for FRA "A" mainline, that is, annual gross tonnages greater than or equal to 20,000,000 tons. Heavy lines in Figure 5-5 represent case study segments meeting the threshold, while light lines represent the remaining universe of lines actually meeting "A" mainline levels. Similarly, Figure 5-6 shows mainlines having annual gross tonnages of 5,000,000 tons or more. In both instances, the case substantially understates tonnage volumes, although predicted mainlines do correspond, with several exceptions to be noted, to actual ones. The underlying reason for the substantially low tonnage can only be the invalid or missing waybill data, especially since volumes are low throughout the study region.

The high tonnage lines shown in Figures 5-3 and 5-4 seems to closely reflect coal traffic patterns. In other words, lines having a heavy volume of coal in the actual system show high densities in the case. Lines having a high volume of merchandise traffic are substantially below actual tonnage levels in the case. This would seem to indicate that coal waybills are more complete than waybills for other commodities.

As an example, consider the actual and predicted flows for railroads in the Knoxville, Tennessee area as shown in Figure 5-7. Although actual volumes shown in this figure are for various years during the period 1989-1993, these should be still be fairly representative of 1992 case. Norfolk Southern's north-south mainline, shown on the western edge of the figure, is a major trunk line linking Cincinnati and St. Louis with

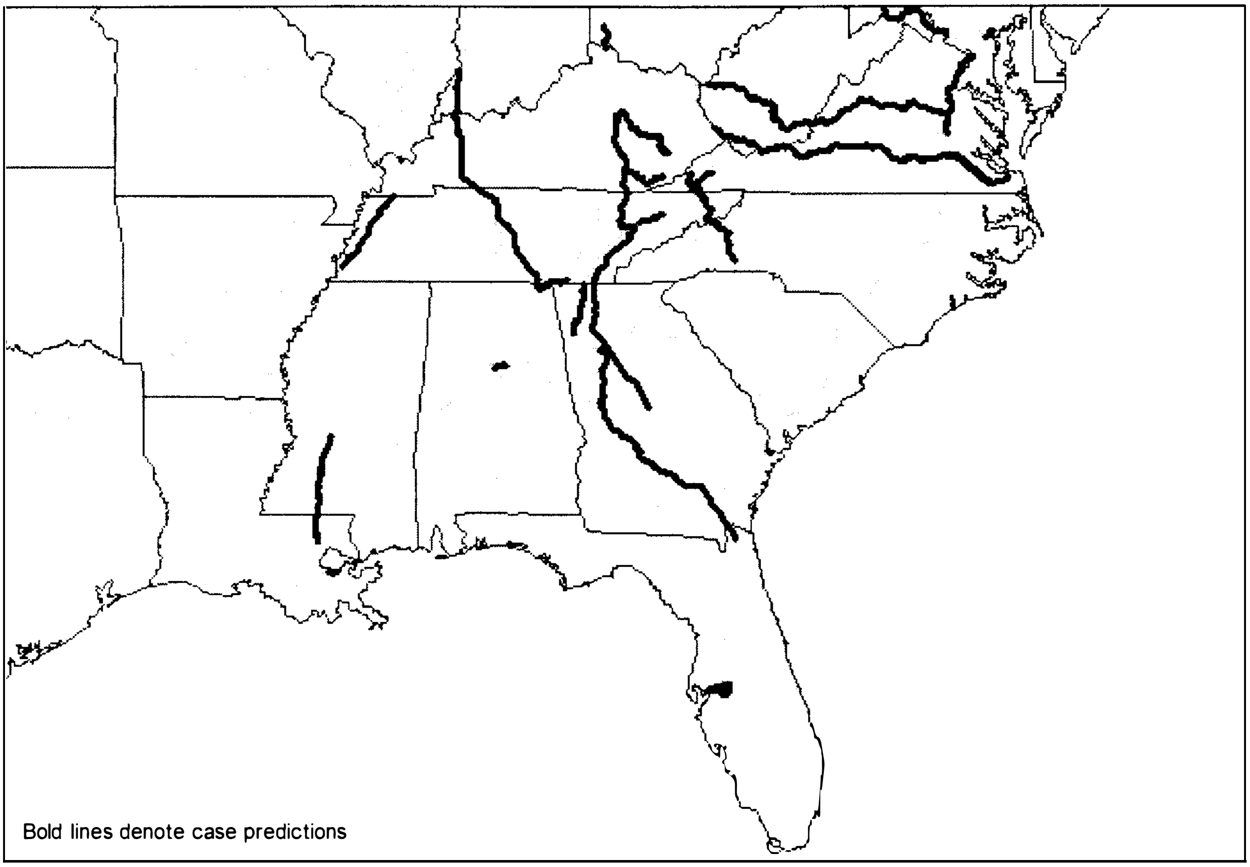

Figure 5-5. FRA "A" Mainlines—Actual Versus Case.

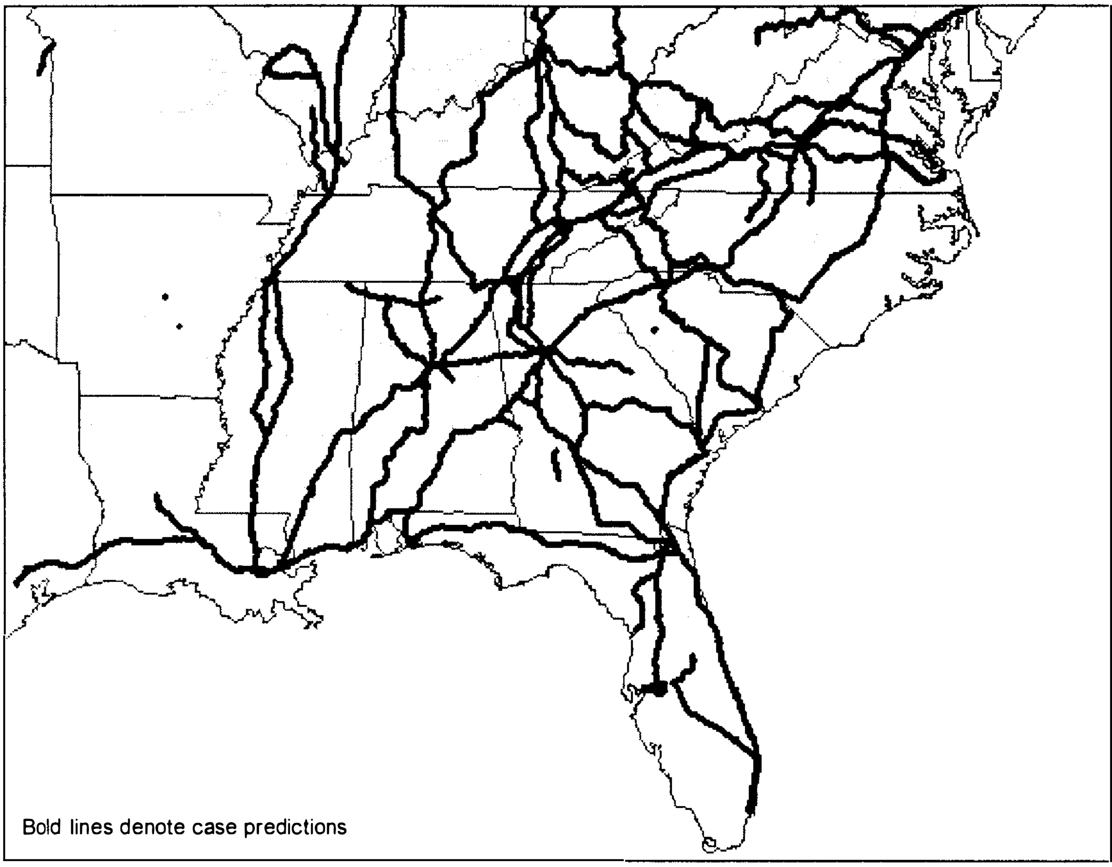

Figure 5-6. All Mainlines—Actual Versus Case.

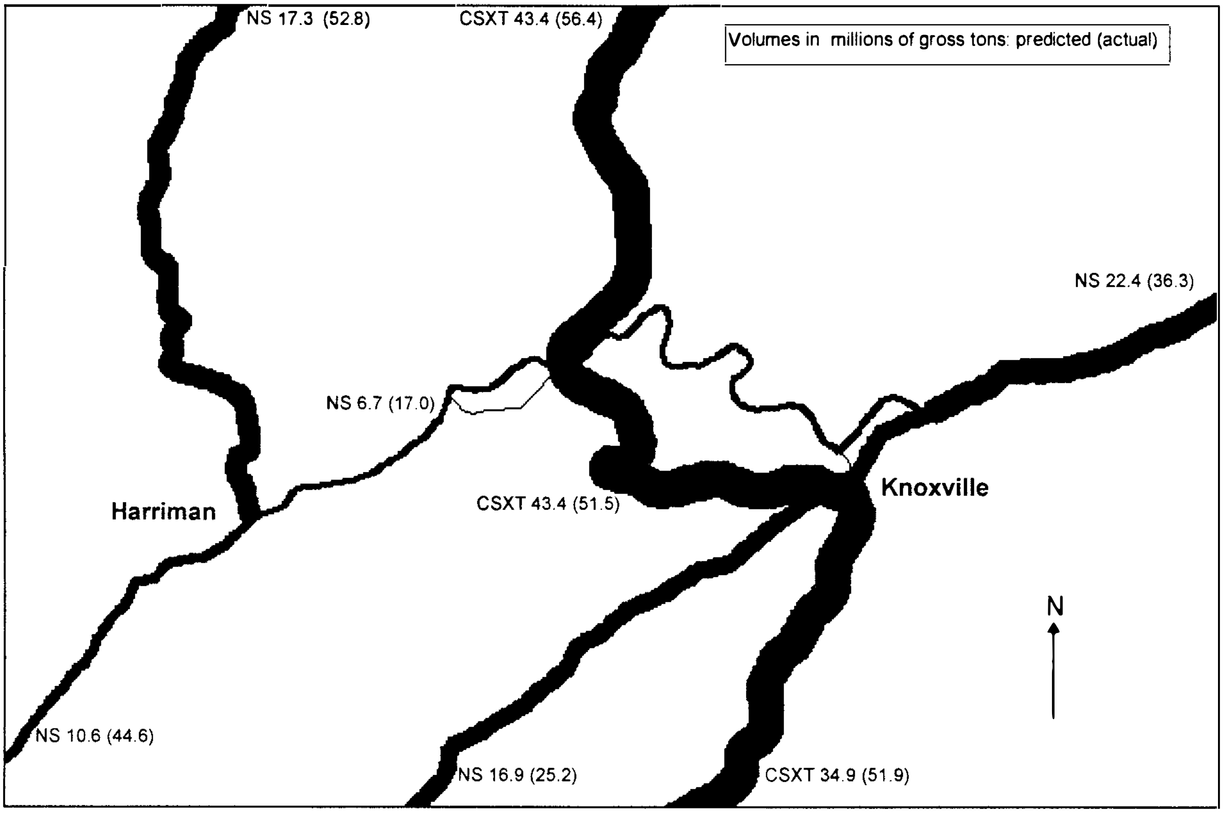

Figure 5-7. Predicted Versus Actual Volumes—Knoxville, Tennessee Area (1992).

Atlanta and points south. This route handles little coal, but is a major merchandise corridor. The model assignment underrepresents traffic by up to 76 percent on the displayed portion of the route. The north-south CSX Transportation route through Knoxville also connects Cincinnati and Atlanta, but this line has substantial coal traffic because it is the primary rail corridor serving the Eastern Kentucky coal fields. Predicted traffic volumes on this line are also low, but only by a maximum of 32.8 percent. Norfolk Southern's Bristol-Chattanooga mainline through Knoxville also has a substantial coal volume. The model underpredicts traffic on this route by a maximum of 38.3 percent.

Given that the model appears to severely underrepresent gross annual tonnages on most of the rail network, what are the likely reasons? Five possibilities may be advanced:

- The CWS data set does not adequately reflect rail traffic;
- Through rail traffic within the study region is significant;
- Tonnage of empty cars is understated;
- Model flow assignments do not reflect actual railroad practice; and
- Tonnages accrued by local and switching trains are significant.

The first reason seems to have the greatest likelihood of being correct, and arguments can be made to refute the others as general causes. Through rail traffic should have little or no influence on flows on the eastern and southern edges of the region, yet tonnages on routes in these areas are no closer to actual levels than in any other portion of the network. Tonnage of empty cars may be somewhat understated, but under no

circumstances can the author envision empty cars accounting for such severe deficiencies in tonnage. Links receiving significant quantities of flow do reflect actual mainlines, as indicated by the FRA line class data, so the model flow assignments, in general, appear reasonable. Finally, the model does not account for switching and local train tonnage. Such tonnages should, however, be small except in major urban areas. In the case study, they cannot reasonably explain the missing volume.

Local Analysis

Because the CWS was not able to support realistic tonnage volumes in much of the network, the case study could not address actual capacity issues, including intermodal service levels. With the low traffic volumes, very few facilities experienced any significant delay. The maximum average delay, for example, was only 33.3 minutes.

CSX Coal Corridor

The model did, however, yield some interesting results in specific sections of the network which had heavy coal flows. Sizable volumes of coal originate on CSX Transportation and Norfolk Southern lines in southern West Virginia, western Virginia, and eastern Kentucky coal fields. In the study region, much of this coal flows in volume to electric utilities in the Sun Belt states of South Carolina, Georgia, and Florida. A high volume north-south coal flow corridor, with multiple route possibilities for each carrier, lies between the mining areas and central Florida.

CSX Transportation operates two essentially parallel mainlines between Eastern Kentucky and Jacksonville, Florida. A direct route connects the coal center of Corbin, Kentucky with Jacksonville, Florida via Knoxville, Tennessee and Atlanta, Georgia. This line extends north to Cincinnati, Ohio and also serves as a potential merchandise corridor. The more circuitous CSX Blue Ridge division mainline to the east links Shelby, Kentucky with Spartanburg, South Carolina, where other mainline routes funnel traffic south to Savannah and Jacksonville. Both of these mainlines have high percentages of coal traffic, although they do not serve the same mines. Recently, the CSX Corbin mainline has experienced increased demands for merchandise freight, including high priority automobile and intermodal traffic. In response, CSX now diverts coal east through connecting routes to the Blue Ridge line so that the Corbin line can better handle the priority traffic.

The model seems to reflect CSX practice in routing coal in this corridor. Figure 5-8 shows traffic volumes on the coal corridor routes. The model divides coal tonnage between the two routes, with a modest amount of coal tonnage proceeding to Florida from Eastern Kentucky via the east route.

An interesting flow condition was noted between Spartanburg and Savannah. The model predicts, again as shown in Figure 5-8, that coal flows from the Blue Ridge division will travel Spartanburg–Laurens–Columbia–Savannah. Current flows predominantly use a Spartanburg–Greenwood–Augusta–Savannah route because the junction at Laurens is incorrectly oriented for through coal movements to Columbia.

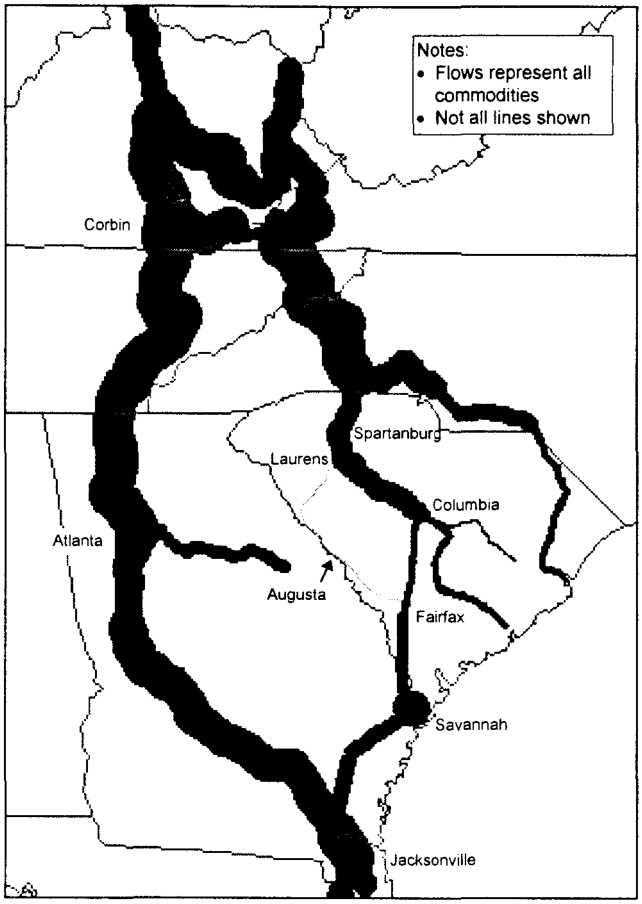

Figure 5-8. CSX Coal Corridor Flows.

CSX is currently engaged in a multi-million dollar line construction project at Laurens to correct this problem. In addition, the Laurens–Columbia–Savannah line has already received extensive signal and siding work to accommodate increased traffic. The model might, therefore, support the CSX investments.

Virginia Tidewater Coal Flows

Examination of the flow patterns in Virginia revealed a set of anomalies. Coal flows east from coal fields in West Virginia to tidewater at Newport News and Norfolk. Norfolk Southern and CSX Transportation each have several east-west lines across Virginia. Traditionally, each railroad has used one route for manifest traffic and coal empties, while the other route handles primarily loaded eastbound coal. The coal route typically is flat or has descending grades, often following watercourses.

The model did poorly in selecting routes in Virginia which reflected tidewater flow patterns. On the Norfolk Southern, for example, coal is routed east from Roanoke via the low grade, ex-Virginian Railway route to Abilene. As Figure 5-9 shows, the model did not reflect this, choosing instead to route most flow via the ex-Norfolk & Western Railway line to the north through Lynchburg. In practice, the Blue Ridge grade just east of Roanoke is a substantial impediment to heavy eastbound trains on this route.

The routing of CSX traffic was no better. Between Clifton Forge and Richmond, CSX routes almost all traffic over the water level James River line via Lynchburg. The more northerly North Mountain and Piedmont subdivisions connect these same endpoints via Charlottesville. Earlier in the century, the latter route hosted most merchandise and

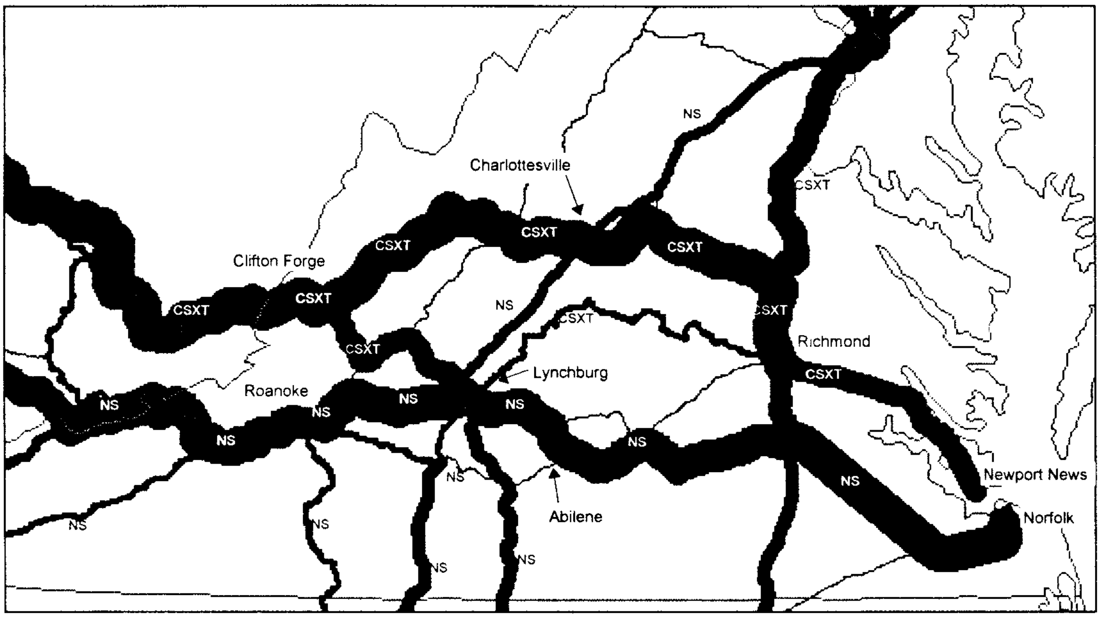

Figure 5-9. Virginia Tidewater Coal Flows.

passenger trains and today carries Amtrak's *Cardinal.* Equipped with signals and possessing a generally favorable alignment, the route is suitable for high speed service. The James River line, in contrast, is some 40 miles longer because it follows its namesake river. Although signaled, its numerous curves limit train speeds for many miles to 45 mph or less.

The model consistently routed traffic over the northerly CSX route east of Clifton Forge, placing only a small amount of traffic on the James River route. In contrast, CSX presently routes almost no through traffic over the northern route, choosing instead to use the James River line exclusively.

The reason for these discrepancies is unclear. It is likely, however, that the model calibration does not accurately reflect the costs of the low grade lines. The cost advantage of hauling loaded cars continuously downgrade must be significant, even in the face of slower transit times and greater distances as on the James River line. In addition, the alternate lines generally have a less favorable alignment for heavy trains. The calibration process for the case was not sophisticated enough to reflect these differences. The author does not believe that the Virginia case reflects any underlying flaw in the model logic.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

Research Summary

The research conducted in this study resulted in the formulation and application of a multicommodity, multicarrier network model for predicting equilibrium flows within a railroad network. Designed for strategic planning with a short term horizon, the model assumes fixed external demand, but attempts to replicate the economic decisionmaking process used by carriers in developing flow patterns. The predicted flows meet the conditions for Wardropian system equilibrium. Congestion functions which predict link travel time as a function of link characteristics and traffic volume are incorporated into the objective function. At equilibrium, the solution algorithm predicts the expected delay per train on each link, allowing the analyst to identify areas of congestion.

The need for this model was identified as a result of numerous concerns expressed in the railroad and shipper trade press during the past several years. The railroads have had difficulties handling the recent dramatic increase in service sensitive intermodal traffic along with a general overall increase in carload traffic. Many industry observers feel that long-term reductions in the size and extent of the national rail network have eliminated capacity margins. Rhetoric from both the railroads, their customers, and competing modes cloud the issue.

Rail freight network models suitable for evaluating capacity issues are not widely available. Most of the network assignment tools in wide use are oriented to the prediction of passenger automobile flows. The auto passenger assignment problem is structurally very different from the railroad freight problem. First, the behavior of automobile drivers is assumed to follow Wardropian user equilibrium, where individuals attempt to select routes minimizing their individual travel costs. In railroad systems, traffic flow is controlled by central authorities that presumably wish to minimize total costs. Second, automobile drivers are assumed to be equally influenced by cost in selecting their routes. In the railroad freight problem, commodities have individual cost functions and service characteristics. This makes the problem much more complex to formulate and solve. Third, the railroad system consists of a series of carriers having distinct markets and service characteristics and operating over a subnetwork. Exchange between carriers is frequently required to take a shipment from origin to destination. These exchanges can only occur at designated points in the network. This differs from the highway problem where a driver may move through the entire network without restriction.

The RAILNET model was formulated and implemented to address the specific aspects of rail freight transportation. The computer codes were developed using FORTRAN and compiled on an IBM-compatible personal computer. The test case was developed using the rail network in the southeastern U.S., with the ICC Carload Waybill Sample providing traffic flow data for this region. The case study revealed that the model

performance met all expectations, but that the waybill file was a poor source of commodity flow data.

Model Performance

The case study provided a test of the implemented model using a large problem. The study network contained 670 nodes, 954 links, and 444 interchanges. Twelve carriers operated on the 34,251.2 route miles of track represented by the network. The problem examined the flows of 12 commodity types moving between 128 demand centroids. Although the demand data obviously underrepresented flows of certain commodities, the model appeared to give reasonable results for the supplied volumes. The specific results of the study are described in a subsequent section.

Execution of the case study problem on a personal computer equipped with a 40 MHz 80386, a math coprocessor, and 8 megabytes of random access memory took just under two hours. An analyst could evaluate several scenarios each day at this rate. Use of faster 80486 or Pentium processor equipped personal computers which are now widely available increase productivity drastically, reducing the case study execution time to about eleven minutes. Thus, the model appears to be quite practical for use in a production environment.

As might be expected, increasing the number of commodities increased the length of time required to execute the model for the case network. This rate of increase appeared to be approximately linear. It is expected that changes in network size and

configuration would have a polynomial effect on performance due to the behavior of the path algorithm. This hypothesis was not tested, however, since in a typical case the network configuration will remain essentially static while demand and cost data are varied.

Use of the polynomial function, which with its exponentiation requirements is more complex to compute than the hyperbolic function, did not seem to affect the model performance. Thus, the polynomial function seems worthwhile where the problem formulation requires a continuous, twice differentiable congestion function.

Case Study Methodology and Results

The case study exercise examined flows with the rail network in the southeastern states. Several comments and findings associated with this case study are in order.

Flow Data Limitations

The case study required the development of flow matrixes using the public use CWS. In general, this data source seems more oriented to the development of traffic statistics rather than to flow analysis. The aggregation of waybill origins and destinations to the BEA centroid level is less than ideal when dealing with detailed networks, although it certainly can be endured. The major problem, however, is that almost 50 percent of the waybills have a masked origin and/or destination, rendering them useless for flow studies. It is obvious that these missing bills greatly reduce the available tonnage volume in the CWS. Without accurate flows, the actual operating conditions in the network

cannot be predicted. Users of the full version of the CWS will not, of course, experience this problem. The majority of academic users, however, will only have the public use file.

The greater question, however, regarding the CWS is whether the dataset represents a reasonable portion of the actual universe of rail shipments. The author feels that the results of the case raise a reasonable doubt about the completeness of the dataset. The waybills with missing origins and destinations account for only 48 million tons. This seems insufficient to explain the persistently low volumes throughout the network.

Another problem with the CWS is its handling of intermodal traffic. Intermodal trailers and containers move most frequently under individual waybills. A railcar, however, typically handles multiple intermodal units. It is not possible to determine from the waybill file how many other loads accompanied the sampled container on the railcar. The true average payload for intermodal cars cannot be estimated. This problem is especially thorny for doublestack equipment.

Case Results

Given the flow data limitations, predicted traffic throughout much of the network fell well below known levels. In many cases, flows were as low as 20 percent of known tonnages. Graphically, the flows appear to have the expected pattern, but at an incorrect magnitude. The immediate implication is that congestion seems minimal, especially on corridors without a significant volume of coal.

The bright spot in this scenario was the general adequacy of coal flow data in the file. Supplemental sources for coal demand even allowed enhancement of the waybill information. In certain corridors in the region, coal is the major rail commodity, with merchandise, other bulk materials, and intermodal accounting for a small fraction. The behavior of these corridors could be realistically studied. Here, the model results proved especially interesting and relevant.

Coal originated on CSX Transportation and Norfolk Southern lines in southern West Virginia, western Virginia, and eastern Kentucky flows in large volumes to electric utilities in Sun Belt states. A high volume north-south corridor for these flows exists between mining areas in the aforementioned region and central Florida. This corridor has multiple route possibilities for each carrier. The model decreased coal volumes on the direct CSX mainline between Corbin, Kentucky and Atlanta, Georgia via Knoxville as high priority automobile and intermodal flows were introduced on the line. The coal was diverted east to the more circuitous CSX Blue Ridge division mainline between Shelby, Kentucky and Spartanburg, South Carolina and then over the Florence Division to Savannah and Jacksonville. This exactly reflects a similar operating decision made recently by CSX corridor managers in an effort to expedite the high value traffic. Another interesting route pattern was noted between Spartanburg and Savannah. The model predicted most flow should take the route Spartanburg–Laurens–Columbia–Savannah. Current flows predominantly use a Spartanburg–Greenwood–Augusta–Savannah route because a junction at Laurens is incorrectly oriented for through coal

movements. Interestingly, CSX is currently engaged in a multi-million dollar line construction project at Laurens to correct this problem. The Laurens–Columbia–Savannah line has already received extensive signal and siding work to accommodate through traffic. The model might, therefore, be validating the CSX investments.

A contrasting situation where the model did not reproduce actual flow patterns occurred in Virginia. Tidewater coal flowing east over both CSX Transportation and Norfolk Southern follows low grade routes, with other commodities taking paralleling, but less favorable routes for both carriers. The model placed most flow on these alternate routes and dramatically understated flow on the low grade routes. This appears to be a calibration problem, since the generalized calibration procedure did not consider the highly favorable cost structure for loaded eastbound movements on these specific routes.

The case study did not reveal any line segments having critical levels of delay in the study region. Given the artificially low flow levels, however, this finding is inconclusive.

One notable omission from the case study was the inclusion of terminals. The model formulation accommodates terminals and is sensitive to terminal congestion. The research revealed, however, that calibration of terminal delay functions, unlike those of line segments, is best done on a case by case basis because of the individuality of the facilities. Given the number of railyards in the study area, it was not feasible to undertake the extensive effort required to calibrate individual terminal functions. The RAILNET

program structure could have been modified to yield a fixed delay for terminal arcs, but, given limitations on time, this was not done.

Opportunities for Future Research

The research study revealed several avenues for future research. The model formulation uses a link capacity function which can be calibrated for a number of fixed conditions. The model is not sensitive, however, to operational characteristics such as traffic mix. Line capacity functions are affected by the mixture of trains having differing performance characteristics. The travel time curve will shift if the traffic mix on a line changes, for example, from 100 percent unit coal trains to 50 percent fast intermodal trains and 50 percent unit coal trains. This shift occurs even if the total traffic volume remains constant. In an assignment model, such shifts are certainly possible during the solution of the problem. Development of a travel time function sensitive to traffic mix would enhance the overall formulation.

Another avenue for investigation is the development of solution algorithms capable of attacking the problem of finding the global minimum given a nonconvex objective function. Such an algorithm would allow the use of U-shaped cost functions demonstrated by economies of density. Many transportation enterprises have such functions. Current research in nonconvex optimization might provide avenues for enhancing the assignment program to allow the use of such typical cost functions.

REFERENCES

REFERENCES

______ (1993), 1993 Intermodal Index, Intermodal Association of North America and The National Industrial Traffic League, December, 1993.

______ (1995a), "MOVING COAL still a basic function of RAILROADING," Progressive Railroading, vol. 38, no. 7, pp. 35-40.

______ (1995b), "Utility coal car maintenance costs rising," Progressive Railroading. Volume 38, no. 7, pp. 47-50.

______ (1994), Railway Line Clearances, vol. 204. New York: K-III Directory Corp.

______ (1994), The Official Railway Equipment Register®, vol. 110, no. 1. New York: K-III Directory Corp.

American Railway Engineering Association (1993). Manual for Railway Engineering.

Association of American Railroads (1995). Railroad Facts-1995 Edition.

Bronzini, M.S. and R.C. Miller (1978). Freight Transportation Energy Use, Volume II - Methodology and Program Documentation (Draft for project DOT-TSC-1252). U.S. Department of Transportation, Transportation Systems Center.

Bronzini, M.S. and D.M. Sherman (1986). Railroad Routing and Costing. Report EA-4716 prepared by the University of Tennessee Transportation Center and CACI, Inc.—Federal for the Electric Power Research Institute.

Chang, C.-J., Miles, R.D., and Sinha, K.C. (1981). "A Regional Railroad Network Optimization Model for Coal Transportation," Transportation Research, vol. 15B, no. 4, pp. 227-238.

Chatterjee, Arun, Clarke, D.B., Sink, H.L., Rutner, S.M., and Stamatiadis, N. (1995). Intermodal Freight Transportation and Highway Safety, final report prepared by the University of Tennessee Transportation Center for the Southeastern Transportation Center under the University Transportation Centers Program, U.S. Department of Transportation.

Clarke, David B. (1982). A Simulation Model for Railroad Line Capacity Analysis. M.S. Thesis, Department of Civil Engineering, University of Tennessee, Knoxville, Tennessee.

Crainic, T.G., Florian, M., and Léal, J. (1990). "A Model for the Strategic Planning of National Freight Transportation by Rail," Transportation Science, vol. 24, no. 1, pp. 1-24.

Dafermos, Stella C. (1971). "The Traffic Assignment Problem for Multiclass-User Transportation Networks," Transportation Science. pp. 73-87.

Daughety, A., and M.A. Turnquist (1979). Development of Hybrid Cost Functions from Engineering and Statistical Techniques: The Case of Rail. Prepared by Northwestern University, Evanston, Ill. for U.S. Department of Transportation.

Eash, R.W., Janson, B.N., and Boyce, D.E. (1981). "Equilibrium Trip Assignment: Advantages and Implications for Practice," Transportation Research Record. pp. 1-8.

Friesz, Terry L. et al. (1981). The Northeast Regional Environmental Impact Study: Theory, Validation and Application of a Freight Network Equilibrium Model. Report ANL/ES-120 prepared by Argonne National Laboratory, Argonne, Ill. for U.S. Department of Energy.

Gallo, G. and Pallottino, S. (1988). "Shortest Path Algorithms," Annals of Operations Research, vol. 13, pp. 3-79. ed. Bruno Simeone et al., Basel Switzerland: J.C. Baltzer AG Scientific Publishing Company.

Greenberg, B.S., Leachman, R.S., and Wolff, R.W. "Predicting Dispatching Delays on a Low Speed, Single Track Railroad," Transportation Science, vol. 22, no. 1, pp. 31-38.

Guélat, J., Florian, M., and Crainic, T.G. (1990). "A Multimode Multiproduct Network Assignment Model for Strategic Planning of Freight Flows," Transportation Science, vol. 24, no. 1, pp. 25-39.

Harker, Patrick T. (1986). Predicting Intercity Freight Flows. Utrecht, The Netherlands: VNU Science Press.

Hay, William W. (1982). Railway Engineering, Second Edition. New York: John Wiley & Sons.

Hu, T.C. (1969). Integer Programming and Network Flows. Reading, MA: Addison-Wesley Publishing Co.

Loureiro, Carlos F.G. (1994). Modeling Investment Options for Multimodal Transportation Networks. Ph.D. Dissertation, Department of Civil Engineering, University of Tennessee, Knoxville, Tennessee.

Janic, Milan (1984). "Single Track Line Capacity Model," Transportation Planning and Technology, vol. 9, pp. 135-151.

Janic, Milan (1988). "A Practical Capacity Model of a Single Track Line," Transportation Planning and Technology, vol. 12, pp. 301-318.

Johnson, P.E., Joy, D.S., Clarke, D.B., and Jacobi, J. (1993). Interline 5.0—An Enhanced Railroad Routing Model: Program Description, Methodology, and Revised User's Manual, Report ORNL/TM-12090 prepared by Oak Ridge National Laboratory for U.S. Department of Energy.

Kresge, David T. and Paul O. Roberts (1971). Techniques of Transport Planning Volume Two: Systems Analysis and Simulation Models. ed. John R. Meyer, Washington, D.C.: Brookings Institution.

Lansdowne, Zachary F. (1981). "Rail Freight Traffic Assignment," Transportation Research, vol. 15A, pp. 183-190.

Margiotta, Richard A. (1995). Telephone conversation. Oak Ridge, TN, September 22, 1995.

Mosher, W.E., Jr. (1963). "A Capacity Restraint Algorithm for Assigning Flow to a Transportation Network," Highway Research Record, no. 6, pp. 258-289.

Norris, Bahar (1994). Intermodal Freight: An Industry Overview. Report PM-42-BBNI prepared by Volpe National Transportation Systems Center for Federal Highway Administration, U.S. Department of Transportation.

Patton, Edwin P. (1992). Railroad Costing. Unpublished course notes, Department of Marketing, Logistics, and Transportation, University of Tennessee.

Petersen, E.R. (1974). "Over-the-road transit time for a single track railway," Transportation Science, vol. 6, pp. 65-76.

Petersen, E.R. (1977a). "Railyard modeling: part I. Prediction of put-through time," Transportation Science, vol. 11, no. 1, pp. 37-49.

Petersen, E.R. (1977b). "Railyard modeling: part II. The effect of yard facilities on congestion," Transportation Science, vol. 11, no. 1, pp. 50-59.

Poole, Ernest C. (1962). Costs—A Tool for Railroad Management. New York, NY: Simmons-Boardman Publishing Co.

Prokopy, J.C., and R.B. Rubin (1975). Parametric Analysis of Railway Line Capacity. Report No. DOT-FR-4-5014-2, U.S. Department of Transportation, Federal Railroad Administration.

Sheffi, Yosef (1985). Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Turnquist, M.A. and M.S. Daskin (1982). "Queuing Models of Classification and Connection Delay in Railyards," Transportation Science, vol. 16, no. 2, pp. 207-230.

Wardrop, J.G. (1952). "Some Theoretical Aspects of Road Traffic Research," Proceedings, Institution of Civil Engineers, vol. 1, part III, no. 2, pp. 325-378.

Welty, Gus (1995). "Coal comes on strong. Can we cope?," Railway Age, February 1995, pp. 25-29.

APPENDIXES

APPENDIX A

DERIVATION OF THE ARC MARGINAL COST FUNCTION

RAILNET's optimization algorithm minimizes the total system generalized cost of commodity flows over network arcs and through transfer nodes:

$$\text{Minimize } Z = \sum_{p \in P} \left(\sum_{a \in A} s_a^p(v)\, v_a^p + \sum_{t \in T} s_t^p(v)\, v_t^p \right) \qquad \text{(A.1)}$$

where:

P,A,T	=	sets of commodities, arcs, and transfers respectively in the problem;
(v)	=	vector of product flows within the network;
$s_a^p(v)$	=	average cost function for the flow of commodity p on arc a, given flow (v);
$s_t^p(v)$	=	average cost function for the flow of commodity p through transfer t, given flow vector (v);
v_a^p, v_t^p	=	flow volume of commodity p over arc a and transfer t, respectively.

As discussed in the body of the dissertation, flows are subject to the usual conservation and non-negativity constraints.

For each iteration over a commodity, p, the algorithm selects candidate paths between origin-destination pairs for flow enhancement by solving a shortest path algorithm. The arc and transfer costs used in the solution of these shortest path problems

represent the sum of the current unit cost and the marginal cost for p based upon (v). The marginal cost, well known in economics, is the incremental change in cost $\partial C / \partial (v_p)$ incurred by adding an additional unit of p to the flow. In this discussion, we will refer to the flow dependent unit cost of p as the marginal cost. We have two types of facilities of interest: arcs and transfer nodes. In general, the marginal cost $c_{\bar{a}}^{\bar{p}}$ for transporting product $\bar{p}$ on arc $\bar{a}$ is:

$$c_{\bar{a}}^{\bar{p}} = s_{\bar{a}}^{\bar{p}}(v) + \sum_{p \in P} \left(\sum_{a \in A} \frac{\partial s_a^p(v)}{\partial v_{\bar{a}}^{\bar{p}}} v_a^p + \sum_{t \in T} \frac{\partial s_t^p(v)}{\partial v_{\bar{a}}^{\bar{p}}} v_t^p \right). \quad \text{(A.2a)}$$

The equivalent function for transfer facility $\bar{t}$ is:

$$c_{\bar{t}}^{\bar{p}} = s_{\bar{t}}^{\bar{p}}(v) + \sum_{p \in P} \left(\sum_{a \in A} \frac{\partial s_a^p(v)}{\partial v_{\bar{t}}^{\bar{p}}} v_a^p + \sum_{t \in T} \frac{\partial s_t^p(v)}{\partial v_{\bar{t}}^{\bar{p}}} v_t^p \right). \quad \text{(A.2b)}$$

In practice, the following simplifying assumptions can be made:

a) The cost function for a given transfer is not affected by the flows at other transfers or by arc flows;

b) The cost function for an arc is not affected by transfer flows; and

c) The cost function for an arc is only affected by flows on arcs which represent the same physical link. There is no interaction between flows on separate physical links.

These do not seem to conflict with real world behavior of the railroad system.

We define $\bar{A}$ as the set of logical arcs representing a physical link, $l = (i; j)$, $l \in L$, connecting nodes i and j. Arc $a = (i, j, m)_l$ then represents a service of mode m using l. In general, l is an undirected link, so that for each arc $a = (i, j, m)_l$, there is a corresponding reverse arc $\acute{a} = (j, i, m)_l$. We may have any number of modes using l, each represented by corresponding logical arcs. The set $\bar{A}$ is, therefore

$$\left\{a \in A \mid a = (i,j,m)_l \ or\ a = (j,i,m)_l, m \in M\right\}.$$

The load pattern for $\bar{A}$ is denoted by $v_{\bar{A}}$.

If an arc $a \notin \bar{A}$, then by assumption (c), $\partial s_a^p / \partial v_a^{\bar{p}} = 0$. This said, the marginal cost function for arcs can be simplified to:

$$c_{\bar{a}}^{\bar{p}} = s_{\bar{a}}^{\bar{p}}(v_{\bar{A}}) + \sum_{p \in P} \sum_{a \in \bar{A}} \frac{\partial s_a^p(v_{\bar{A}})}{\partial v_{\bar{a}}^{\bar{p}}} v_a^p. \tag{A.2c}$$

For transfers, the marginal cost becomes:

$$c_t^{\bar{p}} = s_t^{\bar{p}}(v_t) + \sum_{p \in P} \frac{\partial s_t^p(v_t)}{\partial v_t^{\bar{p}}} v_t^p. \tag{A.2d}$$

We further assume, however, that transfers are uncapacitated using the rationale that railroads will dispatch trains to handle interchange traffic as necessary. The capacities of the adjacent arcs will then govern transfer volumes. This leads to the conclusion that $\partial s_t^p / \partial v_t^{\bar{p}} = 0$ and, therefore:

$$c_t^{\bar{p}} = s_t^{\bar{p}}. \tag{A.2e}$$

We now address the problem of deriving a working form of the arc marginal cost function. The line haul unit cost function for moving commodity p on arc a is defined as:

$$s_a^p(v) = m_a^p l_{\bar{A}} + T_{\bar{A}}(v) f_a^p h_a^p \qquad \text{(A.3)}$$

where:

m_a^p = cost per gross ton-mile for hauling commodity p on arc a;

$l_{\bar{A}}$ = length in miles of link l, i.e. all arcs in $\bar{A}$, including a;

h_a^p = cost per train-hour for hauling commodity p over arc a;

f_a^p = conversion factor, trains/ton, for commodity p on arc a;

$T_{\bar{A}}$ = travel time in hours, given (v), on link l, i.e. all arcs in $\bar{A}$, including a.

The first term in the cost function covers that portion of the arc costs related to ton-mile volume, including such factors as track and equipment maintenance, overhead costs, fixed charges, etc. The second term describes the contribution of factors influenced by time, including, for example, crew wages, energy costs, and inventory carrying costs of transported goods. This cost function assumes that the mileage based coefficient in the first term is constant over all volumes. In the second term, the link travel time is, of course, a direct function of the total volume, in trains, on the link. This is determined using the polynomial link travel time function.

The polynomial link travel time function has the desirable properties of being continuous, convex, monotone increasing, and differentiable. The function is easy to

work with, especially compared with the analytic travel time functions discussed earlier. Furthermore, the same basic polynomial form, with the appropriate selection of constants, can be used to estimate terminal delay as a function of volume. This allows terminals to be modeled as a special class of link. The polynomial function form employed in RAILNET is:

$$T_{\bar{A}} = R_{\bar{A}}\left[1 + k_1 t_{\bar{A}} + k_2\left(\frac{V_{\bar{A}}}{C_{\bar{A}}}\right)^{\gamma}\right] \tag{A.4}$$

where: $R_{\bar{A}}$ = free flow travel time, hours, for arcs in $\bar{A}$;

k_1, k_2, γ = empirical constants;

$V_{\bar{A}}$ = total daily train volume for arcs in $\bar{A}$;

$C_{\bar{A}}$ = total capacity, trains per day, for arcs in $\bar{A}$.

The total train volume over the link, i.e. the arcs in $\bar{A}$, is:

$$V_{\bar{A}} = \sum\nolimits_{p\in P} \sum\nolimits_{a\in \bar{A}} f_a^p v_a^p . \tag{A.5}$$

Substituting, the arc cost function then becomes:

$$s_a^p = m_a^p l_{\bar{A}} + R_{\bar{A}} f_a^p h_a^p \left[1 + k_1 \sum_{p\in P} \sum_{a\in \bar{A}} f_a^p v_a^p + k_2\left(\frac{\sum\limits_{p\in P} \sum\limits_{a\in \bar{A}} f_a^p v_a^p}{C_{\bar{A}}}\right)^{\gamma}\right]. \tag{A.6}$$

For a given arc $\bar{a}$ and commodity $\bar{p}$, we are faced with the partial differentiation of this function with respect to the volume $v_{\bar{a}}^{\bar{p}}$ in computing the arc marginal cost. This may be done most easily by considering the separate terms in the equation, as follows:

a) $$\frac{\partial m_a^p l_{\bar{A}}}{\partial v_{\bar{a}}^{\bar{p}}} = 0, \quad \forall\, a \in \bar{A}, p;$$

b) $$\frac{\partial R_{\bar{A}} f_a^p h_a^p}{\partial v_{\bar{a}}^{\bar{p}}} = 0, \quad \forall\, a \in \bar{A}, p;$$

c) $$\frac{\partial R_{\bar{A}} f_a^p h_a^p k_1 \sum_{p' \in P} \sum_{a' \in \bar{A}} f_{a'}^{p'} v_{a'}^{p'}}{\partial v_{\bar{a}}^{\bar{p}}} = k_1 R_{\bar{A}} h_a^p f_a^p f_{\bar{a}}^{\bar{p}}, \quad \forall\, a \in \bar{A}, p;$$

d) $$\frac{\partial R_{\bar{A}} f_a^p h_a^p k_2 \left(\dfrac{\sum_{p' \in P} \sum_{a' \in \bar{A}} f_{a'}^{p'} v_{a'}^{p'}}{C_{\bar{A}}} \right)^{\gamma}}{\partial v_{\bar{a}}^{\bar{p}}} =$$

$$\frac{\gamma}{C_{\bar{A}}} k_2 R_{\bar{A}} h_a^p f_a^p f_{\bar{a}}^{\bar{p}} \left(\frac{\sum_{p' \in P} \sum_{a' \in \bar{A}} f_{a'}^{p'} v_{a'}^{p'}}{C_{\bar{A}}} \right)^{\gamma - 1}, \quad \forall\, a \in \bar{A}, p.$$

The full marginal cost equation for the arc, commodity combination then becomes:

$$c_{\bar{a}}^{\bar{p}} = m_{\bar{a}}^{\bar{p}} l_{\bar{a}} + R_{\bar{A}} f_{\bar{a}}^{\bar{p}} h_{\bar{a}}^{\bar{p}} \left[1 + k_1 \sum_{p \in P} \sum_{a \in \bar{A}} f_a^p v_a^p + k_2 \left(\frac{\sum_{p \in P} \sum_{a \in \bar{A}} f_a^p v_a^p}{C_{\bar{A}}} \right)^{\gamma} \right] +$$

$$k_1 R_{\bar{A}} f_{\bar{a}}^{\bar{p}} \sum_{p \in P} \sum_{a \in \bar{A}} h_a^p f_a^p v_a^p + k_2 \frac{\gamma}{C_{\bar{A}}} R_{\bar{A}} f_{\bar{a}}^{\bar{p}} \sum_{p \in P} \sum_{a \in \bar{A}} h_a^p f_a^p \left(\frac{\sum_{p' \in P} \sum_{a' \in \bar{A}} f_{a'}^{p'} v_{a'}^{p'}}{C_{\bar{A}}} \right)^{\gamma - 1} v_a^p. \quad \text{(A.7a)}$$

We recognize that, from equation (A.5), $\sum_{p\in P}\sum_{a\in\bar{A}} f_a^p v_a^p = V_{\bar{A}}$, the total train volume over the link. The terms within the parenthesis in equation (A.7a) are then recognizable as the volume/capacity ratio for the link. Rewriting equation (A.7a) yields:

$$c_{\bar{a}}^{\bar{p}} = m_{\bar{a}}^{\bar{p}} l_{\bar{a}} + R_{\bar{A}} f_{\bar{a}}^{\bar{p}} h_{\bar{a}}^{\bar{p}} \left[1 + k_1 \sum_{p\in P}\sum_{a\in\bar{A}} f_a^p v_a^p + k_2 \left(\frac{V_{\bar{A}}}{C_{\bar{A}}} \right)^{\gamma} \right] +$$

$$k_1 R_{\bar{A}} f_{\bar{a}}^{\bar{p}} \sum_{p\in P}\sum_{a\in\bar{A}} h_a^p f_a^p v_a^p + k_2 \frac{\gamma}{C_{\bar{A}}} \left(\frac{V_{\bar{A}}}{C_{\bar{A}}} \right)^{\gamma-1} R_{\bar{A}} f_{\bar{a}}^{\bar{p}} \sum_{p\in P}\sum_{a\in\bar{A}} h_a^p f_a^p v_a^p . \qquad \text{(A.7b)}$$

Further reorganizing the terms, we obtain:

$$c_{\bar{a}}^{\bar{p}} = m_{\bar{a}}^{\bar{p}} l_{\bar{a}} + R_{\bar{A}} f_{\bar{a}}^{\bar{p}} h_{\bar{a}}^{\bar{p}} + k_1 R_{\bar{A}} f_{\bar{a}}^{\bar{p}} \sum_{p\in P}\sum_{a\in\bar{A}} [f_a^p v_a^p (h_{\bar{a}}^{\bar{p}} + h_a^p)] +$$

$$k_2 R_{\bar{A}} f_{\bar{a}}^{\bar{p}} \left(\frac{V_{\bar{A}}}{C_{\bar{A}}} \right)^{\gamma-1} \left[\left(\frac{V_{\bar{A}}}{C_{\bar{A}}} \right) h_{\bar{a}}^{\bar{p}} + \frac{\gamma}{C_{\bar{A}}} \sum_{p\in P}\sum_{a\in\bar{A}} h_a^p f_a^p v_a^p \right] . \qquad \text{(A.7c)}$$

A final reorganization yields the working form of the equation:

$$c_{\bar{a}}^{\bar{p}} = m_{\bar{a}}^{\bar{p}} l_{\bar{a}} + R_{\bar{A}} f_{\bar{a}}^{\bar{p}} \left[h_{\bar{a}}^{\bar{p}} + k_1 \sum_{p\in P}\sum_{a\in\bar{A}} [f_a^p v_a^p (h_{\bar{a}}^{\bar{p}} + h_a^p)] + \right.$$

$$\left. k_2 \left(\frac{V_{\bar{A}}}{C_{\bar{A}}} \right)^{\gamma-1} \left[\left(\frac{V_{\bar{A}}}{C_{\bar{A}}} \right) h_{\bar{a}}^{\bar{p}} + \frac{\gamma}{C_{\bar{A}}} \sum_{p\in P}\sum_{a\in\bar{A}} h_a^p f_a^p v_a^p \right] \right] . \qquad \text{(A.8)}$$

This form applies to line haul arcs, where volume and capacity are measured in trains. For freight terminal arcs, volume and capacity are traditionally measured in cars.

In this case, the same functional form can be used if we redefine the variables to reflect cars processed rather than trains. For terminal arcs, the cost function is defined as:

$$s_a^p(v) = \hat{f}_a^p \hat{m}_a^p + \hat{T}_{\bar{A}}(v)\, \hat{f}_a^p \hat{h}_a^p \qquad \text{(A.9)}$$

where:

$\hat{m}_a^p$	=	cost per car for processing commodity *p* through arc *a*;
$\hat{f}_a^p$	=	conversion factor, cars/ton, for commodity *p* on arc *a*;
$\hat{h}_a^p$	=	cost per car-hour for processing commodity *p* through arc *a*;
$\hat{T}_{\bar{A}}$	=	processing time in hours, given (v), for link *l*, i.e. all arcs in $\bar{A}$, including *a*.

Except that the $m_{\bar{a}}^{\bar{p}} l_{\bar{a}}$ term is replaced by $\hat{m}_{\bar{a}}^{\bar{p}} \hat{f}_{\bar{a}}^{\bar{p}}$, the function form of the marginal cost equation is identical. Variables are defined as follows:

$R_{\bar{A}}$	=	free flow car processing time, hours, for arcs in $\bar{A}$;
$V_{\bar{A}}$	=	total car volume for arcs in $\bar{A}$;
$C_{\bar{A}}$	=	total capacity, cars per day, for arcs in $\bar{A}$.

APPENDIX B

SHORTEST PATH ALGORITHM

The heart of the RAILNET solution procedure is a shortest path algorithm (SPA), which must be executed at worst once for each origin-destination pair during a commodity subiteration. Since it could be run (*WP*) times during each master iteration, the SPA must be efficient. This appendix describes some unique characteristics of the SPA.

Recall from Chapter 3 that the complete network is represented by $G = (N, A)$, where N is the set of nodes and A is the set of directed arcs which connect these nodes. Each carrier ***m*** operates a subnetwork G_m which consists of N_m nodes and A_m directed arcs. The complete network therefore consists of the union of the carrier subnetworks, with $N = \bigcup_{m \in M} N_m$ and $A = \bigcup_{m \in M} A_m$. The set T of transfers defines connections where flows may pass between the subnetworks.

The basic problem solved by the SPA is the generation, for the given network, of a directed spanning tree based at a source node, *s*. For each vertex, *v*, in the tree, the path *s-v* is a minimum path, based upon the sum of costs associated with arcs and, possibly, nodes, in the path. In RAILNET, costs are the arc marginal costs for a given commodity, and must be non-negative. The tree extends to all other demand nodes in the network.

A variety of implementations of the SPA have been reported in the literature. Gallo and Pallottino (1988) describe many of the more common versions, and the reader is referred to this reference for a general description of the procedure. The SPA employed in RAILNET is based loosely upon their SHEAP algorithm, which employs a binary heap to efficiently order the priority queue of candidate elements for scanning.

Railroad network routing adds a wrinkle to the standard SPA, however, in that path generation must respect the subnetworks of the individual carriers. Paths may move between carrier subnetworks only at transfer points. Johnson et al. (1993) describe the structure of a two-stage SPA which solves the rail routing problem. This procedure, while not employed directly, provided insight into the development of RAILNET's SPA. The principal advantage of the two-stage procedure is computer memory conservation, which is not a major issue with modern personal computers.

During the network indexing process, NETBLD builds, for each node, a list of the arcs emanating from the node. This list, called the forward star, is sorted by carrier. The SPA can quickly locate the subset of arcs for a given carrier in any forward star.

NETBLD also constructs a list of transfers for each node in the network. The SPA can check this list to quickly determine whether any given node is a transfer point. If so, it can determine, via indexing, the allowable set of transfers at the node. This set is sorted by "from" carrier and then by "to" carrier.

The SPA builds the path tree using an arc chain representation rather than the typical node chain representation usually employed. Paths are described as a series of

consecutive arcs linking the source and destination. This simplifies transfer handling and path traceback logic. A dummy arc with the origin as its head node forms the root of the path tree.

The SPA labels arcs in much the same way as a traditional algorithm labels nodes. As a candidate arc is selected, the forward star of its head node is scanned to determine potential candidates for the heap. The node is first examined to determine whether it is a transfer point. If so, then all emanating arcs of carriers for which transfers are allowed from the current carrier enter the scan set. If not, only arcs matching the current arcs carrier will be scanned. The ordering of the transfer vectors makes it simple to select appropriate arcs.

For path traceback, arcs maintain pointers to predecessors. If the predecessor arc has a different carrier, then the traceback algorithm knows a transfer occurred at that point in the path.

Demand nodes in the network have a label and a predecessor arc pointer. This allows the cost of reaching that node from the source to be readily determined. In addition, traceback of the path is initiated by starting with the arc pointer and proceeding backwards until the origin arc is reached.

APPENDIX C

ICC PUBLIC USE WAYBILL RECORD LAYOUT

This appendix describes the contents of the public use version of the ICC Carload Waybill Sample (CWS). Much of the information provided here is taken directly from the description file on the CWS CD-ROM provided by the U.S. Department of Transportation, Bureau of Transportation Statistics.

Cols. 1-6 Waybill Date (Month, Day, Year) (3I2)

The waybill date is the date the origination railroad prepares the waybill.

Cols. 7-10 Accounting Period (Month, Year) (2I2)

The accounting period is the month and year during which the study waybill was entered into the railroad's revenue accounting system. This information is subsequently reflected in the net income statement of the company for the specified account month.

Cols. 11-14 Number of Carloads (I4)

The total number of carloads on the sampled waybill.

Col. 15 Car Ownership (A1)

This field contains one of the following codes indicating car ownership:

(P) Privately-owned car

(R) Railroad-owned Car

Cols. 16-19 AAR Car Type (A4)

Alpha-numeric code giving a general physical description of the type of car. For more information, refer to Section VI, Exhibit D, of the Uniform Machine Language Equipment Register (UMLER) Specification Manual.

Cols. 20-23 AAR Mechanical Designation (A4)

Mechanical designation is dependent on AAR car type. (Refer to Section V, Item F of UMLER Specification Manual).

Cols. 24-25 ICC Car Type (I2)

The ICC car type is inferred from the AAR car type, described in columns 20-23. This number corresponds to the line number on ICC Form 710 for the type of car.

Cols. 26-28 TOFC/COFC Plan (A3—last position always blank)

The TOFC/COFC plan code must be entered in the first position of the field. If possible, when different TOFC/TOFC plans are used during the course of the movement, the code for the applicable plan at termination in the first digit of the field and the code for the applicable plan at the origination is entered in the second position of the field. For example, '24' indicates that the TOFC movement started on Plan 2-1/2 and terminated on Plan 2. In cases where this delineation is not possible, Code 9 (indicating a combination of TOFC/COFC plans) is entered.

If the waybill covers multiple trailers/containers with different plans, the plan code used for the first trailer/container is entered. Valid TOFC/COFC plan number codes are listed below:

(0) Not a piggyback shipment;

(1) TOFC/COFC Plan 1;

(2) TOFC/COFC Plan 2;

(3) TOFC/COFC Plan 2-1/4;

(4) TOFC/COFC Plan 2-1/2;

(5) TOFC/COCC Plan 3;

(6) TOFC/COFC Plan 4;

(7) TOFC/COFC Plan 5;

(8) All other TOFC/COFC/COFC plan numbers;

(9) Combination of TOFC/COFC plan numbers; or

(X) Unknown.

Cols. 29-32 Number of TOFC/COFC units (I4)

The total number of TOFC/COFC units reported on the sampled waybill.

Column 33 Trailer or Container Ownership (A1)

The field contains one of the following codes:

(P) Privately-owned Trailer/Container; or

(R) Railroad-owned Trailer/Container.

Column 34 Trailer or Container Type (A1)

The field contains one of the following codes indicating the type of intermodal package:

(T) TOFC Trailer;

(C) COFC Container; or

(U) Unknown.

Column 35 Hazardous/Bulk Material in Boxcar (A1)

The field contains one of the following codes:

(B) Bulk, non-hazardous material (STCC 50 series), moved in a boxcar;

(H) Hazardous material (STCC 49 series) moved in any type of car; or

( ) neither of the above.

Cols. 36-40 Commodity Code (STCC-Non HAZMAT) (I5)

The Standard Transportation Commodity Code (STCC) identifies the product designation for the commodity being transported. This field includes the first five digits of the seven-digit STCC; however, STCC 19 series commodities are reported only at the 2-digit level.

The field does not include Hazardous materials (series 49xxx) or Bulk materials in Boxcars (series 50xxx). All STCC 49 and 50 series codes have been translated to actual product commodity codes.

Cols. 41-47 Billed Weight in Tons (I7)

The billed weight of lading, calculated in tons.

Cols. 48-54 Actual Weight in Tons (I7)

The actual weight of lading (if provided), calculated in tons.

Cols. 55-63 Freight Revenue (I9)

The total line-haul freight revenue from origin to termination, shown in dollars for the study waybill.

Cols. 64-72 Transit Charges (I9)

Transit charges, where applicable, shown in dollars.

Cols. 73-81 Miscellaneous Charges (I9)

The total of all miscellaneous charges (excluding transit charges and freight revenue), shown in dollars.

Column 82 Interstate/Intrastate Charges (I1)

Normally, an Intrastate routing is inferred if the origin and termination states are the same. However, an Interstate routing is inferred in cases where the origin and termination stations are within a state but the customary routing exits and re-enters the state. Interstate movements also include import, export, ex-lake and lake cargo movements. The field contents are as follows:

(1) Interstate;

(2) Intrastate; or

(3) Unknown.

Column 83 Type of Move (I1)

This field contains one of the following codes:

(0) Neither import nor export;

(1) Imported commodity;

(2) Exported commodity;

(3) Commodity imported and exported, e.g., land bridge type traffic; or

(9) Unknown.

Column 84 All Rail/Intermodal Code (I1)

This field contains one of the following codes denoting whether a shipment is intermodal:

(1) All rail;

(2) Intermodal—a continuous movement involving at least one railroad and another mode; or

(9) Unknown.

Column 85 Type of Move Via Water (Inferred) (I1)

The field contains one of the following codes:

(0) Not a water movement;

(1) Ex-lake (from Great Lakes to reporting railroad);

(2) Lake cargo (rail to Great Lakes);

(3) Intercoastal: a continuous movement by U.S. rail which is part of an Atlantic Ocean (or Gulf) and Pacific Ocean movement—either direction;

(4) Coastwise: a continuous movement involving rail at either end of a coastwise movement between ports on the East Coast (including Gulf) or between ports on the West Coast; or

(5) Inland waterways: a rail movement in combination with a barge movement on rivers and canals other than the Great Lakes that is not considered a part of the rail movement, e.g., rail car ferry.

Column 86 Outbound Transit Code (I1)

The field contents are as follows:

(0) Not a transit movement;

(1) Transit—indicates that the shipment is the outbound movement from a transit point where some service has been performed to the termination point (which can be another transit point); or

(9) Unknown.

Column 87 Substituted Truck-for-Rail Service (I1)

Field contents assume one of the following values:

(0) Not a substituted truck-for-rail service;

(1) Study movement involves substituted truck-for-rail service. (For example, a rail carrier may be authorized by the commission to institute truck-for-rail service when rail service is abandoned or a track is closed for various reasons); or

(9) Unknown

Column 88 Rebill code (I1)

Field contents are as follows:

(0) Not a rebill;

(1) Rebill indicates that the shipment is rebilled at a portion of the through rate from origin to termination and involves non-through billing railroad(s); or

(9) Unknown.

Cols. 89-92 Estimated Short Line Miles (I4)

The short line miles (shortest rail distance between origin and termination), rounded up to the nearest 10 miles.

Column 93 Stratum Identification (I1)

This fields contains information which describes the population and sampling rate for which the record was obtained. The term MRI refers to computerized waybill information. The values used are listed in the following table.

Code	Medium	Carloads per Waybill	Sampling Rate
(1)	MRI	1-2	1 of 40
(2)	MRI	3-15	1 of 12
(3)	MRI	16-60	1 of 4
(4)	MRI	61-100	1 of 3
(5)	MRI	over 100	1 of 2
(6)	Hardcopy	1-5	1 of 100
(7)	Hardcopy	6-25	1 of 10
(8)	Hardcopy	over 25	1 of 5

Column 94 Subsample Code Number (I1)

For MRI waybills, this coding (1, 2, 3, or 4) identifies the individual subsamples obtained under the computerized sampling procedure. This field is initialized to a blank for hardcopy waybills, but a replicate subsample code is added after completion of the master file, using the following formula:

Code = Serial Number - ((Serial Number /4) * 4) + 1 (truncated integer)

These subsample code numbers may be used in statistical analysis of the dataset.

VITA

David B. Clarke was born in Columbia, South Carolina, on October 11, 1957. He grew up in that city and was graduated from Dreher High School in June 1975. The following September he entered The University of Tennessee, Knoxville to major in Civil Engineering with an emphasis in transportation. In March 1979, he received the Bachelor of Science Degree in Civil Engineering with Honors.

Following graduation, David entered the graduate program at The University of Tennessee and began studies for the Master of Science degree in Civil Engineering, again with an emphasis in transportation. During this period, he served as a Graduate Teaching Assistant in the Department of Civil Engineering and as a Graduate Research Assistant in the Transportation Center and the Geography Department. He received the Master of Science degree in Civil Engineering in May 1982.

From 1981 to 1982, David was employed as a Civil/Structural Engineer by Bechtel Power Corporation in Gaithersburg, Maryland. From 1982 to 1989, David served as a project manager and supervisor at Science Applications International Corporation in Oak Ridge, Tennessee, working on various transportation programs for the Department of Energy, the Strategic Highway Research Program, and other SAIC clients. In 1990, David joined the staff of the University of Tennessee Transportation Center, where he is an Assistant Director.

David married the former Farzana Shima Najem in July 1983. They reside in Knoxville with their three children: Robert, Christopher, and Kathleen.