TRANSPORTATION RESEARCH RECORD

No. 1341

Rail

Railroad Issues

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1992

Transportation Research Record 1341
Price: $21.00

Subscriber Category
VII rail

TRB Publications Staff
Director of Publications: Nancy A. Ackerman
Senior Editor: Naomi C. Kassabian
Associate Editor: Alison G. Tobias
Assistant Editors: Luanne Crayton, Norman Solomon
Production Coordinator: Karen W. McClain
Office Manager: Phyllis D. Barber
Production Assistant: Betty L. Hawkins

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Railroad issues : a peer-reviewed publication of the Transportation Research Board.
p. cm.—(Transportation research record ISSN 0361-1981 ; no. 1341)
ISBN 0-309-05207-6
1. Railroad engineering. I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 1341
[TF155]
388 s—dc20 92-25493
[385] CIP

Sponsorship of Transportation Research Record 1341

GROUP 1—TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Chairman: Sally Hill Cooper, Virginia Department of Transportation

Multimodal Freight Transportation Section
Chairman: Douglas P. Smith, CN Rail

Committee on Freight Transportation Planning and Marketing
Chairman: Anne Strauss-Wieder, Port Authority of New York and New Jersey
Secretary: Peter M. Montague, Federal Railroad Administration
W. Bruce Allen, A. Don Bourquard, Russell B. Capelle, Jr., Leland S. Case, Garland Chow, John Glover, Paul C. Hanappe, Susan Henriques-Payne, J. Augustine Hsu, L. Michael Kenny, Bernard J. La Londe, Ah-Chek Lui, Robert M. Landry, George F. List, Ronald A. Mauri, Thomas H. Maze, Edward K. Morlok, Stephen C. Nieman, Howard E. Olson, Edwin P. Patton, Jossef Perl, Charles D. Sanft, John R. Scheirman, Anthony E. Shotwell, Judith L. Truax, Joe Tsai, Robert K. Whitford, Marcus Ramsay Wigan

Public Transportation Section
Chairman: James C. Echols, Tidewater Regional Transit

Committee on Intercity Rail Passenger Systems
Chairman: Emmanuel S. Horowitz, Amtrak
Secretary: S. David Phraner, Port Authority of New York and New Jersey
Murthy V. A. Bondada, Ross Capon, Robert J. Casey, John A. Dawson, William W. Dickhart III, Raymond H. Ellis, Sharon M. Greene, George Haikalis, Ata M. Khan, James T. McQueen, Alex E. Metcalf, Jolene Moritz Molitoris, Aad Ruhl, Ronald C. Sheck, George M. Smerk, Charles H. Smith, Cheryl D. Soon, Louis S. Thompson, Merrill L. Travis, Warren D. Weber, Walter E. Zullig, Jr.

GROUP 2—DESIGN AND CONSTRUCTION OF TRANSPORTATION FACILITIES
Chairman: Charles T. Edson, New Jersey Department of Transportation

Railway Systems Section
Chairman: Scott B. Harvey, Association of American Railroads

Committee on Railroad Track Structure System Design
Chairman: Alfred E. Shaw, Jr., Amtrak
Secretary: William H. Moorhead, Iron Horse Engineering Company, Inc.
Ernest J. Barenberg, Dale K. Beachy, Harry Breasler, Ronald H. Dunn, Amir N. Hanna, S. P. Heath, Crew S. Heimer, Thomas B. Hutcheson, Ben J. Johnson, David C. Kelly, Amos Komornik, John A. Leeper, Mohammad S. Longi, Philip J. McQueen, Myles E. Paisley, Gerald P. Raymond, Jerry G. Rose, Charles L. Stanford, David E. Staplin, W. S. Stokely, John G. White, James W. Winger

Committee on Electrification and Train Control Systems for Guided Ground Transportation Systems
Chairman: Paul H. Reistrup, Monongahela Railway Company
Secretary: Robert A. Keith, Robert Keith Associates
Kenneth W. Addison, J. Hampton Barnett, Richard U. Cogswell, Mary Ellen Fetchko, Robert E. Heggestad, Ram Krishna, Stephen B. Kuznetsov, Robert M. Leilich, Thomas E. Margro, Robert W. McKnight, Howard G. Moody, Gordon B. Mott, Per-Erik Olson, William A. Petit, John A. Reoch, Blair A. Ross, Joseph J. Schmidt, Peter L. Shaw, Richard A. Uher, Robert B. Watson, Jeff D. Young.

Committee on Railway Maintenance
Chairman: William C. Thompson, Union Pacific Railroad Company
Secretary: William B. O'Sullivan, Federal Railroad Administration
David M. Coleman, Bernard R. Forcier, Amir N. Hanna, Anne B. Hazell, Thomas B. Hutcheson, David Michael Johnson, Sue McNeil, Daniel B. Mesnick, Albert J. Reinschmidt, Elias H. Rmeili, Ernest T. Selig, Thomas P. Smithberger, Harry E. Stewart, Vincent R. Terrill, Donald R. Uzarski

Elaine King, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1991.

Transportation Research Record 1341

Contents

Foreword

The first part of this Record consists of four papers related to intercity rail passenger services. Hopkins and Harrison report the results from a study of alternative potential fixed-plant improvements to both commuter and intercity rail passenger service between Boston and New York. The improvements ranged from projects needed to maintain safety and preserve existing infrastructure to more substantial capital investments that could dramatically reduce trip times.

Brand et al. summarize advantages and disadvantages of various approaches for forecasting high-speed rail ridership and recommend a forecasting approach involving the use of separate relationships to estimate diversion from each existing mode to high-speed rail.

The analysis of three alternative high-speed rail technologies and their effect on projected air traffic in the Minneapolis – St. Paul (MSP) to Madison, Milwaukee, and Chicago corridor is described in Buckeye's paper. Although the diversion of air traffic from MSP would be relatively small, the analysis identified other potentially substantial benefits of high-speed rail service that could justify the large capital investment involved.

Parolin and Harrington's study of the changes in long-distance passenger markets in two corridors in Australia in which competition between bus and rail has been allowed only in recent years and in which long-distance bus services were also deregulated found that deregulation has had significant effects through lower fares, increased service, better modal choice for consumers, and increased industry efficiency through rationalization.

The papers in the second part of the Record deal with rail freight service issues. Brown addresses the issue of rail service quality, emphasizing the economic interaction among carrier operating variables, service quality variables, and carrier and shipper costs.

Mundy et al. also address the issue of railroad service quality, particularly variances in transit times, wrongful charges, and the like, that result from yard procedures and information processing. They describe the application of statistical process control quality tools to freight classification yard operations.

Quality of service is also related to the maintenance of facilities. Automated systems for acquiring sequential data on the condition of transportation facilities generate a huge volume of data that needs to be processed. Alfelor and McNeil describe the problems of data aggregation and the techniques that can be adopted to divide linear structures into homogeneous segments for modeling deterioration and assigning maintenance actions.

The railroad industry's search for productivity gains has resulted in the movement of larger volumes of heavier loads, leading in turn to increased rail wear. Moser and Pointner describe the manufacturing parameters and rail selection criteria for one type of wear-resistant rail: head-hardened rail produced from rolling heat.

Rose emphasizes the importance of providing adequate trackbed support to maximize turnout life through a relatively new procedure that involves placing a layer of bituminous (hot-mix asphalt) concrete below the ballast to provide proper support, impermeability, and positive drainage for the track structure.

PART 1

Passenger Rail Issues

Boston–New York Commuter and Intercity Rail Improvement Potential

JOHN B. HOPKINS AND JOHN A. HARRISON

Results are presented of a study of potential fixed-plant improvements to commuter and intercity rail passenger service between Boston and New York that was conducted by the Volpe National Transportation Systems Center under the direction of a Departmental Task Force established by the Secretary of Transportation. A range of alternative improvement programs were identified and characterized. A hierarchy of five alternative programs was defined, ranging from a basic set of projects needed to maintain safety and rehabilitate the existing infrastructure (yielding modest performance gains), to programs that also include substantial capital investment in track work, signaling, electrification, and potential curve and route alignments, which could trim up to 1½ hr from the current best scheduled trip time of just under 4 hr. The study found that an investment of about $1.1 billion is needed during the next decade to bring the existing infrastructure to a state of good repair. More extensive programs, reducing trip time to between 2½ and 3 hr, would cost an additional $500 million to $2.5 billion (exclusive of rolling stock). Potential Amtrak ridership gains and societal benefits were estimated in the study, but are not addressed here.

Results of a study of potential system improvements to benefit commuter and intercity rail passenger service in the Boston–New York corridor are presented. Fixed-plant projects needed for system safety, rehabilitation and service improvement were identified. Five alternative overall programs were defined, and the potential Boston–New York trip times for each program under various rolling stock assumptions were estimated. These findings are summarized here.

The study was conducted for the U.S. Department of Transportation (DOT) by the DOT Volpe National Transportation Systems Center (VNTSC) under the direction of a Task Force established by the Secretary of Transportation. Parsons Brinckerhoff Quade and Douglas, Inc., provided technical support to the VNTSC study team in the areas discussed here.

BACKGROUND

The Northeast Corridor (NEC) serves a populous and heavily traveled region. Extensive commuter rail passenger service on the corridor is critical to the metropolitan areas served. Seven transportation authorities and railroads use more than one-half of the 231 route miles between Boston and New York to provide commuter rail services for more than 100,000 riders every weekday. This represents more than 90 percent of NEC riders and two-thirds of total passenger-miles on the corridor.

The corridor has long had a major role in intercity passenger travel, currently carrying 2.3 million riders annually between Boston and New York. Growth of airport and highway congestion has contributed to increased interest in improving passenger rail performance on the northern half of the NEC. The $2.5 billion Northeast Corridor Improvement Program of the 1970s and 1980s resulted in a reliable trip time under 3 hr for rail travel between New York and Washington, which in turn contributed to a high level of ridership. The shortest Boston–New York rail travel time is currently just under 4 hr, which has not proven to be competitive with air transport for many time-sensitive travelers on this route. The High Speed Rail Task Force of the Coalition of Northeastern Governors has been active for several years in identifying improvements that would yield a trip time of 3 hr or less and encouraging implementation of such improvements.

Much of the corridor's fixed plant, such as bridges and catenaries, is 80 years old or older. As a result, major rehabilitation and replacement have been identified as necessary simply to ensure safety and bring the railroad to a state of good repair. The Northeast Corridor Commuter Rail Authorities Committee has developed an extensive list of work identified by NEC operating authorities as needed for rehabilitation and service improvements. The responsible agencies are planning and conducting programs to meet those needs, but funding constraints are such that the necessary rehabilitation will take many years, and new needs continue to accumulate. Similarly, investments are being made to shorten trip times, but there is no assurance that funds will be available.

The multiple services that the corridor supports are reflected in a complex institutional structure that would shape the implementation of any major improvement program. Table 1 presents the division of responsibilities among various organizations for the Boston–New York portion of the NEC.

PURPOSE

The purpose of the study was to identify and characterize costs and benefits of improvements that could be achieved in commuter and intercity rail service on the Boston–New York portion of the corridor. Reported in this paper is the central effort of that study. The focus is on three basic questions:

1. What improvements are needed to ensure safety and continued reliable operations on the corridor?
2. What could be done to the NEC fixed plant infrastructure to achieve substantially faster and more reliable commuter and intercity rail service?

J. B. Hopkins, U.S. Department of Transportation, Volpe National Transportation Systems Center, Kendall Square, Cambridge, Mass. 02142. J. A. Harrison, Parsons Brinckerhoff Quade & Douglas, Inc., 120 Boylston Street, Boston, Mass. 02116.

TABLE 1 INSTITUTIONAL ROLES AND RESPONSIBILITIES FOR THE NEC BETWEEN BOSTON AND NEW YORK

From	To	Distance (miles)	Owner	Maintenance	Dispatching	Commuter Service	Commuter Authority	Freight Service
Penn Station	Harold Interlocking	4	Amtrak	Amtrak	Amtrak	LIRR	MTA	--
Harold Interlocking		--	LIRR	LIRR	LIRR	LIRR	MTA	--
Harold Interlocking	Shell Interlocking	15	Amtrak	Amtrak	Amtrak	--	--	Conrail
Shell Interlocking	NY-CT State Line	10	MTA	MNCR	MNCR	MNCR	MTA	Conrail
NY-CT State Line	New Haven	46	CDOT	MNCR	MNCR	MNCR	CDOT	Conrail
New Haven	Old Saybrook	33	Amtrak	Amtrak	Amtrak	Amtrak	CDOT	Conrail
Old Saybrook	RI-MA State Line	86	Amtrak[a]	Amtrak	Amtrak	--	--	P&W
RI-MA State Line	Boston	38	MBTA	Amtrak	Amtrak	Amtrak	MBTA	Conrail

a. Rhode Island DOT owns approximately 1/4-mile of track through and adjacent to Providence Station.

Abbreviations:
CDOT = Connecticut Department of Transportation
LIRR = Long Island Rail Road
MBTA = Massachusetts Bay Transportation Authority
MNCR = Metro-North Commuter Railroad
MTA = Metropolitan Transportation Authority
P&W = Providence & Worcester Railroad

3. What degree of rail service improvement is attainable for various levels of capital investment?

The study was neither a benefit-cost analysis nor a program plan for the improvement of the corridor. No recommendations were made. Rather, the purpose was to clarify the nature and cost of the primary potential improvements that could be made to the Boston–New York rail infrastructure and to provide estimates of potential trip-time reductions likely to result. As such, the study could provide a basis for developing the necessary consensus among owners, operators, and all levels of government for policy formulation and decision making concerning any future corridor improvements. It brings together, in a consistent and comprehensive manner, the results of previous studies, analyses, and estimates by the involved public agencies, operating railroads and others, as well as independent assessments by the study team.

APPROACH

The approach followed in the study was to identify major infrastructure rehabilitation and improvement projects and organize them into a logical hierarchy of overall programs. Project identification was based primarily on previous studies and extensive interaction with the involved parties, particularly the organizations shown in Table 1. Potential savings in intercity trip times from each of the five programs were calculated for various types of rolling stock using a proven Train Performance Calculator (TPC) computer program. The study was focused on fixed plant improvements; rolling stock was considered only in selecting a broad range of alternatives for trip-time calculations. Rolling stock investments and normal operating and maintenance costs were not analyzed.

Key assumptions of the study were as follows:

- Time frame: Project implementation and funding allocation is assumed to occur between 1991 and 2000.
- Route/right-of-way: Improvements considered are primarily within the existing right-of-way, with the exception of a new inland route segment recently studied by Amtrak.
- Rolling stock: Performance projections assume equipment now available or fully developed and tested so that it would be available for revenue service. Rolling stock costs would depend on the level of service and other operational variables.
- Speed on curves: The study assumes that with modern rolling stock and rehabilitated and reconfigured track higher curve speed limits will be acceptable in terms of safety and passenger comfort.

IMPROVEMENT PROJECTS

Eighteen candidate improvement projects were identified and characterized. Five involve safety or basic rehabilitation; thirteen others would improve trip time, reliability, and capacity. Based on initial estimates of time savings and cost, the projects were grouped into programs representing a hierarchical succession of trip-time reductions and total cost. Trip time was calculated for the speed limit profile appropriate to each of the improvement programs and repeated for several categories of rolling stock. Existing travel demand models were used to assess the ridership expected to result from the calculated trip times under reasonable assumptions concerning fare and departure frequency. A brief description of the identified projects affecting specific segments of the corridor follows. A map indicating project location is shown in Figure 1; track work, bridge rehabilitation, signaling, and curve straightening are too distributed to show.

Penn Station (New York) to Shell Interlocking (New Rochelle)

A major safety-related effort is required at Pennsylvania Station. Emergency egress from platforms must be increased, which will be a major undertaking. Similarly, the East River tunnels require new ventilation shafts and equipment, including evacuation stairways. At Harold Interlocking, a grade separation (flyover) where eastbound Amtrak trains cross Long Island Rail Road commuter tracks would prevent delays—

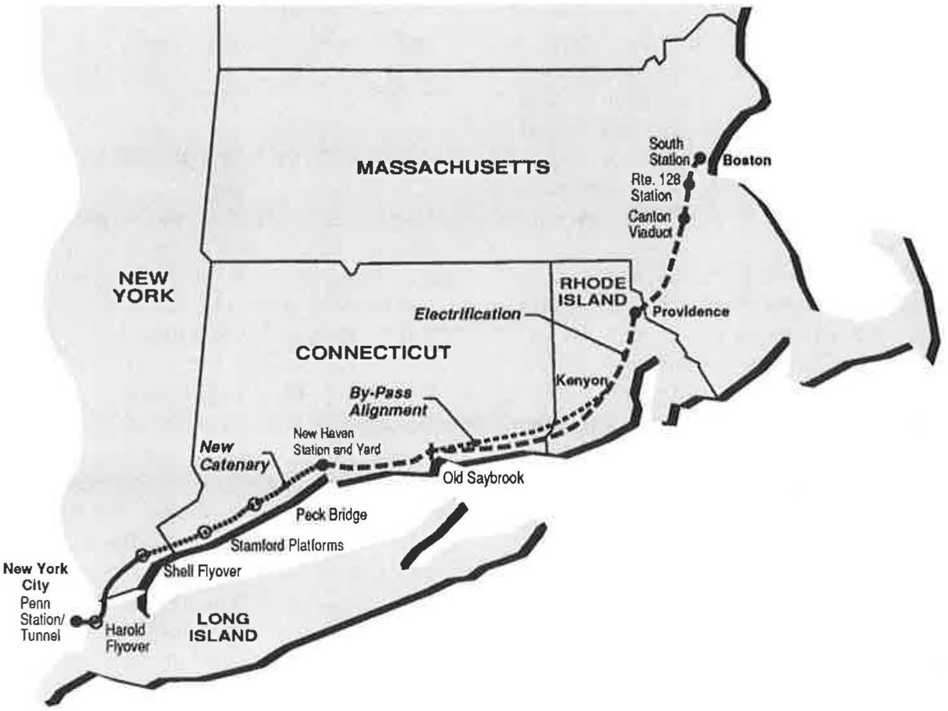


FIGURE 1 Location of potential improvements.

likely to be much more serious at future, higher traffic levels—but will be difficult to construct while bearing full traffic. Beyond Harold Interlocking on the Hell Gate Line, rehabilitation of the Pelham Bay movable bridge and fixed bridges, improved track, and a potential curve realignment with signal modifications for higher speed limits, are needed. Overhead catenary wire was recently rehabilitated, but the catenary supports still need to be repaired.

Shell Interlocking to New Haven

The Metro-North Commuter Railroad (MNCR) New Haven Line contains a major share of substantial location-specific projects. Shell Interlocking, where eastbound Amtrak trains merge with MNCR traffic, is a significant source of delay for both railroads, and low-speed turnouts limit operating speeds. Two approaches are being considered at Shell to reduce the time required to traverse the interlocking plant and eliminate a major bottleneck:

1. Flyover: Depression of the two eastbound MNCR tracks and elevation of the Hell Gate Line tracks on an overpass, and
2. At-grade: Changes to track configuration and turnouts in the vicinity of New Rochelle, which would increase speeds through the area and reduce conflicts.

Design alternatives are being evaluated by Amtrak, MNCR and FRA for this project.

Operationally closely linked to Shell, island platforms, and related track reconfiguration at Stamford are needed to increase platform access and avoid delays, which quickly propagate to New Rochelle. The catenary from the Connecticut–New York line to New Haven is approximately 80 years old and is well beyond normal service life. It now constrains speed and imposes an excessive maintenance burden; replacement is necessary. A project to replace Peck Bridge, a nominally movable bridge over the Pequonnock River at Bridgeport, has been initiated and must be completed; in addition to preventing a future safety problem, this will permit somewhat higher speeds. Four other movable bridges requiring major work, which is well under way, are those over the Saugatuck and Norwalk rivers and at Cos Cob and Devon. New track configuration at and leading into the New Haven station area would result in significantly faster speeds and improved operations through that area. Although not needed at present traffic levels, the fourth track between New Haven and Norwalk will require rehabilitation or replacement early in the next decade. All of these specific projects would be accompanied by ballasting of open deck bridges and track work and signal modification to permit higher speeds along the line.

New Haven to South Station (Boston)

Electrification of the entire route segment would be accompanied by track work and signaling to support higher speeds. In addition to conversion of open deck fixed bridges to ballasted deck, electrification would require that overhead bridge clearances be increased at many locations. There is also a potential for significant curve straightening, particularly between New Haven and Providence. The movable bridge span at Groton and over the Niantic River requires replacement. The viaduct in Canton, Massachusetts is more than 150 years old and needs substantial modification to allow high speed for certain types of commuter cars. High level platforms at Route 128 station would significantly reduce dwell time at that stop. Nine public and eight private at-grade crossings remain on the New Haven–Boston line—one each in Massachusetts and Rhode Island and the remainder in Connecticut. The Rhode Island crossing is scheduled for grade separation. Other crossings will require separation closure or added protection depending on the level of all service improvements to be implemented.

The Shore Line alignment between Old Saybrook, Connecticut, and the Connecticut/Rhode Island state line contains the most restrictive series of curves on the corridor, as well as five movable bridges over Shaw's Cove and the Connecticut, Niantic, Thames, and Mystic rivers. Although much of the alignment is rural and lends itself to curve realignment projects, achieving meaningful 150-mph stretches in this territory is precluded by various "hard spots" (such as the movable bridges), and 100 to 110 mph maximum speeds are the best that can be reasonably obtained. The movable bridges require substantial expenditures for rehabilitation or replacement and are a source of ongoing maintenance requirements and operating delays.

Amtrak has recently explored bypassing the most heavily curved segment of the corridor—the Connecticut/Rhode Island shore line east of New Haven. The project would consist of construction of a new right-of-way along a different alignment for a distance of more than 40 mi, plus major straightening on an additional 10 mi, with a speed limit of 150 mph for the entire 50 mi. Because of the substantially reduced curvature, the new alignment would permit 150-mph operation throughout its length. This by-pass alignment offers significant trip-time savings.

Summary

In summary, a list of the eighteen candidate projects identified, some aggregate in nature, follows.

System Rehabilitation

- Penn Station and tunnels,
- Catenary replacement,
- Peck Bridge replacement,
- Other movable bridges, and
- Fixed bridges.

System Improvement

- Harold Interlocking,
- Shell Interlocking,
- Stamford island platforms,
- New Haven terminal,
- New Haven – Norwalk 4th Track,
- Canton viaduct,
- Track improvements,
- Signal system upgrades,
- Grade crossings,
- Station improvements,
- Electrification,
- Curve realignments, and
- Bypass alignment.

ALTERNATIVE IMPROVEMENT PROGRAMS

Conceptually, a set of alternative programs could be defined by ranking improvement projects in order of cost-effectiveness in reducing trip time, with a hierarchy of programs resulting from working down the list. In practice, three considerations limit the rigor with which that approach can be followed. First, many of the projects are not well defined at present in scope or design, limiting the precision of both cost and time-savings estimates. This renders highly uncertain any explicit calculation of minutes saved per million dollars expended. Second, many improvements provide benefits only in conjunction with other projects. For example, the speed gains from simultaneous signal improvements, track work, and electrification cannot be allocated uniquely to any one of those projects. Third, for projects that address trip-time reliability or system capacity, there is no straightforward way to convert the benefit into minutes; they are simply necessary to creating an improved system.

In spite of these limitations, cost-effectiveness in trip-time reduction remains a useful measure. Ten projects—or appropriate clusters of projects, such as track work and signaling—were found to buy reduced trip time (through a combination of higher speeds and prevention of delays) at an approximate rate of $10 million to $20 million per minute saved. The next most attractive improvement, electrification, was found to be somewhat more expensive in terms of direct time savings, but it offers important additional advantages such as efficient Boston–Washington run-through service, fleet rationalization, and reduced locomotive maintenance expense. The remaining two projects—a program of curve realignments and a segment of new right-of-way—are significantly more costly (per minute saved) than the other projects identified. Each represents a sufficiently large increment in cost and performance to be embodied in a separate program in the hierarchy.

A hierarchy of five alternative programs was defined. All programs include a basic set of five projects needed to maintain safety and rehabilitate the existing infrastructure. The first (System Rehabilitation) consists only of these five projects. The other four include concurrent implementation of system improvement projects, and offer shorter trip times but require higher levels of funding. Each successive program includes all of the projects in the less ambitious programs. Most of the projects that constitute the System Rehabilitation program are already in progress.

Program 1: System Rehabilitation

The System Rehabilitation Program consists of five projects necessary for improved safety and for replacement of major system elements that have exceeded their normal service life. This program represents a continuation of a process now in progress. More than half of the projects are at least partially funded. More than $100 million has been obligated to date. The various responsible agencies are developing long-term plans covering most of the projects that constitute Program 1, although funding constraints limit the pace of implementation.

The rehabilitation projects would be needed for continued safe and efficient operation, in essentially the same form, in the absence of any speed and reliability improvement efforts. Thus, they are necessary elements of all system improvement programs, but need not be completed before initiation of system improvement projects. Program 1 includes two safety projects: replacement of Peck Bridge, and fire safety ventilation and other improvements to Penn Station and the East River tunnels. It provides a necessary framework for substantial improvements in speed.

Program 1 yields improved reliability and reduced trip time for commuter and intercity services. Boston–New York schedules would be shortened by several minutes, primarily by greater speeds at some movable bridges and use of two diesel locomotives (instead of one) between Boston and New Haven. Maximum operating speed is 110 mph. The currently unfunded portion of the cost of this program is estimated to be $1.1 billion (in 1991 dollars). Approximately one-third of this sum has been programmed by the various operating authorities, based on expected funds availability during the next decade.

Program 2: Basic System Improvements

The Basic System Improvement Program includes the five projects in the System Rehabilitation Program as well as ten projects to improve service reliability and speed. More than 30 min can be cut from intercity running time by track work and signaling in conjunction with higher allowed speeds on curves, which increases running speeds to a maximum of 130 mph. Modernization of the New Haven terminal area will eliminate an extended region of slow speeds, cutting an additional 5 min from the trip. Other projects are necessary for capacity enhancement and grade crossing improvements. Flyovers (grade separations) at Harold Interlocking (Queens) and Shell Interlocking (New Rochelle) would ensure service reliability and avoid serious delays at critical points where intercity and commuter lines merge or cross. Island platforms at Stamford would greatly improve flow and capacity at peak hours.

The rehabilitation projects need not be completed, nor some even initiated, before beginning speed and reliability improvements. Most of the system improvement projects in Program 2 are already at least in the preliminary design phase. The best Boston–New York average speed attainable with this program is 75 mph. The program adds a cost averaging $50 million per year over 10 years to the system rehabilitation program, but yields a trip time approaching 3 hr for Boston to New York. Significant time savings and greater service reliability are also achieved for the many rail commuters in the New Haven–New York area.

Program 3: Basic System Improvements and Electrification

Program 3 adds electrification of the route from Boston to New Haven to the projects of Program 2. Electrification, for which initial design funds have been appropriated, eliminates the engine change in New Haven, a saving of almost 9 min, and allows use of electric locomotives for the Boston–New Haven segment. The electric units, with higher acceleration, operating at up to 130 mph, will further reduce trip time by almost 6 mins. Electrification, including associated signal upgrade and bridge clearance projects, also facilitates run-through operation between Boston and Washington, necessary for improving Pennsylvania Station and East River Tunnel capacity and providing high speed service to and from points south of New York. Program 3 yields an average speed for express service, depending on rolling stock, of slightly above 80 mph, with a projected trip time slightly less than 3 hr. Significant time savings are achieved for commuters in the New York area and potentially in the Boston area.

Program 4: All System Improvements and Electrification

This program includes all projects in Program 3 and adds 27 clusters of curve realignments, primarily between Providence and New Haven; maximum speed is 130 mph. The realignments would generally be within existing right-of-way or require only modest excursions from it. They would yield an average speed of about 85 mph. These improvements provide an additional reduction in trip time of about 11 min; the Boston–New York trip could be completed in less than 2¾ hr. Candidate curve realignments are discussed in more detail later.

If the Boston–New Haven line were electrified before implementing realignments, the cost of subsequent curve straightening would be substantially increased. Thus, a choice between Programs 3 and 4 must be made before implementation; Program 4 would not be as practical as a later upgrade from Program 3.

Program 5: Shore Line Bypass

Program 5 adds to Program 4 a new routing to avoid the most curve-intensive portion of the route. The Shore Line Bypass, recently examined by Amtrak, is a 50-mi-long, 150-mph right-of-way to replace the most curved section of the route along the Connecticut and Rhode Island shore east of New Haven. This could yield an average speed of approximately 95 mph. Some elements of Program 4, such as certain curve realignments and bridge rehabilitations, would not be needed if a bypass were constructed. The Boston–New York trip time could be 2½ hr or better, depending on the operating equipment.

In summary, the four system improvement programs yield projected Boston–New York trip times of from slightly more than 3 hr to less than 2½ hr, depending on the level of investment and the rolling stock used, along with substantial speed and reliability benefits for commuter rail service.

PERFORMANCE AND COST OF THE PROGRAMS

The projected minimum running time between Boston and New York City for express (Metroliner-type) service and the

TABLE 2 PROJECTED MINIMUM RUNNING TIME BETWEEN BOSTON AND NEW YORK FOR ALTERNATIVE IMPROVEMENT PROGRAMS

Rolling Stock:	1. System Rehabilitation	2. Basic System Improvements	3. Basic System Improvements and Electrification	4. All System Improvements and Electrification	5. Shore Line Bypass
Current Diesel/ Electric (NEC)[a]	3:47	3:07	System Fully Electrified		
Current Diesel/ Electric with Tilt	3:46	3:02	Diesel-Electric and Gas Turbine Not Applicable		
Current Turbo (Empire Line)[b]	3:48	3:21			
Electric[c]			2:52	2:41	2:29
Electric/Tilt	System Not Fully Electrified		2:47	2:37	2:28
High-Speed Electric[d]	Electric Propulsion Not		2:46	2:35	2:22
High-Speed Electric/Tilt	Usable Over Full Route		2:41	2:33	2:21
Total Program Cost ($B)	$ 1.1 B	$ 1.6 B	$ 2.0 B	$ 2.7 B	$ 3.6 B[e]

a. 2 F40P diesel-electric locomotives Boston-New Haven; AEM7 electric New Haven-New York; 10 min. change.
b. Gas Turbine-powered equipment comparable to that used for current Amtrak Empire Line service.
c. 1 AEM7 locomotive, modified for 150 MPH for Program 5; use of 2 AEM7's improves time by 5 minutes.
d. Lightweight, high-powered equipment comparable to TGV or ABB trainsets.
e. Estimate includes adjustment for movable bridge and curve projects made unnecessary by the bypass.

All trains consist of six coaches and make 1 ¼-min. stops at Back Bay, Route 128, Providence and New Haven. Computed times are increased by 5% to allow for operational variability and uncontrollable delays. All programs assume acceptability of higher speeds on curves than are now allowed (6" superelevation, 6" unbalance for conventional coaches and 8" for tilt suspensions).

estimated cost (in 1991 dollars) for each program are summarized in Table 2. Run times are based on computer simulation plus a 5 percent schedule allowance for normal variations and delays. The engineering cost estimates exclude any funds already available for projects in each program. The contribution of specific improvement projects and groups of projects to trip-time reduction, improved reliability, and other service benefits are presented in Table 3.

The four intermediate stops of Amtrak's present New England Express schedule (Back Bay, Route 128, Providence, New Haven) were assumed in travel time estimates. Six-coach trains were selected for train performance calculations, as is consistent with proposed future express service. The trip times in Table 2 are judged to be the best that might be achieved. Reliable attainment of those values would require full validity of all assumptions as well as railroad operations that meet the highest standards of precision and reliability. Practical scheduled running times could be several minutes greater than the values presented in Table 2. An additional stop in Stamford, which is likely for many trains, would add approximately 3 min.

Four motive power alternatives were considered in establishing the range of trip time that fixed plant improvements

TABLE 3 APPROXIMATE TRIP-TIME REDUCTION AND OTHER SERVICE BENEFITS FOR SPECIFIC SYSTEM IMPROVEMENTS

Improvement Project	Time Saving due to Higher Speeds[a]	Necessary for Capacity and Reliability	Program 2. Basic System Improvements	Program 3. Basic System Improvements and Electrification	Program 4. All System Improvements and Electrification	Program 5. Shore Line Bypass
Canton Viaduct Track Signal System Grade Crossings Station	33 min.	X	These projects, taken together, permit higher speeds--up to 130 MPH, rather than 110 MPH as for Program 1. Approximate trip time reductions: Boston-New Haven, 19 min.; New Haven-New Rochelle, 12 min.; New Rochelle-NYC, 2 min.			
New Haven Terminal Area	5 min.		Permits substantially higher approach and exit speeds in both directions, reducing trip time by approximately 5 minutes, as well as facilitating terminal-area operations.			
Stamford Island Platforms	1 min.	X	Essential to reduce delays, relieve congestion, prevent conflicts at New Rochelle, and increase capacity for commuter service. Also permits higher speed for through trains, gaining 1 minute.			
Shell Interlocking	1 min.	X	Necessary for future capacity and to prevent serious delays to intercity and commuter trains at a high-traffic merge point; it also permits a higher speed and saves approximately 1 minute.			
New Haven-Norwalk 4th Track		X	Required in order to provide adequate capacity for the higher future levels of intercity and commuter traffic.			
Harold Interlocking		X	Major merge and intersection point between intercity and very high commuter rail traffic; separation required to avoid frequent lengthy delays.			
Electrification	15 min.			Electrification saves 9 min. in New Haven by eliminating the engine change. Higher top speed and greater acceleration gains 6 min. between Boston and New Haven.		
Curve Realignments	11 min.				Realignment of curves is projected to shorten trip time by 11 minutes.	
Bypass Alignment	12 min.					150 MPH; saves about 12 minutes.

a. Approximate time saving with respect to trip times for program 1; does not include reduction in pad or improved reliability.

could yield (Table 2). Assessment of the suitability of specific rolling stock to actual corridor operations was not within the scope of the study. The current-technology diesel and electric units on which trip-time projections are based are assumed usable with either conventional coaches or with cars having a tilting suspension, which would permit somewhat higher speed on curves and typically yields about a 5-min time saving. The high-speed electric equipment represents advanced technology now in use in Europe.

The trip-time estimates for turbine power are patterned after equipment now in service on Amtrak's Empire Line, for which two power cars together have a net of 2,280 hp. A version that would use twin turbines of newer design on each power car, with a total power of 5,800 hp, has been proposed. If this equipment were successfully developed and tested, trip time would be substantially improved. However, even an advanced turbine train would be likely to have weaknesses in NEC service. It would not be suited for Boston–Washington run-through service, and concerns about third-rail operation in tunnels and operational reliability and flexibility would have to be resolved.

Rolling stock cost and operating and maintenance expenses were not analyzed in any detail. However, a rough estimate of rolling stock capital cost is possible. Trainsets, consisting of two power units and six coaches, are expected to cost about \$20 million each. As many as 15 to 20 such trainsets might be needed to augment the existing fleet, depending on the program selected and the resulting ridership levels. This implies an incremental cost of \$300 million to \$400 million over several years.

The improvement program to be implemented, which defines the overall rail system of which each project is a part, must be defined before detailed design of that project and sequencing of construction can be completed. Some projects have direct logistic connections with one another, as with track work, signaling and electrification. Others are linked operationally, such as Stamford platforms and improvements at Shell Interlocking, or are connected through the need to minimize disruption of traffic during construction. Improving the corridor one project at a time, without clear definition of the planned end state, would be inefficient and yield poor results.

CURVE SPEED LIMITS AND REALIGNMENTS

Overcoming the basic physical constraint of a highly curved route alignment is central to reducing Boston–New York trip time. The improvement programs described previously approach this challenge in steps. In addition to projects to raise track classification, provide signaling appropriate to high speed, improve bridges, and remove grade crossings, Program 2 includes increasing superelevation to 6 in. wherever possible, and trip-time computations are based on the assumption of acceptability of a 6-in. unbalance. Program 4 goes a step further by incorporating numerous changes in alignment, largely within the existing right-of-way, to reduce curvature. Program 5 is still more aggressive, replacing approximately 50 mi of the route with an alignment sufficiently straight to permit continuous 150 mph operation. The increased superelevation and unbalance, as well as the curve realignments, are critical to the projected trip-time reductions for Programs 4 and 5 and warrant further discussion.

The VNTSC study included examination of each of the approximately 238 curves between Penn and South stations. Along with the curves, other speed restrictions along the existing route were identified. No detailed field examinations were possible in this preliminary feasibility study. Rather, each curve or restriction was examined through each of the following sources:

- Railroad track charts and curvature listings,
- Railroad track geometry car measurements,
- U.S. Geological Survey topographic mapping,
- Railroad valuation maps, and
- Video recordings of right-of-way, which were provided by Amtrak.

The first finding of the investigation was that track superelevation is significantly less than the existing alignment could support on both Amtrak and MNCR territory. Thus, the first step was to determine the reduction in trip time if full superelevation were restored.

A theoretical track data set was developed as input for the TPC computer program that consisted of maximum speed limits or maximum authorized speeds for each section of the existing New York to Boston alignment. Several assumptions were embodied in this process:

Six in. of actual track superelevation (*Ea*) was assumed to be present at restricting curves in the route, with certain exceptions, as on the Hell Gate Bridge and in terminal and station areas. Although this amount of superelevation can generally be achieved, isolated instances will occur in which it will prove more complex (and hence more costly) to achieve than is practicable. This may be the case, in particular, at certain MNCR locations at which open deck bridges, curves, existing high-level station platforms, and track centers interact. However, full superelevation appears to be achievable in large measure.

Superelevation

Achievement of full superelevation requires adjustment to the transition curves (spirals) associated with the existing curves. Engineering specifications for spirals generally address two concerns: passenger comfort, which is speed dependent, and car twist, which is not speed dependent but depends on the degree of equalization provided by the rolling stock. Many of the specifications still in use today have their origins in the distant railway past, when truck equalization was relatively primitive.

Amtrak and MNCR specifications are relatively conservative in comparison to those used elsewhere in the world and are more stringent than FRA safety standards. The use of these specifications increases required spiral lengths beyond those used by other passenger carriers and limits the amount of superelevation placed in curves. It was beyond the scope of the VNTSC study to develop standards for spirals, but it appears that future testing, analysis, and refinement of these standards will show that, to a large degree, existing spiral transitions and runoffs can be adequate for full superelevation.

Allowed Unbalance

Unbalanced elevation (*Eu*), or cant deficiency, is a measure of the lateral force on a passenger caused by traversing curves at speeds in excess of equilibrium, or balanced, speed. (Equilibrium speed is the speed at which track superelevation exactly balances this lateral force, and $Eu = 0$). Unbalance affects passenger comfort, but in the range in question does not affect train safety or track stability. Railways in Europe operate priority passenger trains with at least 5-in. unbalance and also allow track elevations in excess of 6 in.; total elevation ($Ea + Eu$) can approach 12 in. on conventional (non-tilt) trains.

Recent tests in the NEC have shown that conventional Amfleet cars can operate at 4 to 5-in. unbalance with acceptable passenger comfort. Amtrak has successfully petitioned FRA for permission to operate over designated NEC curves at 5-in. unbalance under specified conditions. The VNTSC curve analysis assumed that with advanced-technology rolling stock (TGV coaches are a well-tested example) a total elevation of 12 in. can be achieved in the NEC. Tests will be required to prove the validity of this assumption, but there is precedent for it internationally. It is an important element of maximizing performance in the existing NEC.

Curve Realignments

A graphic plot of speed versus distance for TPC runs was analyzed to determine the potential for time savings through curvature reduction. In particular, individual curves were examined in the context of their neighbors, and clusters of curves were isolated that which would need to be realigned together in order to achieve significant time savings in a coordinated manner.

The TPC speed limit input data set was modified to reflect increased radius of curvature, and revised travel times were obtained. For each curve, the amount of track shift required was calculated. On the basis of the data sources described previously, adjacent development, wetlands, and terrain were identified, and basic feasibility, cost, and likelihood of environmental constraints were estimated.

Several realignments require modest shifts that are within the existing rail right-of-way. For several other curves the land on either side of the railroad is owned by the same landholder, with the possibility of land swaps. Because the right-of-way is up to 250 ft wide in some locations, the possibility of releasing land from railroad use in conjunction with a realignment is also possible, theoretically allowing an increase of wetlands.

Of the 34 realignments projects examined, 33 are clusters of one to six curves, having estimated costs ranging between \$0.5 million and \$88 million per project. One cluster consists of 11 curves shifted to a largely new alignment over 18 mi between Westerly and Kingston, Rhode Island, estimated to cost \$262 million.

Time savings were estimated using the TPC, assuming TGV 1-6-1 equipment, and 1 AEM-7 locomotive with 6 Amfleet coaches. Time savings for individual clusters range from several seconds to 3½ min for the TGV and up to 2¼ min for the AEM-7 consist. Aggregate time savings of approximately 12 and 10 min are achieved by all realignment projects for TGV and AEM-7, respectively.

The cost-effectiveness of specific realignments ranges from approximately \$31 million per minute saved on the Hell Gate Line to \$142 million per minute saved on the Stamford-Bridgeport segment. The average cost per minute saved is \$70 million overall, or \$59 million excluding Shell-Bridgeport.

CONCLUSIONS

1. Rehabilitation Program 1, with a cost of about \$1.1 billion, is needed to ensure safety and maintain the present level of intercity and commuter rail service between Boston and New York. Some of this work has been initiated by the responsible agencies. These projects will contribute to corridor safety and reliability well into the next century.
2. Trip time can be improved substantially using existing technology and with little or no excursion beyond the existing NEC right-of-way. The time for a trip from Boston to New York could be reduced to approximately 2½ to 3 hr, depending on rolling stock and level of investment.
3. Much of the investment would be in segments heavily used by commuter rail passengers. These commuters would experience long-term service improvements comparable to those for intercity riders, as well as increased system capacity.
4. The currently unfunded cost of the improvements necessary for substantially reduced trip time, in addition to the \$1.1 billion for rehabilitation, would range from \$500 million to \$2.5 billion in 1991 dollars, depending on the level of trip-time improvement sought. Initial work is being undertaken on many of the needed projects, although only a small part of the needed funding has been identified and no coordinated overall program exists. The programs could be implemented within 8 to 10 years; service improvements could be apparent within 5 to 6 years. The necessary additional rolling stock (15 to 20 trainsets) is estimated to cost approximately \$300 to \$400 million.
5. Commuter and intercity schedules and service reliability will suffer during implementation of any major improvements; the degradation of commuter service between New Haven and New York could be significant for several years. A concerted effort will be required to design and sequence the improvements in a manner that minimizes disruption of service.
6. Commuter railroads will be subject to new operating constraints, costs, and requirements concerning track maintenance, compatibility of rolling stock, and dispatching. This will require arrangements for equitable sharing of costs, responsibilities, and access between commuter and intercity operations.

OTHER CONSIDERATIONS

Rolling Stock

The selection of a rolling stock alternative depends not only on the trip time it makes possible, but also on capital, operating, and maintenance costs; reliability; suitability for run-through operation between Boston and Washington; and other characteristics and operational considerations.

The performance of advanced-technology high-speed foreign trainsets in the U.S. railroad environment remains to be evaluated. Demonstrations, trial use, and testing of a variety of motive power and rail car suspension technologies during the lengthy period of fixed-plant improvements would provide a good foundation for future long-term fleet acquisition decisions.

Electrification

Electrification between Boston and New Haven has important benefits and implications beyond travel time. Operationally, electrification harmonizes operations in the north and south ends of the corridor, making it possible to use high-performance electric trainsets running between Boston and Washington, with few trains being turned around in New York. This provides needed capacity at Pennsylvania Station and in the tunnels serving it.

Corridor Capacity

Corridor capacity was not explicitly examined in this study. On the basis of the improvements defined in Program 2 as a minimum, capacity appears to be adequate for anticipated commuter and intercity traffic through 2010. At Pennsylvania Station and the East River tunnels operational improvements or changes may be required to avoid serious impacts, particularly on commuter operations. At other locations the system will be near or at its limit, and a concerted and integrated effort will be required to maximize corridor capacity for all services.

Operating Standards

The projected higher speeds in all programs are based on the assumption that FRA, MNCR, and Amtrak will approve higher speeds on curves and define standards for rolling stock and inspection and maintenance procedures necessary for safe and comfortable operation at those speeds.

Institutional Coordination and Integration

Successful implementation of any major improvement program and practical attainment of the trip times estimated in this study will require a reinvigorated institutional and procedural framework. The direct responsibilities and objectives of the several owning and operating organizations differ significantly. The specific form of some projects, as well as the manner of implementation and cost allocation, can only be determined through compromise based on full consideration being given to all viewpoints. All parties—railroads, government agencies at all levels, and transportation authorities—will need to work in a highly coordinated and cooperative manner to define and realize a common vision of integrated NEC rail services with equitable distribution of all capital and operational costs.

Accessibility of Railroad Stations

The Americans with Disabilities Act of 1990 established specific accessibility standards for physically handicapped passengers for intercity and commuter rail stations and passenger cars. The station improvements project in this study includes an estimate for provision of high level platforms and pedestrian overpasses at those Amtrak stations between Boston and New York not currently so-equipped. However, the special nature of the requirements of this act is considered beyond the general scope of the study, particularly insofar as commuter stations and rolling stock is considered.

NEXT STEPS

Possible alternative improvement programs were identified and characterized in the study described here. No recommendations were made, and no specific plan for upgrading the NEC was presented. Any program of the complexity and magnitude associated with the NEC would have to be preceded by attainment of consensus among the many involved private and public bodies as to goals, funding, and implementation process. Were a program to be initiated, other topics would need to be addressed to support design, construction, and scheduling decisions for any improvements. Logical next steps for any major improvement program should include the following:

1. Testing and analysis to confirm the acceptability of higher speeds on curves and to define standards necessary for safe and comfortable operation at those speeds;
2. Analysis of long-term operating and maintenance costs of alternative improvement programs and rolling stock choices;
3. System capacity and traffic conflict analysis, addressing both long term outlook and impact on phasing of construction projects;
4. Data collection and analysis to refine ridership projections and expected commuter and intercity benefits; and
5. Examination of the future role of rail freight transportation along the corridor and the freight railroad impacts and benefits associated with corridor improvements.

ACKNOWLEDGMENTS

The study described here was carried out under the overall direction of a DOT task force consisting of the UMTA (now FTA) and FRA administrators and the Counselor to the Secretary of Transportation. The program manager was S. Barsony of UMTA, and the VNTSC effort was directed by R. Madigan. J. Hopkins was responsible for technical coordination of the study team and integration of its efforts. Key VNTSC staff members on the team included P. Mattson, M. Safford, and D. Pickrell. Extensive technical support was provided by Parsons Brinckerhoff Quade & Douglas, Inc., under the direction of J. Harrison; K. Ullman conducted the curve analysis, a critical element of the study. Estimates of future ridership, not presented in this paper, were developed by Charles River Associates. The study benefited greatly from the active cooperation of many involved organizations and individuals; appreciation is expressed particularly to R. Rathbun (Connecticut Department of Transportation), E. Courtemanch (Amtrak), and H. Permut (MNCR).

Publication of this paper sponsored by Committee on Intercity Rail Passenger Systems.

Forecasting High-Speed Rail Ridership

DANIEL BRAND, THOMAS E. PARODY, POH SER HSU, AND
KEVIN F. TIERNEY

Advantages and disadvantages of various high-speed rail (HSR) ridership forecasting approaches are summarized, and a recommended forecasting approach is presented. The recommended approach involves the use of separate relationships to estimate the diversion from each existing mode to HSR. This approach makes use of the behavioral information travelers have already provided by their revealed preferences to use existing modes for intercity trips. The choice of current modes for specific trip purposes reveals a great deal about how individuals value the attributes of that mode relative to other modes. This information is also of use in estimating induced demand. The approach presented here has been used in forecasting HSR ridership and revenue in Florida, Texas, and the Northeast Corridor. To illustrate how different factors influence the demand for HSR, model results are presented along with implied values of time and selected demand elasticities. The variation between market segments for the various components of travel time and cost is strong evidence that this approach is necessary for forecasting HSR ridership. The resulting models are also shown to be transparent in providing design information for new mode applications that can be used to maximize ridership, revenue, or the public benefits that justify public subsidies for the new modes.

A particular approach that has been used to forecast ridership for proposed high-speed rail (HSR) lines between a number of cities in the United States is described in this paper. (The use of the term "rail" here does not preclude maglev systems, which technically do not operate on rail tracks but otherwise share certain common characteristics with HSR systems.) A review of the procedures used most recently to make projections of HSR ridership in the United States reveals a wide variety of approaches. In the first section of this paper, these approaches and their strengths and weaknesses are reviewed to derive a recommended approach for forecasting HSR ridership.

ALTERNATIVE APPROACHES FOR FORECASTING HSR RIDERSHIP

Description

Although details may differ, three basic approaches have been used recently to forecast ridership for HSR systems. The first approach involves projecting total origin/destination (O/D) travel for the forecast year(s) and using a multinomial mode choice model to determine the share, and thus the number, of trips that would be made by each existing mode and by the new HSR mode. The share of trips by all modes would sum to 100 percent. Typically, a multinomial logit functional form is used for this mode choice step. However, because of the independence from irrelevant alternatives (IIA) property of the logit (or any multimodal share) model, individuals traveling on the new mode are automatically forecast to be drawn from other modes in direct proportion to the share of trips made on the existing modes.

The second approach begins in the same manner as the first, with forecasts of total O/D travel by all modes. To minimize potential IIA problems, a nested choice modeling (e.g., logit) procedure is used to separate automobile trips from common carrier trips. A subsequent choice model is used to separate common carrier trips into those made by air versus HSR. Sometimes the bus mode is also included in this latter choice set. However, in most intercity corridors, bus trips are small in number, and individuals who use buses are much more sensitive to price than to time. Therefore, these trips can usually be ignored in the final analysis of HSR demand.

In both of these approaches, HSR is sometimes treated as a new mode, whereas in other instances, HSR trips are estimated by assuming that the existing rail mode will simply go faster (along with accompanying changes in fares or frequencies or both).

The third approach to forecasting HSR ridership involves projecting trips that would be made by each existing mode, and then determining with separate mode choice models, the share of trips by each mode that could be expected to shift to the new HSR mode as a function of relative service characteristics and other factors found to be important. Just as it is commonly accepted that individual behavior varies by trip purpose (e.g., business versus nonbusiness), this approach recognizes explicitly that individuals traveling on different existing modes exhibit different behaviors when confronted with the choice or opportunity to use HSR. This is because individuals traveling on existing modes have widely divergent values of time and demand elasticities and place different values on the convenience and flexibility attributes of travel by automobile versus common carrier modes.

Assessment

Perhaps the most significant disadvantage of the first approach is the IIA property of multinomial mode choice models mentioned previously. Unless otherwise ameliorated (through use of the second or third approach, for example), the IIA property of models of this type will indicate that the share of riders diverted to HSR will come from other modes in direct proportion to the share of trips made on these other modes. For example, if 80 percent of the trips between any two O/D zones

---

Charles River Associates, 200 Clarendon Street, Boston, Mass. 02116-5092.

are made by auto, then the first approach will indicate that approximately 80 percent of the HSR trips made will be diverted from auto. Analysis of survey results in Florida and Texas indicate that this is not likely to be true (*1,2*).

The main problem with the second approach is determining how the level-of-service characteristics of the existing air and proposed HSR modes (i.e., the high-speed common carrier modes) should be combined for use in the choice model that interacts with auto. This same situation also occurs in the analysis of urban travel demand. Theoretically, it should be dealt with through the use of inclusive prices (or log-sum terms). However, even this approach has certain problems, which are outlined later. In some cases, a rather simplistic all-or-nothing approach is assumed.

In the all-or-nothing approach [used, for example, in the forecasts of HSR ridership for the Texas TGV (*3*)], forecasts of the diversion of trips from automobile to HSR do not incorporate both HSR and air into the level of service for the high-speed mode. Instead, this method assumes first that only air exists, and second that only HSR exists. The highest share is used in each instance. Therefore, if HSR is just slightly less attractive than air, this approach would indicate that zero HSR trips are diverted from auto. All else being equal, this approach is likely to underestimate HSR ridership.

The problem with the use of log-sum or inclusive price terms is that the nesting coefficient on these terms does not change the basic relationship (or set of trade-offs) between the values of the various components of travel time and cost that determine (or derive from) individual preferences for different modes. Because individuals who choose to travel by auto, air, bus, or conventional rail trade off the times, costs, and convenience of those modes differently, separate relationships (models) must be used to forecast the diversion of travel from each existing mode to the new high-speed mode.

The third approach, which is recommended here, recognizes that existing travelers have already revealed, or exhibited, their preferences for the available modes by the choices they have made. Thus, it is necessary only to determine, for each mode and trip purpose market segment, what percentage of travelers will divert to HSR for the service levels assumed. (As indicated later, induced demand is treated separately.) With this approach it is possible to examine functional forms and variable specifications that differ for each market segment. This is not the case for the first two approaches.

RECOMMENDED FORECASTING APPROACH FOR HSR RIDERSHIP

The recommended approach for forecasting HSR ridership (Approach 3) can be described as a three-step process:

1. Estimate demand for travel between O/D pairs by each of the existing modes (and market segment/trip purpose);
2. Quantify the diversion from each existing mode to HSR (by market segment); and
3. Compute the amount of induced travel on the HSR mode.

In summary, the total travel market is broken down into a number of mutually exclusive and readily definable mode and trip-purpose market segments that exhibit distinct patterns of travel behavior. Overall ridership forecasts are prepared by summing across market segments. This approach avoids the forecasting of completely arbitrary diversions of travel from existing modes that do not account for the great variation in the substitutability of the new mode for the various current modes. It also allows for differences in the trade-offs among time, cost, and comfort that characterize travel behavior in different market segments.

Each of the three steps is described in more detail in the following section.

Step 1

In the first step, the total volume of trips by each of the existing modes for a particular time period can be estimated using a direct demand model with the following functional form:

$$T^{m}_{\mathrm{OD}} = f(P_{\mathrm{OD}}, I_{\mathrm{OD}}, LOS_{\mathrm{OD}}) \tag{1}$$

where

T^{m}_{OD} = number of trips by mode *m* made between O and D,
P_{OD} = population levels in O and D,
I_{OD} = income of travelers between O and D, and
LOS_{OD} = level of service on existing modes between O and D.

Given total demand models of the type shown in Equation 1, projections are made of the number of trips on existing modes in future years (i.e., in the absence of HSR) given projected changes in various input variables (such as the population, income, and level-of-service terms shown in Equation 1).

Step 2

The share of total trips (by trip purpose) made by each existing mode that can be expected to divert to HSR can be estimated using the following functional relationship (a separate relationship is estimated for each existing mode and purpose market segment):

$$S^{m,\mathrm{HSR}}_{\mathrm{OD}} = f(Time^{m,\mathrm{HSR}}_{\mathrm{OD}}, Cost^{m,\mathrm{HSR}}_{\mathrm{OD}}, Frequency^{m,\mathrm{HSR}}_{\mathrm{OD}}, Constant^{m,\mathrm{HSR}}) \tag{2}$$

where

$S^{m,\mathrm{HSR}}_{\mathrm{OD}}$ = share of existing mode m trips between O and D that will divert to HSR;
$Time^{m,\mathrm{HSR}}_{\mathrm{OD}}$ = components of access, egress, and line-haul travel time for mode *m* and HSR;
$Cost^{m,\mathrm{HSR}}_{\mathrm{OD}}$ = components of access, egress, and line-haul travel cost for mode *m* and HSR;
$Frequency^{m,\mathrm{HSR}}_{\mathrm{OD}}$ = measures of the frequency and terminal processing times for mode *m* and HSR; and
$Constant^{m,\mathrm{HSR}}$ = effect of the unobserved characteristics of HSR relative to mode *m*.

As discussed previously, this approach makes use of the critical finding that people who travel by air, rail, and automobile exhibit different behavior when confronted with the choice or opportunity to use HSR. This means that current and future air, auto, and rail users will divert to HSR in different proportions when offered the same HSR option.

For example, people who choose to drive 4 or 5 hr between cities that are 200 to 250 mi apart can be expected to place a lower value on line-haul time than those who take a 1-hr flight to cover the same distance. Conversely, it is expected that automobile users place a high value on the privacy and convenience of their car, which allows them complete departure time flexibility, control over the rest of their travel schedule (such as making stops along the way), and the ability to take children and extra luggage at no additional cost.

Travelers who have already revealed, or exhibited, these different values will therefore respond quite differently to the travel time, fare, and comfort levels offered by HSR service relative to the mode they currently use. Disaggregating the market in this manner yields results that represent how individuals actually behave in making intercity travel decisions. Of course, the actual diversion to HSR from air, rail, and automobile in any corridor will depend on the actual speeds, fares, frequencies, station locations, and amenities of the new rail service.

Estimating the share of trips by HSR for each market segment using Equation 2 allows an empirical examination of a wide range of explanatory variables. For example, separate line-haul time, access and egress time, wait time, and travel cost variables can be specified. Alternatively, various combinations or transformations of these (or other) terms can be identified. In modeling urban travel behavior, a typical observation is that out-of-vehicle time has about twice the effect of in-vehicle time. When modeling intercity travel behavior, however, this relationship is likely to vary—at least by trip purpose (given different travel party sizes) and trip distance. For instance, when the length of a trip is relatively short, access and egress travel times (or more generally, impedances) may be significantly more important than line-haul time. Conversely, for relatively long trips, the value of access and egress times (as a percentage of line-haul times) is reduced. This result has been noted in two recent major studies (*1,2*).

Because HSR does not yet exist in the United States, it is not possible to use revealed preference techniques to determine how travelers actually trade off characteristics between existing modes and HSR. Solving this problem involves the use of stated-preference survey techniques to measure traveler perceptions and preferences for new modes. The first application of this approach using ordered logit models was the estimation of the demand for electric vehicles (*4*). Subsequently the approach has been used to examine a wide range of issues, from rail station choice (*5*), to telecommunications (*6*), to transit capital improvements (*7*). In the present instance, surveys administered to individuals making relevant trips by current modes can be used to obtain stated preference information on how travelers make trade-offs among different components of time, cost, and technology. This approach has been used to develop models (as illustrated by Equation 2) and estimates of HSR use in Florida, Texas, and elsewhere (*1,2*).

Step 3

Induced travel is estimated in the third step by incorporating the mode choice model utility functions into Equation 1. (Induced travel demand can be defined as trips not currently being made either on other existing modes or to alternative destinations. Induced travel does not include future new trips made because of normal population or employment growth.) In practice, induced travel for a new mode should be closely tied to its market share or its attractiveness relative to existing modes. A new mode that captures 30 to 40 percent of an existing market will probably induce its own trips. If, alternatively, it attracts only 1 percent of existing trips, it is unlikely that much induced travel can be expected.

Because of the relationship between induced travel and modal choice, the methodology for forecasting induced travel must be consistent with the models for forecasting mode choice. This means that the values travelers place on the level-of-service variables from the mode choice models are incorporated into the total demand models (Equation 1). This is done using the mode choice model coefficients that equate the level-of-service improvements on the new mode to the effects on total travel of service improvements on the existing modes. Using this approach will guarantee that the induced travel calculations consistently reflect intermodal trade-offs among service measures across all the travel choices (e.g., trip frequency and mode choice).

Thus, to forecast the induced travel associated with the introduction of the new HSR mode, it is necessary only to calculate how much of a reduction in the equivalent price of travel results from that introduction. The introduction of a new mode that captures a large share of the market will result in a large improvement in the ease of travel. Introduction of a small share mode, in contrast, may have little effect.

MODELING RESULTS

The coefficients and *t*-statistics for various mode choice models that were estimated using data from Texas are presented in Table 1. All the coefficients are statistically significant, except the HSR constant in the air business model. This means that the HSR constant in the air business model cannot be shown statistically to be significantly different from zero (the value of the air mode constant). A value of exactly zero for

TABLE 1 ESTIMATED MODE CHOICE MODEL COEFFICIENTS BY MARKET SEGMENT (FOR INTERCITY TRAVEL IN TEXAS) (*2*)

	Market Segment			
	Air		Automobile	
Variable	Business	Nonbusiness	Business	Nonbusiness
Cost	−0.0379	−0.0609	−0.0283	−0.0321
(1990$)	(−4.5)	(−4.2)	(−2.2)	(−3.3)
Time (composite)	−1.3444	−1.7230	−0.5636	−0.2817
(hours)	(−6.4)	(−5.3)	(−3.4)	(−2.5)
HRS Constant	−0.0599	−0.3325	−0.7710	−1.1967
	(−0.4)	(1.7)	(−1.2)	(−2.3)

NOTE: (*t*-statistics in parentheses).

the mode-specific constant would imply that, if all times and costs of the two modes were equal, air business travelers would be indifferent between air and HSR (i.e., 50 percent would choose one mode, and 50 percent would choose the other). A negative (positive) value of the HSR constant implies that, all else being equal, the share of current travelers who would prefer HSR is less (more) than 50 percent.

Cost and travel time coefficients of all the models presented in Table 1 are negative, implying that increases in travel time or cost of a mode will reduce use of that mode.

Based on the mode diversion models estimated (Table 1) for a proposed HSR system in Texas, it is possible to compute how intercity air and automobile travelers value line-haul time. Table 2 illustrates how the value of time for individuals traveling between Houston and Dallas varies by trip purpose and current mode. As expected, the values of time for air travelers are much higher than for automobile travelers. Also, as expected, the values of time for nonbusiness travelers are lower than for business travelers currently traveling on a given mode. In studies of urban fixed-route and schedule (common carrier) transit travel competing with automobiles, the value of time for access and wait time is commonly observed to be much greater than the value of time for line-haul transit. However, this result is not transferable to the intercity air models because there are two competing common carriers modes, plus a scale difference in the length of the intercity trip. In addition, current travelers are willing to pay dearly for the high line-haul speed of air travel (or they are willing to have their companies pay dearly for business travel, considering the value of their own time to the company and its clients).

In the case of intercity automobile travel, models based on data from Florida indicate that as trip distance (and hence line-haul time) increases, both business and nonbusiness travelers place less importance on access and wait time, and more importance on line-haul time (*1*). As trip lengths increased to nearly 200 mi, the values of HSR line-haul time became greater than the values of access, egress, and wait time. (Houston and Dallas are about 240 mi apart.)

Described in the following sections are the ways in which the values of time, direct elasticities, and modal constants vary by four main market segments: air business, air nonbusiness, automobile business, and automobile nonbusiness travelers.

TABLE 2 IMPLIED VALUES OF TRAVEL TIME BY MODE AND TRIP PURPOSE IN TEXAS (*2*)

	Trip Purpose			
	Business		Nonbusiness	
Current Mode	Line-Haul Time	Access/Egress Time	Line-Haul Time	Access/Egress Time
Air				
Value of Time (Fraction of Hourly Wage Rate)	$35 (1.3)	$24 (0.9)	$28 (1.5)	$19 (1.0)
Automobile				
Value of Time (Fraction of Hourly Wage Rate)	$20 (1.0)	$13 (0.7)	$9 (0.5)	$6 (0.3)

NOTE: Dollar values are per hour in 1990 dollars.

Business Travel by Air

As expected, air business travelers are very sensitive to line-haul travel time. The value of time for business air trips is about $35 an hr. This figure is equal to 1.3 times the average hourly wage rate of the intercity travelers surveyed in Texas, a result that falls squarely in the range reported in a recent FAA comprehensive literature review of air travel demand models (*8*). The FAA range of 1.0 to 1.5 times the average wage rate is based on 17 models of air business travel demand.

The value of access and egress time for air business travelers presented in Table 2 is $24/hr. This value also reflects the premium this segment places on time. Adjusted for inflation, this value is similar to the mid-1980s value of $17/hr reported by FAA, based on airport access data from San Francisco. It is slightly higher than the 1989 value of $16/hr for this market segment in Florida, but reported traveler incomes in this segment are also higher in Texas than in Florida.

In logit mode choice models such as this, direct elasticities are not constant. Instead, they vary with both the values of the independent variables and the resulting mode share. Consequently, they depend on the assumed fare and service characteristics and the O/D pair. For example, as shown in Table 3, the air business HSR line-haul time elasticity for travel between the areas served by the proposed downtown Houston and downtown Dallas stations at two-thirds of air fare is about −0.86, whereas the HSR fare elasticity is −0.81. The latter falls within the range of −0.8 to −1.2 reported by others (*9*). This −0.81 value was found to increase to above −1.0 as HSR fares are set equal to air fares, indicating that the HSR revenue-maximizing fare is less than the air fare for the proposed HSR service in this corridor.

Finally, the HSR access and egress time elasticity for air business travelers was −0.36. This is a much lower value than for the air business line-haul time elasticity of −0.86, indicating the reduced relative importance of access and egress time for common carrier modes at the 240 mi distance between Houston and Dallas.

TABLE 3 HIGH SPEED RAIL ELASTICITIES BY MODE AND TRIP PURPOSE IN TEXAS (*2*)

	Level-of-Service Component		
Mode and Trip Purpose	Line-Haul Time	Access/Egress Time	Fare
Air			
Business	−0.86	−0.36	−0.81
Nonbusiness	−0.85	−0.37	−0.74
Automobile			
Business	−0.61	−0.21	−1.02
Nonbusiness	−0.38	−0.14	−1.05

NOTE: Elasticities calculated for characteristics between Houston and Dallas assuming that high speed rail fares are two-thirds the air fare.

TABLE 4 IMPLIED VALUE OF HSR CONSTANTS BY MARKET SEGMENT IN TEXAS (*2*)

	Trip Purpose	
Current Mode	Business	Nonbusiness
Air	$1.58	($5.46)
Automobile	$27.24	$37.28

NOTE: Values are in 1990 dollars and are equivalent to the fare advantage of existing mode over HSR, keeping all times and costs equal for competing modes.

The implied values of the HSR constants presented in Table 4 strongly support the findings that air and HSR are quite similar in the net effect of the unobserved (or unmeasured) attributes of each mode on ridership. That is, controlling for all the conventional level-of-service attributes included in the mode choice model (cost, line-haul time, access and egress time, and wait time), travelers perceive the air and HSR fixed route and schedule common carrier modes as essentially equal. Automobile travel, on the other hand, is valued quite highly relative to HSR if all the travel times and costs are held equal. Of course, the travel times of HSR and automobile are not equal between Dallas and Houston. Nevertheless, the HSR constants in the automobile mode choice models mean that certain attributes of automobile are valued highly relative to HSR (and presumable to air, although that was not measured explicitly in these models).

The implied value of the HSR constant indicates that if the cost and travel times of air and HSR are equal, business travelers will have a slight preference for air. A HSR fare reduction of less than $2 (or about 3 min reduction in HSR line-haul travel time) is needed to make this group of travelers feel indifferent between the two modes. As noted earlier, however, this is the only coefficient in all the individual market segment models that was not statistically significant. This confirms the hypothesis that business travelers regard air travel and HSR as similar competing common carrier modes.

Nonbusiness Travel by Air

As expected, individuals traveling by air for nonbusiness purposes are less sensitive than business travelers to line-haul time relative to cost. Their implied value of line-haul time is estimated at $28/hr in the Texas corridor (Table 1). This is slightly less than 1.5 times the average wage rate of travelers observed in the survey of air travelers, and within the range reported in the FAA study previously mentioned (*9*). The line-haul time elasticity is about the same for nonbusiness air travelers as for business air travelers (Table 3). Because a high proportion of nonbusiness travelers pay for their own air trip, they clearly value the time savings of the high-speed mode very highly.

The value of the HSR fare elasticity for a HSR fare equal to two-thirds of the nonbusiness air fare is -0.74. This is slightly less than the previously reported range of -0.8 to -1.2 and indicates that at two-thirds of the already lower nonbusiness air fare, the HSR fare is too low in this (proposed) private air and HSR competitive marketplace. The HSR fare elasticity was found to increase to -1.0, its farebox revenue maximizing value, at a HSR fare of about 90 percent of air fare for this market segment and O/D pair.

The value of access and egress time for air nonbusiness trips is about $19/hr (Table 2). This is higher than the mid-1980s value of $10/hr reported by FAA for nonbusiness airport access travel to San Francisco's airport, and by a study of Las Vegas nonbusiness airport access travel. However, the higher values of access and wait times for this market segment relative to Las Vegas reflect the higher incomes of these air travelers relative to the nonbusiness air travelers included in the Las Vegas survey.

The HSR constant is statistically significant, and its implied magnitude (equivalent to $5.46) suggests that, all else being equal, nonbusiness air travelers are somewhat more likely to use HSR than air business travelers. The constant represents how much lower than HSR the air fare would have to be to make travelers indifferent between air and HSR if all travel times were equal. Similarly, if fares were equal, HSR would enjoy a greater than 10-min inherent time advantage over air for nonbusiness travelers.

The difference in the HSR modal constants (relative to air) between business and nonbusiness travelers is reasonable, given the potentially greater comfort of HSR (such as bigger seats, more leg room, and the ability to look out the window and walk between cars). These additional comfort characteristics are likely to be more highly valued by nonbusiness travelers than by business travelers. In the future, it is possible that both types of travelers will value these attributes more highly than when they are actually provided and marketed in revenue service.

Business Travel by Automobile

Travel time is a less important determining factor for individuals traveling on business by automobile than by air. As discussed earlier, individuals who fly would be expected to place a high value on their travel time, whereas individuals who use automobiles place lower values on time but much higher values on the other attributes of automobile travel—flexibility, privacy, and the ability to make multiple stops, for instance. The value of line-haul time for the relatively high-income automobile travelers making business trips between Houston and Dallas is $20/hr (Table 2). This equals the average wage rate of the intercity travelers in this market segment. There are no comprehensive studies of the value of intercity automobile business travel time in the literature. However, this value falls logically between the values of time supported in the literature for both air travelers (referred to previously) and automobile nonbusiness travelers (discussed next). The value of access and egress time for automobile business travelers is again less than the value of line-haul time for this market segment for the 240 mi trip in this corridor.

Note that the values of time (in Table 2) and the demand elasticities for HSR time (in Table 3) are consistent across market segments and modes. That is, because automobile time is not as valuable as air travel time, the demand elasticities are lower for automobile than for air. An hour of saved line-haul time on HSR does not divert as many travelers from automobiles as from air.

Conversely, HSR fare elasticities for automobile travelers are higher than for air travelers. As expected, automobile travelers value saving money more highly than saving time relative to air travelers. Note that at the two-thirds air fare for which these fare elasticities are calculated, both business and nonbusiness automobile travelers turn fare-elastic (that is, lost revenue from lost riders due to a fare increase is greater than added revenue gained from the remaining riders).

The HSR constant in the automobile business model is worth $27.24 of fare reduction to make a traveler indifferent between automobile and HSR if all times and costs explicitly included in the model are equal. Thus, intercity travelers who already have selected a common carrier mode (for example, air) over travel by automobile are much more likely to switch to another common carrier mode, such as HSR, all else being equal. All things considered, automobile business travelers are much less likely to switch to HSR than are air business travelers.

Nonbusiness Travel by Automobile

The value of line-haul time for automobile nonbusiness trips was found to be the lowest among all four market segments, reflecting (again) the discretionary nature of nonbusiness trips relative to business trips (Table 2). The value of $9/hr equals about one-half of the average wage rate of automobile nonbusiness travelers between Houston and Dallas. It is consistent with a large English value-of-time study (*10*), which reported about $6/hr for nonbusiness long-distance automobile trips by the highest income group surveyed (but lower than the income of automobile nonbusiness travelers in the Texas market segment discussed here). The English study did not report actual trip lengths, but a review of the survey methodology suggests that fairly short trips (100 mi) constituted most of the sample.

Again, the value of access and egress time is less than line-haul time for the reasons discussed previously. Indeed, in the intercity Florida corridors (*1*), where trip lengths varied greatly, automobile nonbusiness access times were valued higher than for any other market segment for short intercity trips (85 mi). This result is to be expected because these travelers, who often have children and extra luggage, do not want to divert from automobile travel for short trips on a common carrier mode that involves additional access and egress times.

The elasticities for automobile nonbusiness travelers presented in Table 3 exhibit a similar pattern to that for business travelers who travel by automobile. The lower values of time result in lower time elasticities and higher HSR fare elasticities. The HSR fare elasticity would be even higher than that shown in Table 3 if the nonbusiness air fare were not already one-third lower than the business air fare.

The automobile nonbusiness market segment has the largest negative mode-specific HSR constant (equivalent to $37.28) among the four travel market segments reported here. This result is to be expected because nonbusiness travelers (e.g., individuals on vacation) most need the features of an automobile. Therefore, if times and costs are held equal, this group of intercity travelers is the least likely to switch to HSR.

Forecasting Model Applications

Most travel on HSR systems will be diverted from existing modes in the high-volume intercity markets where they are being proposed. These corridors are already served by Interstate highways, frequent air service, and (sometimes) conventional rail service. The modeling results described here provide important information for designing HSR applications that maximize ridership and passenger fare revenue. For example, the modeling results show that the passenger revenue maximizing fares that may be charged for HSR are very sensitive to whether the new mode's utility function (i.e., weighted travel time) is less or more than air. A shorter corridor (200 mi instead of 300 mi) will allow exploitation of the ability of HSR to offer multiple on-line stations in the areas served, reducing access and egress time without increasing waiting time. Air travel does not offer this feature, but the trade-off is extra line-haul travel time for the ground mode that may only be affordable for the shorter, 200-mi intercity travel distances.

Integrating HSR stations into local and regional transportation systems is therefore extremely important. Local access is a key variable in forecasting HSR ridership. Many private and public benefits can be obtained from facilitating access to and from the HSR system. If these private benefits are captured through the farebox, passenger revenue on HSR can be maximized. Conversely, HSR may be priced to maximize ridership and the benefits it provides from the reduced air and highway congestion, energy consumption, and air pollution that justify the public capital subsidies that most likely will be needed to build and operate HSR. The trade-offs between the private benefits of HSR captured through the farebox and the public benefits from foregoing farebox revenue are important outputs from using these models. They provide improved market understandings for system design and evaluation purposes.

The models presented in this paper therefore facilitate an understanding of how travelers on existing intercity modes value the potential travel-time savings offered by HSR and what the effects on demand and revenue are of the possible access and egress and waiting and terminal processing time advantages of the new modes.

CONCLUSION

The following three-step approach for forecasting HSR ridership is recommended.

1. Total air, automobile, and conventional rail volumes are each modeled separately using revealed preference (behavioral) data.
2. Separate air, automobile, and conventional rail (where relevant) mode choice models are estimated using stated preference methods. These models are applied to forecast the diversion of trips from each existing mode to HSR by trip purpose.
3. Induced travel is forecast on the basis of the behavioral relationships in the first two models.

A great advantage in forecasting ridership is that most travel on HSR will be diverted from existing modes in the corridors in which it is being seriously considered. It allows use of the behavioral information travelers have already provided by their revealed preferences to use these modes for their intercity travel. A critical finding is that persons who travel by air, automobile, and conventional rail exhibit different behavior when confronted with the choice or opportunity to use a new high-speed mode.

The resulting models have attractive properties in their ability to forecast the different rates of travel substitution between HSR and the existing modes and to incorporate the different values of time and other (nonquantifiable) factors that determine the mode choice of current intercity travelers. The models are also quite transparent in the way they reveal market driven information for HSR design and evaluation purposes. Finally, the Texas and Florida values of time and demand elasticities presented in this paper show that the modeling results have considerable face validity and conform well to the results of earlier studies.

REFERENCES

1. *Development of Mode Choice Models for High Speed Rail Travel in the Miami-Orlando-Tampa Corridor*. Prepared for the Florida High Speed Rail Corporation. Charles River Associates, Boston, Mass., March 1990.
2. Ridership Estimation and Marketing Plan. In *Franchise Application to Finance, Construct, Operate and Maintain a High Speed Rail Facility in the State of Texas*, Vol. 1. Presented to Texas High Speed Rail Authority by Texas FasTrac Inc. Charles River Associates, Boston, Mass., Jan. 16, 1991.
3. Ridership Estimates and Marketing Plan. In *Franchise Application to Construct, Operate, Maintain, and Finance a High-Speed Rail Facility*. Presented to Texas High-Speed Rail Authority by Texas TGV. Wilbur Smith and Associates, Jan. 16, 1991.
4. *The Demand for Electric Automobiles*. Prepared for the Electric Power Research Institute. Charles River Associates, Boston, Mass., Oct. 1981.
5. *Long Island Rail Road Consumer Preference Analysis*. Prepared for Long Island Rail Road. Charles River Associates, Boston, Mass., June 1, 1983.
6. *Telecommunications Preference and Usage by Clustered Small and Medium-Sized Firms: Implications for Bypass*. Prepared for Southwestern Bell. Charles River Associates, Boston, Mass., Sept. 1983.
7. *Setting a Ridership Goal for New York City's Subways*. Prepared for Metropolitan Transportation Authority. Charles River Associates, Boston, Mass., Feb. 1987.
8. *Economic Value for Evaluation of Federal Aviation Administration Investment and Regulatory Programs*. FAA, U.S. Department of Transportation, June 1989.
9. R. de Neufville and C.R. King. Access, Fares, Frequency: Effects on Airport Traffic. *Transportation Engineering Journal*, Vol. 105, No. TE2, ASCE, March 1979.
10. The MVA Consultancy, Institute for Transport Studies (University of Leeds), and Transport Studies Unit (University of Oxford). The Value of Travel Time Savings. In *Policy Journals*. Berks, England, 1987.

Publication of this paper sponsored by Committee on Intercity Rail Passenger Systems.

Implications of High-Speed Rail on Air Traffic

KENNETH R. BUCKEYE

Three high-speed rail (HSR) technologies were examined in the corridor connecting Minneapolis–St. Paul to Madison, Milwaukee, and Chicago. Travel characteristics in the existing market show the predominance of the air and automobile modes for nonbusiness purposes. Projected travel times for the 185-mph Train á Grand Vitesse and 300-mph magnetic levitation trains show that rail could be competitive with air depending on origination and termination points of a given traveler. An analysis was conducted of the various HSR alternatives and their effect on projected air traffic from Minneapolis–St. Paul International Airport (MSP) to Madison, Milwaukee, and Chicago. The results indicate that although HSR could divert 17 to 33 percent of air passenger traffic in the corridor and 12 to 21 percent of aircraft operations, total impact on passenger movements and aircraft operations at MSP would range from 1 to 2 percent. Whereas capital costs for an HSR system range from nearly $1 billion to more than $5.4 billion depending on the technology, consumer and community benefits range from $8 billion to $10 billion. Projected annual ridership and revenue would cover operating and maintenance costs for any of the rail technologies considered. Despite apparently low impacts on airport operations at MSP, cumulative benefits of HSR to the Minneapolis – St.Paul metro area and the upper midwest should be weighed against the socioeconomic, environmental, and financial costs of airport relocations. Airport and rail operations concepts and strategies may work in concert to significantly extend the life of MSP.

The success of high-speed rail (HSR) systems in Europe, Japan, and the Northeast Corridor of the United States is prompting many officials, organizations, and individuals in the United States to take a serious look at the application of these technologies in several high-demand corridors. HSR proposals are being promoted as a way of alleviating airport and highway congestion, reducing infrastructure costs, and, through private-sector capitalizing, possibly reducing some of the funding problems facing the nation's transportation systems. Experts have suggested that these rail technologies might be appropriate in reducing dependence on the automobile and easing pressure at crowded airports in corridors of 100 to 400 mi in length.

One proposal currently under consideration is the study of HSR service connecting Minneapolis–St. Paul, Milwaukee, and Chicago. This corridor, which is 430 mi long, is served by highways, air connections, bus service, and the Amtrak Empire Builder. Travel in the corridor is dominated by automobile and air travel.

One concern of transportation officials in Minnesota is the impact that HSR might have on the need for a new major airport in the Minneapolis–St. Paul metropolitan area. The question of a new airport has resurfaced in recent years amid controversies surrounding airport expansion, noise, and capacity problems (particularly in relation to economic implications for the region) at Minneapolis–St. Paul International Airport (MSP). Since deregulation of the airline industry, demand at the airport has increased significantly.

The MSP master plan projects an increase of nearly 30 percent in aircraft operations and 75 percent in passenger movements by 2010 (*1*). By some national estimates, revenue passenger miles could double by 2010 (*2*).

With all of the inherent uncertainties about growth projections, the need for a new airport has not been clearly determined. Because of this, the 1989 state legislature established the dual-track planning process, which lays out the requirements of the Metropolitan Airports Commission to plan for major improvement options of runways and terminal facilities at MSP while the Metropolitan Council conducts a broad search for a new airport location. Both agencies are to report to the legislature, which is scheduled to make a decision on airport expansion or relocation by 1996.

The legislation establishing the dual-track planning process did not require consideration of alternatives such as HSR and the impact it might have on the need for a new airport or the relief it could provide at the existing airport. Examined here is the issue of HSR and its potential effect on reducing air demand in the Minneapolis–St. Paul to Madison, Milwaukee, and Chicago corridor; landside and airside congestion at MSP; and socioeconomic considerations.

EXISTING TRAVEL CHARACTERISTICS

The Departments of Transportation in Minnesota, Illinois, and Wisconsin conducted a preliminary feasibility study of HSR as a means to determine whether the corridor among Minneapolis–St. Paul, Milwaukee, and Chicago could support an HSR system. The Tri-State Study of High Speed Rail Service was conducted by the consultant team of TMS/Benesch (*3*). Requirements of the study were to evaluate three technology alternatives in two study corridors, identify the existing market shares for all modes of travel, and project the ridership and revenue for three selected HSR technologies given the various performance characteristics. In addition, an analysis was conducted of potential routes, engineering and environmental considerations, and the financial and economic impacts of such a project.

The study determined that if an HSR system were to be constructed, the southern corridor, with the cities of Madison,

---

Minnesota Department of Transportation, Office of Railroads and Waterways, Suite 925, Kelly Annex, 395 John Ireland Blvd., Transportation Building, St. Paul, Minn. 55155.

Milwaukee, and Chicago, would achieve the greatest ridership and revenue and the greatest net consumer surplus. Therefore, for the purpose of this analysis, only the southern corridor will be assessed in determining the impact of HSR on air traffic.

Modal Preference

The study revealed that the air and automobile modes constitute the majority of person trips in the corridor, accounting for more than 97 percent of the existing travel. The automobile alone has nearly 63 percent of the market share; air accounts for 35 percent. Existing bus and passenger train modes serve a small portion of the passenger movements in the corridor, accounting for less than 3 percent. Although the bus and existing train modes are not expected to grow substantially or increase their respective market shares in the next 10 to 20 years, projections for air and automobile modes indicate increasing demand and associated congestion (*3*).

Annual person trips for air, automobile, and rail or bus between Minneapolis–St. Paul and Madison, Milwaukee, and Chicago total nearly 3.7 million today (Figure 1). The trips represent travel in both directions and have been annualized from FAA and Illinois and Wisconsin departments of transportation survey travel data. The figure indicates that 2.3 million automobile person trips were made in 1989 between Minneapolis–St. Paul and the three origin and destination cities. Rail or bus person trips amounted to less than 75,000.

Whereas total air travel amounted to nearly 1.3 million trips for the corridor in 1989, the number of air trips originating and terminating in the Minneapolis–St. Paul and Madison, Milwaukee, and Chicago city pairs is estimated today at 1 million annually (Department of Transportation, Origin-Destination Survey). Chicago's two major airports (O'Hare and Midway) and MSP are currently served by 98 commercial carrier flights per day. Madison, Milwaukee, and MSP are currently served by 22 flights per day.

Trip Purpose

The category of other travel, which includes personal business, pleasure, and vacation, is the major reason for travel in the corridor, representing 80 percent of all trips (Figure 2).

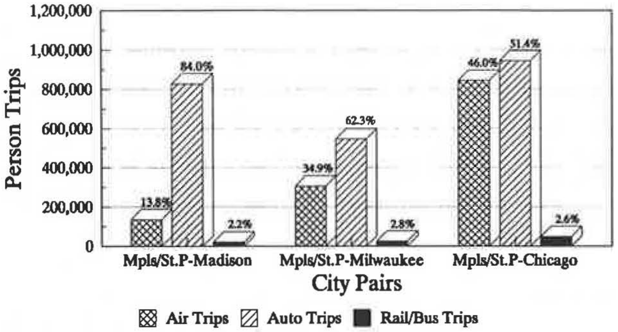


FIGURE 1 Trips by air, automobile, and rail or bus to and from Minneapolis–St. Paul, 1989 (source: FAA, Illinois DOT, Wisconsin DOT, Southeastern Wisconsin Regional Planning Commission).

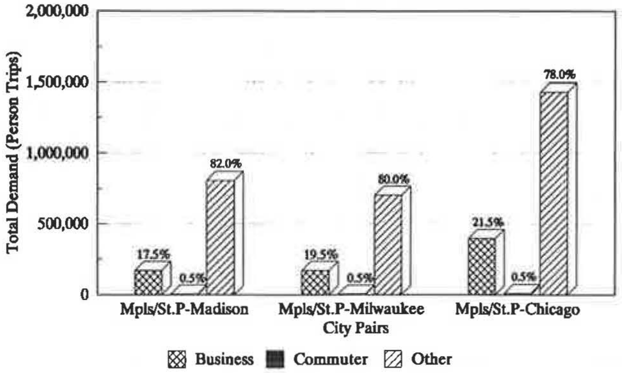


FIGURE 2 Travel purpose by city pair (data are from 1989) (*3*).

Cumulatively, 19 percent of all travel between the major metropolitan areas is considered business. Commuting accounts for less than 1 percent of travel purpose.

HSR POTENTIAL IN CORRIDOR

It was concluded in the tri-state study that significant potential existed for developing an HSR system in the southern corridor, achieving operating times, ridership, and revenue that could make it viable and competitive with other modes. The following summary presents the major findings of the Tri-State High Speed Rail Study in relation to the southern corridor. HSR technologies, a route evaluation, timetables, costs, and ridership and revenue estimates are presented in the following tables and discussion. Although Minneapolis–St. Paul is an important traffic generator, only a portion of the costs as well as ridership and revenue can be directly attributed to that metropolitan area or the state of Minnesota.

HSR Technologies

The HSR technologies that were evaluated represented three performance categories: high speed (125 mph), very high speed (185 mph), and super speed (300 mph). Within each performance category, existing rail systems, or prototypes, were selected in order to determine timetables, costs, and physical performance requirements (such as track curvatures) for routing considerations. Selected for study were the British High Speed Train (HST), representing the 125 mph category; the French Train á Grande Vitesse (TGV), representing the 185 mph category; and the prototypical Japanese magnetic levitation (maglev) train, representing the 300 mph category. Steel wheel on steel rail systems are proven performers under conditions similar to the Midwest, but maglev systems have not yet been put into commercial service anywhere.

Route Evaluations

Within the two general corridors identified for the tri-state study, potential routes were analyzed on the basis of field inspection of right-of-way, environmental constraints, population and employment distribution, and topographic and geologic conditions. To evaluate the alternatives it was necessary to establish route objectives to form a basis for comparisons. The objectives were as follows:

- To minimize travel times among the major cities to be served,
- To minimize the impact of natural topography on the rail system,
- To maximize regional accessibility, and
- To minimize the impact of construction and operation of HSR on the environment.

On the basis of a comparative engineering and environmental analysis, one route in each corridor was selected. For purposes of this analysis, only the South Route Modified is considered for the very-high-speed and super-speed options. For the high-speed option, the Amtrak Route was selected (Figure 3). The termini of the routes were Minneapolis–St. Paul and Chicago with a stop in Milwaukee. In the southern corridor, the route included Rochester, La Crosse, and Madison. The Amtrak upgrade option included only cities on the existing line.

Specific station stops were not identified in any of the cities. However, using a unit cost assessment, consideration was given to possible stops in suburban Chicago and downtown Chicago or O'Hare Airport, as well as a suburban Minneapolis–St. Paul station (possibly at MSP) and a downtown stop in one of the two cities. It was assumed that all other cities identified along the selected routes were to have one station stop.

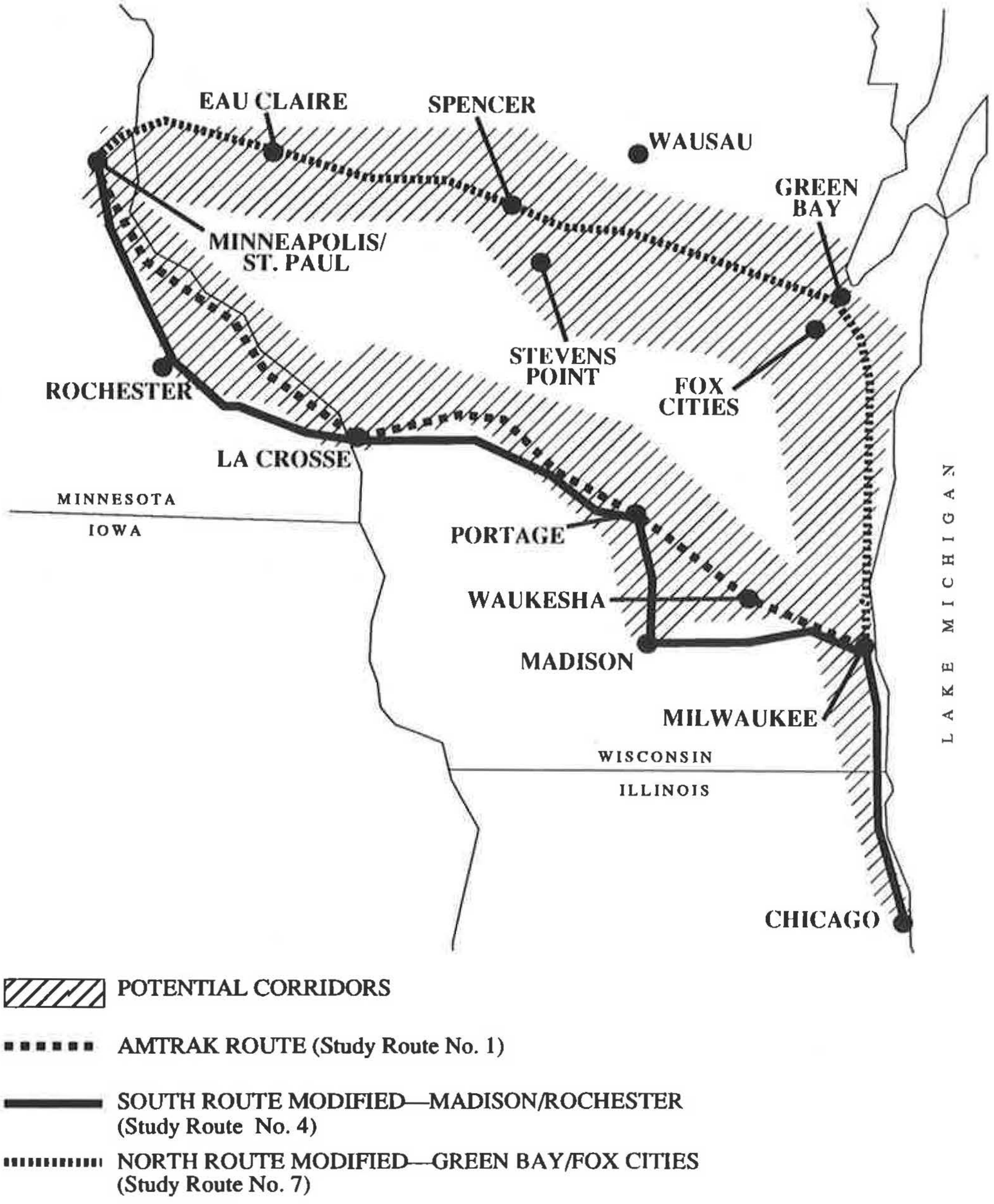


FIGURE 3 Routes selected for study of HSR service.

Operating Timetables

Operating times for the various HSR technologies were calculated using the LOCOMOTION© Train Performance Calculator (*2*) for each route option. Train running times were computed using performance data, including acceleration and deceleration and horizontal curve speed capabilities for each technology. Also considered were physical track condition data on a milepost-by-milepost basis, curve radii, station locations, and track speed limitations. Travel times for terminal-to-terminal and terminal-to-downtown trips by the various modes are compared in Table 1.

Essential to achieving projected rail timetables was the assumption that the technologies would be operated as modern railroads. This means high platforms for ease of passenger entry and exit, and 2-min station stops. Travel times for the 185-mph TGV and 300-mph maglev technologies show that rail will compete with air for terminal-to-downtown business trips, assuming rail stations are located downtown. Although total access, egress, and transfer costs have not been compared, the terminal-to-terminal rail trip also could be competitive with air, depending on the origination and destination points for a given traveler, as well as the location of suburban rail stations.

Capital and Operating Costs

Capital cost estimates were assembled for each route and technology using a unit cost data bank generated from previous HSR feasibility studies and local information on rail infrastructure costs (Table 2). Cost estimates were made by assessing the physical requirements of each route on a milepost-by-milepost basis and estimating the construction quantities. Rolling stock costs were added to infrastructure costs. The estimates assume that the 125-mph option would operate on the existing Amtrak alignment, and approximately one-third of the crossings would be grade separated, one-third closed, and one-third fully gated. For the 185- and 300-mph technologies all crossings would be grade separated or closed.

Estimates of the capital costs for development of an HSR system ranged from nearly $1 billion for the 125-mph high-speed technology on the existing Amtrak route to more than $5.4 billion for the 300-mph maglev option.

Annual operating and maintenance costs were generated from an understanding of the life cycle and maintenance costs of rolling stock combined with the proposed levels of service. Operating unit costs were estimated from the data obtained on operational high-speed railroads in Europe and Japan and from estimates of previous HSR studies in the United States and Canada. Operating costs include track, signaling, rolling stock and other equipment maintenance costs, train control, station and administrative staff costs, and energy requirements.

Ridership and Revenue

The ridership estimates for each route/technology were developed using the COMPASS© demand forecasting system, which provided specific behavioral analysis of travel characteristics in the study corridors (*3*). Ridership estimates reflect the central case economic scenario, which assumes continuation of current trends and the implementation of currently planned air and highway improvements. As expected, ridership is projected to increase with the increasing speed of the technology (Table 3). The models for total, induced and diverted demand were calibrated using value of time from original data, the origin and destination data base developed from existing data sources, and the network information (Tri-State Study). Train frequencies were set at 12 per day for the 125-mph technology, 18 per day for the 185-mph technology, and 24 per day for the 300-mph technology.

In the southern corridor for the year 2000, 5.8 million riders are forecast on the entire system for the 125-mph technology. Of these riders, 1.03 million passengers would travel between the Minneapolis–St. Paul and Madison, Milwaukee, and Chicago city pairs. Passenger movements increase with the 185-mph and 300-mph technology to 1.64 million and 2.10 million, respectively. By the year 2020, passenger movements between Minneapolis–St. Paul and Madison, Milwaukee, Chicago for the 125-mph technology would total 1.70 million, for the 185-mph technology 2.72 million, and for the 300-mph technology 3.50 million.

Like the ridership projections, revenue estimates increase with the increasing speed of the technologies. The revenue estimates were not optimized, but were based on a reasonable level of fare, set at 65 percent of 1989 full business class air fare. An analysis of the competitive response from the air and bus modes was not conducted for this report, although in reality it is fair to assume that such a response would occur.

TABLE 1 TRAVEL TIMES FROM MINNEAPOLIS–ST. PAUL TO CHICAGO

Technology	Terminal-to-Terminal (hours, minutes)	Terminal-to-Downtown[a] (hours, minutes)
Air	1:15	2:00-2:30
125 mph - HST	4:20	4:20
185 mph - TGV	3:15	3:15
300 mph - Maglev	2:15	2:15
Amtrak	9:30	9:30
Auto	8:00	---
Bus	10:00	---

[a] Terminal-to-Downtown assumes 45 minute ground travel time for Midway and MSP and 1 hour and 15 minute ground travel time for O'Hare.

TABLE 2 CAPITAL AND ANNUAL OPERATING AND MAINTENANCE COSTS FOR THE SOUTHERN CORRIDOR (*3*)

Technology Options	Capital	Annual Operating and Maintenance
125 mph - HST	**940.0**	**90.9**
185 mph - TGV	**3,020.0**	**101.3**
300 mph - Maglev	**5,450.0**	**123.3**

NOTE: Values are in millions of 1989 dollars.

Summary of Major Findings

The Amtrak Route and South Route Modified were considered for analysis in this study. Although the application of any HSR technology in the Midwest is yet unproven, steel wheel on steel rail systems operate in somewhat similar conditions in Europe and Japan. However, maglev systems, still in prototype development, have not yet been proven. The 185-mph very-high-speed technology, with a running time of 3 hr and 20 min, may have a slight advantage over the 125-mph high-speed technology for service to the Minneapolis–St. Paul market. The analysis indicates that the 185-mph technology can be competitive with air travel in the corridor, especially for terminal-to-downtown trips.

Revenue and ridership increase with increasing speed of the technology in the market. Using a unit cost assessment in the estimating procedure, revenue achieved for all technologies will cover the operating and maintenance costs for the various technologies, suggesting that an HSR system could potentially operate without a public subsidy.

PROJECTED TRAVEL MARKET

HSR Projections

Ridership estimates projected for HSR comprise trips generated by natural growth of the rail mode, trips diverted from other modes, and travel that is induced because of the existence of more and better travel options. If ridership estimates projected in the Tri-State High Speed Rail Study could be achieved, the market share for high-speed train travel between Minneapolis–St. Paul and Chicago would increase with increasing speeds (Figure 4). Whereas the 125-mph HST technology could expect to capture 15 percent of the market in 2020, the 185-mph TGV might gain as much as 22 percent. At 300 mph, the maglev system might gain as much as 26 percent. The various systems would all divert a significant number of trips from both air and automobile modes. At the same time, a large portion of the ridership would come from induced travel (i.e., travel that occurs because of the existence of a new mode and the improved travel opportunities for the business and social market). The existence of improved travel opportunities for one mode not only induces travel for that mode but for all other modes as well. Therefore, the implementation of HSR in the Minneapolis–St. Paul to Chicago corridor would create additional demand for air, automobile, and bus.

Growth Projections for Air Traffic

Enplanements and Deplanements

In 1989 enplanements and deplanements of passengers at MSP on major, regional, and charter carriers totaled 19.4 million

TABLE 3 FORECASTS OF RAIL RIDERSHIP AND REVENUE FOR 2000 AND 2020 (*3*)

		Year 2000[a]	Year 2020[a]
125 mph - HST	**Ridership**	**5.8 (1.0)**	**7.7 (1.7)**
	Revenue	**226.6**	**320.6**
185 mph - TGV	**Ridership**	**7.5 (1.6)**	**10.1 (2.7)**
	Revenue	**336.1**	**480.0**
300 mph - Maglev	**Ridership**	**8.5 (2.1)**	**11.6 (3.5)**
	Revenue	**409.3**	**586.0**

NOTE: Ridership forecasts are in millions of trips; revenue forecasts are in millions of 1989 dollars.

[a] Numbers in parentheses indicate ridership in millions: Mpls-St. Paul to/from Madison/Milwaukee/Chicago, including Minneapolis-St. Paul suburban to downtown trips.

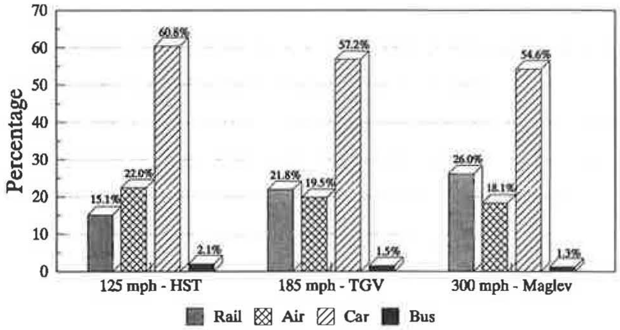


FIGURE 4 Projected market shares by mode, implementation of HSR, 2020 [source: Tri-State High Speed Rail Study (*3*), COMPASS Strategic Transportation Planning Model].

(*4*). By 1990 passenger activity surpassed 20.3 million, representing a 5 percent increase and a continuation of the trend in the 1980s, which saw a doubling of passengers. Passenger activity forecasts from the MSP Comprehensive Plan show a significant increase during the next 10 to 30 years, assuming average economic growth. By 2000, more than 30 million enplanements and deplanements could be expected. By 2020, enplanements and deplanements could reach 39 million annually (*1*).

Air passenger traffic from Minneapolis–St. Paul to Madison, Milwaukee, and Chicago in 1989 totaled 1.3 million and represented 6.5 percent of all passenger movements at the airport. It is estimated that nonconnecting passengers (i.e., those who may be more likely to use HSR) totaled more than one million. If current trends continue, by 2000 there may be nearly two million air passengers in the corridor. By 2020, 2.5 million air passengers could be expected in the corridor.

Aircraft Operations

Aircraft operations (landings and take-offs) at MSP have grown proportionally with passenger activity and also are expected to increase significantly during the next 10 to 30 years. There were nearly 380,000 operations of the scheduled air carriers, regional and air taxi, charters, air freight, general aviation, and military in 1990. That number is projected to reach 453,000 by 2000, reflecting a 19 percent growth. By 2020, operations could grow an additional 16 percent, reaching 527,000 annually (*3*).

On the average, 110 flights per day (120 per average week day) serve the Minneapolis–St. Paul to Madison, Milwaukee, and Chicago corridor, representing nearly 40,000 annual flights, or more than 10 percent of all aircraft operations at MSP. Chicago's O'Hare and Midway airports alone account for 88 flights per day, or 32,000 flights annually. Assuming continuation of current trends, more than 58,000 operations could be expected in the corridor by 2020, representing 158 daily operations.

Impacts of HSR on Air Passenger Movements

A limiting factor facing airport development is the ability to expand landside capacity. Landside capacity addresses passenger facilities essential to the airport operations such as terminal and concourse space, parking, gates, baggage claim, car rental, and roadway capacity, all of which limit the volume, comfort, and ease with which passengers pass through an airport. As situated, the ability to expand passenger facilities at MSP is constricted, and significant capacity improvements will require major construction.

Issues facing HSR in this corridor are its competitiveness with the air mode and its impact on passenger movements through the airport, thus reducing landside and airside congestion.

Timetables for the 185-mph TGV and 30-mph maglev HSR technologies indicate that they can compete with air for terminal-to-downtown trips given the transfer costs and delays experienced at most airports. Although the 125-mph technology attracts a significant number of passengers, it is less competitive with air for total trip time.

The Tri-State High Speed Rail Study concluded that as many as 1.8 million passengers may be attracted to the train in the Minneapolis–St. Paul to Madison, Milwaukee, and Chicago corridor by 2000. This number could increase to 2.8 million by 2020 if the maglev technology were implemented. However, only those passengers diverted from the air operations would have any effect on capacity at MSP. As a percent of rail ridership, the COMPASS© demand forecasting model predicted that total diversions from all modes (air, automobile, and bus) ranged from 60 percent for the 300-mph maglev technology, to 68 percent for the 185-mph TGV technology, to 72 percent for the 125-mph HST technology (*3*). Total rail ridership and diversions increase with increasing speed. However, trip diversions as a percentage of total ridership decrease with increasing speed. Natural growth and induced travel accounts for the added volume of passengers attracted to the progressively higher speed technologies.

As formulated, the COMPASS© model does not segregate diversions from individual modes. That is, it is not possible to say precisely the percentage of trips diverted from air, automobile, or bus, only total collective diversions. In this analysis, a conservative estimate was made that one-half of the trip diversions predicted by the COMPASS© model will be attributed to air.

Table 4 presents projected rail ridership for 2020 by technology, total passengers diverted from air, and net percentage effect that HSR will have on air passenger movements in the Minneapolis–St. Paul to Madison, Milwaukee, and Chicago corridor and total percent reduction at MSP. In 2020, with implementation of the 125-mph HST technology, 493,000 trips will be diverted from air. This could reduce passenger movements in the corridor by nearly 17 percent. The net effect of total passenger movements at MSP, however, will be less than 1.3 percent. With the 185-mph TGV technology, 758,000 trips will be diverted from air, reducing passenger movements in the corridor by almost 30 percent. The effect on total passenger movements at MSP will be under 2 percent. With implementation of the 300-mph maglev technology, 851,000 passengers will be diverted from air, reducing air passenger movements in the corridor by more than 33 percent. The net effect at MSP will be slightly more than a 2 percent reduction in total passenger movements.

Although these diversions appear to be modest, they are compared against airport projections, which "double count"

TABLE 4 AIR PASSENGER DIVERSIONS WITHIN MINNEAPOLIS–ST. PAUL TO MADISON, MILWAUKEE, AND CHICAGO CORRIDOR, 2020

	Projected Rail Ridership in Corridor[b]	Passengers Diverted from Air (% rail ridership)	Air Passengers Diverted in Corridor	Reduction of Air Passengers at MSP
125 mph - HST	**1,368,455**	**492,644 (36%)**	**16.95%**	**1.26%**
185 mph - TGV	**2,229,853**	**758,150 (34%)**	**29.90%**	**1.94%**
300 mph - Maglev	**2,837,041**	**851,112 (30%)**	**33.57%**	**2.18%**

[a] **Assumptions: 39 million passengers projected at MSP for 2020. Minneapolis-St. Paul to Madison/Milwaukee/Chicago Corridor accounts for 6.5% of MSP passenger movements.**

[b] **Ridership in corridor counts only trips to/from Minneapolis-St. Paul and Madison, Milwaukee and Chicago. This forecast excludes internal zone trips i.e., Mpls-St. Paul suburban station to downtown station trips.**

on-line and inter-line (transfer) passengers. By conservative estimates, these account for more than 30 percent of passenger movements at MSP (*4*). If airport projections were factored down accordingly, the effect of HSR diversions would be proportionally greater.

Impacts of HSR on Aircraft Operations

A more significant issue at MSP is the effect HSR might have on aircraft operations. Physical airside capacity is a concern because MSP is essentially affected by urban and suburban development, which prevents the physical expansion of airport property and the addition of runway capacity. Although it may be possible to develop additional runways on existing airport property, severe environmental and safety concerns must first be addressed. Aside from aircraft safety, the most significant concern is that of excessive noise, which affects many residents of Minneapolis, St. Paul, and surrounding suburbs. Proposals for physical expansion of the airport, which permit significant increases in airside capacity, continue to run into major opposition from home owners. As well, the second track of the dual-track process, involving siting of a new airport, is running into major opposition from residents in the identified search areas. For some, however, relief may be in sight with enactment of the new federal law requiring upgrading of airline fleets with Stage III aircraft by 2000.

As with passenger movements, there would be a concordant reduction of air carrier operations in the Minneapolis–St. Paul to Madison, Milwaukee, and Chicago corridor, as well as total operations at MSP with the implementation of HSR (Table 5). Assuming no competitive response from the airline industry, development of the 125-mph technology between 2000 and 2020 could reduce operations in the corridor by 12 percent, with a net reduction on total traffic at MSP of 1.33 percent. This assumes an average passenger load factor of 70 per plane, which appears to be somewhat higher than typical loadings in the corridor. For the 185-mph technology, aircraft operations could decrease by 10,800, or 18.66 percent, in the corridor and slightly more than 2 percent of total operations at MSP. At 300-mph, HSR could reduce total operations at MSP by 12,159, or 20.9 percent of air carrier flights in the corridor. Total operations at MSP could be reduced by 2.3 percent.

The impact of these reduction estimates also includes the categories of regional and air taxi, cargo, express, general aviation, and military, which currently account for about 25 percent of all operations. Although these categories of operations are projected to increase modestly by 2020, implementation of measures that might reduce their numbers would slightly enhance the effect of HSR on total operations. Although not explored in this paper, some potential does exist in the corridor to reduce small package express operations by means of an HSR system.

SOCIOECONOMIC AND ENVIRONMENTAL CONSIDERATIONS

An important aspect of the HSR and airport issue is the benefit/cost analysis and the consumer surplus generated by such projects. Estimates for construction of an HSR system range from nearly \$1 billion to almost \$5.5 billion, depending on the technology. Some of these costs would likely be shared by the three states, with a portion coming from private-sector investors. The consumer surplus and community benefits generated by such a project may surpass \$10 billion for the region, with as many as 16,000 to 19,000 person years of employment created during the 25-year life of the project (*3*). Costs to the federal government, the state of Minnesota and the Minneapolis–St. Paul metropolitan area in developing a new airport range from \$2.0 to \$4.5 billion for construction. Improved highway access to the site and business relocation costs are additional associated expenses. Other costs will be severe in terms of land acquisition and access and egress times for travelers if a new airport would be developed 20 to 30 mi or more from the downtowns as currently proposed (*5*).

TABLE 5 AIR CARRIER REDUCTIONS WITHIN MINNEAPOLIS–ST. PAUL TO MADISON, MILWAUKEE, AND CHICAGO CORRIDOR, 2020

	Diverted Air Passengers	Flight Equivalent (70 pass/flt)	% of Corridor Reduction	% Reduction MSP Operations
125 mph - HST	492,644	7,037	12.13%	1.33%
185 mph - TGV	758,150	10,830	18.66%	2.05%
300 mph - Maglev	851,112	12,159	20.95%	2.30%

[a] Assumptions: 527,450 flights projected for year 2020 (MSP Comprehensive Plan): Minneapolis/St. Paul to Madison/Milwaukee/Chicago accounts for 11% of all flight operations at MSP (or 58,020 operations)

Environmentally and socially, large public infrastructure projects have major problems to surmount. A new airport however, consuming 15,000 acres of land in one large tract, may face stiffer public resistance than the development of an HSR system consuming 5,000 acres in a linear corridor, much of which may be on existing right-of-way. HSR with direct access to the downtowns could help to cohere the Minneapolis–St.Paul metropolitan area and reduce the costs of urban sprawl. Development of a remote major airport will encourage sprawl and require the costly extension of urban services.

The impacts of an HSR network cannot be considered in isolation to a single airport of metropolitan area. The benefits, as well as costs, of HSR in reducing passenger movements and air carrier operations at MSP would extend to Chicago, Milwaukee, and Madison in varying degrees. Viewed from a national and regional transportation system perspective, HSR may offer the opportunity to delay or eliminate the need for a new major airport in the upper Midwest, particularly if it were designed as an element of an air and rail hub concept. (HSR has also been proposed as a remote airport link in Minnesota and is under study in the Chicago metropolitan area.) Benefits of an HSR system to the tri-state area could be especially attractive when considering capitalization by the private sector or multi-state cost sharing on various portions of a project. Although both the development of new airports and HSR systems will create jobs, only HSR will diversify transportation opportunities, reduce the region's reliance on imported petroleum single fuel source, and provide considerable overall energy savings.

CONCLUSIONS

This paper presented an analysis of the impact that various HSR technologies could have on passenger movements and air carrier operations in the Minneapolis – St. Paul to Madison, Milwaukee, and Chicago corridor. By 2020 development of an HSR system in the southern corridor would offer relief in passenger movements ranging from 16 percent for the 125-mph HST technology to 33 percent for the 300-mph maglev technology. However, total effect of HSR on landside capacity at MSP would be significantly less, ranging from 1.3 to 2.2 percent.

In the same time frame, a reduction in aircraft operations for the corridor could range from 12 percent for the lower speed rail technology to 21 percent for the highest speed rail technology. Total relief on airside capacity at MSP would range from 1.3 to 2.3 percent, depending on the technology implemented.

On the basis of this conservative analysis, it can be concluded that HSR alone will have a modest effect on total passenger movements and aircraft operations at MSP. An investment in HSR cannot be justified as a means of extending the life of MSP unless it is done in conjunction with other air traffic reduction measures. However, if HSR were coupled with other management strategies, such as reducing general aviation and military operations, the short- to midterm viability of MSP may be extended. It is possible that if HSR were developed in conjunction with expansion of MSP, or linking with a remote hub, long-distance travel needs in the Twin Cities metropolitan area could be satisfied for decades to come.

The socioeconomic benefits of an HSR system must be weighed against the costs of airport expansion or relocation. Viewed from a regional transportation system perspective, HSR will create jobs, cohere the urban area, and conserve land resources with minimal social and environmental disturbance. Diversification of transportation opportunities and reducing energy reliance may provide a substantial economic savings for the region.

ACKNOWLEDGMENTS

The author wishes to acknowledge the invaluable assistance of Annette Swanson of the Office of Railroads and Waterways, Minnesota Department of Transportation (MnDOT). Additional assistance was provided by William Newstrand of Ports and Waterways, and Kenton Hoeper of the Office of Aeronautics at MnDOT. The author also wishes to thank

TMS/Benesch, consultant to the Tri-State Study of High Speed Rail Service.

REFERENCES

1. *Minneapolis–Saint Paul International Airport, Long Term Comprehensive Plan, Volume 3, Activity Forecasts.* Howard Needles Tammen and Bergendoff, Minneapolis, Minn., March 1990.
2. *Future Development of the U.S. Airport Network.* TRB, National Research Council, Washington, D.C., 1988.
3. *Tri-State High Speed Rail Study: Chicago-Milwaukee-Twin Cities Corridor.* Prepared for Illinois, Minnesota, and Wisconsin departments of transportation. TMS/Benesch High Speed Rail Consultants, Chicago, Ill., 1991.
4. *Minneapolis–St. Paul International Airport, Summary of Operations.* Minneapolis–St. Paul Metropolitan Airports Commission, Minneapolis, Minn., December 1990.
5. *Selecting a Search Area for a New Major Airport, Part Two: Draft Data Analysis.* Metropolitan Council of the Twin Cities, St. Paul, Minn., 1991.

Publication of this paper sponsored by Committee on Intercity Rail Passenger Systems.

Changing Long-Distance Passenger Markets in a Deregulated Environment

BRUNO PAROLIN AND ADRIAN HARRINGTON

In 1987 competition between bus and rail services in New South Wales was permitted for the first time on two transport corridors: the Sydney-Canberra and Sydney-North Coast. In 1988 long-distance bus services in the rest of the state were deregulated. Deregulation effectively ended the 57-year monopoly of intrastate passenger transport by the railways. The legislative framework of long-distance passenger travel in New South Wales between 1920 and 1988 is outlined, changes in passenger profiles are described, and structural changes on the Sydney-Canberra and Sydney-North Coast corridors following deregulation are examined. The conclusions from this study indicate that deregulation has had significant and positive effects on the intrastate long-distance travel market through lower fares, increased service, better modal choice for consumers, and increased industry efficiency through rationalization.

In the last decade a worldwide trend toward deregulation of transport services has occurred, and Australia has been no exception. Between 1986 and 1988 the New South Wales (NSW) State Government followed the overseas lead of deregulation by relaxing the strict regulations governing the entry of bus operators into the intrastate long-distance bus passenger market. This heralded a major change in government policy as long-distance bus services were for the first time allowed to compete with intrastate rail services.

The purpose of this paper is, first, to briefly outline the regulation of intrastate transportation in the State of New South Wales before 1986 and the events leading up to the final stages of deregulation in 1988. Second, the paper provides a base of information on changes to the long-distance intrastate passenger market. Specifically, changes in passenger markets, travel demand, and service characteristics along the Sydney-Canberra and Sydney-North Coast corridors are examined (Figure 1). In this context the aim is to provide a preliminary assessment of the impacts of deregulation on the intrastate long-distance passenger market.

BACKGROUND

Before 1987 long-distance transport in New South Wales was controlled by the 1931 State Transport (Co-ordination) Act. The purpose of the act was "to provide for the improvement and for the co-ordination of means of and facilities for locomotion and transport." However, under section 17(3) of the act, the Commissioner of the Department of Motor Transport (DMT) was required to consider whether the service would lead to unnecessary and wasteful competition before issuing a license. There is little doubt this legislation was used to restrict the operation of the long-distance bus industry and effectively prevent competition between bus and rail services.

The first long-distance intrastate bus service within NSW commenced in 1956 between Sydney and Canberra. Following a long court case, the Ansett-Pioneer bus company was given permission to commence services on this corridor even though rail services existed between Sydney and Canberra. Ansett-Pioneer was given permission to commence operations, but with strict conditions restricting the arrival and departure times of services (*1*). Ansett-Pioneer remained the sole operator on the route until 1986 despite receipt by the DMT of numerous applications to operate services.

The next significant intrastate bus service in NSW did not commence until 1970. The Sydney-Broken Hill service began operating in response to political pressures from isolated country electorates in far west NSW. The license was issued on the condition that the operator not convey passengers traveling between Sydney and Narromine, where rail services were already operating. Between 1970 and 1986 15 additional intrastate bus services were granted licenses to convey passengers between isolated rural regions and Sydney. These services were not considered to be competitive threats because most did not operate close to rail services (*2*).

In 1985 the government initiated a review into the long-distance bus industry as a result of growing recognition that existing transport services in NSW were inadequate. The major thrust of complaints came from rural residents who were dissatisfied with existing bus and rail services. They were being forced to travel long distances to join bus or rail services while interstate bus services passing through their towns were prohibited from picking-up and discharging passengers. Further pressure came from the tourism industry. Tourist operators were concerned the existing policy was constraining the ability of both domestic and international tourists to travel in NSW (*2*).

The review concluded that existing transport services were not meeting the needs of the traveling public and that the protection of existing transport services at the expense of the traveling public and other state interests could not be justified (*3*). Despite recognition that reduced regulation would procure substantial benefits to the public, there was concern within the government and the State Rail Authority (SRA) about the effect of increased bus competition on rail revenue (*1,4*). Therefore, instead of pursuing full-scale deregulation immediately, it was found in the review that "[I]t was necessary to demonstrate that the expected benefits were actually

---

B. Parolin, School of Geography, University of New South Wales, P.O. Box 1, Kensington, NSW, 2033 Australia. A. Harrington, Building Owners and Managers Association Australia, Ltd., 10th Floor, CML Building, 14 Martin Place, Sydney, NSW, 2000, Australia.

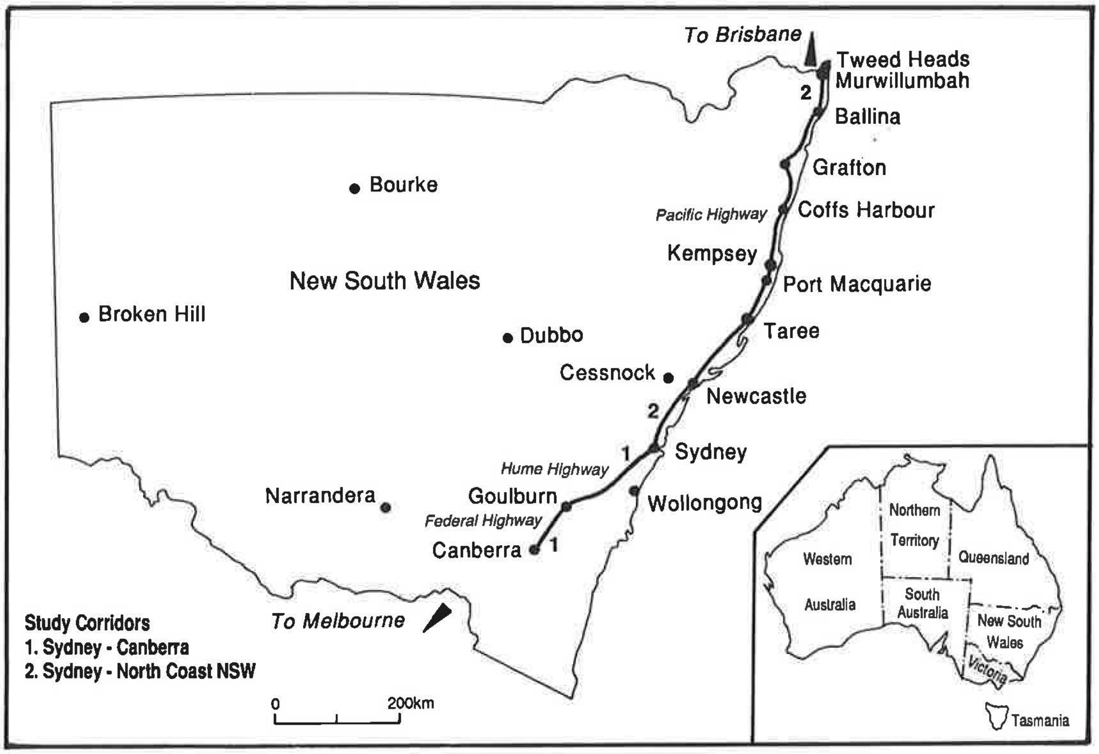


FIGURE 1 Study Corridors.

achievable before making a recommendation for a more general change in policy. . . ." and the best possible way to achieve this was to conduct a trial liberalization on two transport corridors within NSW: Sydney-Canberra and Sydney-North Coast (Figure 1). Unlike governments in the United Kingdom and United States, the NSW government was cautious in its approach to easing regulation of the intrastate bus industry.

The Sydney-Canberra and Sydney-North Coast corridors were selected for the trial by the DMT because both corridors had high demand and therefore could support competition, and the SRA would not suffer significant revenue losses (*3*). The trials commenced in November 1986 and were to last for a period of 6 months. However, they were extended four times between November 1986 and November 1988, and another corridor, Sydney-South Coast, was added. During this period, long-distance services were defined as those carrying passengers further than 160 km in NSW. After 6 months of trial liberalization, the Bureau of Transport Economics (*5*) concluded as a result of increased intermodal and intramodal competition, "[T]he major beneficiaries of the trial were the travelling public due to significant improvements in bus frequencies and reduced fares. . . ." (see Table 1).

Deregulation of the intrastate long-distance bus industry in NSW was completed in November 1988 with the passing of the State Transport (Co-ordination) Amendment Act.

The new act significantly eases entry conditions into the market. In the consideration of a license application from a long-distance bus operator emphasis is now placed on whether the service is necessary and desirable in the public interest. Because licenses are still required it would be more appropriate to use the term partial deregulation instead of deregulation, whereas in Britain deregulation of intercity bus services in 1980 entailed the abolition of licencing for routes (*6*).

The government also examines bus timetables to ensure that they are structured in such a way that the bus travels within the speed limits set for heavy vehicles. This occurred following several horrific road crashes involving long-distance buses.

The easing of entry conditions has seen a proliferation of intrastate long-distance bus services in NSW. In the 12 months following deregulation, 40 bus operators were licensed to operate more than 84 routes throughout NSW.

DATA SOURCES AND METHOD

The statistical analysis reported in this paper is based on surveys of bus and rail passengers undertaken by the Bureau of Transport Economics (BTE) to monitor and evaluate the impact of trial liberalization on the traveling public and competing modes of transport along two of the trial corridors: Sydney-Canberra and Sydney-North Coast. Other data sources are used for descriptive purposes only.

The BTE used on-board random surveys to elicit information from bus and rail passengers traveling on both corridors during the peak period (January 14–21, 1987) and the off-peak period (April 2–8, 1987). The sample frame consisted of bus operators who were granted permits to convey passengers over the two corridors and rail services that competed directly with these services.

A required sample size of 900 for both the peak and off-peak surveys was based on the assumption that 10 percent of

TABLE 1 CHANGES IN INTRASTATE SERVICE LEVELS

Centre	Bus Frequency Pre Trial	Bus Frequency During Trial[1]	Bus Frequency 1989[1]	Bus Frequency 1991[1]	Bus Fares During Trial	Bus Fares Pre Trial	Economy Rail Fares[2]	Range of Air Fares[2]
Taree	-	28	25	15	$20-$25	n.a.	$25.70	$84-$98
Port Macquarie	3	18	21	11	$19-$25	$32.70	n.a.	$96-$135
Kempsey	-	26	26	16	$25-$30	n.a.	$33.70	n.a.
Coffs Harbour	-	26	25	17	$25-$30	n.a.	$37.90	$107-$159
Grafton	-	21	20	15	$25-$32	n.a.	$37.90	$125-$128
Ballina	2	19	26	14	$28-$35	$39.80	n.a.	n.a.
Tweed Heads	-	21	27	16	$30-$41	n.a.	n.a.	$152
Canberra	2	20	40	37	$15-$18	$20.00	$21.30	$67-$90

1. Daily bus frequency in each direction including interstate services.
2. Economy fares prior to and during trial liberalisation.
- Interstate services passed through these centres but were not permitted to carry intrastate passengers.

n.a. Not applicable.

Source: Department of Motor Transport (1988) and Bus Company Timetables.

TABLE 2 NUMBER OF PASSENGERS SURVEYED (7)

Corridor	Peak[a] Bus	Peak[a] Rail	Off-Peak[a] Bus	Off-Peak[a] Rail
Sydney-Canberra	210	125	137	116
Canberra-Sydney	201	156	233	25
Sydney-North Coast	272	270	99	197
North Coast-Sydney	220	263	154	191
Total	903	814	623	529

[a] Both Peak and Off-Peak surveys are used in the analyses reported in this paper.

bus and rail passengers diverted from other modes of transport (7). The lower-than-required sample size for the off-peak survey is associated with lower-than-anticipated patronage levels and limited resources available for follow-up field work. Table 2 provides a breakdown of the number of usable questionnaires that were returned by passengers on each corridor traveling by bus and rail. All the peak and off-peak surveys were used in the analyses reported in this paper.

The actual percentages given in this paper (Tables 3 and 4) differ slightly from those in the BTE report (7) because the data base used for statistical analysis in this study was a modified version of the original and contained fewer data coding errors.

The questionnaires sought information on the travel characteristics of passengers, their attitudes to particular aspects of the mode on which they were traveling and their socioeconomic characteristics.

It should be stressed that in this analysis there was no control over statistical aspects of questionnaire design or survey methodology. Data collected by the BTE were used together with data from other sources to examine the nature and extent of changes following liberalization and deregulation.

RESULTS

Pretrial Liberalization

Information on base long-distance services, the market, and general travel demand on the two corridors before trial liberalization is inadequate and limited to published data from the National Travel Survey of 1977–78 (*8*) and reports from the Bureau of Transport Economics (*5*). The data sets are not directly comparable given different sampling and estimation methods; they were simply used to highlight trends. No complete time series data were available for this purpose.

The National Travel Survey provides sample data on market share on the two corridors based on a sample of households (Table 5). On both corridors automobile trips dominate long-distance travel, followed by air, rail, and bus trips respectively. More bus trips are commonplace on the Sydney-Canberra corridor, which is a 5-hr journey, compared with a 15-hr journey on the Sydney-North Coast corridor (extending to Brisbane). The number of services on both corridors was not reported.

TABLE 3 SOCIAL CHARACTERISTICS OF BUS AND RAIL PASSENGERS

	Sydney-Canberra		Sydney-North Coast	
	Bus	Rail	Bus	Rail
	(n=774)	(n=421)	(n=737)	(n=911)
Sex	(percentage)			
Male	41	41	40	40
Female	59	59	60	60
X^2 (df=7)	0.002*		0.005*	
Age	(percentage)			
> 15	5	8	11	5
15-19	15	14	20	12
20-29	32	21	21	21
30-39	16	15	13	9
40-49	11	17	8	13
50-59	9	11	9	11
60+	12	14	20	29
X^2 (df=6)	23.69**		65.35**	
Occupation	(percentage)			
Student	23	26	26	15
H/hold Duties	10	15	15	13
Clerical	14	7	8	4
Plant/Mach. Op.	1	1	2	1
Salesperson	4	1	2	1
Professional/ Tech	21	14	9	8
Tradesperson	4	5	7	4
Semi-Professional	2	2	4	3
Labourer	2	4	4	4
Manager/Admin.	7	7	3	9
Unemployed	3	3	4	6
Retired	16	10	16	32
X^2 (df=11)	45.75**		120.85**	

* Not Significant

** Significant at the 0.01 level

Source : Calculated from BTE Passenger Surveys, 1987.

In 1984 the BTE estimated that long-distance bus passenger numbers had been increasing by about 60 percent nationally between 1980 and 1984. Furthermore, between 1980 and 1984 the relative position of air, rail, and bus fares changed, with bus fares declining after the entry of new bus operators. In the same time period, economy rail and air fares increased by more than 50 percent (*5*).

On the Sydney-Canberra corridor the BTE estimated that 48,000 bus passenger trips were completed in 1984, a 108 percent increase over figures for 1977 to 1978 (Table 5). Air transport statistics indicated that a total of 484,814 passengers traveled by air on this corridor during 1983 and 1984 (*9*).

On the Sydney-North Coast corridor figures are only available for Sydney to Brisbane interstate services. It was estimated that 451,000 passenger trips were completed by bus in 1984, compared with 1,198,298 air passenger journeys at the same time (*9*).

Corresponding data for rail travel are not available, but it has been reported that intercity rail (XPT) services had undergone fare reductions during 1983 and 1984, which resulted in increased patronage. The highest percentage increase was on the Sydney-Canberra corridor (67.2 percent), whereas pa-

TABLE 4 TRAVEL CHARACTERISTICS OF BUS AND RAIL PASSENGERS

	Sydney-Canberra		Sydney-North Coast	
	Bus	Rail	Bus	Rail
	(n=754)	(n=420)	(n=739)	(n=893)
Trip Purpose	(percentage)			
Holiday	22	24	28	35
Visiting Friends	48	37	38	35
Personal Business	10	15	8	5
Business/Work	7	18	7	8
Holiday/Visiting	13	16	19	17
X^2 (df=4)	48.09**		13.29**	
Length of Stay	(percentage)			
1 - 3 days	54	35	31	21
4 - 14 days	31	40	55	63
15 +	15	25	14	16
X^2 (df=2)	9.87**		18.72**	
Size of Party	(percentage)			
Number of Family Members				
None	44	44	56	50
1+	56	56	44	50
X^2 (df=1)	0.01*		6.69***	
Number of Friends				
None	57	61	67	61
1 +	43	39	33	39
X^2 (df=1)	2.03*		5.25**	

* Not Significant
** Significant at the 0.05 level
*** Significant at the 0.01 level

Source: Calculated from BTE Passenger Surveys, 1987.

TABLE 5 PASSENGER DATA, PRETRIAL LIBERALIZATION

	National Travel Survey 1977–78[a] ('000 Passenger Trips)[b]		Bureau of Transport Economics 1984 ('000 Passenger Trips)[b]	
	Sydney-Canberra	Sydney-Brisbane	Sydney-Canberra	Sydney-Brisbane
Air	85	235	552	1,198[c]
Rail	13	15	NA	NA
Bus	34	8[d]	48	451

NOTE: NA = not applicable.
[a] Based on national sample of households. Corridor trips are based on sample households who reported trips on these corridors.
[b] All passenger figures are two-way (round) trips.
[c] Department of Aviation statistic for financial year 1984–85.
[d] The survey provided one-way trips for bus on this corridor. This figure has been doubled.

tronage increased by 13.6 percent on the North Coast XPT rail service (*10*). These increases occurred in the context of long-term declines in country rail travel in NSW.

Passenger profiles for the study corridors are not available for the pretrial liberalization period. The National Travel Survey and BTE surveys of air and bus travelers in 1983 to 1984 only provide a national profile of the long-distance passenger market, which can be compared with emerging profiles during trial liberalization in NSW.

Generally, these surveys show that more bus than rail travelers come from lower income categories than do automobile and air travelers. Both bus and rail show a high proportion (more than 60 percent) of passengers in the young and old categories. A large proportion of bus and rail travelers are retired, students, or unemployed. The sex of passengers also distinguishes the travel market. Female passengers account for 60 percent of bus and rail passengers and slightly more than 20 percent of the total airline market (*5*).

Travel by bus, rail, and automobile for private reasons is similar and accounts for 80 percent of trips on each mode. The air market is different, with only 53 percent of journeys being for private purposes and 40 percent for business (*5*).

In summary, the period before trial liberalization on the study corridors was characterized by increasing competition among long-distance bus operators in NSW, lower fares, and a growing bus passenger market relative to rail and air.

Effects of Trial Liberalization

As Table 1 indicates, trial liberalization on both corridors produced significant changes in service levels for the traveling public. On both corridors the frequency of bus services increased significantly. On the North Coast corridor the change in service frequency was substantial, and all the more significant, as interstate bus companies were now able to pick up and discharge passengers in centers along the corridor. In addition, trial liberalization enabled better accessibility to bus services for people from many North Coast centers. Direct services were also available, whereas previously passengers had to transfer. Travelers now enjoyed a greater choice of operator and type of vehicle.

Table 1 also shows reduced fare levels on both corridors during liberalization. By comparison, rail and air fares remained unchanged, as did service levels. Toward the end of the trial period, East-West airlines introduced discounted fares on the North Coast corridor. Failure to respond to increased bus competition had a significant impact on the revenue of rail and air operators in both corridors. On the Sydney-Canberra corridor 64 percent of total revenue loss was attributed to the fall in rail passenger revenue, as compared with 40 percent of total revenue loss on the Sydney-North Coast corridor. Revenue loss to regional airlines on the Sydney-North Coast corridor was higher compared with rail losses: 60 and 40 percent, respectively (7).

Table 6 shows the estimated demand for bus, rail, and air services over the two trial corridors during trial liberalization. On the Sydney-Canberra corridor intrastate bus patronage (134,000 passengers) is more important relative to rail, and represents more than one-quarter of all passengers traveling between Sydney and Canberra. The increase in demand for bus travel during the trial on the Sydney-Canberra corridor compared with the same period 12 months earlier was 80 percent (74,500 passengers). Compared with BTE data for 1984 (Table 5), the increase was in the order of 65 percent.

On the Sydney-North Coast corridor the intrastate passenger market is dominated by air, rail, and bus, respectively, and services to highlight the presence of different modal characteristics and market shares as compared with the Sydney-Canberra corridor. Of the total passengers over this corridor (1,090,000), intrastate rail patronage accounted for 18 percent of the market. However, of the combined bus and rail market, nearly 70 percent (197,000) traveled by rail transport. The estimated increase in demand for bus travel during the trial compared with the same period 12 months earlier was 130 percent (42,000 passengers).

Of interest are estimates of passengers diverted to bus services from other modes and the number of generated trips. BTE calculations suggest that between 35 and 42 percent of the 134,000 estimated bus passengers diverted from other modes of transport to bus services on the Sydney-Canberra corridor. The largest percentage diverted from rail, followed by automobile, then air (7). On the Sydney-North Coast corridor, fewer passengers diverted from rail, air, and automobile to buses. Between 26 and 37 percent of estimated bus patronage (97,000) diverted from other modes.

The BTE study indicated that, on the Sydney-Canberra corridor, fare reduction was the factor that most influenced the decision of passengers to divert to bus travel from competing modes of public transport. More flexible departure and arrival times also influenced the decision to travel by bus. Similar reasons also influenced bus travelers on the Sydney-North Coast corridor, however, fare reductions were not as influential on this corridor.

The trends identified during trial liberalization clearly had their beginnings in the preliberalization period. Bus patronage continued to increase at the expense of rail, especially on the Sydney-Canberra corridor. Rail patronage, although declining at this time, was still more important than bus travel on the Sydney-North Coast corridor because of longer distances and the availability of overnight sleeper accommodation on Sydney-to-Brisbane trains. However, since completion of trial liberalization the declining long-distance passenger market for rail has led to a major rationalization of country and intrastate rail services in NSW.

The sex distribution of passengers does not vary statistically across modes on both corridors (Table 3). Females constitute a higher proportion of bus and rail users than males on both corridors and appear to be a considerable segment of the traveling public—a trend already evident before liberalization. There is a distinct market segmentation based on the age distribution of passengers. The bus market on both corridors is dominated by young travelers. On the Sydney-Canberra corridor approximately one-third of bus passengers are between 20 and 29 years old, compared with 21 percent of rail passengers on the Sydney-North Coast corridor. Approximately one-third of bus passengers are under 20 years of age.

The rail market on both corridors is dominated by passengers older than 50 years old. Twenty-five percent of rail passengers on the Sydney-Canberra corridor are older than 50, and the proportion increases to 40 percent on the Sydney-North Coast corridor.

The quite distinct market segmentation based on age can be partly attributed to the concession fares offered by the railways to pensioners. Also, people with limited mobility,

TABLE 6 ESTIMATED PASSENGERS DURING TRIAL LIBERALIZATION

	Thousands of Passengers			
	Sydney-Canberra	Canberra-Sydney	Sydney-North Coast	North Coast-Sydney
Air	150	153	209	183
Rail	35	40	95	102
Bus	63	71	49	48

particularly the aged, are discouraged from using conventional bus services because it is difficult for them to use steps.

The proportion of middle-aged travelers on both modes is quite low but higher on the Sydney-Canberra corridor because of business travelers. Compared with passenger profiles in preliberalization times, the young and old continue to be the major customer base in the long-distance passenger market.

On the Sydney-Canberra corridor, the majority of bus and rail passengers are students (Table 3). The next two largest occupational groups differ between the modes. The next largest occupational groups on bus services are passengers employed in professional occupations followed by retired people, whereas on rail services, the next two largest occupational groups are passengers engaged in household duties and professional occupations.

On the Sydney-North Coast corridor, a different market segmentation emerges. Students dominate the bus market (26 percent), whereas retired persons make up the largest segment of rail passengers (32 percent). The next largest occupational group on bus services is the retired (16 percent), and the next largest group on rail are students (15 percent). Note that bus and rail travel on both corridors is not associated with lower occupational status groups or with the transportation disadvantaged.

Table 4 shows that discretionary travel (e.g., to visit friends and relatives) and holiday travel were the most important reasons for travel on both modes.

The only significant difference in trip purpose between the modes occurred on the Sydney-Canberra corridor. Business travelers make up a significantly larger proportion of rail passengers than bus passengers. This is quite surprising given the 10-fold increase in bus services and wider selection of arrival and departure times that occurred after the relaxation of restrictions on buses (Table 5). Buses are now an ideal mode for business travel between Sydney and Canberra because business travelers now have a same-day return service, allowing up to 8 hr in either city. The rail timetable does not offer such flexibility, and one would expect bus travel to be more popular with business travelers.

On the North Coast corridor, discretionary travel is the most important reason given for making the journey, and obviously the tourist functions of the region are reflected in this result. Very few passengers on either mode make personal or business trips.

There is a distinct market segmentation between bus and rail services on both corridors in terms of the length of stay at destination of bus and rail passengers (Table 4). On the Sydney-Canberra corridor, the bus is more popular for stays of less than four days (54 percent), whereas the train is more popular for longer duration stays, particularly those between 4 and 14 days (40 percent).

On the Sydney-North Coast corridor, close to two thirds of rail passengers and more than 50 percent of bus passengers stay at their destination between 4 and 14 days. Buses on this corridor are also more popular than rail services for short duration stays (less than 4 days), whereas rail services are more popular than buses for stays of longer than 2 weeks.

On the Sydney-Canberra corridor bus and rail passengers are more likely to be traveling either with the family or alone (Table 6). There is no difference between modes in the proportion of bus and rail passengers traveling with family (56 percent), and the bus is slightly more popular for traveling with friends. The most interesting result to emerge from this table is that group travel on both modes is more likely to be in family groups than in groups of friends.

The Sydney-North Coast corridor differs from the Sydney-Canberra corridor in that passenger segmentation based on size of travel party is slightly more distinct on this corridor. A higher proportion of rail passengers on this corridor either travel with family (50 percent) or friends (39 percent) than on the buses (44 percent and 33 percent, respectively). Presumably, the large proportion of elderly travelers on rail services would be more sensitive to comfort than factors such as departure and arrival times.

Interestingly, despite the introduction by bus companies of high-quality luxury coaches with reclining seats, toilets, air conditioning, and videos in an effort to improve the public perception of bus services, comfort has been the dominant factor for the railways in attracting passengers (*7*).

This factor becomes even more evident when considering passengers' reasons for diverting to buses. Only 17 percent of bus passengers on the Sydney-Canberra corridor diverted to the bus because of the comfort of coaches (*11*).

Effects of Deregulation

Deregulation of the intrastate long-distance bus market in NSW occurred in 1988 following the success of trial liberalization. Although a proliferation of intrastate bus services occurred across the state, events during the past 2 years have led several bus operators on the study corridors to leave the market, with resulting changes to service schedules. However, the contestable market for long-distance services remains sustainable, with continued growth in the bus passenger market and consolidation of market share positions.

The most significant changes to have occurred in the long-distance passenger market relate not to bus services but to rail and air services. As mentioned earlier, rural rail services in NSW had experienced declining patronage and revenue over many years, with significant losses occurring during trial liberalization. As a result, country rail services underwent a major reorganization that involved substitution of buses for rail services, formation of a long-distance government bus service called Countrylink, and concentration of rural rail passenger services on trunk line intercity (XPT) train services.

Travelers on the Sydney-Canberra corridor were able to benefit from rationalization in that Countrylink bus service was added to the market to replace the one bus company that left the market. Several Countrylink bus services also commenced on the Sydney-North Coast corridor. Countrylink now competes with remaining rail services and with other long-distance bus operators.

In addition, the NSW government has partially deregulated the intrastate airline system, thereby allowing increased competition on main passenger corridors and commuter routes. However, during the past 12 months some commuter airlines have withdrawn services from low-volume routes as a direct result of increased competition from buses.

The government's policy of deregulation, rationalization, and privatization of transport supply has created a competitive market on the major trunk routes where efficiency and market

TABLE 7 PASSENGER DATA SINCE DEREGULATION

		Sydney-Canberra	Sydney-North Coast
1988-89	AIR[a]	621,471	218,244
	RAIL[c]	220,698	796,310
	BUS	na	na
1989-90	AIR[a]	344,252	153,080
	RAIL[c]	148,221	671,529
	BUS	na[b]	na
1990-91	AIR[a]	556,168	204,667
	RAIL[c]	70,051	606,532
	BUS	na[b]	155,000[d]

[a] Department of Aviation statistics.
[b] Bus passenger figures are available for Countrylink services.
[c] Figures obtained from State Rail and Countrylink.
[d] Estimate only based on sample data supplied by the Department of Transport.
na Not Available.

share are fundamental for sustained operation. Bus companies have also consolidated their position on the lower volume intrastate routes following cessation of rail service to many small centers.

It is difficult to assess the impacts of deregulation on the passenger market in the study corridors because no follow-up studies have been undertaken. Comparative statistics on travel demand and market share are also limited. Complete statistics are available only for air and rail services (Table 7). However, it seems reasonable to suggest that bus patronage on the Sydney-North Coast corridor increased significantly in 1988 due to demand generated by EXPO 88, which was held in Brisbane. The examination of 1988 bus timetables on this corridor showed the same number of bus operators as during the trial period, but an increase in the number of services provided to certain centers (Table 1). Note that services doubled on the Sydney-Canberra corridor, whereas the number of bus operators remained constant. Bus patronage on both corridors was also boosted by rail service cutbacks and lower fares relative to rail and air.

In 1989 bus patronage on both corridors would have increased once again due to the national pilots' strike, which grounded all intercity services. Tourism and business traffic was diverted to bus, rail, and automobile transport (Table 7). On the Sydney-North Coast corridor a series of horrific road accidents involving long-distance buses affected patronage levels for certain operators (both operators have since ceased operation), and the NSW government came under increased pressure to improve rail services (especially overnight sleeper services to Brisbane), enforce speed limits for buses and trucks (legislation has since been enacted), and commence major road improvements on the main highway linking Sydney and Brisbane. Three bus operators left the Sydney-North Coast market in late 1989, with a resulting decline in service frequency (Table 1).

The dynamics of the long-distance passenger market were once again highlighted in 1991 when a new airline (Compass Airlines) commenced intercapital services following national deregulation of the domestic airline system. Compass commenced services between Sydney and Brisbane (1 hr) with heavily discounted airfares. It is now quicker to reach the North Coast of NSW by flying to Brisbane and traveling south by bus. The entry of a new air carrier into the long-distance passenger market most certainly affected travel demand and market share among bus and rail operations. However, Compass went into liquidation on December 21, 1991, and attempts are under way to revive the airline. At least two new airlines are scheduled to commence intercapital service in 1992.

Deregulation has continued to provide major benefits to the traveling public as a result of improvements in service frequencies, lower fares, and improved quality of service. Improved access to tourist centers on the North Coast of NSW and access to larger centers for residents of small communities are additional benefits. Furthermore, it appears that the intrastate long-distance passenger industry is more efficient and responsive to changing circumstances.

CONCLUSION

Strict regulations governing the operation of long-distance intrastate bus services in NSW between 1931 and 1987 prevented competition between bus and rail services. Without the competitive threat of buses, the railways had an artificial monopoly of the intrastate land passenger market. Deregulation of the intrastate bus industry created an environment in which bus services now represent a significant threat to rail services. Further, deregulation of the NSW and national airline system has produced lower air fares, which, in turn, now represent a threat to intrastate long-distance bus and rail services. Passenger markets and travel demands are continually responding to changes within the intrastate long-distance passenger industry.

This discussion has shown that trial liberalization and deregulation have been beneficial to both the industry and the traveling public. The authors have also shown that the industry is dynamic and responsive to changing circumstances. The future scenario is of an even more dynamic and competitive intrastate long-distance passenger industry.

ACKNOWLEDGMENT

The authors would like to thank the Bureau of Transport Economics, Canberra, for providing the trial liberalization data.

REFERENCES

1. State Rail Authority. *Submission to the Department of Transport: Review of Intrastate Long Distance Bus Services in New South Wales.* State Rail Authority, 1986.
2. Department of Motor Transport. *Review of Intrastate Long Distance Bus Services.* Department of Motor Transport, 1988.
3. M. Carnahan. *An Economic Analysis of the Canberra-Sydney Coach Route Following Deregulation.* Paper presented to Competition and Ownership of Bus and Coach Services: An International Conference. Thredbo, Australia, May 1989.
4. D. Saffron. *Policy Relating to Long Distance Bus Services in New South Wales.* Paper presented to Regional Science Association, Australia and New Zealand Section 11th Annual Conference, University of New South Wales, Sydney, 1986.
5. Bureau of Transport Economics. *Australian Long Distance Coach Industry Review.* Australian Government Publishing Service, Canberra, 1985.
6. A. K. Cross and R. P. Kilvington. Deregulation of Inter-City Coach Services in Britain. *Transport Reviews,* Vol. 5, No. 3, 1985, pp. 225–246.
7. Bureau of Transport Economics. *Intrastate Bus Services in New South Wales.* Australian Government Publishing Service, Canberra, 1987.
8. N. A. Hirsch and D. A. Russell. *National Travel Survey 1977–78: Statistical Adjustments and Final Results.* Bureau of Transport Economics, 1981.
9. Department of Aviation. *Domestic Air Transport 1983–84.* Commonwealth of Australia, Canberra, 1985.
10. State Rail Authority of New South Wales. *Annual Report.* 1984.
11. A. Harrington and B. Parolin. Factors Affecting the Use of Bus and Rail Services in a Partially Deregulated Environment. *Transportation,* Vol. 18, 1991, pp. 175–193.

Publication of this paper sponsored by Committee on Intercity Rail Passenger Systems.

PART 2

Freight Rail Issues

Freight Service Quality Cost Economics and a Hypothetical Railroad Example

David G. Brown

Service quality is an integral aspect of the freight transportation process. Variations in service quality have a direct effect on shipper resource consumption, which in turn affects both the demand and supply aspects of freight transportation. In this paper, the cost (supply) aspects of freight service quality are examined. Emphasis is on the implications for carriers and transportation analysis, with the primary goal of providing some concrete understanding of these issues. This goal is achieved by using a hypothetical railroad example to explore and illustrate a freight service quality model. The example is based on an engineering model of railroad operations that captures the essence of run-through train operations. The example also serves a secondary goal of fostering general understanding of railroad service quality. After the freight service quality and railroad operating models are reviewed, a more detailed discussion of service quality economics, focused on the interaction between carrier operating variables, service quality variables, and carrier and shipper cost, is presented. In particular, the implications of three conclusions are examined: (a) the most economically efficient service quality level is obtained by minimizing the full cost of transportation—the sum of carrier costs and service sensitive shipper costs; (b) this service quality level is also profit-maximizing for the carrier; and (c) transportation cost analysis should be based on the full cost of transportation instead of carrier costs alone. The negative consequences of disregarding these conclusions are illustrated with the railroad example.

"In a railroad system ideally organized from the economic point of view we should expect the transit time of a particular service to be cut down whenever the consequent reduction of inventory costs exceeds the costs of providing this faster service," (*1*, p. 116).

"Shippers are faced with ordering costs, inventory costs, stockout costs, and a variety of other costs which vary with the type and quality of the transportation purchased. These costs are equally relevant to the analysis of the efficiency of the freight transport system but are often ignored or treated incorrectly," (*2*, p. 500).

Service quality is an integral aspect of the freight transportation process. Variations in service quality have a direct effect on shipper resource consumption; one example is the inventory cost associated with goods-in-transit. In turn, this resource consumption affects both the demand and supply aspects of freight transportation. Therefore, shippers, carriers, and transportation analysts in general should all be concerned with the impact of freight service quality. However, of these three groups, only shippers have begun to fully recognize its significance and the direct impact it has on their cost structure and ultimately on their profit. Examples of this recognition include the use of mechanisms such as just-in-time delivery.

Carriers have reacted to this increased shipper sensitivity by paying greater attention to the demand or marketing aspects of freight service quality. Individually, they have improved overall service quality and introduced new services, with the goal of attracting new traffic and retaining current traffic in an increasingly competitive transportation market.

However, there is little evidence that carriers have paid significant attention to the supply or cost aspects of service quality. That is, carriers generally have not explicitly considered the efficiency factors associated with service quality in the joint utilization of carrier and shipper resources. Like other production efficiency considerations, this may have a direct impact on carrier profit. Traditional transportation economic analysis has also largely ignored the effects of freight service quality.

A general model of freight service quality and carrier economics presented by Brown (*3*) addresses both the demand and supply considerations. Three of the major conclusions in the paper concern transportation cost and supply:

- At any given volume level, the most economically efficient service quality is one that minimizes the full cost of transportation—the sum of carrier costs and service sensitive shipper costs. This optimal service level is called quality efficiency (QE).
- QE is also profit maximizing for the carrier (at the given volume level).
- No matter what service quality level is implemented, transportation cost analysis should be based on the full cost of transportation instead of only carrier costs.

These points are reflected in the quotations at the beginning of this paper.

In Figure 1, the two dashed cost curves indicate how shipper inventory and carrier expenses might respectively vary with a single service quality variable; the solid full cost curve is the sum of these expenses. Beckmann et al. (*1*) indicated that the most efficient service quality level minimizes the full cost of transportation; this QE service quality level is indicated by Z^* in Figure 1. The Beckmann quotation also suggests that it is in the carrier's self-interest to implement this service

---

Code AS/Bz, Department of Administrative Sciences, Naval Postgraduate School, Monterey, Calif. 93943-5000.

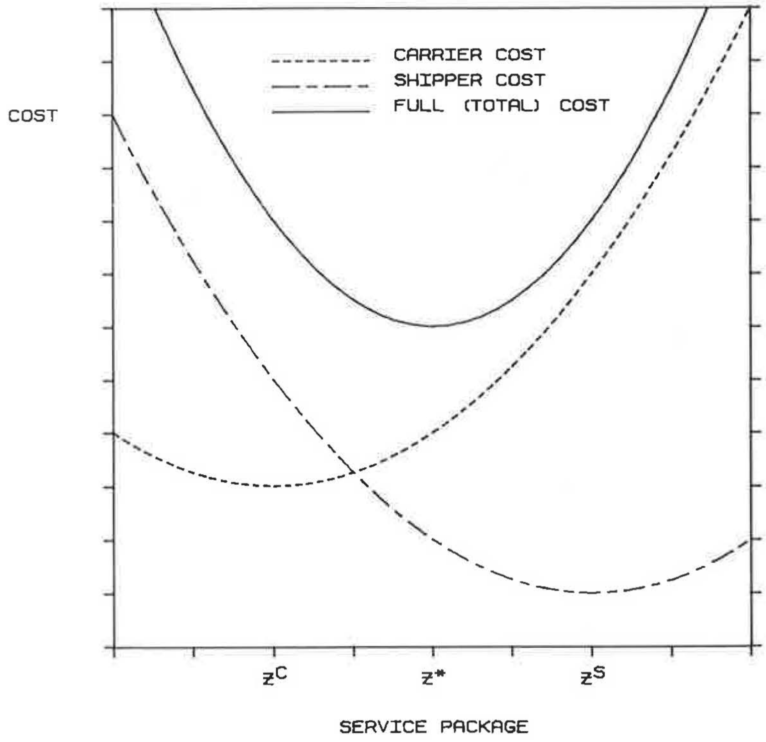


FIGURE 1 Illustration of freight service quality cost economics.

quality level (i.e., QE maximizes carrier profit). However, QE is not always implemented. In particular, if the carrier effectively ignores shipper satisfaction and concentrates on controlling its own costs, then pure carrier cost minimization (CCM) is a likely alternative to QE; this quality policy is indicated by Z^c in Figure 1. In this paper, CCM provides a useful reference point for comparison with QE.

The Nason and Kullman quotation (*2*) indicates that, in addition to defining the optimal service quality level, full cost is also appropriate for general transportation economic analysis, such as carrier profit, cost, and regulatory policy analysis. Nason and Kullman specify several internal shipper cost elements that vary with freight service quality. These cost elements reflect resources that are as integral to the provision of transportation service as are the resources consumed by the carrier and should be considered in any broad-based economic evaluation of freight service. By including these elements, the full cost of transportation accounts for all resource consumption associated with the provision of freight service (ignoring externalities). Transportation economic analysis based on carrier cost alone is incomplete and potentially misleading.

The primary goal of this paper is to provide a more concrete understanding of the cost (or supply) aspects of the freight service quality model. This is accomplished by establishing a relationship between specific service quality variables and carrier operating variables through a simple operations model based on railroad technology. Thus, the service quality model and its implications may be discussed with respect to tangible attributes such as dollars, miles per hour, transit time, train length, and horsepower. Fostering general understanding of railroad freight service quality economics is a secondary goal of this paper. The limited example presented here captures the essence of run-through train operations.

The cost aspects of the service quality model are introduced next, followed by an overview of the railroad operations model. The railroad model is then implemented to illustrate and support a more detailed discussion of freight service quality cost economics. This paper and another by the author (*3*) are based on the author's doctoral dissertation (*4*).

FREIGHT SERVICE QUALITY COST MODEL

Production cost functions are used by economists and others to address issues such as basic questions of production efficiency and industrial organization. As discussed previously, the production of freight service requires the consumption of both carrier and shipper resources. The costs associated with both sets of resources are internalized in the service production process and are affected by the service quality. In this section, the cost portion of the freight service quality model is summarized and discussed. In the model presented here, a single carrier provides a single service to one or more shippers; shippers are treated as a relatively homogeneous aggregate group. This relatively simple version of the service quality modeling framework allows one to focus on and clarify some central issues.

The freight service quality model includes three cost functions that correspond to the three curves in Figure 1. Each specifies cost (per unit time) as a function of annual freight volume v and service quality z. The functions are carrier cost

$C(v, z)$, shipper cost $S(v, z)$, and the full cost of transportation:

$$T(v, z) = C(v, z) + S(v, z)$$

For this simple application, freight volume is the total annual volume summed over all shippers. The service package z is a vector of one or more service quality variables; a two-variable service package is specified later.

A service quality variable is an observable characteristic of the freight service, which affects both the carrier's cost and the shipper's cost. In addition to average transit time, other possible service quality variables include transit-time variability, loss and damage, and shipment size. These variables are distinguished from related carrier operating variables, such as speed and frequency, that are not always observable by the shipper.

For this application, the carrier provides the same service quality to all shippers [a multiple service package extension is presented elsewhere (*4*)]. The carrier cost function explicitly acknowledges the functional relationship between carrier cost and service quality level. For example, improving reliability by increasing train frequency may require a larger labor expense.

Shippers are simply consumers of freight transportation and active customers of the carrier. Shipper cost $S(v, z)$ includes the opportunity costs of all shippers, which directly vary with the service quality variables; these are typically inventory costs. The volume argument allows these internal shipper cost elements to also vary in a generalized way with annual volume. The shipper cost function does not include the freight charge paid to the carrier; the freight bill does not (in this model) directly vary with the service quality variables, and more important, it does not represent additional resource consumption by the combined shipper-carrier entity. (Instead, the freight bill may be viewed as a transfer payment within the combined entity.)

Formally, a quality policy is a rule that specifies the quality level implemented by the carrier as a vector function of the volume level. The QE and CCM quality policies are respectively indicated by an asterisk and a superscript c and defined by minimizing full cost and carrier cost with respect to service quality:

$$Z^*(v) = \underset{z}{ARG\ MIN}\ T(v, z) \tag{1}$$

$$Z^c(v) = \underset{z}{ARG\ MIN}\ C(v, z) \tag{2}$$

where *ARG MIN* denotes the service quality vector (service package), which minimizes the objective function.

The QE policy is fundamentally important because it specifies the socially optimal service quality level for every volume level (with respect to technical efficiency). Furthermore, QE also maximizes carrier profit. This should provide the carrier with sufficient motivation for implementing this quality policy. However, successful implementation requires a management structure sufficiently fine-tuned to discern and fully exploit all quality-related profit opportunities. No management is perfect, and carriers sometimes fail to explore these opportunities.

The CCM policy provides an alternative for comparison with QE. CCM generally does not maximize carrier profit. However, a carrier will tend toward CCM if it is overly concerned with internal cost control and relatively insensitive to shipper satisfaction. For example, this is a likely outcome with the classical railroad management structure, in which the viewpoint of the operations department eclipses input from the marketing (traffic) department (*5*).

DESCRIPTION OF MODEL UNDERLYING RAILROAD EXAMPLE

The primary focus of this paper is a hypothetical railroad example that provides a framework for discussing the service quality economic concepts introduced above, in the context of realistic freight operations. The example is based on a model that was originally designed to examine relatively detailed considerations, such as the tradeoff between unit crew cost and train frequency in terms of the total labor cost when implementing QE. In the present context, the detail adds to the realism and thus increases the illustrative power of the example. In this section, the transportation context is described, shipper cost and transit time are discussed, railroad cost elements are described, and model implementation issues are addressed. A complete description of this model is available elsewhere (*4*).

Transportation Process

In this model, a railroad provides through train service from a single origin yard to a single destination yard. The railroad has two operations decision variables: annual train frequency and train cruising speed. Shippers are responsible for getting shipments to the origin yard and picking up at the destination yard (like some types of railroad intermodal service).

The scheduled headway between successive train departures is inversely proportional to train frequency. It is assumed that shipments arrive at a constant rate, are loaded on identical freight cars, and leave on the next train scheduled for departure. The locomotive horsepower per revenue ton assigned to each train is a function of cruising speed. Both the number of cars and the number of locomotives per train are continuous variables in this model. A lag between the scheduled and the actual departure allows time for loading the last shipments, attaching power, and other unexpected delays. On actual departure, trains accelerate up to the cruising speed.

The only stops between the origin and destination yards are for crew changes, where the trains brake to a stop, wait for the actual crew change and any required tests and inspections, and accelerate again up to cruising speed. The number of crews per train along the route is a continuous variable determined by the average crew district length and trip distance. The route is assumed to be entirely level and straight. At the destination yard, the locomotives are disconnected from the train, and after a short unloading process, the shipments are available to the shippers.

Shipper Cost and Transit Time

For this application, the service package is defined by two service quality variables, average transit time in hours (μ_t), and transit-time standard deviation (σ_t):

$$z = (\mu_t, \sigma_t)$$

Shipper cost is linear on both service quality variables and proportional to the total annual volume.

$$S(v, z) = v \cdot (U_\mu \cdot \mu_t + U_\sigma \cdot \sigma_t) \tag{3}$$

This shipper cost formulation is based on work documented elsewhere (*4*). The shipper unit-cost coefficients U_μ and U_σ are parameters of the railroad model.

Shippers may vary in their individual sensitivities to transit-time characteristics (because of different commodity values and other factors). Equation 3 imposes a loose restriction on the shipper mix by fixing the aggregate unit-cost characteristics when averaged across the individual shippers on a per-unit-volume basis.

Transit time begins when the shipment arrives at the origin yard and ends when it is available to the shipper at the destination yard. This includes five types of time segments: (a) waiting for scheduled departure, (b) delay before actual departure, (c) interval of continuous train movement (within a crew district), (d) crew change delay during which the train is stationary, and (e) unloading and other delay at destination yard.

Each of these time segment types is stochastic. The waiting time at the origin yard has a uniform probability distribution over the scheduled headway between trains. The standard deviation for each of the other four time segment types is determined by multiplying an average time value by an externally specified coefficient of variation. Exogenous average time values are also specified for three of these time segment types: delay before actual departure, crew change delay, and destination yard delay.

Each interval of continuous train movement includes one acceleration period, one cruising speed period, and one braking period. The time and distance required to accelerate from a stop to cruising speed, and to brake to a stop from cruising speed, are all functions of the cruising speed. The average interval of continuous movement is thus a function of cruising speed and average crew district length.

Railroad Cost

Four railroad cost elements are included in the carrier cost function. These are the annual cost of train crew labor (LC) and diesel fuel (FC), and the implied annual rentals associated with both railroad cars (CC) and power (PC):

$$C(v, z) = LC + FC + CC + PC$$

The formulation of each cost element includes a unit-cost parameter. For example, annual train labor cost is the product of train frequency, number of crew districts, and unit crew cost. On the basis of empirical studies (*6*), annual fuel consumption is assumed to be proportional to work performed by locomotives during acceleration and while cruising, which is primarily a function of annual volume and cruising speed. Annual car rental is the product of annual volume, average transit time, and cost per available car hour, divided by the net capacity of each car. Locomotive rental is similarly based on the available horsepower hour between actual departure and arrival.

These four cost elements account for most of the above rail operating costs, which have a direct relationship with the two service quality variables. However, other railroad cost elements, such as track maintenance and investment, also have a significant relationship with freight service quality. For example, poor quality track may impact service quality directly through loss and damage, or indirectly through speed restrictions. Furthermore, improving service quality through increased train frequency reduces the usable time windows available for track maintenance activities, and thereby may increase track maintenance costs (*7*). Therefore, the analysis presented in this paper is primarily illustrative, and although incomplete, still offers some understanding of railroad service quality economics.

Implementation Issues

The horsepower, acceleration, and braking functions required for the railroad cost elements are implemented with a simple train performance calculator (*4*). Horsepower per net ton is determined by equalizing the available tractive effort of the locomotives with the total train resistance at cruising speed. An iterative algorithm is used to simulate train acceleration, whereas train deceleration is approximated with a constant braking rate.

Most of the parameter values for this example are presented in Table 1; others include such train performance calculator parameters as car resistance coefficients. The derivation or rationale for all of the parameter values is discussed elsewhere (*4*). For the purposes of this hypothetical example, these values only need to be realistic. Most are based on empirical data or expert opinion. The shipper unit-cost coefficients are the most arbitrary; they were calibrated so that the model would yield reasonable and illustrative results. An extensive sensitivity analysis with respect to these two coefficients indicated that they have no effect on the general observations presented here.

The cost-minimization problems associated with the QE and CCM policies (Equations 1 and 2) were solved with a basic pattern-search type vector optimization procedure in conjunction with a golden section line-search (*8*), based on the two operating variables.

DISCUSSION OF HYPOTHETICAL RAILROAD EXAMPLE

In this section, a hypothetical example based on the railroad model is examined with respect to the freight service quality cost model. The discussion focuses on efficiency and cost analysis implications for carriers and transportation economists. The relationships among operating variables, cost, and service quality for fixed annual volume are examined first, followed by the effects of annual volume variability.

TABLE 1 PARAMETER VALUES FOR RAILROAD EXAMPLE

LOCOMOTIVE PARAMETERS		
A_P	Axles/Locomotive	4
E_P	Efficiency	0.83
HPL	Horsepower/Locomotive	3,000
R	Fuel Consumption Rate (Gallons/Million Foot-Pounds)	0.028
U_P	Unit Cost ($/Available Horsepower Hour)	0.006
W_P	Weight (Tons/Locomotive)	130
f	Coefficient of Adhesion between Wheel and Rail	0.25
RAILROAD CAR PARAMETERS		
A_C	Axles/Car	4
TW_C	Tare Weight (Tons/Car)	30
U_C	Unit Cost ($/Available Car Hour)	0.75
v_C	Net Capacity (Tons/Car)	70
TRAIN CREWS		
d_L	Crew District Length (Miles)	150
U_L	Unit Cost ($/Crew)	700
OTHER RAILROAD PARAMETERS		
b	Braking Rate (MPH/Second)	2.2
U_f	Unit Fuel Cost ($/Gallon)	0.5
SHIPPER UNIT-COST COEFFICIENTS		
U_μ	Average Transit-Time Unit-Cost ($/Ton-Day)	2.0
U_σ	Transit-Time Standard Deviation Unit-Cost ($/Ton-Day)	2.0
TRANSIT-TIME PARAMETERS		
μ_2	Average Delay before Actual Departure (Hours)	2.0
V_2	Actual Departure Delay Coefficient of Variation	0.1
V_3	Continuous Movement Coefficient of Variation	0.1
μ_4	Average Crew Change (Hours)	0.25
V_4	Crew Change Coefficient of Variation	0.5
μ_5	Average Unloading and Other Destination Delay (Hours)	2.0
V_5	Destination Delay Coefficient of Variation	0.1
INDEPENDENT VARIABLES		
d	Trip Distance (Miles)	1,500
v	Annual Volume (Tons)	500,000

Operating Variables, Cost, and Service Quality

In this section, service quality and cost are presented as functions of the operating variables, carrier and full cost are examined as functions of service quality, and the QE and CCM quality policies are compared. The discussion is focused on the contour surfaces presented in Figures 2–9. These surfaces were developed by implementing the railroad model for each of 1,116 operating variable ordered pairs, with annual volume fixed at 500,000 tons (Table 1). The train frequency range (Figure 2) implies that train length will vary from 18 to 143 cars, and as a consequence of the cruising speed range, horsepower per net ton will vary from 0.29 to 5.07.

Service Quality and Cost as Functions of Operating Variables

A carrier generally cannot implement the service package directly. Rather, service quality variables are functions of the operating variables, and the carrier implements the service

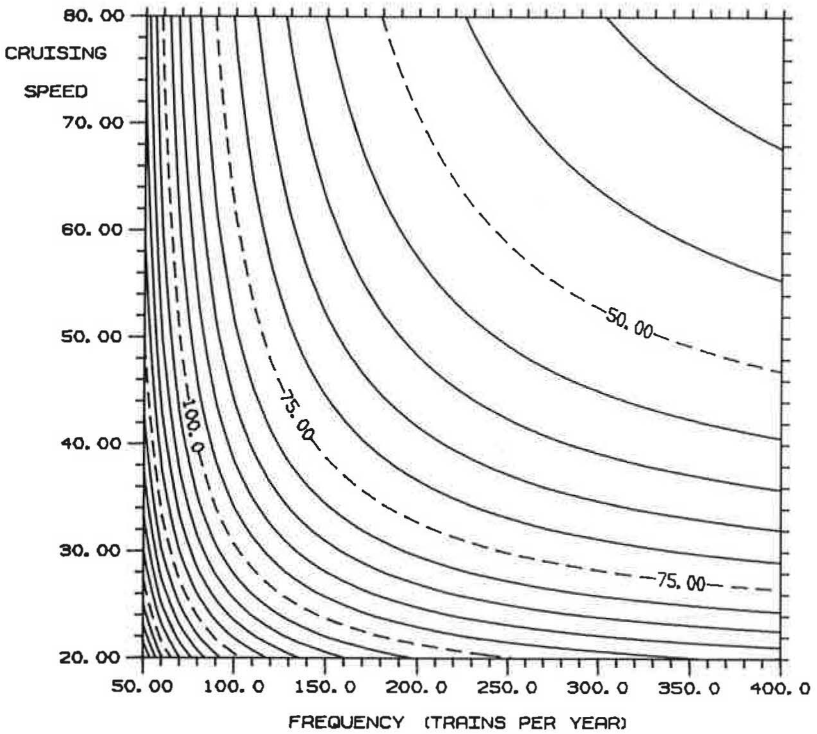


FIGURE 2 Average transit time (hours) as a function of operating variables.

package indirectly by choosing values for the operating variables (train frequency f and cruising speed s):

$$z = z(f, s): \mu_t = \mu_t(f, s) \text{ and } \sigma_t = \sigma_t(f, s) \quad (4)$$

In this model, service quality is not a function of annual volume, except through the indirect medium of quality policies. In other applications, service quality may be modeled as a direct function of annual volume; for example, this may reflect the effects of congestion.

The two service quality variable functions [$\mu_t(f, s)$ and $\sigma_t(f, s)$] are depicted in Figures 2 and 3, respectively. In this instance, average transit time is strongly affected by both operating variables with diminishing returns (Figure 2) (i.e., further increases in either operating variable will yield smaller decreases in average transit time). In Figure 3, it is evident that transit-time standard deviation is almost solely a function of train frequency (with diminishing returns). This reflects the effect that train frequency has on the wait before scheduled departure. Cruising speed would have a greater impact in Figure 3 if V_3 were set at a higher value in Table 1. For more typical railroad operations with mixed-freight trains, missed connections in intermediate yards are the main cause of transit-time variability (*9,10*). Figures 2 and 3 are conceptually important because they illustrate the link between the carrier's operating variables and the quality of its product.

With engineering models such as the one used for this paper, carrier cost is more naturally presented as a function of the operating variables than as a function of the service quality variables. Equation 4 may be used to specify either carrier, shipper or full cost as a function of annual volume and the operating variables. For example, with carrier cost:

$$C(v, z) = C[v, z(f, s)] = C(v, f, s) \quad (5)$$

The shipper cost and full cost functions, $S(v, f, s)$ and $T(v, f, s)$, are similarly derived; all three functions are respectively depicted in Figures 4–6 (with fixed annual volume).

The importance of shipper satisfaction is often publicly recognized by carrier representatives, and it is systematically addressed in the day-to-day operations of some carriers. However, in practice, many carriers may find this issue elusive because of the difficulty of specifying a relationship between carrier operating variables and shipper satisfaction. Shipper cost is one measure of shipper satisfaction, and Figure 4 is a graphical depiction of such a relationship. Although such a specific relationship may be impossible to delineate in the real world, a working approximation could still be useful. In this example, shipper cost strictly decreases with both operating variables; in particular, cost appears to be approximately inversely proportional to both variables. As might be expected, this cost surface is quite similar to the average transit-time surface presented in Figure 2.

In Figures 5 and 6, both carrier cost and full cost appear to be convex functions of the operating variables. The two cost-minimizing points (L) correspond respectively to the CCM and QE policies. Annual carrier cost is minimized in Figure 5, with 58 trains per year and a 28 mph cruising speed; further reduction in either operating variable will result in equipment cost increases greater than the savings in labor or

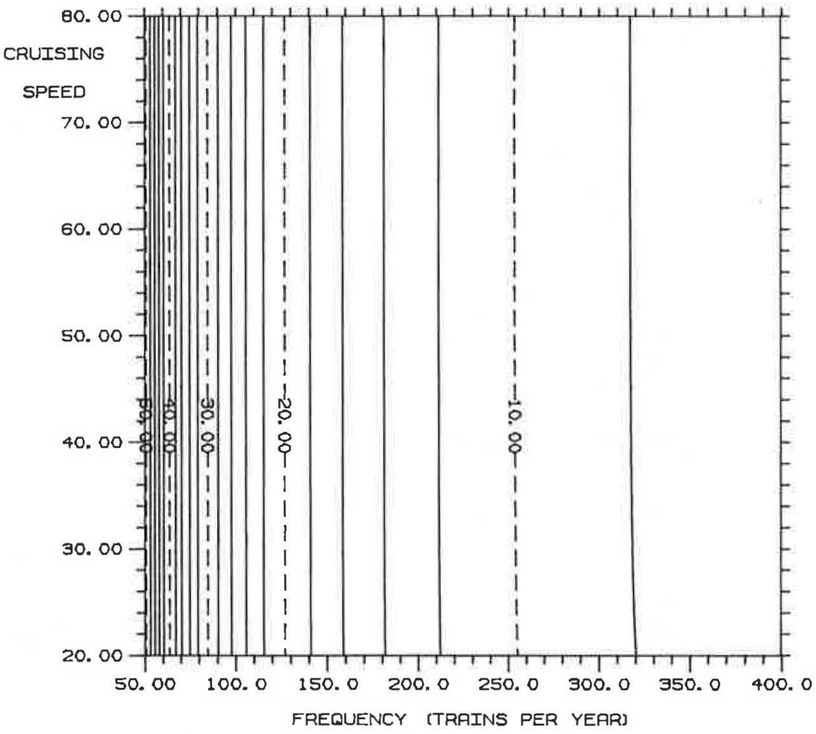


FIGURE 3 **Transit-time standard deviation (hours) as a function of operating variables.**

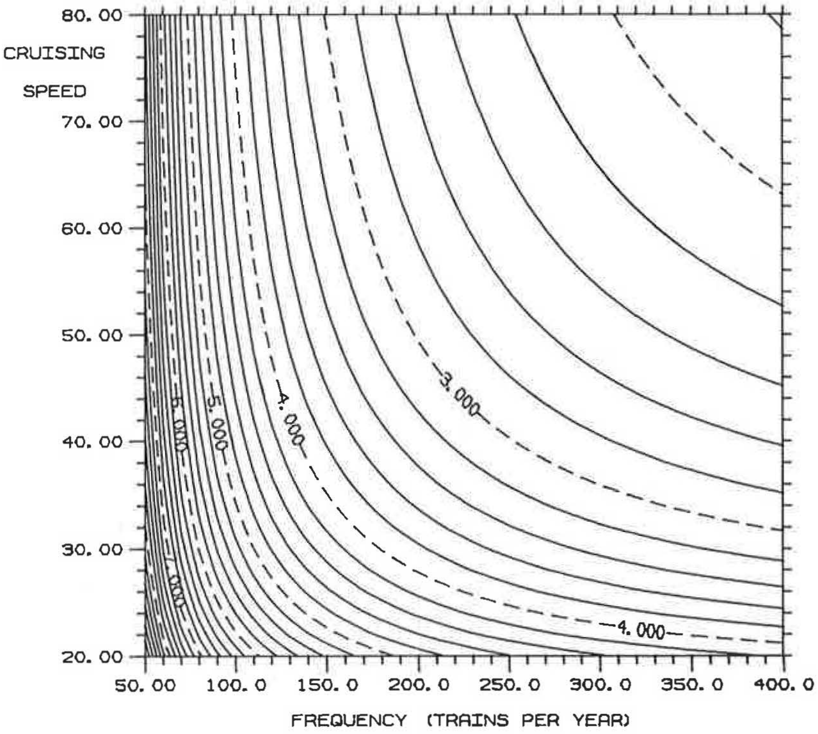


FIGURE 4 Shipper cost surface ($million/year) as a function of operating variables.

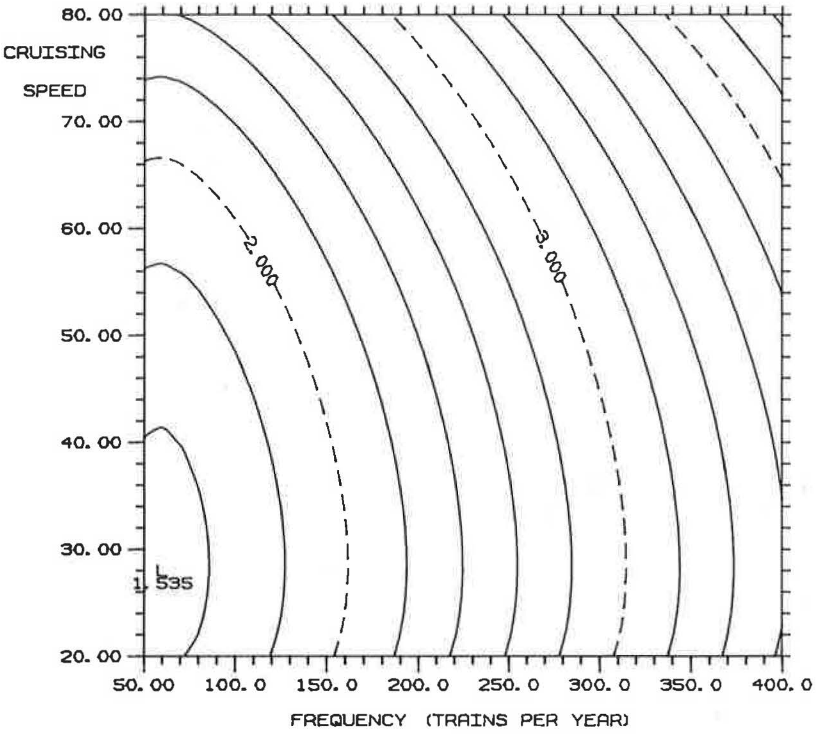


FIGURE 5 Carrier cost surface ($million/year) as a function of operating variables.

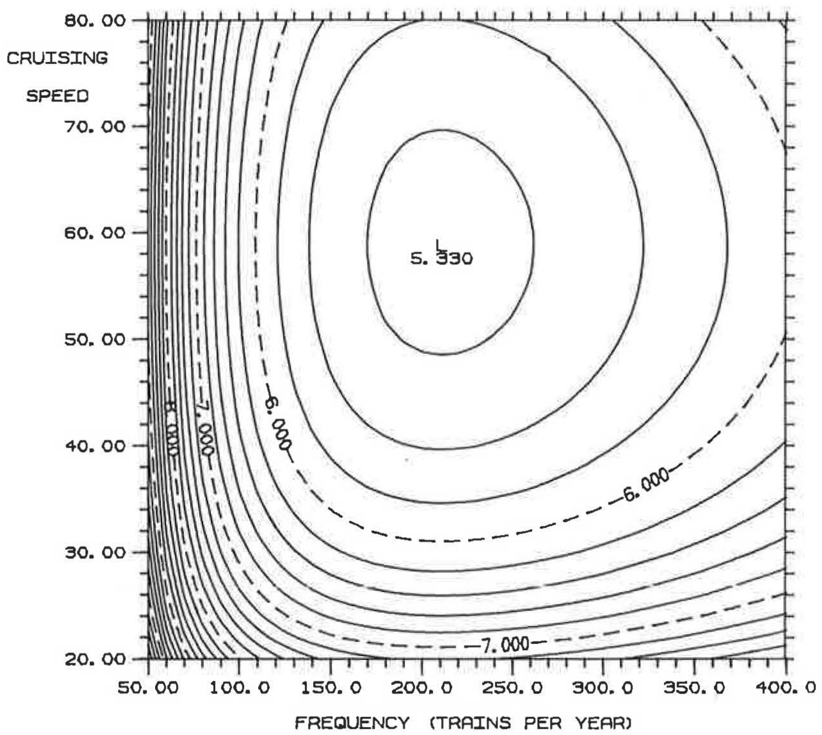


FIGURE 6 Full cost surface ($million/year) as a function of operating variables.

fuel costs. As expected, significantly larger operating variable values are required to minimize full cost (Figure 6). Both cost-minimizing points are discussed further in the following two sections.

Costs as Functions of Service Quality

In the service quality model, carrier, shipper, and full costs are all conceived as functions of service quality variables—not carrier operating variables. The conceptual distinction is that service quality is observable by both carrier and shipper, whereas operating variable values might be only known to the carrier. This distinction has practical implications; for example, contractual service quality commitments by a carrier to a shipper must be based on criteria that can be verified by both shipper and carrier.

Shipper cost is already fully specified as a function of the service package (Equation 3). Carrier cost and full cost are respectively presented as functions of the two service quality variables in Figures 7 and 8 (the bordered stairstep domain area was obtained by superimposing Figure 2 on Figure 3). Each of these cost surfaces is a transformation of one of the previously discussed relationships between cost and carrier operating variables. For example, the carrier cost surface depicted in Figure 7 is a transformation of the surface presented in Figure 5.

The functional relationship between carrier cost and service quality may be as difficult to specify as the relationship between shipper cost and carrier operating variables, and it is equally important. From the shipper's perspective, service quality variables define the carrier's product and its value. Therefore, for effective product management, the carrier must have some understanding of the relationship between its cost structure and the service quality it provides. Equation 5 indicates how such a relationship may be established by combining current knowledge of carrier cost and service quality (as functions of the operating variables). This is essentially how Figure 7 was obtained from Figures 2, 3, and 5.

The natural expectation is that carrier cost should rise with better service quality, and hence decrease with both elements of this service package. However, in Figure 7, carrier cost increases with at least one of the service quality variables for more than three-fourths of the domain area (and with both variables in the region directly northeast of the CCM point *L*). Because shipper cost strictly increases with both quality variables, it does not make sense for the carrier to offer a service level where carrier cost also increases with either variable (because it could otherwise simultaneously reduce its own costs and increase the quality of its service). In Figure 1, the service quality region to the left of Z^c is similarly irrational.

It cannot be assumed that carriers will never operate in this irrational service region. For example, the railroad industry has sought to decrease the time cars spend in intermediate switching yards, with the twin goals of improving overall service quality and reducing carrier operating costs through higher rolling stock utilization (*9*). If both goals are simultaneously accomplished, the carriers must have previously provided this type of irrational service quality.

Full cost as a function of the service quality variables is presented in Figure 8. The cost minimizing point *L* indicates

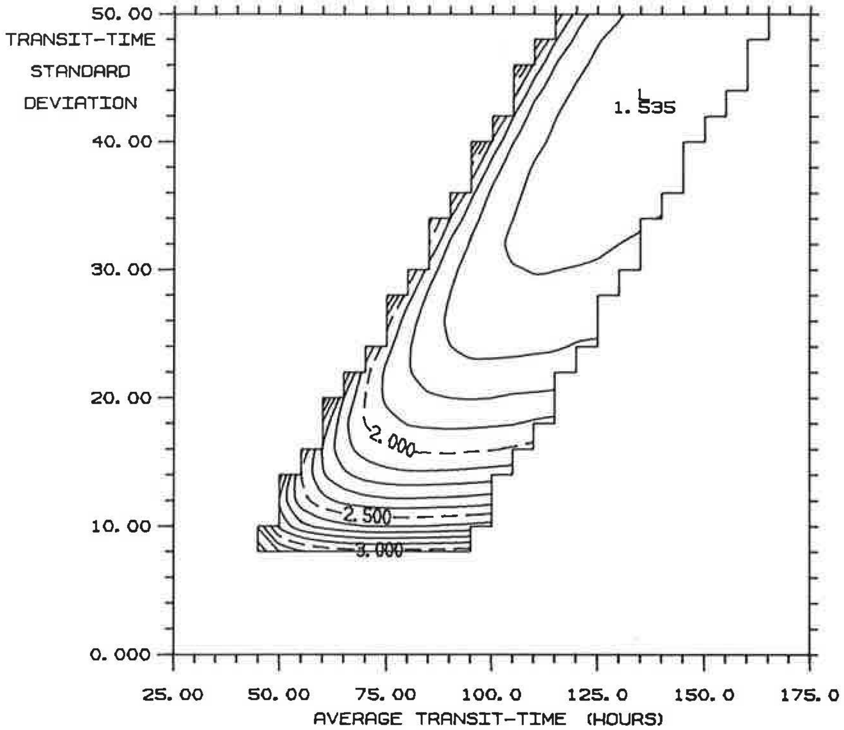


FIGURE 7 Carrier cost surface ($million/year) as a function of service quality variables.

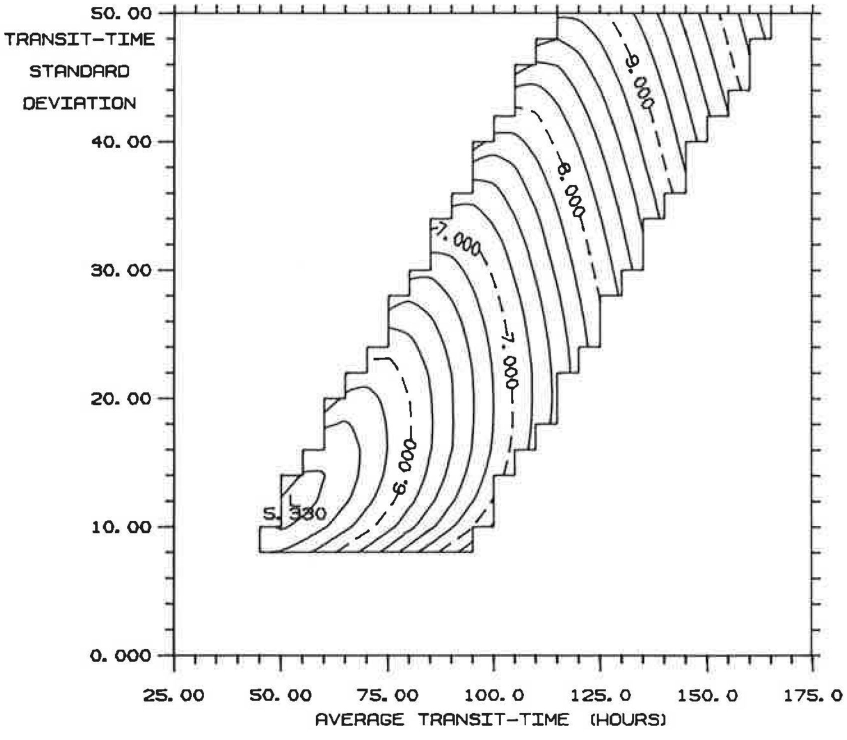


FIGURE 8 Full cost surface ($million/year) as a function of service quality variables.

quality efficiency. As expected, QE offers significantly better service quality than the CCM service package. Further improvements in service quality from this point would result in carrier cost increases greater than the shipper cost savings.

Quality Policy Comparison with Fixed Annual Volume

The full-cost minimizing point in Figures 6 and 8 and the carrier-cost minimizing point in Figures 5 and 7 suggest a comparison of the QE and CCM policies. The pertinent information is summarized in Table 2. The difference in profit between any two quality policies is equal to the full-cost difference (*3*), and as previously discussed, QE is both socially optimal and carrier profit maximizing. In this case, by implementing CCM instead of QE, the carrier would sacrifice 3.683 million dollars in annual profit.

As discussed at the beginning of this paper, implementing QE is not easy and CCM may be tempting; after all, CCM minimizes all carrier costs. In this example, the carrier can reduce the specified operating costs almost in half by running trains once instead of four times a week and cutting the cruising speed in half. However, in order to keep the same traffic level, the carrier would have to compensate the shippers for the lower service quality with a rate cut equal in value to the \$4.769 million increase in shipper costs. When the carrier cost savings of \$1.077 million are combined with this lost revenue, carrier profit is diminished by the specified increase in full cost. In this way; profit maximization compels the railroad to consider the most efficient joint use of carrier resources and relevant shipper resources.

A carrier implements a given quality policy via operating variables. In the real world, it may be difficult to precisely determine and enforce the QE operating variable values. This raises the question of potential full cost (and hence profit) sensitivity to errors in these variables. From the *L* in Figure 6, full cost increases faster for lower operating variable values than it does for higher values. Thus, for each of the two operating variables in this example, it would be better to err on the higher side than by the same amount on the lower (CCM errs on the low side). This effect is caused by the shape of the shipper cost surface (Figure 4), which may be typical for a large class of shippers (*4*).

TABLE 2 QUALITY POLICY COMPARISON

	Quality Policies		Difference	
	QE	CCM	Value	Percent
ANNUAL COST IN MILLIONS				
Carrier Cost	\$2.612	\$1.535	-\$1.077	-41%
Shipper Cost	2.718	7.478	4.760	175%
Full Cost	\$5.330	\$9.013	\$3.683	69%
CARRIER OPERATING VARIABLES				
Frequency (Trains Per Year)	210.84	57.90	-152.94	-73%
Cruising Speed (mph)	58.77	28.36	-30.41	-52%
TRAIN CHARACTERISTICS				
Cars	33.88	123.37	89.49	264%
Locomotives	1.75	1.48	-0.27	-15%
Horsepower Per Revenue-Ton	2.217	0.515	-1.702	-77%
SERVICE QUALITY VARIABLES				
Average Transit-Time	53.22	135.77	82.55	155%
Standard Deviation	12.01	43.69	31.68	264%
Coefficient of Variation	0.226	0.322	0.096	43%

Effects of Annual Volume

Transportation analysis, and economic analysis in general, is often concerned with the effects of annual volume variability. Discussed in this section are the effects of annual volume on costs, service quality, and operating variables. Quality policies are used to reduce the complex interaction among these parameters, and thus facilitate an indirect examination. In this section, the preceding comparison of the QE and CCM quality policies is extended, followed by a discussion of the use of full cost in transportation economic analysis, focusing on returns-to-scale.

Further Quality Policy Comparison

By definition, a quality policy specifies each service quality variable as a function of volume. Furthermore, any quality policy may be used to define new cost functions whose only independent variable is volume. For example, with the QE policy and the carrier cost function:

$$C^*(v) = C[v, Z^*(v)]$$

The other QE cost functions, $S^*(v)$ and $T^*(v)$, and CCM cost functions $C^c(v)$, $S^c(v)$ and $T^c(v)$ are similarly defined. Quality policies also allow carrier operating variables to be presented as functions of volume.

Both the QE and CCM policies were analyzed and compared with respect to annual volume for eight trip distances from 50 to 3,000 mi. For each trip distance, 31 annual volume values were examined, spread approximately logarithmically from 10,000 to 10 million revenue tons. Some of the results for a 1,500-mi trip are shown in Figures 9–13. The 500,000 net-ton tic in the middle of each graph indicates the fixed volume situation addressed previously. The graphs developed for the other seven trip distances are generally similar to these figures, and support the observations presented here. Some of the results for the two smallest distances (50 and 100 mi) were relatively extreme, but these trip distances are outside the normal railroad service market.

The cost functions of the two quality policies are compared in Figure 9. These percent values correspond to those presented in Table 2. For all three costs, the absolute percentage impact decreases as volume increases. These curves are virtually identical for each of the six larger trip distances. For the two smaller distances, the percentage impact becomes smaller significantly more quickly with increasing volume.

The impact on full cost (and profit) associated with implementing CCM instead of QE was discussed previously with respect to Table 2. Figure 9 augments that discussion by allowing one to extend the conclusions over a broad volume range. Over this range, the carrier is able to significantly reduce its own costs by implementing CCM instead of QE. However, the result is much higher shipper and full costs, and hence lower carrier profits. The full cost curve in Figure 9 indicates that the relative impact declines with volume. However, the decline is gradual and the relative impact is still significant (49 percent) at the highest volume level considered. Furthermore, the absolute dollar impact significantly increases with annual volume.

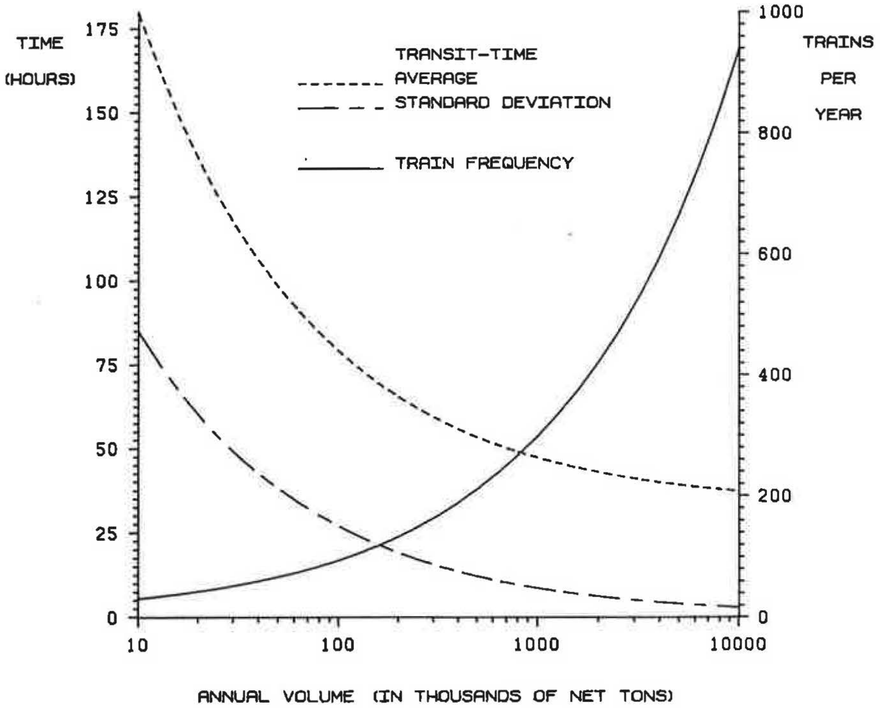


FIGURE 10 Train frequency and service quality with QE policy.

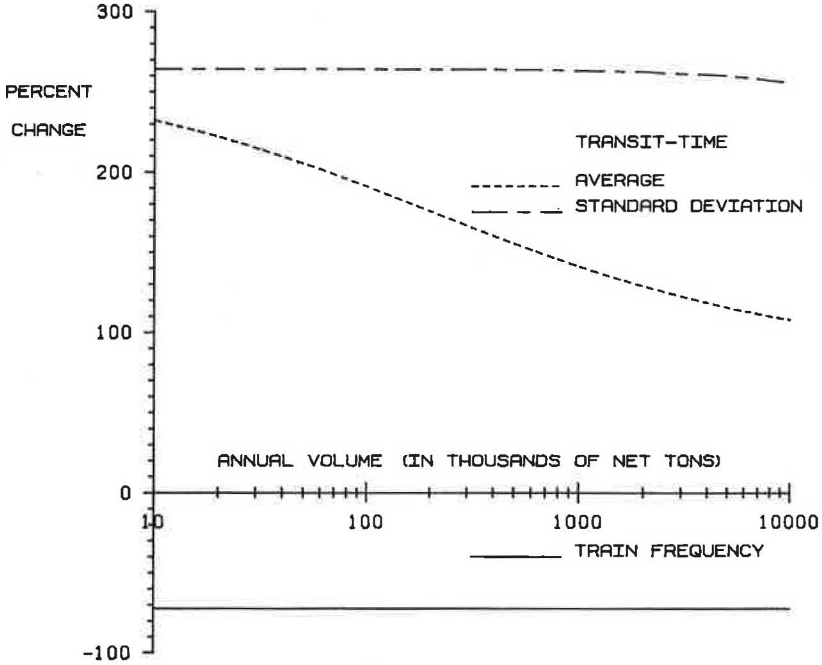


FIGURE 11 Train frequency and service quality percent differences of CCM with respect to QE.

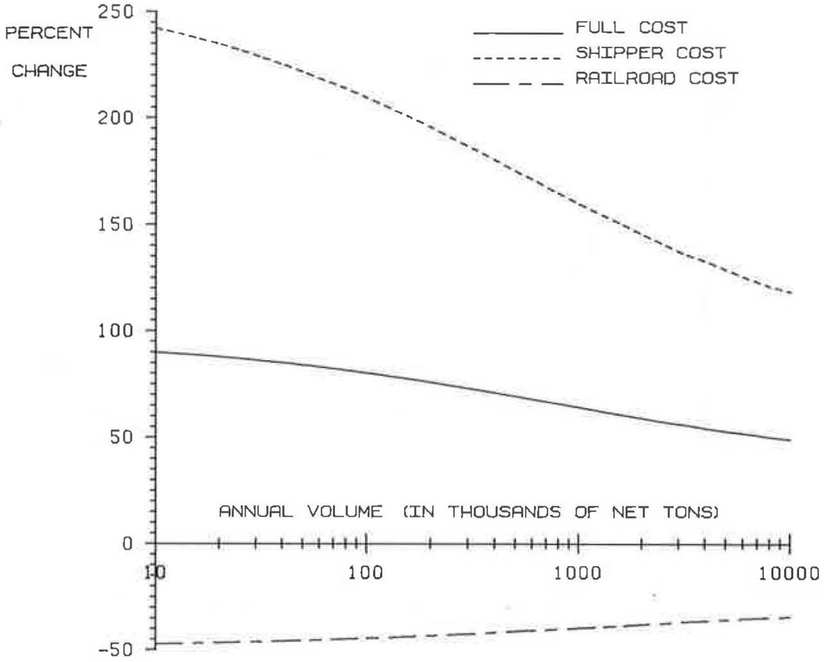


FIGURE 9 **Cost percent differences of CCM with respect to QE.**

QE train frequency and service quality variables are presented in Figure 10. Cruising speed is not presented because it is substantially unaffected by volume; for all trip distances and volumes considered, the QE cruising speeds are within a 2-mph range, and the CCM cruising speed range is less than 1 mph (typical cruising speed values are presented in Table 2). The shape of the three curves in Figure 10 is almost exactly the same for all eight trip distances and both quality policies, but with different vertical scales.

The relative differences between QE and CCM, in train frequency and service quality variables, are presented in Figure 11. These percent values also correspond to those presented in Table 2. The train frequency percent difference is within 0.5 percent of −72.25 percent, and the cruising speed percent difference is approximately −52 percent for all volumes and trip distances considered. These relatively constant values were unexpected and merit further investigation. The range of the average transit-time curve is larger for smaller trip distances, and it is smaller for larger distances. The range of the standard deviation curve in Figure 11 is only 9 percentage points; it is significantly larger only for the two smallest distances. This relative flatness might be explained by the strong relationship between the standard deviation and train frequency (see Figure 3). From Figures 10 and 11, it is clear that both service quality variables had a significant impact throughout the volume range on the shipper cost increase depicted in Figure 9.

Transportation Cost Analysis—Returns-to-Scale

In conventional microeconomic models, the cost function of the firm is responsible for specifying the efficiency of the production process. However, in the production of freight service, full cost, not carrier cost, captures the essence of carrier efficiency in creating value (independent of market demand considerations). Therefore, at least theoretically, transportation economic analysis should be based on the full cost (*3*,*4*).

The examination of average and marginal cost geometry with respect to variations in quantity is central to traditional microeconomic cost analysis. The interaction between these two cost curves may be summarized by the degree of scale economies, defined here as the ratio of average to marginal cost. This measure may be applied to the full, carrier, and shipper cost functions (*3*), as presented in Figures 12 and 13 for QE and CCM, respectively. Returns-to-scale are said to be increasing, constant, or decreasing as the degree of scale economies is greater than, equal to, or less than unity.

With the service quality cost model, economies-of-scale should be based on the full cost of transportation, instead of the traditional carrier cost measure. For example, with the numeric example presented by Brown (*3*), the full cost returns-to-scale are constant for the perfectly competitive carrier, whereas the carrier-cost based measure indicates increasing returns (a perfectly competitive firm has constant returns-to-scale). Figures 12 and 13 indicate the difference between the full cost and railroad cost measures, and hence the potential error of using carrier cost returns instead of full cost returns.

With both the numeric and railroad examples, the magnitude of this error is greater if CCM is implemented instead of QE. It is ironic that when the carrier ignores the effects of service quality and the importance of full cost (by implementing CCM instead of QE), it becomes even more important for the outside economist or cost analyst to consider these effects (by using full cost returns instead of carrier cost returns). An explanation for this phenomenon may be based on the fact that full cost degree of economies is equal to an

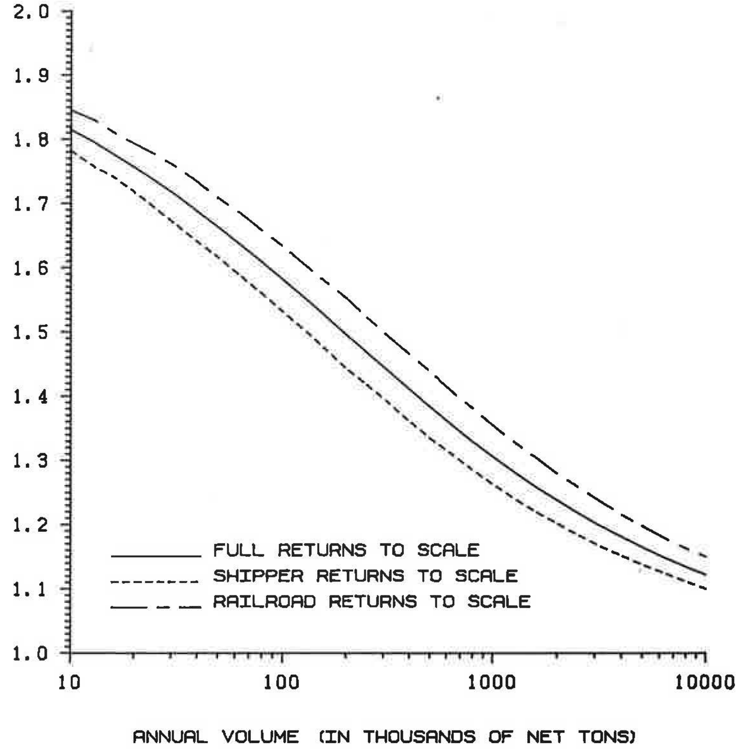


FIGURE 12 Degree of scale economies with the QE policy.

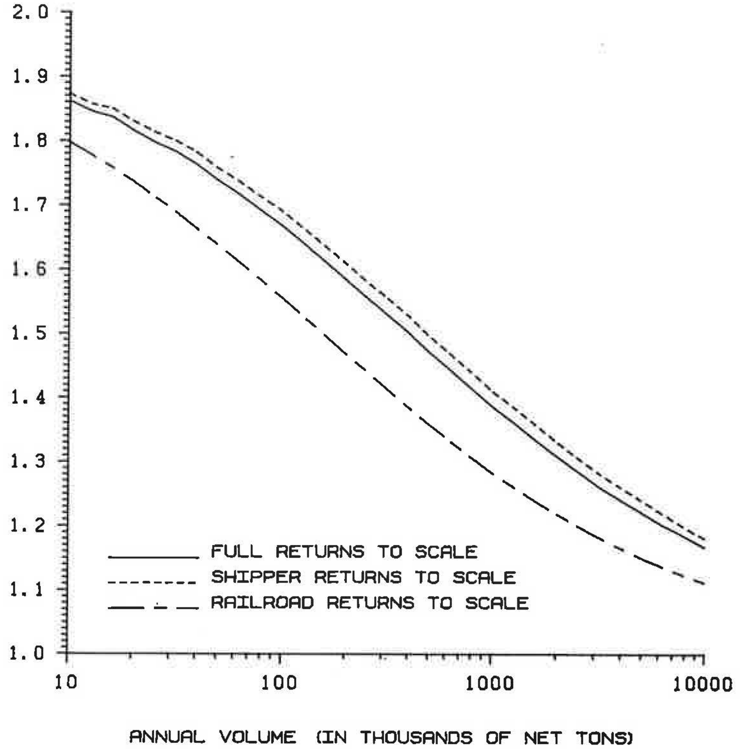


FIGURE 13 Degree of scale economies with the CCM policy.

average of the carrier cost and shipper cost economies, weighted by their respective marginal cost functions (*3*). The CCM carrier marginal cost is generally smaller than the QE carrier marginal cost, whereas the CCM shipper marginal cost is greater than the QE shipper marginal cost. Therefore, the full cost economies curve will generally lie between the carrier and shipper economies, and closer to the carrier cost curve with QE (Figure 12) than with CCM (Figure 13).

In trying to address the difference between the true (full cost) economies and the value obtained from carrier costs alone, the analyst cannot make an a priori assumption about the sign of this error. For instance, in this railroad example, the error leads to an overestimation of returns-to-scale if QE is implemented, and an underestimation if CCM is implemented; in the numeric example, returns-to-scale are overestimated with both quality policies. The sign is determined by the relative magnitudes of the shipper and carrier returns within the previously discussed weighted average formulation of full cost returns.

It should not be concluded from Figures 12 and 13 that railroading is characterized by increasing returns-to-scale. As previously discussed, this analysis has been limited to a few above rail cost elements directly related to service quality. These elements do not include any capacity constraint and, given the nature of the operating variables, inherently imply a declining marginal cost curve. If other cost elements, such as maintenance-of-way, are properly included, then average costs may increase at some point (*7*).

CONCLUSION

This examination of freight service quality cost economics, with the railroad example, fosters a practical understanding of the service quality model. In particular, it demonstrates the importance of quality efficiency and the full cost of transportation. Implications for the carrier and transportation analyst are emphasized, and the negative consequences of disregarding these implications are illustrated with the railroad example. For the carrier, this is reflected in lost profits; for the analyst, it means potentially erroneous conclusions. The examination also augments our general understanding of railroad service quality economics.

This paper, and other work by the author (*3,4*), indicate the need for further inquiry in three areas. The first area concerns the theoretical structure of the freight service quality model. This would include more rigorously establishing the underlying aspects of the model in terms of welfare theory and extending the model to explicitly consider multiple shippers and carriers, with separate service packages specified for individual carrier-shipper pairs.

The second area includes efforts by individual carriers to understand their full cost structure and then implement quality efficiency. This may be a challenging undertaking, especially for railroads, for which the same trains and fixed facilities are used for shipments with very different sensitivities to service quality. Tools such as the MIT Service Planning Model (*10*) may help rail carriers in this endeavor, and shipper cost functions might be based on commodity characteristics or shipper surveys or both.

The third area involves further use of detailed operating models to explore the service quality model and its implications. This last area would provide a bridge between the theoretical efforts and practical applications. For example, a more detailed model of branchline and mainline railroad operations could be developed in conjunction with a model of truck-load trucking to examine the competitive and value creation aspects of service quality and related public policy issues. Whereas this paper uses a simple model of freight operations to facilitate an understanding of the service quality model, these future research efforts would focus on using the service quality modeling framework to further the understanding of freight operations.

ACKNOWLEDGMENT

The author would like to thank dissertation advisor David Boyce for his help and guidance; Caroline Fisk, who provided an introduction to the modeling principals underlying the hypothetical railroad example; the anonymous referees and others who provided useful comments and constructive evaluation; and the author's family and friends for their support.

REFERENCES

1. M. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the Economics of Transportation.* Yale University Press, New Haven, Conn., 1956.
2. S. D. Nason and B. C. Kullman. The Economic Effects of Freight Rate Changes: An Analysis of Total Logistics Costs. *Proc., 19th Annual Meeting, Transportation Research Forum,* 1978, pp. 499–503.
3. D.G. Brown. Freight Service Quality and Carrier Economics. *Journal of the Transportation Research Forum,* Vol. 30, No. 1, 1989, pp. 141–152. Errata: No. 2, pp. 558–561.
4. D. G. Brown. Optimal Freight Service Quality and Implications for Economic Analysis, Inventory-Theoretic Shipper Cost Functions and Run-Through Train Operations. Ph.D. dissertation. University of Illinois at Urbana-Champaign, 1988.
5. D. D. Wyckoff. *Railroad Management.* Lexington Books, D. C. Heath and Co., Lexington, Mass., 1976.
6. E. C. Poole. *Costs—A Tool for Railroad Management.* Simmons-Boardman, Omaha, Nebr., 1962.
7. M. B. Hargrove. Track Maintenance Cost Analysis: An Engineering Economics Approach. In *Transportation Research Record 1029,* TRB, National Research Council, Washington, D.C., 1985, pp. 64–70.
8. S. S. Rao. *Optimization: Theory and Applications.* John Wiley and Sons, New York, N.Y., 1979.
9. A. S. Lang and C. D. Martland. *Reliability in Railroad Operations, Studies in Railroad Operations and Economics,* Vol. 8, MIT Center for Transportation Studies, Cambridge Mass., 1972.
10. J. R. McCarren and C. D. Martland. *The MIT Service Planning Model, Studies in Railroad Operations and Economics,* Vol. 31, MIT Center for Transportation Studies, Cambridge, Mass., 1980.

The views of the author do not necessarily reflect the position of the Department of the Navy or the Department of Defense.

Publication of this paper sponsored by Committee on Freight Transportation Planning and Marketing.

Applying Statistical Process Control Methods in Railroad Freight Classification Yards

RAY A. MUNDY, RANDY HEIDE, AND CHARLES TUBMAN

Quality experts and rail customers have long admonished rail management for the need to improve service reliability and consistency. Investigation has discovered that most of the variance in rail transit times, wrongful charges, and so forth stem from origin, intermediate, and destination yard procedures and information processing. Although there are known problems, little constructive assistance is typically offered in the literature. The use of statistical process control (SPC) quality tools to address this common railroad problem is discussed. A brief explanation of SPC is followed by an examination of a typical rail freight classification yard and discussion of how these tools can be used to identify and prioritize problems. Special emphasis is placed on the need to bring these yard operations under control, thereby eliminating special causes of variation. With yard operations under control and predictable, rail operations can work on common causes to improve service delivery. Management can then redesign procedures to structurally improve the systems process. Both approaches are necessary to attract quality-conscious shippers. Procedures include the use of flowcharts, control charts, and Pareto analysis. Implications for management are also discussed.

Accompanying the widespread renewal of interest in quality management by North American businesses has been increased interest in ways to apply these procedures to service industries as well as more traditional manufacturing applications.

The idea of measurement of service levels in the railroad industry is not a new one. Landow and Wharton advocated regular measurement of individual car movements as a way to effectively compete with trucks (*1*). Landow, in particular, advocated the adaptation of a service reliability index to measure service performance and to allow adjustments for tariffs to reflect the costs of inventory for the customer in light of poor reliability.

While not specifically advocating statistical process control (SPC) charting techniques, a 1974 Harvard Business School case (*2*) listed the primary causes of delay to individual freight car movements, as well as the problems encountered in the operation of a classification yard. This study listed a number of potential solutions, such as more frequent and shorter train movements, but did not seek to improve the process of freight car classification itself.

Mundy et al. (*3*) noted the importance of service industries adopting SPC techniques, including control charts, Pareto analyses, and histograms. They cited the example of a limousine company for which Pareto analysis was used to pinpoint reasons for service failures, including lateness, reservation errors, client failures, and uncontrollable factors.

Deming (*4*) suggested a number of potential applications for SPC techniques in the railroad industry. These applications include monitoring of transit times for freight car movements in specific corridors, reducing errors in interline settlements and local billing, decreasing idle time of freight cars, studying specific delays in transit times, studying time spent repairing freight cars by type of repair, elapsed time between a customer's call for an empty or the pickup of a load, probability sampling to determine sections of roadbed to be examined, and determining future needs for parts and general maintenance.

Although some companies, including Ford, Tennessee Eastman, and Holley Carburetor, use SPC to monitor their rail carrier performance, only recently have rail carriers begun to use SPC techniques to monitor their own freight car movements for continuous improvement.

The traditional approach toward measuring service levels in the railroad industry has been to monitor on-time train performance. Although this approach has relevance to terminal trainmasters and district superintendents, it has little, if any, meaning to the individual shippers. Unless they are using dedicated unit trains, shippers are concerned only with whether shipments are on their sidings or team tracks when needed.

The first comprehensive study to examine the issue of reliability (as compared with on-time performance) was conducted at the Massachusetts Institute of Technology in 1972 (*5*) under a contract from FRA. Researchers studied service levels on several different railroads and analyzed in-depth the operations on the Southern Railway. They concluded that the greatest barrier to reliable operation was in the classification yards. The authors not only suggested making reliability a chief corporate goal, but also listed a series of intermediate steps that could be undertaken to bring it about. These steps included the following.

1. Provide sufficient motive power at terminals to avoid yard and road delays.
2. Increase the number of run-through trains to avoid blocking at intermediate yards.

R. A. Mundy, University of Tennessee, College of Business Administration, Department of Marketing, Logistics and Transportation, 320 Stokely Management Center, Knoxville, Tenn. 37996. R. Heide, Santa Fe Railway, Intermodal Division, 1700 E. Golf Road, Schaumburg, Ill., 60173-5860. C. Tubman, Norfolk Southern Railroad, Commodities Marketing, 3 Commercial Place, Norfolk, Va. 23510.

3. Adjust train schedules to reflect actual, not ideal, performance.
4. Allow sufficient time at yards for cars to make important connections without excessively disrupting yard operations or delaying outbound trains.
5. Limit the number of train and block cancellations to emergencies requiring management approval.

Although a number of trucking companies have successfully marketed their ability to facilitate just-in-time deliveries, it has only been in the past several years that railroad companies have shown an interest in going after this type of traffic. Some railroad companies are now developing corresponding SPC techniques (primarily X-bar and R charts) to monitor and guide their efforts for specialized service, such as intermodal and some automotive trains.

Based on interviews with shippers (*6*), those who (still) use rail services are as much, if not more, concerned with consistency as they are with origin to destination delivery times. Thus, if railroads are to compete for traffic they must continually improve reliability and consistency while lowering unit costs. Just as SPC techniques have permitted manufacturing organizations to do this, so may they for transportation service industries including railroad operations. Although the use of SPCs will not directly bring about improved transit times and consistency, it will allow carriers to pinpoint areas for improvement and give clear direction to the improvement efforts.

INTRODUCTION TO SPC AND ITS IMPLICATIONS

Before the discussion of the application of SPC techniques at railroad classification yards, it should be clarified what SPC is and what it is not. Basically, SPC is a management tool. It is not a miracle cure or a panacea.

SPC is generally taught as one part of a larger overall management philosophy, the same philosophy that spawned such concepts as just-in-time, quality circles, participative management, and continuous improvement. It can be adopted alone, by an individual manager or to solve individual problems, but without a change in overall corporate management style its potential and effectiveness will be extremely limited. This is because SPC does not solve problems. Instead it provides clues as to what causes problems and requires a management team that is committed to the philosophy of continuous, gradual improvement.

To make constructive use of SPC data, one must first understand the basic statistical principles behind it.

Measurement

To use statistics, one must have both something to measure and a means of measurement. The something should be thought of as the output of a process. In manufacturing, units coming off a production line are measured for conformance to specifications or standards. In a service industry, one may not have a physical good to measure, but one will have a service produced by some sort of process, with some set of defined specifications for that service.

In manufacturing, measurement of output is usually in terms of physical description, such as size, weight, thickness, or color. These types of measurements produce a continuous distribution. When measuring the production of a service (e.g., the delivery of a package from point A to point B), one may use either continuous or binary (yes or no) distributions. The time elapsed between pickup and delivery would be a continuous variable, whereas the occurrence of a loss or damage would be binary. Often, it is most appropriate, or at least easiest, to test a service product against a preset standard of success, resulting in a binary success or fail variable.

Variation

In any statistical population sampled, variation will exist. Expressed more simply, "No two things are alike. They will always vary," (*7*, p. 1). Deming describes two types of causes of variation: common causes and special causes.

Common causes can be thought of in terms of variation due to chance. These are the causes of variation in the results of an experiment performed over and over under identical conditions, such as rolling a die. They arise out of the process or out of the way the process is organized and operated.

Special causes are sources of variation that do not belong to the system. Often they will be specific to a certain operator, machine, or batch of material. In the die-rolling experiment, by occasionally substituting a chipped die for the regular one, a special cause of variation has been introduced.

Note that whereas a continuous variable measurement, such as elapsed time, can capture the subtle difference from one product to another, a binary measurement cannot. If one records a failure, one does not have a record of the amount of variation present. This is why continuous measurement is generally preferred whenever possible.

Four Areas of Statistics

The four areas of statistics are descriptive statistics, probability theory, statistical inference, and SPC (*7*, pp. 22–24).

Descriptive statistics involve the summarizing of information contained in a data set. This includes basic measurements of a population, such as the mean, median, and standard deviation.

Probability theory is the mathematical modeling of random phenomena. Generally, probability theory allows one to describe the future outcomes of a system that is completely known, even though individual outcomes are random.

Statistical inference involves an attempt to infer the properties of an unknown population based on a randomly drawn sample. Basically, the goal is to determine whether the sample drawn is representative of the total population.

SPC involves an attempt to determine whether a series of data sets came from the same population, or resulted from the same process. If it did, statistical inference can be used to draw conclusions about the underlying population, such as a prediction of variance in future output. If not, then the questions of inference are moot.

The application of SPC in industry, therefore, allows determination of whether the output, or thing being measured, is the result of a single, identical process. Given that there will be variation in the output, SPC allows determination of whether variation results entirely from common causes (factors within the system), or from special causes (factors outside the system). If variation is shown to be entirely from common causes, the process is said to be in control.

Many managers, after applying SPC and learning that their process is in control (a bit of often misunderstood SPC jargon), smile, pat themselves on the back for doing a good job, and go no further. They fail to understand that SPC does not pass judgment on their process. It simply shows how much variation to expect in output, given that the process is left to operate the same way each day, with no unusual or outside factors influencing it. It is up to the individual or the firm (and ultimately, the customer or end user of the output) to determine whether the level of variation inherent in the process is acceptable. The value of having a process in control, that is, with all variation in output being due to the system or process itself, is that it allows the manager to detect the causes of variation within the process and eliminate them. By watching the control charts, the manager can determine whether changes made in the system, such as a new track layout, more frequent locomotive maintenance, or a change in switching schedules, have truly changed the system's capabilities and whether the change is for the better. Likewise, if a process that has been operating in control with no changes enacted begins to send out-of-control signals, the manager receives early warning that a special cause has crept in and changed the process.

Philosophy Behind SPC

Deming's much-talked-about management philosophy, which revolutionized Japan's post–World War II economy, is based on the idea that variation, while inevitable, is the root of all evil. Deming argued that reducing variation leads to lower cost and increased productivity, regardless of whether existing variation conforms to a set of expectations or standards. Furthermore, he contended that the notion that quality and cost are incompatible and represent a trade-off is completely wrong. By reducing variation, quality and productivity go up, and cost goes down (*4*, p. 3).

Another significant argument of Deming's is that only a fraction of the total variation in output can be corrected by workers doing their best. By working harder and studying their own work for causes of mistakes, workers can only remove about 15 percent of the variation within the system.

The remaining 85 percent of the variation is caused by the system factors put in place and controlled directly by management, such as equipment or standard operating procedures. (Recently Deming suggested that this ratio may actually be 5 percent labor to 95 percent management.)

It is the observation of this principle in practice that has led early SPC practitioners such as Deming, Juran, and others to a new management philosophy. This philosophy is essentially the same as that of the Japanese, whose approach actually springs from their application of SPC methods on a widespread basis in the early 1950s.

Vaughn Beals, chairman and chief executive officer of Harley-Davidson, studied his Japanese competitors thoroughly and concluded, "It is not robotics or automation that gives them their competitive edge. It is not substandard wages. It is not culture. And, it is not the morning calisthenics and company songs. What it is, is management—no more, no less," (*8*, p. 9).

Therefore, although SPC is only a management tool, it is a tool whose effectiveness is greatly related to the management philosophy of the user. Companies that have attempted to adopt it in isolation, just as with those that have tried to utilize just-in-time or quality circles without adapting the organization to fit the new methods, have had limited success. These implications regarding management philosophy and corporate culture for railroads are beyond the scope of this paper, but should be taken into consideration.

DESCRIPTION OF RESEARCH STUDY AND OPERATIONS AT TENNESSEE YARD

The Burlington Northern's (BN's) Tennessee Yard facility in Memphis (built for the St. Louis–San Francisco Railway Company, which was merged into BN in 1980) is the newest and largest railroad classification yard in the Memphis area. It was built in 1959, and is the only hump or gravity yard in the Memphis area.

The Tennessee Yard was designed to classify up to 2,400 cars on 61 tracks. Ten of these tracks have since been removed to make way for a trailer-on-flat-car facility, and current managers believe the yard could work efficiently handling up to 2,000 cars daily. During the initial phase of the University of Tennessee study from August to the first part of December 1987, the yard handled from 7,778 to 9,730 cars weekly and 596 to 1,817 daily. The average number of cars handled daily is between 1,260 and 1,300, with 5:00 a.m. being the busiest and 2:00 p.m. being the least busy. With the exception of Expediter trains (BN's scheduled, time-sensitive intermodal trains) and through unit coal and grain trains, all trains arriving at Tennessee Yard are broken up and reclassified. Sixteen arrivals and departures daily were scheduled.

The present set of standards calls for through trains to be reclassified and on their way within 8 hr. Cars bound to or from local industries within the Memphis Terminal area are to be placed locally or placed on a through train within 24 hr of their pickup. Cars being sent to connecting railroads are to be moved within 14 hr.

DATA COLLECTED

Managers at Tennessee Yard used a daily terminal performance report, (the TPC report), generated by BN's central information service. This report contained a variety of detailed information on car movements through the terminal. Three sections of the report were used to collect data for SPC analysis.

Daily Cars Over Standard Report

This section of the TPC report provided detailed information on each car within specific movement groups that left the terminal in more than standard time during the day. An example of this report is shown in Table 1.

TABLE 1 DAILY CARS OVER STANDARD REPORT (C-1 SUNDAY, SEPTEMBER 20, 1987)

Exhibit A

DAILY CARS OVER STANDARD REPORT C-1 SUNDAY SEPTEMBER 20, 1987

CAR		L	CR	CON-	DEST.		ARRIVED-RELEASED			STANDARD GOAL		DEPARTURE-PLACED			HRS
UNIT & NRR		E	KD	TENTS	JCT	TAG	MODYHR	TRB/RD/ETC		MODYHR	TRN/RD/ETC	MODYHR	TRN/RD/ETC		OVR
TRAIN ARR/MH - THRU CRS/120															
HTTX	98887	L	FE	MACHRY	BIMICS	328	092005	01120	15	092010	120	092012	95636	20	2
HTTX	98447	L	FE	MACHRY	BIMICS	328	092005	01120	15	092010	120	092012	95636	20	2
									13 TOTAL CARS. 15% OVER STANDARD. $2						
TRAIN ARR/MH - THRU CRS/101															
TGAX	131425	L	TS C	ACID	20712	536	091823	01181	18	091905	181	091913	01793	19	8
TGAX	131362	L	TS C	ACID	20712	536	091823	01181	18	091905	181	091913	01793	19	8
									26 TOTAL CARS. 7% OVER STANDARD. $9						
INDUSTRY/MH - NORTHTOWN/MH															
BNFE	19306	E	R7	836001	00673	966	091814	RLSE	132801	091910	143	092007	01143	20	20
OTTX	90442	E	FE	830	01600	966	091814*	RLSE	120210	091910	143	092007	01143	20	20
OTTX	90394	E	FE	840	01600	966	091815*	RLSE	120210	091910	143	092007	01143	20	20
BN	630485	E	FE	830100	01600	966	091815*	RLSE	120210	091910	143	092007	01143	20	20
									4 TOTAL CARS. 100% OVER STANDARD. $49						
TRAIN ARR/MH - GALESBURG/MH															
MILW	4416	E	A5	8488222	CHIBR	716	091901	01792TU	18	091910	143	091918	01247	19	7
									16 TOTAL CARS. 6% OVER STANDARD. $4						
INDUSTRY/MH - GALESBURG/MH															
USLX	13135	E	R5	838	CHICR	716	091814	RLSE	063404	091910	143	092007	01143	20	20
CNW	172216	E	C6	856	RISCI	716	091816	RLSE	142020	091910	143	092007	01143	20	20
GTW	598346	E	XF	838222	CHIGT	716	091816	RLSE	131599	091910	143	092007	01143	20	20
									7 TOTAL CARS. 42% OVER STANDARD. $37						
TRAIN ARR/MH - NE SHORTS/MH															
MP	650327	L	GS	STLBAR	98237	220	091809	01120	13	091910	143	091918	01247	19	7
BN	457292	E	C6	856001	98237	220	091819	95635	18	091910	143	091918	01247	19	7
BN	456422	E	C6	856001	98237	220	091819	95635	18	091910	143	091918	01247	19	7
ATSF	350004	E	C5	854	98131	210	091823	01181	18	091910	143	091918	01247	19	7
BN	448847	E	C6	856001	98045	200	091823	01181	18	091910	143	091918	01247	19	7
BN	455837	E	C6	856001	98045	200	091823	01181	18	091910	143	091918	01247	19	7
ATSF	350133	E	C5	853	98131	210	091823	01181	18	091910	143	091918	01247	19	7
BN	454841	E	C6	856001	98237	220	091823	01181	18	091910	143	091918	01247	19	7
BN	390158	E	A6	848112	98237	220	091823	01181	18	091910	143	091918	01247	19	7
BN	455785	E	C4	856001	98237	220	091823	01181	18	091910	143	091918	01247	19	7
BN	437230	E	C5	854007	98037	200	091823	01181	18	091910	143	091918	01247	19	7
BN	419259	E	C5	853007	98037	200	091823	01181	18	091910	143	091918	01247	19	7
BN	418578	E	C6	856001	98045	200	091823	01181	18	091910	143	091918	01247	19	7
BN	450828	E	C6	856001	98045	200	091823	01181	18	091910	143	091918	01247	19	7

The first two cars listed, HTTX 93887 and 93447, were part of a movement group of 13 through cars on Train 120. These particular cars were loaded with machinery bound for Birmingham. They arrived at Tennessee Yard on September 20 at 0500 hr and, according to standards, were to be released the same day on the same train at 1000 hr. They were actually released at 1200 hr on Train 95636, 2 hr over standard.

The next two cars have a similar story. They were among 26 cars scheduled to pass through the terminal on Train 181; however, they were delayed and sent out on Train 793, 8 hr over standard.

Daily Group Performance

The daily group performance report shows all the activity in the yard by movement group. In Table 2, the first two movement groups listed are through cars for Trains 120 and 181. In each case, two cars out of the total group were over standard—in the first case by 2 hr each, in the second by 8 hr each.

The report has separate columns for cars that were less than or more than 24 hr over standard, as well as a total column. In addition to showing the percentage of the movement group over standard, the report provides a theoretical cost in dollars of the failure to meet standards.

Note that whereas the daily cars over standard report only shows cars exceeding standards, the group performance report provides information on all cars moving through the terminal. For example, the fifth group listed shows that all 24 cars received in interchange from the Southern and scheduled to leave on Train 073 made the connection and were within standard.

Daily Terminal Performance Summary

This report, shown in Table 3, provides a summary of daily performance during a 4-week period. For example, on September 20, 749 cars were in the terminal, requiring 1,270 car movements. The report provides a good deal of cost and productivity measurements, such as car movements per engine hour, but in terms of monitoring variation and process capability, the most useful figure is the total over standard, which was 12 percent on September 20.

TABLE 2 WEEKLY GROUP PERFORMANCE AT MEMPHIS TERMINAL (B2: SEPTEMBER 14–20, 1987)

MOVEMENT GROUP	TOTAL MVMNTS	TOTAL CAR HOURS	CAR HOURS PER MVT	CARS OVER STD	HOURS OVER STD	0-24 OVER PCT	HRS STD DLRS	GR 24 OVER PCT	STD DLRS	OVER PCT	STD DLRS
TRAIN ARR/MH-THRU CRS/120	116	739	6.3	3	27	2	$17			2	$17
TRAIN ARR/MH-THRU CRS/181	208	1579	7.5	5	110	1	$40		$29	2	$68
TRAIN ARR/MH-CONN 021/143	15	53	3.5								
TRAIN ARR/MH-CONN 021/181	13	21	1.6								
I/C RECD /MH-CONN SOU/073	114	541	4.7	2	54		$ 9		$24	1	$33
EXPEDITER/MH-EXPEDITER/MH	31	104	3.3								
TRAIN ARR/MH-EXPEDITER/MH	31	97	3.1								
TCF RAMP /MH-EXPEDITER/MH	223	606	2.7								
TRAIN APR/MH-NORTHTOWN/MH	82	1399	17.0	12	105	14	$65			14	$65
I/C RECD /MH-NORTHTOWN/MH	105	1479	14.0	2	3	1	$2			1	$2
TCF RAMP /MH-NORTHTOWN/MH	1	12	12.5								
INDUSTRY /MH-NORTHTOWN/MH	7	288	41.1	7	143	100	$89			100	$89
OTHER /MH-NORTHTOWN/MH	5	68	13.7								
TRAIN ARR/MH-GALESBURG/MH	150	1890	12.6	6	69	49	$43			4	$43
I/C RECD /PI-GALESBURG/MH	1	20	20.0								
I/C RECD /MH-GALESBURG/MH	28	374	13.3								
AUTO RAMP/MH-GALESBURG/MH	9	146	16.3								
INDUSTRY /MH-GALESBURG/MH	57	1801	31.6	26	441	45	$273			45	$273
OTHER /MH-GALESBURG/MH	16	316	19.7								
TRAIN ARR/MH-ST. LOUIS/MH	186	2574	13.8	39	362	20	$204		$ 20	20	224
I/C RECD /MH-ST. LOUIS/MH	45	906	20.1	10	75	22	$47			22	$47
AUTO RAMP/YA-ST. LOUIS/MH	2	88	44.4	2	42	100	$26			100	$26
AUTO RAMP/MH-ST. LOUIS/MH	34	694	20.4	7	63	20	$39			20	$39
INDUSTRY /MH-ST. LOUIS/MH	13	402	30.9	9	166	61	$74	7	$ 29	69	$103
OTHER /MH-ST. LOUIS/MH	4	69	17.3								
TRAIN ARR/MH-NE SHORTS/MH	118	2171	18.4	63	483	53	$299			53	$299
I/C RECD /MH-NE SHORTS/MH	41	883	21.5	25	188	60	$117			60	$117

TABLE 3 DAILY TERMINAL PERFORMANCE (A-1 SUNDAY, SEPTEMBER 20, 1987)

		TOTAL	ENGINE HOURS		CAR MVMTS	C & E	CAR DAY	TOTAL	TOTAL OVER STANDARD	
DATE	CARS IN TERM.	CAR MVMTS	ST	OT	PER ENG HOUR	COST PER MVT	COST PER MVT	COST PER MVT	FRONT	DOLLARS
09-20	749	1270	96			$ 6.34	$0.03	$14.37	12%	$1191
09-19	1085	1523	80			$ 4.40	$6.95	$11.36	9%	$ 964
09-18	1097	1409	128			$ 7.63	$6.73	$14.36	9%	$ 775
09-17	1161	1445	128		11.29	$ 7.44	$8.15	$15.59	16%	$1754
09-16	1145	1195	136	1	8.72	$ 9.65	$7.03	$16.68	9%	$ 919
09-15	1047	1186	128		9.27	$ 9.07	$6.60	$15.67	12%	$ 947
09-14	750	1156	128	2	8.89	$ 9.49	$7.62	$17.11	12%	$ 590
	1006	9184	520	3	17.56	$ 4.79	$7.30	$14.87	11%	$7148
09-13	734	1049	104		10.09	$ 8.33	$9.03	$17.36	15%	$ 944
09-12	836	1404	80		17.55	$ 4.79	$9.11	$13.90	18%	$1749
09-11	1138	1360	128	2	10.46	$ 8.07	$8.92	$16.99	19%	$1326
09-10	1127	1245	128	1	9.65	$ 8.72	$8.17	$16.89	13%	$ 892
09-09	1153	1132	128	1	8.78	$ 9.59	$7.68	$17.27	17%	$1361
09-08	1013	904	128	3	6.90	$12.25	$7.35	$19.60	10%	$ 979
09-07	743	969	48		20.19	$ 4.16	$8.87	$13.03	12%	$ 767
	963	8063	744	7	10.73	$ 7.84	$8.49	$16.33	15%	$8017
09-06	950	1331	96	1	13.72	$ 6.14	$9.01	$15.15	14%	$1182
09-05	1154	1436	80		17.95	$ 4.68	$7.41	$12.09	13%	$ 702
09-04	1092	1378	128	1	10.68	$ 7.88	$8.81	$16.69	15%	$1790
09-03	1207	1321	128		10.32	$ 8.14	$8.77	$16.91	17%	$1510
09-02	1154	1536	128	2	11.82	$ 7.14	$8.17	$15.31	15%	$1363
09-01	1523	1132	128	1	8.78	$ 9.59	$6.32	$15.91	7%	$ 508
08-31	977	984	128	1	7.63	$11.04	$8.27	$19.31	12%	$ 637
	1151	9118	816	6	11.09	$ 7.58	$8.13	$15.71	13%	$7692
08-30	902	1394	96		14.52	$ 5.78	$6.39	$12.17	4%	$ 482
08-29	1340	1605	88		18.24	$ 4.61	$8.12	$12.73	11%	$1041
08-28	1401	1343	128	1	10.41	$ 8.09	$7.49	$15.58	11%	$ 662
08-27	1243	1154	128	1	8.95	$ 9.41	$8.54	$17.95	17%	$1084

STANDARDS OF MEASUREMENT

Each terminal has had its own performance standards developed. These unique standards were built into the TPC report. Basic standards for the Memphis terminal, as outlined earlier, are 8 hr for through movements and 24 hr for local origin/destination traffic.

It is important to note that failure to meet these standards does not always reflect a failure to perform by the terminal. For example, cars on an outbound train cannot be released until the train actually leaves the yard. However, trains may be delayed or even cancelled due to problems outside the terminal's control, such as a derailment on the mainline or a power shortage. In such cases, all affected cars show up on the report as being over standard, even though the terminal has done its job and constructively placed these cars in an outbound consist within the standard time allotted.

Another example involves local industry. In order to gain maximum use of cars delivered to their siding while avoiding demurrage charges, some shippers have been known to release cars immediately after the switch crew has left the area. If the shipper is approaching the point at which demurrage will be charged, this tactic allows the shipper to avoid demurrage while retaining use of the cars until the next scheduled switch. To the terminal, this may mean that the car is released 23 hr before the next day's scheduled switch, making it almost impossible to remain within standards without performing a special switch.

Thus, in looking at the Memphis terminal's performance, or any railroad terminal, the yardstick used for measurement must be kept in mind. The standard used in this case is based simply on the hours elapsed from the time of release to the terminal and time of release from the terminal. These times are influenced not only by the performance of the terminal operation itself, but also the actions of connecting railroads, shippers, and other BN operating groups. Under this standard of measurement, the only process that can be truly measured must be considered to include all of these parties. However, detailed analysis can identify where the causes of variance (i.e., service failure) occur.

A final point to note is that although one is working with a continuous variable (percentage over standard), this is based on a binary test to determine whether each car was or was not over standard. At this summary level, the number of hours over standard is not taken into account.

APPLICATION OF SPC

Flowcharting

The first step in applying SPC methods to the Memphis terminal was to flowchart the process to be studied. As discussed

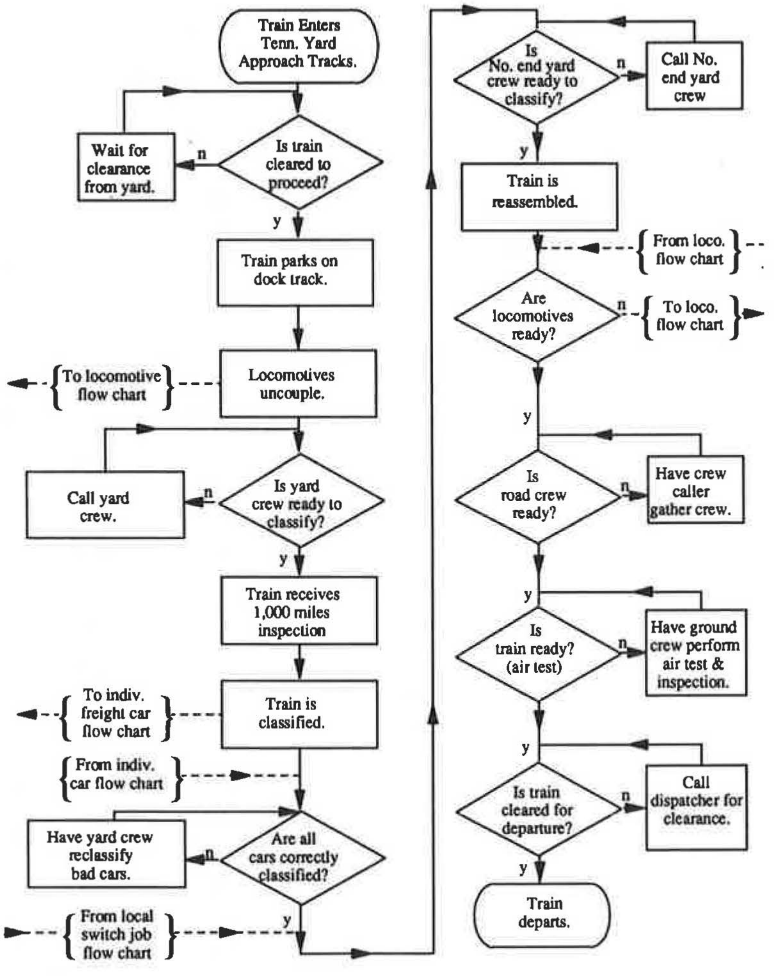


FIGURE 1 Train flow chart (BN Tennessee Yard, Memphis).

earlier, actual flowcharting was done on several subprocesses and not on the process as a whole. Example flowcharts can be found in Figures 1–3.

These flowcharts were prepared by the researchers from interviews with management and were intended primarily as a framework for further discussion and flowchart development by management. They were by no means complete. They were, however, a starting point for further work.

Accurate flowcharting is important to in-depth SPC analysis because it clarifies the boundaries of the process and provides information on points within the process at which to collect sample data. In this study, SPC analysis was performed on existing data because the collection of additional sample data was impractical given the limited scope of the project. However, a more serious long-term effort would require the collection of data at specific points as identified through flowcharting and other methods such as "fishbone" and Pareto analysis.

An important side effect of the flowcharting step is that it can be educational for managers. Asked to develop flowcharts individually, managers will usually not come up with identical versions. As they interact to create a flowchart, their understanding of the process is enhanced. Additional enlightenment is often obtained when line workers are involved in the process. It is common for managers to find that what is actually going on is different from what they think is going on. Also, it may be found that different shifts or work groups are performing the same work through different processes.

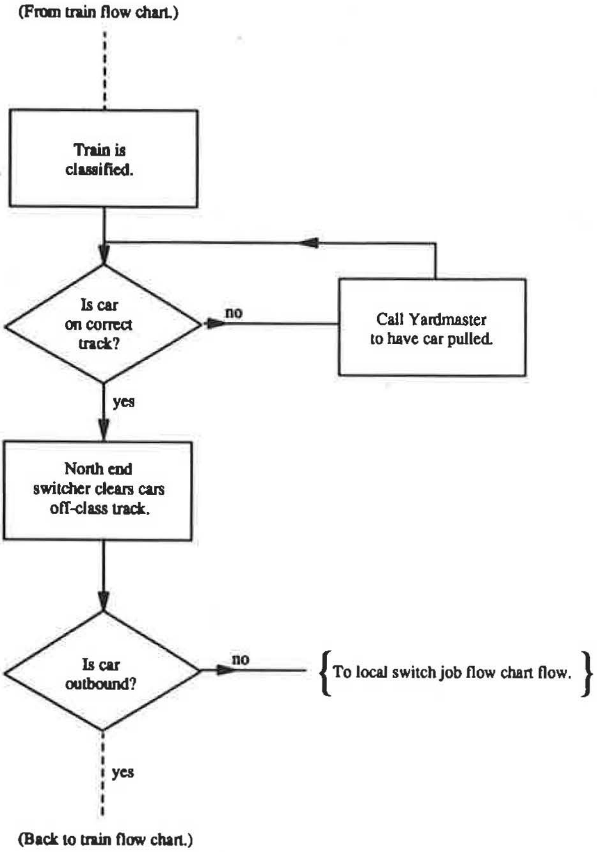


FIGURE 3 Individual freight car flow chart (BN Tennessee Yard, Memphis).

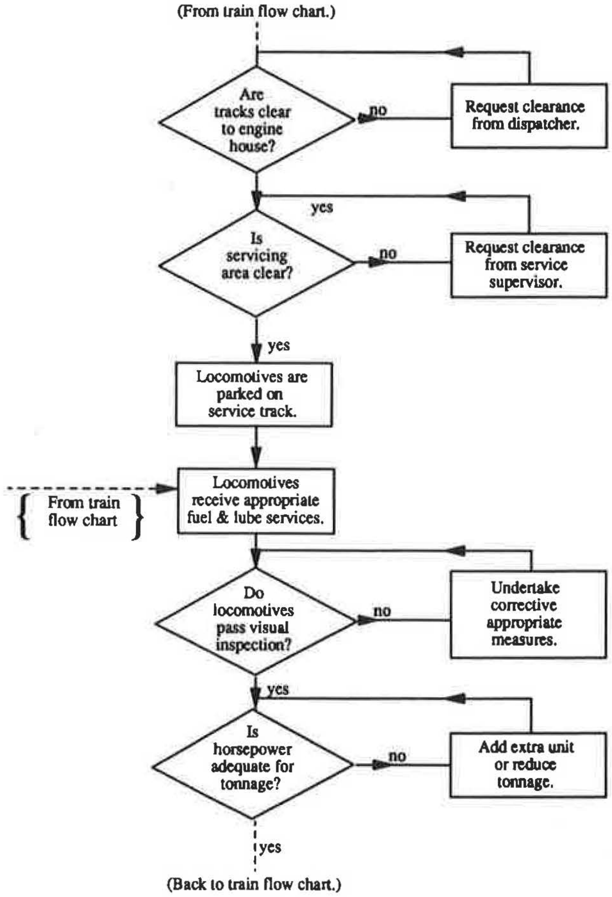


FIGURE 2 Locomotive flow chart (BN Tennessee Yard, Memphis).

In manufacturing, flowcharts identify the points at which inspections are performed on the product, which may lead to acceptance or may destine it for rework or the scrap pile. The goal of reducing variation in the process is to eliminate this scrap or rework and ultimately even the inspection process itself, thereby reducing costs. Data are collected and analyzed at these points in order to determine the causes of variation.

Such decision points in the process are represented by diamond-shaped boxes. Square boxes represent production steps or activities. In the flowcharts presented here, the diamonds contain variables that are decisions in the sense that they may or may not result in rejection of the service performed (meaning that standards were not met).

For example, if a car is found to be incorrectly humped (placed on the wrong track), rework must be performed—a switcher must be sent in to correct the error. The resulting delay may also cause the car to exceed standard time. If management wants to eliminate incorrect humping, it must collect and analyze data at this point in the process to determine the causes of variation present.

Control Charts

As a first step in analyzing the Memphis terminal operation, control charts were created on the basis of aggregate performance data provided in the TPC report. These charts track the weekly average performance of all movements through the terminal (including Expediter trains).

Specifically, the daily percentage of cars over standard, as reported in the daily terminal performance summary, was collected for the period July 24 through November 30. Each day's performance was treated as a single observation in a weekly sample. The *R* chart in Figure 4 charts the weekly range of the observations. The average range between the high and low observation is .117 (i.e., the average variation is 11.7 percentage points). The upper control limit is .226, and the lower control limit is .009. This means that a weekly range of more than 22.6 or less than 0.9 percentage points indicates that the process is out of control—that a special cause has entered. Variation within these limits is due to the process itself and can be expected until the process is changed.

As can be seen from the chart, the process is indeed out of control with respect to range for the week beginning October 13. On the 15th, a bridge fire closed the mainline for several hours, backing up traffic and delaying the departure of outbound trains in the yard. On that day, 29 percent of the cars in the yard exceeded standard. Four days later, only 5 percent exceeded standard, leading to a range of 24 percentage points. Whereas other observations showed performance levels at 6, 5, and even 4 percent, the highest observation excluding the 15th was 26 percent, indicating that the October 15 observation was the outlier signaling a special cause.

The *X*-bar chart plots weekly averages. Average weekly performance was 13.9 percent over standard, with control limits at 18.8 and 9 percent. As can be seen, the *X*-bar chart remains in control, indicating a single process in operation.

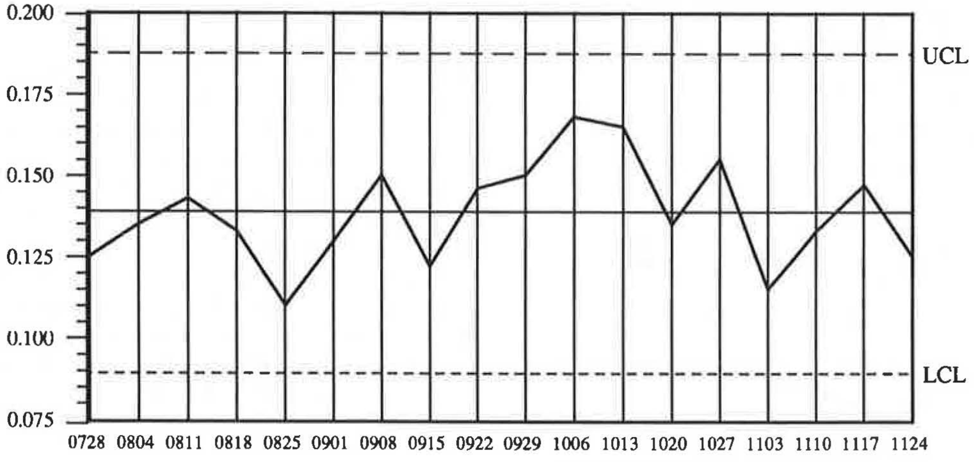
File:	BN1
Company:	Univ. of Tennessee
Plant:	Burlington Northern R.R.
Department:	Memphis Terminal
Machine:	TN Yard
Operation:	Class., Switch
Characteristic:	Weekly cars over standard
Sample frequency:	1 week
Units:	% of total

* * *

1	.12	.13	.11	.14	.08	.07	.10	.12	.13	.10	.10	.15	.06	.15	.11	.13	.08	.11
2	.09	.16	.17	.06	.13	.15	.17	.09	.15	.16	.12	.15	.12	.16	.14	.11	.17	.12
3	.14	.22	.20	.17	.17	.17	.13	.16	.17	.16	.19	.29	.17	.24	.16	.18	.26	.23
4	.13	.12	.21	.15	.11	.15	.19	.09	.17	.20	.20	.17	.21	.14	.13	.18	.15	.21
5	.08	.16	.07	.17	.11	.13	.18	.10	.15	.14	.21	.18	.11	.12	.10	.17	.17	.12
6	.12	.06	.14	.12	.04	.14	.15	.12	.13	.09	.22	.17	.11	.15	.08	.09	.12	.05
7	.20	.11	.10	.12	.12	.12	.12	.12	.12	.19	.15	.05	.17	.15	.13	.09	.09	.06
Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
X-Bars:	.126	.137	.143	.133	.109	.133	.149	.114	.146	.149	.170	.166	.136	.159	.121	.136	.149	.129
Ranges:	.120	.160	.140	.110	.130	.100	.090	.070	.050	.110	.120	.240	.150	.120	.080	.090	.180	.180

X-BAR CHART

LCL = .09 MEAN = .139 UCL = .188
USING HISTORICAL LIMITS, BASED ON 0728 to 1117



NO SAMPLES FAILED ANY OUT-OF-CONTROL TESTS
±1 SIGMA=68% ±2 SIGMA=100% ±3 SIGMA=100%

FIGURE 4 Weekly cars over standard.

The observation for the week of October 13, however, is theoretically in control only by chance, because the range chart shows it to be the result of a different process with unknown average and control limits.

One of the important things about the *X*-bar chart is that it shows clearly the process capability and the variation that can be expected from week to week. The Memphis terminal, when measured against the existing standards, will always have somewhere between 9 and 19 percent of cars over standard. Furthermore, it is known that the average is 14 percent, and that there will always be some weeks above and some below the average. Finally, unless the process is changed, performance better than 9 percent over standard in a given week cannot be expected.

Such information is useful. The size and scope of the process should be considered, including all activities or subprocesses in the terminal. Expediter trains ought to do much better than the overall process, and in fact should be expected to pull the average down. They are really a separate process. Likewise, connecting railroads and shippers' actions are included in this process. It is obvious that interchange, local industry service, and through-train classification are different processes.

Another factor is the cyclical nature of traffic within the week. According to management, and as shown in the daily terminal performance summary, traffic through the terminal is heavier Wednesday through Saturday than Sunday through Tuesday.

Another approach to charting performance on *X*-bar and *R* charts is to treat each day's percent over standard as an individual sample of one, with the range being the difference between the day being charted and the previous day. When the data were calculated in this method, the resulting charts failed out-of-control tests at several points, and a cyclical pattern was evident. This suggests that each day is actually a subprocess that could be analyzed and improved.

Pareto Analysis

The third SPC technique applied was Pareto analysis—the simple assignment of causes to failures, or cars over standard. For this analysis, management tracked down and recorded the causes of delay for all cars more than 24 hr over standard during the month of November, using the daily cars over standard report. The results are shown in Figure 5.

This analysis provides guidance in allocating time and effort to improve performance. Three of the 13 causes discovered were responsible for more than half the service failures. Clearly, management will want to know more about the reporting errors leading to cars missing standard. Are these errors causing actual delays to cars, or are they simply failures on paper? An example is failing to release a car sent to the rip track for repairs.

Such results are typical, however, of a first iteration of the Pareto chart principle. Usually, reporting and other data collection techniques must be refined in order to capture more of the story. In examining reporting errors, it is expected that several causes for these errors will be identified. The reporting process then, is a subprocess, which can be subjected to continued Pareto analysis, control charting, and the like to determine its capabilities or whether changes made have produced a better process.

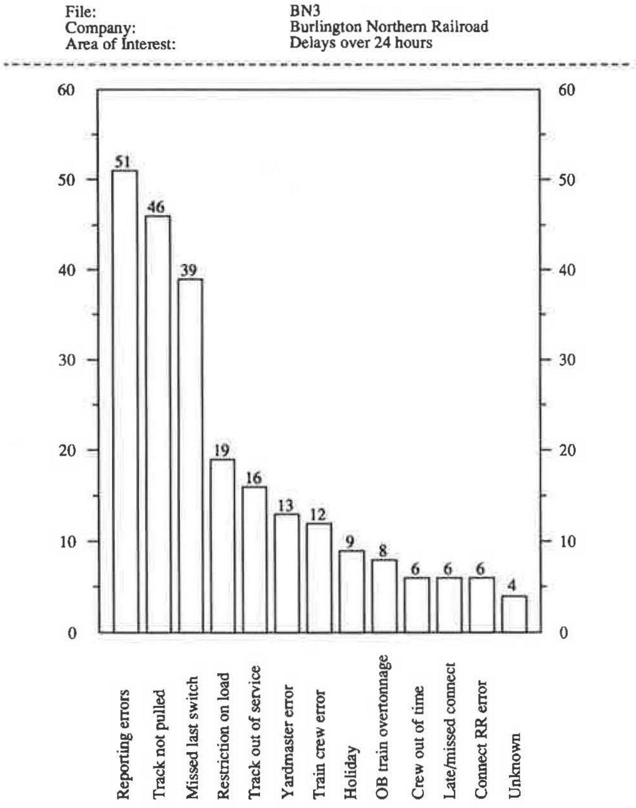


FIGURE 5 Delays longer than 24 hr.

IMPLICATIONS FOR MANAGEMENT

In manufacturing, control charts are traditionally generated using information taken from samples of a larger population, and the sampling procedure used is critical to the interpretation of the charts. In this application, the sample is all-inclusive, but, as noted in the discussion on control charting, this population is really the result of several subprocesses and in fact consists of several products. Thus, the composition of our sample is critical to interpretation of the results.

Figure 6 graphically depicts the sampling procedure. The data utilized to generate control charts are the aggregate of several different types of products produced on different days of the week, which have been identified as being separate processes. There are many different ways to collect and analyze data (or sample), depending on which process or product managers want to know more about.

For example, one could sample weekly aggregate data for just one product line, such as trailer on flatcar service or interchange service. One could go further down the line by sampling just a particular product provided to one customer, such as interchange service with one connecting railroad. Total performance by day of week, or by product by day of week could be charted. One could even track inputs from suppliers, such as cars released by local industry. Note however that the boxes representing separate daily processes are really a composition of numerous subprocesses such as data entry and track maintenance, each of which could be monitored independently.

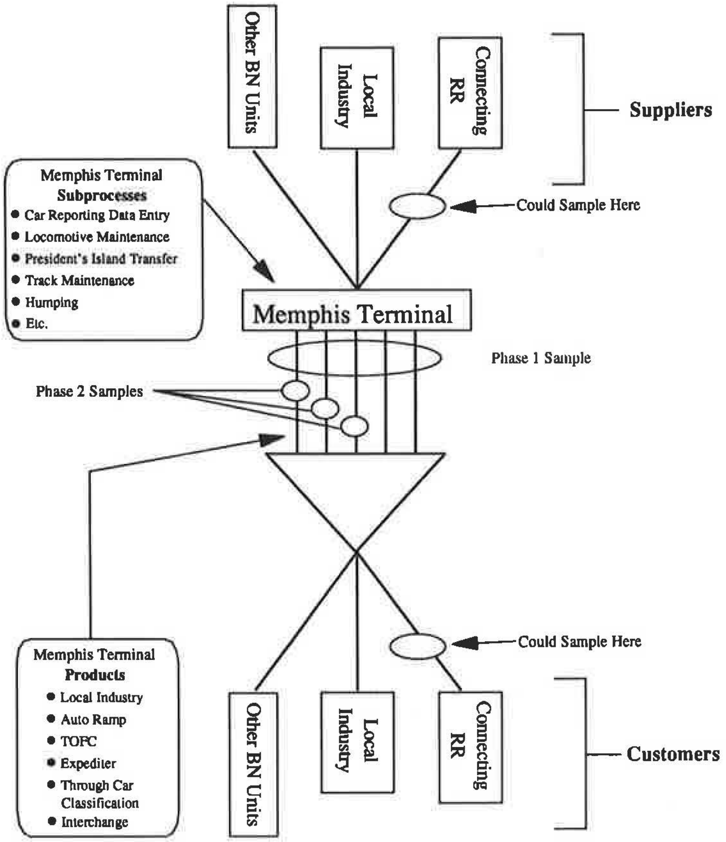


FIGURE 6 Sampling procedure.

The existing data reporting techniques used by most railroads, such as the TPC report, provide sufficient data for some of the suggested sampling techniques; however, other sampling plans would require new and different data collection and reporting. Clearly, it is not feasible to begin charting in all possible ways all at once. It is the task of management, through the use of flowcharting and Pareto analysis, to determine problem areas or possible areas for improvement, and to concentrate on those areas first. Sweeping, systemwide changes may bring in as many new problems as improvements: Deming vigorously argued that continual, incremental improvement and refinement of the process is the true path to superior productivity and quality.

REFERENCES

1. H. T. Landow and R. A. Wharton. The Measurement of Service. *The Matter of Railway System Performance,* Railway Systems and Management Association, Chicago, Ill., 1970, pp. 13–27.
2. P. T. Healy and A. L. Morton. *Note on Train Scheduling and Operating Procedures.* Harvard Business School, Cambridge, Mass., 1974, 9-675-024.
3. R. A. Mundy, R. Passarella, and J. Morse. Applying SPC in Service Industries. *Survey of Business,* University of Tennessee, Center for Business and Economic Research, Spring 1986, pp. 24–29.
4. W. E. Deming. *Out of the Crisis,* Center for Advanced Engineering Study, Massachusetts Institute of Technology, Cambridge, 1986.
5. J. M. Sussman and C. D. Martland. Improving Railroad Reliability: A Case Study of the Southern Railway. *Studies in Railroad Operations and Economics.* Report R74-28. Massachusetts Institute of Technology, Cambridge, 1972.
6. R. A. Mundy and R. Heide. The University of Tennessee Institute for Productivity Through Quality—Alumni Survey. Transportation Management and Policy Studies, Department of Marketing and Transportation, University of Tennessee, Knoxville, 1987, (unpublished).
7. D. J. Wheeler and D. S. Chambers. *Understanding Statistical Process Control.* Statistical Process Controls, Inc., Knoxville, Tenn., 1986.
8. V. L. Beals. Quality and Productivity: The Harley-Davidson Experience. *Survey of Business,* Center for Business and Economic Research, University of Tennessee, Knoxville, Spring 1986, pp. 9–11.

Publication of this paper sponsored by Committee on Electrification and Train Control Systems for Guided Ground Transportation Systems.

Definition of Homogeneous Segments on the Basis of Condition Data: General Approaches and Specific Application to Rail Maintenance

Roemer M. Alfelor and Sue McNeil

Automated systems for acquiring sequential data on the condition of many transportation facilities generate a huge volume of data to be processed. Aggregation of automatically collected and processed data is required in order to make the information appropriate and useful for maintenance management. The problem of data aggregation and techniques that can be adopted to divide linear structures into homogeneous segments for modeling deterioration and assigning maintenance actions are described. The objective is to use the disaggregate sequential condition data from a data acquisition system to determine points of transition from one homogeneous segment to another. The theory and general methodologies are explained in the context of transportation facilities, and an application to maintenance-based aggregation of rail surface condition data is described.

The deteriorated state of the nation's infrastructure has received considerable attention in the past decade (*1,2*). New technologies have been developed to quickly and consistently provide measures of facility condition to enable better quantification of the extent and degree of deterioration and provide an accurate assessment of the actual condition so that funds may be allocated and transportation facilities repaired or rehabilitated effectively. Because transportation facilities such as highways, pipelines, and railroads are linear in structure and extend continuously over several miles, such state-of-the-art automated systems result in large volumes of condition data. Not only does the volume of data create problems for physical storage despite recent advances in storage media, but the data also become incomprehensible.

One way to reduce the quantity of data is to use samples. However, population data are desirable because deterioration does not occur uniformly over the entire length of facilities. Spatial variations in condition occur as a result of variations in traffic loadings, construction quality, subsurface conditions, curvature, and environmental factors to which the facilities are exposed.

An alternative solution is to aggregate the data into physically homogeneous segments. As condition data are used to develop deterioration models and identify defective sections or segments that require maintenance, replacement, or rehabilitation, the aggregation procedure can account for physical constraints and economies of scale in the maintenance process. Maintenance, replacement, and rehabilitation are intended to ensure safe operations and adequate performance of the facility.

In this paper, the concepts and parameters associated with data aggregation are discussed for the two main objectives of data collection: modeling deterioration and planning maintenance. A review of existing approaches to data aggregation is presented, and general steps for maintenance-based aggregation are proposed. A hypothetical example that demonstrates the proposed procedures is also presented. A specific railroad application for determining grinding strategies based on rail surface data is also discussed.

AUTOMATED CONDITION ASSESSMENT

Automation of data collection and condition assessment for transportation facilities has taken a major step during the last decade. For example, pavement condition data can now be obtained using a wide variety of technologies such as video cameras, laser range finders, acoustics, infrared, and many others (*3*). Application of these technologies to highway management has potential for increased productivity and reduced costs. The processed data are used to identify the appropriate maintenance strategies and develop models of pavement deterioration, which could be incorporated in pavement management systems.

Similarly, existing technologies for obtaining railroad defect data are designed to acquire rail flaws including geometry, wear, corrugation, internal cracks, and transverse profile irregularities. This information is used to make decisions on grinding, rail and plug replacement, lubrication, and rail relays. Recently, railroads have experienced unusually high instances of fatigue defects on the surface of the rail. These surface flaws, consisting of cracks, spalls, nicks, slivers, and batters, have been found to increase the dynamic impacts of the wheels on the rail, thereby reducing its service life and requiring premature replacement (*4*). Grinding usually eliminates these surface defects. The growing problem with surface defects becomes apparent as more and more railroads resort to aggressive grinding strategies to remove the flaws. Grinding now constitutes a sizeable amount of rail maintenance budgets (*5*).

R. M. Alfelor, The Urban Institute, Transportation Studies Program, 2100 M St., N.W., Washington, D.C. 20037. S. McNeil, Department of Civil Engineering, Carnegie Mellon University, Schenley Park, Pittsburgh, Pa. 15213.

In most cases, sensors are used to collect the data at high speed. These data are processed and interpreted (either in real-time or at a later date) using signal or image processing to provide condition measures such as amount of head wear on the rail or percentage of pavement covered by alligator cracking. These automatically collected and processed condition data are then used to determine maintenance strategies and understand patterns of deterioration. Because condition data are collected and processed sequentially and are disaggregative, an incomprehensible volume of data is available. For example, every digitized image of the rail surface corresponding to approximately 1 ft in length requires around 200 kilobytes of disk space. Therefore, it is necessary to combine or cluster these data into longer segments having the same level of deterioration or requiring the same level of maintenance in order to be useful. The process is called data aggregation.

DATA AGGREGATION: DEFINITION OF HOMOGENEOUS SEGMENTS

The nature and intended use of the data influence the complexity and approach to the problem of data aggregation. The following issues must be addressed:

- The expected use of the data will determine the method of aggregating the condition data. If the data are used for estimating deterioration models, it is desirable to have segments of similar lengths to prevent aggregation bias in the estimation. If the data are used for specifying maintenance decisions, minimum or maximum lengths may be specified. For example, machines that grind the rail must grind a minimum length. In contrast, if internal defect rates for rail are at such a high level that relay is more economic than plug replacement, a maximum length has been specified for plug replacement.
- The aggregation problem should be able to treat condition data as either deterministic or stochastic. Data collection and processing introduces error into the data (*6*), and the aggregation method should be able to take this uncertainty into consideration. However, analytical procedures that include the impacts of uncertainty in the aggregation process and overall life-cycle cost computations are not trivial.
- The complexity of the problem is also influenced by the dimension of the condition data that require aggregation. In some situations it may be necessary to use a vector of condition measures. For example, the decision to remove a portion of rail in one location and move it to another is based on the amount of wear or loss in cross-sectional area and the number of fatigue defects detected. In the case of highway pavements, resurfacing and overlay are often based on the amount of cracking and rutting on the surface of the road. The approaches for aggregating segments based on multivariate condition indicators are much more involved than those based on single condition measures.
- The complexity of the problem increases with the number of categories of maintenance activities into which the aggregate segments are classified.

These issues are discussed in the following sections.

Aggregated data may be used for modeling deterioration or planning maintenance. In some cases, however, aggregation for both purposes is done by predefining segments that have uniform attributes other than condition. Such attributes include alignment, construction, and traffic loading, which tend to have abrupt changes but remain constant for a significant length. The aggregate condition and maintenance needs of each segment are subsequently determined. This is a special case of aggregation, and although it provides a convenient approach to assigning maintenance or modeling deterioration, it is not efficient because the type and severity of defects within a segment are not homogeneous because of the effects of other variables not taken into consideration. The aggregation problem considered in this research is based on condition and not parameters such as alignment or traffic.

There is also the issue of sensor limitations. For example, an optical rail profile measuring system produces a digitized rail profile at 20 to 30 ft intervals (*7*). This is the minimum unit of measurement for aggregation, even if the rail profile is theoretically nonuniform within the 20 to 30 ft interval. Maintenance and deterioration modeling will then have to be made at least at this level of aggregation. The same is true for measuring rutting or cross profile, roughness, and longitudinal profile of pavements, which cannot be done continuously using existing sensors. However, video cameras can obtain continuous images of the rail or pavement surface, which usually results in disaggregate surface condition data.

Deterioration models are often estimated on a unit length basis (i.e., number of cracks per foot, number of internal defects per mile). This means that, for example, for a frame-by-frame video condition data, each frame represents an individual sample and the deterioration model should be estimated at this level of disaggregation. However, at this level, it may be difficult to relate condition to causal variables, such as traffic or construction. A more typical deterioration model will be based on continuous data, such as percentage defective or a more defect-specific measure. This may be derived from binary defect/nondefect frame-by-frame condition data, which can be aggregated into longer unit segments and the condition data for each unit length calculated as the percentage of frames defective or the sum of continuous data for individual frames. This aggregate condition data can be used to model deterioration as a function of construction, maintenance, and traffic parameters. It is clear that the unit length of segment required to model deterioration may be arbitrarily chosen.

However, if segments must be categorized as either good or bad for defining problematic segments, considering nonuniform traffic, curvature, construction, and maintenance, aggregate condition data based on fixed lengths is no better than aggregate condition data for predetermined segments. For this purpose, the aggregation procedure requires clustering the data samples into segments with or without minimum length requirements, and the segments need not be of uniform length. The problem becomes one of identifying specific locations on the facility where the condition changes from good to bad or from one level or severity of deterioration to another and then finding the explanatory variables to which the condition can be attributed. The aggregation procedure becomes similar to that required for assigning maintenance.

On the basis of maintenance decision-making, the data points or samples can be categorized according to the total percentage of defect (or whatever extent measure), without regard to the type of surface defect. The objective of categorization in this case is to put different segments (made up of adjacent samples) in different categories corresponding to specific types and levels of maintenance. Hence, one does not make a distinction between a spall and a shell that extend with the same depth beyond the rail surface.

The maintenance-based categorization is different from deterioration modeling because there are constraints on the minimum length of a segment. For example, rail maintenance strategies include grinding, plug replacement, rail relay, and rail replacement. Each type of maintenance is appropriate for a particular extent of defect and length of defective segment. In the case of rail grinding, a typical grinding train will grind at least 500 ft of rail surface defects as suggested by one railroad. This is a constraint on the minimum length of rail segments to grind, and therefore, the data should be aggregated to lengths of at least this much. Rail surface grinding is the focus of this research, but aggregation is also a relevant issue for other types of condition data, maintenance, and track renewal.

APPROACHES TO DATA AGGREGATION

Approaches to definition of homogeneous segments may be classified by (a) the way the raw data are analyzed, (b) the type of condition data, or (c) the number of maintenance categories associated with the aggregate data. Based on the way the data are analyzed, aggregation that does not use predefined segments of fixed lengths can be classified simply as either on-line or off-line. An on-line technique sequentially analyzes data from one end to the other end, and transition points between homogeneous segments are determined one by one using a window of fixed or varying size. Off-line search (also called retrospective search) looks at the entire data set and determines the transition points simultaneously. Depending on the length of data to be processed, on-line search has the advantage of not having to deal with a large set of data in a single step. However, it may not lead to the best solutions. The type of condition data aggregated may be binary or multivariate, categorical or continuous, and deterministic or stochastic. The type of data has significant bearing on the complexity of the aggregation procedures. Aggregation of multivariate categorical stochastic data is the most difficult to implement. Finally, the number of categories associated with the data refers to the number of maintenance actions into which the aggregated data are assigned. For example, one might be interested in rail grinding, relay, and plug replacement as maintenance activities.

Existing approaches to aggregation of condition data illustrate the diverse nature of the process. Three aggregation procedures, namely rule-based aggregation (on-line), control chart (on-line), and change-point analysis (on-line or off-line) are described next.

Rule-Based Aggregation Method

A rule-based aggregation method is an on-line procedure through which the transition point is determined by comparing running averages or other derived measures with a given threshold or set of thresholds. These thresholds may be derived analytically or arbitrarily. For example, application of a rule-based system to aggregation of rail condition data was implemented to determine homogeneous wear segments from transverse profiles of the rail from an optical rail wear measurement system (*8*). The algorithm uses a technique whereby a window of specified fixed length is moved along the track and the homogeneous segments are determined from the running averages and standard deviations of rail wear. The bases for combining segments are the differences in the averages and deviations of the wear measurements for adjacent segments. These thresholds are set arbitrarily. The final results consist of segments classified on the basis of wear as either requiring rail relay or not. Rule-based aggregation such as this is convenient for problems for which the thresholds can be determined by expert judgment or analysis or based on common practice. However, in many aggregation problems, the objective is not simply to satisfy the rules but to maximize or minimize a performance function, which is usually an economic measure (e.g., benefits or costs).

Control Charts

Control charts are commonly used in assembly lines to check product quality (*9*). It is an on-line process based on the concept that each product sample should conform to certain specifications in terms of measurable properties. It is assumed that somewhere along the manufacturing process, the quality of production deteriorates and products become rejectable. Using control charts, the point in time at which the process becomes out of control is determined by looking at the distribution of the sample properties with time.

To put the control-chart analysis in the context of the highway or railroad, each section of a specified length might represent one sample. The vector of parameters describing each section consists of the following:

- Defect measure (e.g., number of cracks),
- Curvature,
- Construction quality,
- Previous maintenance, and
- Traffic loading.

Assuming several miles of road that are homogeneous in terms of the above parameters except the defect measure, control charts can be used to determine which sections have experienced unusually high frequencies of surface defects and consequently greater-than-expected deterioration. The idea behind a control chart is that the transition from acceptable to unacceptable samples is gradual. In other words, the transition point does not correspond to an abrupt jump in quality. This may not be the case for many transportation facilities because defects tend to occur randomly and may be found in isolated locations.

Change-Point Analysis

In statistics, problems of inferring transition points in a sequence of data are referred to as change-point problems (*10*). This procedure differs from control charts because the changes are abrupt as opposed to gradual. The simplest illustration of the change-point problem is that of determining one change-point from a series of observations. Change-point analysis can be off-line or on-line, depending on whether the entire segment is analyzed for all change-points (off-line) or a running window is used in the search (on-line).

To illustrate, let $y_1, y_2, y_3 \ldots y_n$ be the series of n observations. Assuming only one change-point at location r, $1 < r < n$, the best location of this change-point is determined by calculating the propensities of the change-point being at all possible locations. The location for which the propensity is maximum represents the most likely location of the change-point. This procedure is equivalent to the maximum likelihood technique or Bayes' ratios for analyzing pairwise change-point models (*11*).

The propensity referred to in the change-point problem could be the likelihood of condition or any other performance function like cost impacts of the aggregate data. The problem of clustering frames into longer segments for rail surface defect data involves many change-points because the sample may consist of miles and miles of track. Literature on change-point analysis states that it is virtually impossible to solve problems involving more than one change-point (*12*). The change-point analysis can also be formulated as a mathematical optimization problem using an objective function, which is explained later.

Change-point analysis and its mathematical programming formulation is a complex optimization procedure, and the solution becomes prohibitive for large problems with many possible change-points. However, it provides the most optimal aggregation strategy if solution procedures allow.

GENERAL STEPS IN AGGREGATING CONDITION DATA

The following steps are proposed for maintenance-based aggregation of condition data.

1. Determine costs and life-cycle impacts of maintenance. These include costs of not performing maintenance or allowing the facility to deteriorate and fail prematurely.
2. Determine the threshold condition value that warrants maintenance. This value may be obtained from maintenance experts on the basis of current standards. Because administrators have different approaches to doing maintenance, the threshold values will likewise vary. An alternative approach is to determine these thresholds analytically using economic analysis. It is apparent that these maintenance standards will dictate the type and measure of condition data to collect.
3. Identify the constraints on maintenance. These constraints may be equipment-related (as the case of minimum rail grinding length of 500 ft because of the length of the grinding train) or system-related (uniformity in the alignment of the road).
4. Collect and process sequential defect data and, if necessary, apply smoothing algorithm to remove noise.
5. Use the threshold values (maintenance standards) described in Step 2 to identify the type of maintenance to use for each data point corresponding to a unit length of facility. This step does not take into consideration the constraints identified in Step 3.
6. Apply decision analysis to the entire facility to satisfy the constraints while minimizing the cost of incorrect maintenance decisions made or maximizing the benefits derived from maintenance. This is an optimization problem that can be solved analytically using a variety of techniques including heuristics and integer programming formulations. The complexity of the solution procedures depends on the type of constraints and the number of maintenance strategies, which are related to the measures of condition. Obviously, the longer the facility being analyzed, the longer it takes to arrive at a solution.

The following generic example will help illustrate the above procedures. It is a more general problem than the one that will be solved later. Consider the hypothetical conditions stated next, which may have resulted from application of Steps 1–3.

- The thresholds for percent defective for doing maintenance to each data sample are as follows:

M_1 = percent defect < 30 percent,
M_2 = 30 percent $\leq$ percent defect $\leq$ 50 percent,
M_3 = 50 percent $\leq$ percent defect $\leq$ 65 percent, and
M_4 = 65 percent < percent defect.

- M_1 corresponds to a particular maintenance action associated with a range of percentage area defective. The level of maintenance is a function of the severity of the defect.
- M_2 uses an equipment whose configuration will correct at least 20 adjacent samples. Therefore, if one sample is maintained by M_2, 20 adjacent samples will be affected and subjected to the same maintenance.
- M_4 requires replacing the sample, but the minimum length for replacement is 50 ft (governed by the standard lengths of replacements).
- M_1 is the do-nothing alternative, and M_3 is a spot-maintenance activity that can be performed on individual samples.

Hypothetical condition data are shown in Figure 1 for every unit of collected data (Step 4). Each data point is subject to error, assuming the data collection technology and the data processing techniques are both imperfect. Noise-removal algorithms in the form of data smoothing eliminate some of these imperfections (Step 4). The idea behind a smoothing algorithm is that sequential data points are spatially dependent. For instance, autocorrelation of pavement condition data is discussed elsewhere (*13*). In the case of rail defects, it has been reported that defects tend to cluster at some locations (*14*). Smoothing not only accounts for minor errors in technology or measurement but also makes the clustering of data into homogeneous segments easier. There are many ways of smoothing data (*15*), but one must be careful in using them because they can smooth out abrupt changes in condition, which may be important in understanding the behavior of the facility.

Figure 2 shows the same data points after applying the 3R running median smoothing procedure, which determines the running median of data points taken three at a time (*15*). The continuous curve in Figure 2 is then used in the succeeding steps. Given the ranges of percentage defective (percent of defect) for which the defined maintenance activities are applied, the entire facility is divided into different segments

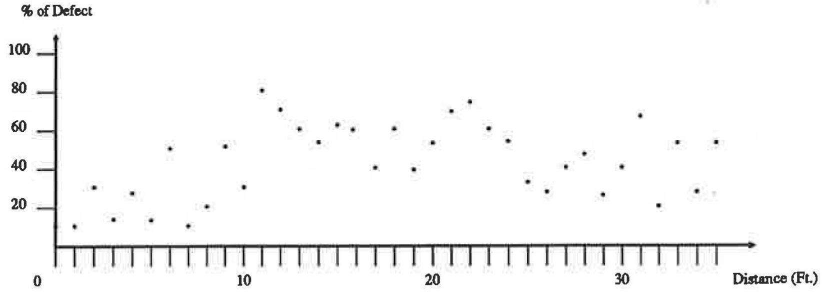


FIGURE 1 Hypothetical raw defect data.

corresponding to different maintenance levels as shown in Figure 3 (Step 5). However, the constraints imposed by M_2 and M_4 may not be satisfied by the segmentation shown in Figure 3. Hence, segments will have to be combined. There are many possible ways of doing this; Figure 4 shows one possible combination (Step 6). The objective is to come up with the best segmentation given a criterion and the set of constraints. The criterion may be to minimize the total cost of performing and not performing maintenance on the different segments of the facility. This is an optimization/decision analysis problem that is equivalent to determining the change-points in a series of observations.

This example can be used to represent the problem of aggregating rail condition data for different maintenance actions; where M_2, M_3, and M_4 may correspond to light grinding, corrective grinding, and plug replacement, respectively, and the constraints on length are changed to their more realistic values.

The general steps for maintenance-based aggregation illustrated in the generic example were adopted in formulating the simpler and more specific problem of using rail surface condition data for one maintenance action: corrective grinding. A description of the rail-grinding problem and the procedures adopted for aggregating rail surface condition data for the purpose of doing maintenance are presented in the following example.

AGGREGATING RAIL SURFACE CONDITION DATA

A prototype optical system using video camera and image processing subroutines was developed at Carnegie Mellon University to acquire continuous images of the rail surface (*16*). The system can identify the presence or absence of rail surface defects on a frame-by-frame basis. An automated defect recognition system was also developed that can process and classify continuous images of the rail surface (*17*). Initial rail surface data were obtained on a 10-mi track and processed in the laboratory using the automated recognition system.

The sequential data obtained from the inspection system were then aggregated to identify segments that needed to be ground. Each frame sample of video data corresponds to approximately 1 ft of actual railhead. The standard identified for performing maintenance is that only defective frames or samples should be ground. However, there is a minimum grinding length that is dictated by the length of the grinding train. This means that a minimum length will be ground. This information is used, along with the costs of incorrect grinding decisions, in formulating the problem as a set-packing optimization problem (*18*).

A set-packing problem is an integer programming problem in which an optimal combination of feasible subsets of a choice set is sought such that an objective is maximized and that each element of the choice set belongs to at most one of the subsets chosen. In the case of the rail grinding problem, the choice set consists of a sequence of video (frame) data, and a subset represents a group or pack of adjacent frames that will either be ground or not. Within each pack, individual frames can be either defective or not, and the higher the proportion of defective frames in a pack chosen for grinding, the better the grinding decision is in terms on maximizing the correct decisions.

The rail grinding problem is a set-packing problem because of the constraint on the minimum grinding length. If such constraint does not exist, then the optimal solution is just to

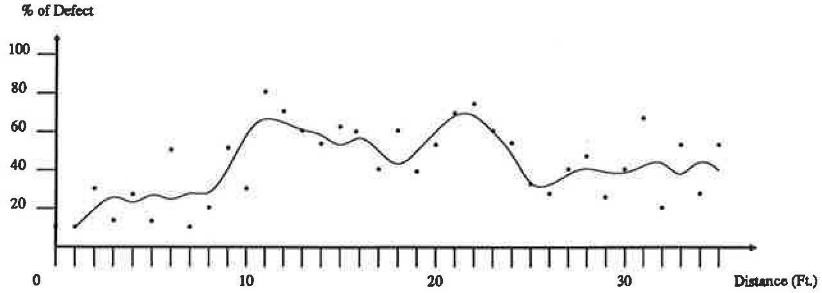


FIGURE 2 Approximate 3R running median smooth.

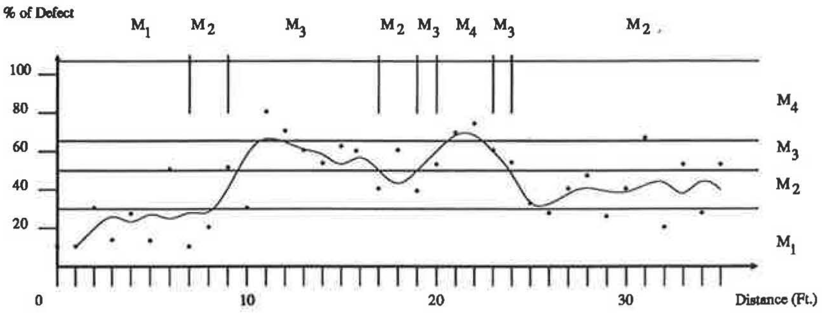


FIGURE 3 Categorized segments based on maintenance level.

grind all the defective frames. However, the constraint becomes a parameter for the minimum size of a pack. Hence, each pack becomes a feasible grinding subset, and the optimal combination of feasible grinding subsets, which maximizes the total benefits due to grinding, is sought. Two aspects of this problem make it different from the conventional set-packing problems. First, it is much, much larger in dimension because of the number of packing combinations and the number of frames involved (again, about 1 ft of rail per frame). Second, the problem has a structure in which each pack is a segment of sequential data.

Solutions exist for solving the set-packing integer programming problem, but then again the dimensions of those problems that have been encountered in operations research literature are much less than the dimension involved in determining rail-grinding strategies.

To solve the problems realistically, two rule-based heuristic algorithms were developed that utilize the structure of the problem. The heuristic solutions can be described as on-line rule-based techniques for solving the set-packing problem. A window that slides from one end to the other end of the track is used in analyzing feasible packs that represent segments of sequential condition data. As the window moves along the rail, segments to grind are determined locally using the criterion for grinding, which is the minimum grinding length, and the minimum threshold for the proportion of defective frames in the pack, which is determined analytically.

The heuristic procedures differ from the general set-packing optimization formulation in that the latter is an off-line procedure whereas the former use a threshold value for the average percentage defective of segments to grind. This threshold value is calculated on the basis of comparisons of cost of grinding and not grinding defective and nondefective rail segments. Therefore, instead of minimizing the overall life-cycle costs of maintenance for the entire track, the algorithms seek on-line local solutions as the window moves along the track. Like most heuristic solution methods, this procedure does not guarantee optimal results. However, although the results obtained by using the heuristics are not necessarily optimal, they are more efficient than the optimal approaches resulting from solving the set-packing problem with respect to execution time and perform reasonably well compared with greedy solutions. Also, the existing approaches for solving set-packing problems are not guaranteed to converge to a solution.

For a lengthy discussion on the aggregation of rail surface condition data, the formulation of the grinding problem as a set-packing problem, and the description as well as application of the heuristic algorithms, the reader is referred to other work by Alfelor (*17*). The effects of uncertainty were also considered in the analysis. The heuristic procedures for rail surface data aggregation were modified to account for uncertainty. Indeed, the solutions to the aggregation problem are shown to be sensitive to uncertainty and imperfect information about the condition of the rail.

CONCLUSION

Maintenance of linear structures requires definition of pieces or segments that represent uniform condition and hence require a particular form of maintenance. Automated inspection of these structures provide the data needed in defining homogeneous segments.

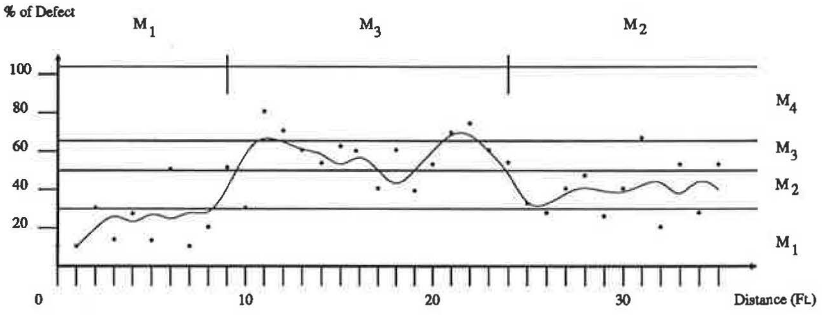


FIGURE 4 Small segments combined.

This research provided a detailed analysis of the problem of defining homogeneous segments for modeling deterioration and assigning maintenance. The general solution procedures are described for any type of linear facilities. An example is illustrated in the case of rail maintenance planning, specifically rail grinding.

Further research and illustrations are necessary for solving the same problem when the condition data are continuous and when the number of categories (maintenance actions) to which the segments are assigned is increased, which is true for most transportation facilities. The same can be said about dealing with the uncertainty in data collection. Moreover, the impacts of making incorrect decisions in terms of overall life-cycle cost and performance of the facility need to be explored.

ACKNOWLEDGMENT

This work was supported by an award from the National Science Foundation.

REFERENCES

1. National Council on Public Works Improvement. *Fragile Foundation: A Report on America's Public Works*. U.S. Government Documents, Washington D.C., 1988.
2. Science, Information and Natural Resources Division, Office of Technology Assessment. *Delivering the Goods: Public Works Technologies, Management and Finance*. Special Report, U.S. Congress, April 1991.
3. D. H. Mendelsohn. Automated Pavement Crack Detection: An Assessment of Leading Technologies. In *Proc., Second North American Conference on Managing Pavements*, Ministry of Transportation, Ontario; FHWA, U.S. Department of Transportation, Nov. 1987.
4. A. M. Zarembski. The Impact of Rail Surface Defects. In *Railway Track and Structures,* Nov. 1984.
5. W. Glavin. *Rail Grinding—The BN Experience*. American Railway Engineering Association, Washington, D.C., 1989.
6. F. Humplick. Theory and Methods of Analyzing Infrastructure Inspection Output: Application to Highway Pavement Surface Condition Evaluation. Ph.D. thesis. Massachusetts Institute of Technology, Cambridge, 1989.
7. C. Martland, J. Sussman, and S. McNeil. *Analytic Techniques for More Effective Rail Scheduling*. Summary Report. Burlington Northern Railroad Company, 1989.
8. D. Acharya. Using Expert System and Network Optimization Technique for Rail Relay Scheduling. Ph.D. thesis. Massachusetts Institute of Technology, Cambridge, 1990.
9. R. V. Hogg and J. Ledolter. *Engineering Statistics*. Macmillan Publishing Company, New York, N.Y., 1987.
10. A. F. Smith. A Bayesian Approach to Inference about a Change-Point in a Sequence of Random Variables. *Biometrika,* Vol. 62, No. 2, 1975.
11. A. F. Smith. *Change-Point Problems: Approaches and Applications*. University of Nottingham, England, 1975.
12. S. D. Silvey. The Lindisfarne Scribes' Problem. In *Journal of the Royal Statistical Society B, 20,* 1957.
13. P. Ullidtz. *Pavement Analysis*. Elsevier Publishing Company, New York, N.Y., 1987.
14. D. D. Davis and T. A. Gudiness. *Rail Planning: A Progress Report of the Track Maintenance Research Committee*. Technical Report WP-128. Association of American Railroads, Washington, D.C., 1986.
15. J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, Reading, Mass., 1977.
16. S. McNeil, B. Motazed, R. Alfelor, T. Short, and J. Eshelby. Automated Collection and Detection of Rail Surface Defects. In *1991 International Heavy-Haul Workshop: Maintenance of Way and Train Operations—Balancing the Conflicts*, International Heavy-Haul Association, 1991.
17. R. M. Alfelor. Analysis of Automatically Collected Rail Surface Defect Data for Rail Maintenance Management. Ph.D. thesis. Carnegie Mellon University, Pittsburgh, Pa., 1991.
18. M. W. Padberg. Covering, Packing and Knapsack Problems. *Annals of Discrete Mathematics, 4*, 1979.

Publication of this paper sponsored by Committee on Railway Maintenance.

Head-Hardened Rails Produced from Rolling Heat

ALFRED MOSER AND PETER POINTNER

A process was developed by which rails are cooled immediately from the rolling heat in such a way that the fine pearlitic structure essential for optimum wear properties is achieved in practically the whole area of the head. Discussed are manufacturing parameters for heat treatment directly from rolling heat; criteria for the selection of rail qualities; structure, mechanical properties, and wear resistance results obtained during service; and production and inspection.

To enable faster rail transport of heavier loads at shorter service intervals, the use of rails with better wear resistance is of great importance. There has been a continuous tendency in the development of rails toward increasing the wear resistance. This is based on the knowledge that the wear resistance of rails with natural mill hardness is directly related to their strength. Substantially higher strength values than 1100 MPa are not attainable in rails with natural mill hardness because they are required to have a pearlitic structure in all their parts and the formation of martensite must be avoided under all circumstances. In the case of alloyed rails the chemical composition of the steel must be adjusted so that during self-cooling of the rail on the cooling bed no martensite is formed, not even in the web and the base of the rails. Therefore, a limitation exists in the strength attainable in this manner in the rail head.

A different approach to develop a higher strength in a comparatively low-alloyed steel has been reached by accelerated cooling from the austenitic range. Because such processes make it possible to perform accelerated cooling, especially in the rail head, an optimum combination of alloy composition and cooling rate can be adjusted, whereby strength levels of 1200 MPa can be obtained in the rail head. Attention must be paid to achieve these high strength levels by a fine pearlitic structure, because a fine pearlitic structure has a considerably higher resistance to wear than a quenched-and-tempered structure of equal strength. This applies in particular to abrasive wear in sharp curves under a high specific load per unit area.

From the metallurgical point of view this behavior is easy to understand. A quenched-and-tempered structure consists of ferrite with embedded, comparatively short, needlelike cementite, whereas the fine pearlitic structure consists of ferrite and cementite plates arranged in layers. If abrasion occurs, the cementite plates of the fine pearlite structure are far more resistant than the isolated cementite precipitations of the quenched-and-tempered structure.

There are in principle two different alternatives for adjusting a fine pearlitic structure in the rail head: by a process in which heat treatment of the rails takes place directly from the rolling temperature or one in which the rail, after normal rolling and cooling on the cooling bed, is reheated and subsequently subjected to accelerated cooling. Suitable coolants are hot water, compressed air, a water-air mixture, and water with synthetic additives. Therefore, in all cases a fine pearlitic structure, rather than a heat-treated structure (which is achieved by quenching and subsequent tempering), is the aim for a rail head. In spite of this fact, the term head-hardened rails is generally used for this product.

HEAT TREATMENT FROM ROLLING TEMPERATURE

Currently, head-hardened rails are generally manufactured by using an additional heat treatment. After reheating the head of rail—usually by induction heating—accelerated cooling is used with a mixture of water and air. By this process, in principle, good results are achieved with regard to structure, strength, and wear characteristics.

For economical reasons, however, a heat treatment directly from the rolling temperature is preferred. The authors have therefore worked intensively on the development of an appropriate technique, for which the prescribed requirements were as follows:

- Heat treatment of the rails shall take place directly from the rolling temperature without causing pollution.
- A fine pearlitic structure shall be obtained, as far as possible, in the entire head area, but at least over a depth of 20 mm.
- The coolant shall have a homogeneous effect over the entire length of the rails, remaining unaffected by contaminations of the installation; it shall not be toxic or flammable.
- The chemical composition of the rail material shall be adjusted to ensure perfect weldability.

Starting from these requirements, first of all, a suitable quenching medium was obtained. Water with synthetic additives was selected. The medium is a high-polymeric compound used in hardening practice. As a result of this synthetic coolant admixture, a layer was formed on the rail and rail head that diminishes the cooling intensity with respect to water. The layer is of uniform thickness and remains unimpaired during heat treatment. It allows uniform cooling across the entire surface and over the entire length of rail and rail head.

Research Department, Voest-Alpine Stahl Donawitz GmbH, Peter Tunnerstrasse 15, A-8700 Leoben, Austria.

TABLE 1 ALLOY COMPOSITION OF RAIL GRADES

Grade	C	Si	Mn	Cr	Ni	Mo	Al
	Chemical Composition (Weight - %)						
S900A[a]	0.60 0.80	0.10 0.50	0.80 1.30	n.s.	n.s.	n.s.	n.s.
S1100[a]	0.60 0.82	0.30 0.90	0.80 1.30	0.80 1.30	n.s.	n.s.	n.s.
≥ 286 BHN Standard Carbon[b]	0.72 0.82	0.10 0.50	0.80 1.10	n.s.	n.s.	n.s.	n.s.
≥ 341 BHN High Strength[b]	0.72 0.82	0.10 0.50	0.80 1.25	max 0.25	max 0.25	max 0.10	n.s.
HSH[c]	0.77 0.79	0.20 0.30	1.12 1.20	max 0.10	max 0.10	max 0.02	max 0.004

[a] Specification according to UIC 860V
[b] Specification according to AREA
[c] Interne VOEST - ALPINE Specification
n.s: not specified

The rail steel selection has to guarantee the perfect weldability as an essential condition. Of course, all rail grades presently in use can be welded. Nevertheless, it is necessary to weld alloyed grades, especially Grade S1100, with relatively expensive and time-consuming measures (e.g., intense preheating and postweld heat treatment.) Otherwise, martensite formation at the coarse grain boundaries is to be expected after subsequent cooling. For this reason, the heat treatment process was based on the unalloyed UIC Grade 900A. Table 1 presents the chemical composition of Grades S1100 and S900A due to UIC 860-V and AREA steel grades and for comparison the upper and lower limits of the most important elements of the steel for the head-hardened rails.

These head special hardened (HSH) rails correspond entirely to Grade 900A; only the concentrations of C and Mn are placed in the upper region of the allowed scatterband.

PROPERTIES OF HEAT TREATED RAILS

The head areas of the rails heat treated by this process showed a fine pearlitic structure without amounts of ferrite, bainite, and martensite. In the area of head-to-web interface, the structure constantly changes to the usual basic pearlitic structure, which is characteristic of UIC Grade A rails after cooling on the cooling bed.

The hardness obtained in the rail-head is shown in Figure 1. The lines of hardness measurements have been executed vertically from the rail top surface, which is exposed to the highest stresses. All hardness values obtained from the HSH rails are positioned between the lower and upper lines in Figure 1. A minimum hardness of 341 Brinell hardness number (BHN) was guaranteed on the running surface. Figure 2 shows a probability plot of hardness values on the running surface, obtained during production series in 1990 for profile UIC 60 and AREA 136.

The minimum hardness of 351 BHN, which is required occasionally by some railway companies, was reached in a number of cases, but the authors are convinced that there will not be ill effects in case of a hardness scatterband on the surface down to 341 BHN.

In addition to dry abrasion wear, one of the most frequent results of wear is the shelling phenomenon, as a result of high axle loads, with subsequent overflow in the flow-stress in deeper parts of the rail head. Therefore, the most important property of a hardness curve should be securing constant high hardness values also in deeper areas of the rail head (Figure 3). The head top and lateral wheel contacts in daily traffic leads to rail shapes with lower heights and partially extreme lateral wear effects. So, these maximum stress-flow areas could reach depths of approximately 20 to 25 mm related to the original rail head profile.

An investigation of newer head hardening techniques revealed that various rail manufacturers all over the world offer deep-head-hardening with high hardness values as a big advantage in regard to service life. A comparison of the hardness curves in Figure 4 reveals similar properties in the deeper rail head regions for this purpose and comparable techniques of other firms. Additionally, there should be an advantage in the properties of the HSH rails discussed here because this hardness distribution was reached with a steel of Grade S900A according to UIC code, which is an unalloyed steel.

Furthermore, one can see in Figure 4 that head-hardened rails produced by off-line treatment show comparatively high hardness values in the surface area, but hardness decreases

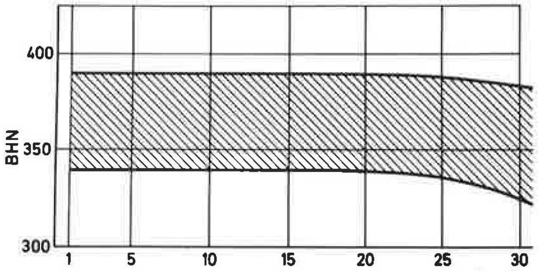


FIGURE 1 Hardness scatterband (measurements executed vertically from the top of the rail head).

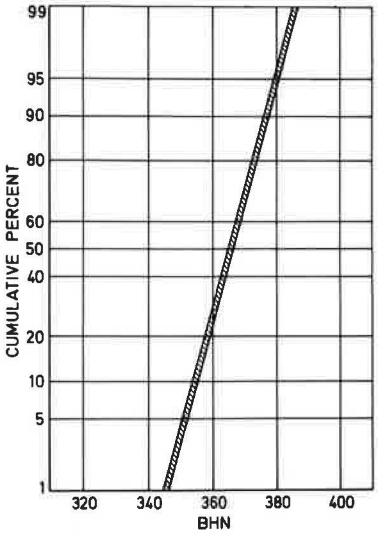


FIGURE 2 Probability plot of hardness on the top of the rail head: production series in June 1990, profiles UIC 60 and AREA 136.

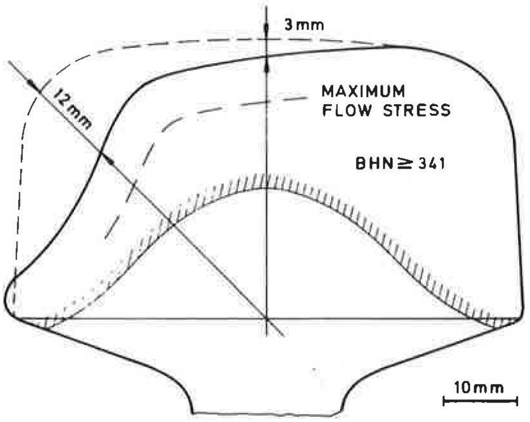
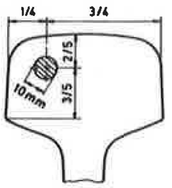


FIGURE 3 Need of an even hardness distribution in the rail head because of the maximum possible wear profile.

with a steep gradient toward the inner area of the rail head. Such rails would cause more problems in the stage of reaching these deeper regions caused by wear. Of course, the position of this hardness scatterband is dependent on the amount of heat, which is transferred into the rail head during heat treatment. More energy will produce deeper austenitic areas; the hardness curves will become more and more similar to those reached by in-line treatment. Nevertheless it is clear that such treatments will be more expensive and therefore perhaps uneconomical.

High-level hardness values that are constant down to deep rail head areas basically correspond with the different kinds of strength values. The strength values are determined on tensile test specimens taken in accordance with UIC specification. The required position of these test specimens is shown in Figure 5. On the basis of the HSH rails with a deep hardness-influence zone it is not necessary to choose special forms of tensile specimens which are normally only taken from rail surface areas and not for inner head controls. The following minimum values were obtained:

Tensile strength: $R_m = 1170$ MPa minimum,
Yield strength: $Rp_{O.2} = 770$ MPa minimum, and
Elongation: $A_5 = 9.5$ percent minimum.



	R_m [MPa] min.	$R_{p0,2}$ [MPa] min.	A_5 [%] min.
HEAD	1170	770	9,5
WEB/ BASE	880	—	10

FIGURE 5 Position of tensile specimen according to UIC 860-V and minimum strength values of HSH rails.

The properties of the untreated parts of the rail (web and base) meet from all viewpoints the conditions according to UIC for the Grade 900A.

To test the toughness characteristics, drop tests were carried out in accordance with UIC specifications. All rails tested have met the specified requirement of withstanding one blow without breaking. In each case, fracture occurred only after two or more blows.

The welding behavior of the rails heat treated by this process also was tested. Tests were conducted with both flash-butt welding and Thermite welding. Welding was carried out in the usual way for UIC Grade A rails. However, after welding, the weld area was rapidly cooled to about 500°C by compressed air, thereby obtaining a favorable hardness variation in the weld area (Figure 6). In the middle of the welded joint, a fine pearlitic structure was observed, and the amount of martensite in all cases was definitely below 1 percent.

Some welding companies recently developed the use of special mixtures for Thermite welding to meet the hardness conditions in the welding area without the necessity of accelerated cooling of the welding joint after welding. The heat affected area is also diminished reasonably by this treatment. Head-hardened rails manufactured by our process can therefore be welded without difficulty.

WEAR CHARACTERISTICS

Of special importance, of course, was the testing of wear characteristics. It is realized that only results obtained during service provide final information in this respect. On the basis

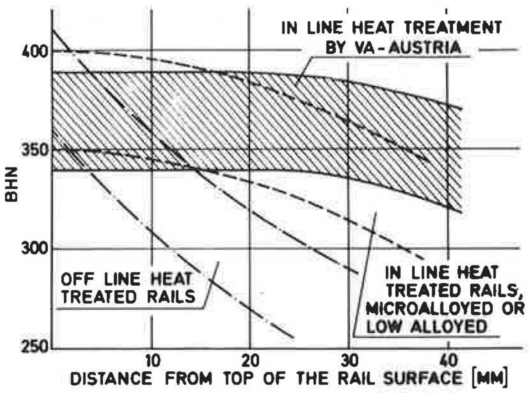


FIGURE 4 Comparison of hardness curves for in-line and off-line heat-treated rails.

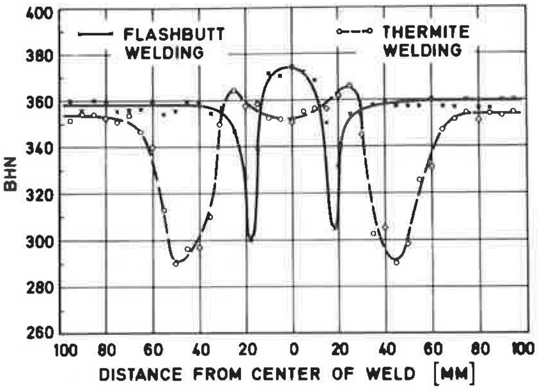


FIGURE 6 Hardness curve in the weld zone of head-hardened rails.

of a great number of wear tests on rails of various strength levels carried out over a long period of time, it is possible to obtain valuable information even from laboratory tests. The tests are abrasion tests (rolling and sliding friction) in which the behavior of rail steel is compared with an ordinary wheelsteel.

There exists a safe correlation between the resistance to wear and tensile strength, respectively hardness, for the common rail steels. The rails heat treated by this process fit in well with this correlation (Figure 7). Therefore, definitely better wear characteristics can be expected from them than from the highly wear-resistant qualities (e.g., S1100) that have been in common use until now.

These results were confirmed by tests in the rolling mill. In a pilot, plant rails up to lengths of 120 ft were heat treated. In each case the rails were taken from a normal rolling of UIC Grade 900A, section UIC 54 E.

In Switzerland these rails were employed in the construction of the outer part of a curve of a radius of about 900 ft on the St. Gotthard line in the valley track section. The curve is lubricated. The line load amounts to 59,000 gross tons per day. Because the lateral surface is exposed to heavy wear, it is necessary to use special rails (at present highly wear-resistant CrMn rails, Grade S1100) for this section of the track.

Figure 8 shows that the lateral wear of a rail in Grade S1100 already amounts to 8.5 mm after having been subjected to a total load of 38 MGT; at this time the rail was taken off the track. The HSH rail was built in at exactly at the same location. Recently this HSH rail was taken off with a lateral wear of 10 mm after a total load of 89 MGT; this corresponds to a wear rate of 22.4 mm/100 MGT for the S1100 rail and 11.2 mm/100 MGT for HSH rails. On this basis one can expect that the HSH rails approximately double the service life of UIC Grade S1100 rails.

The results were confirmed by a significantly different test run. In the Switzerland tests, abrasion was the dominant kind of wear. The tests in the special test track of Stscherbinka in the Soviet Union were executed to receive clear results of the fatigue behavior of the HSH rails. The test was performed on a circular test track that is subjected to a monthly load of approximately 30 MGT with axle loads of 27 tons. Rail wear is only noted to an negligible extent because of heavy lubrication. Therefore, fractures of the running edges as a result of internal defects were the dominant kind of wear. These tests demonstrated that the behavior of HSH rails was by at least the factor 1.6 more favorable than rails of Grade S1100.

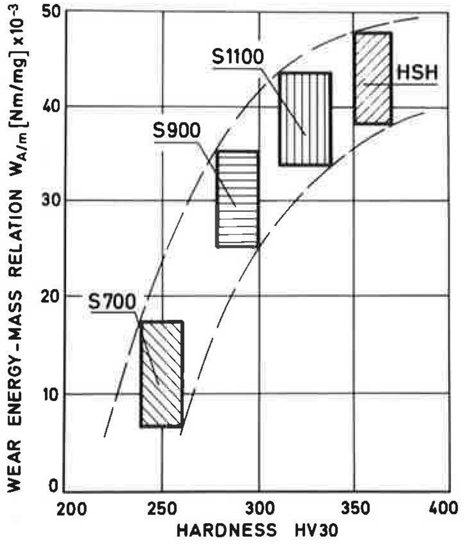


FIGURE 7 Rolling-sliding wear of rail material.

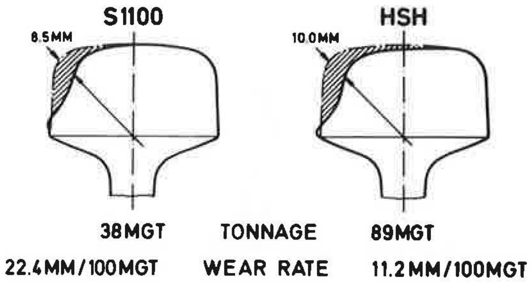


FIGURE 8 Wear of rails S1100 and HSH.

MANUFACTURE AND TESTING OF RAILS

The rail steel discussed here was manufactured by means of the basic oxygen furnace process. Before casting in a three-strand-bloom-caster, a vacuum treatment in an RH installation is carried out, thus ensuring that a hydrogen content well below a maximum of 2.5 ppm can be maintained. This practically excludes the appearance of flakes. Furthermore, this installation makes it possible that the deoxidation is carried out in such a way that the Al content is no more than 0.004 percent and the steel cleanness (concerning the contents of aluminates) can be improved considerably.

The heat treatment of the rails was carried out on a new installation that was put into operation early 1990 together with a new flow line finishing shop. It was designed in such a way that the whole strand up to a length of 400 ft can be heat treated. The strand is automatically dipped into the cooling bath immediately after the last rail mill pass. The position of the rail in the bath is schematically shown in Figure 9.

The depth of immersion is adjusted so that the level of the bath reaches to edge of the head-web interface. This prevents the thinner area of the web from being cooled too drastically, which could lead to the formation of bainite and martensite.

Apart from tee rails, crane rails and special rail sections for turnouts can be heat treated on this installation. In the case of these special sections, the depth of immersion is changed, thus obtaining a fine pearlitic and consequently highly wear-resistant structure in all areas important for the special intended use.

After an immersion time of approximately 2.5 min the entire 400 ft rail is automatically transferred to a walking beam type cooling bed. From there the cooled rail passes through a roller-straightening machine. As again the whole strand is straightened in one piece, the proportion of unstraightened ends without output loss is reduced, and a rail strand with constant height over its total length is maintained. The rails are cut to lengths of no more than 200 ft after straightening

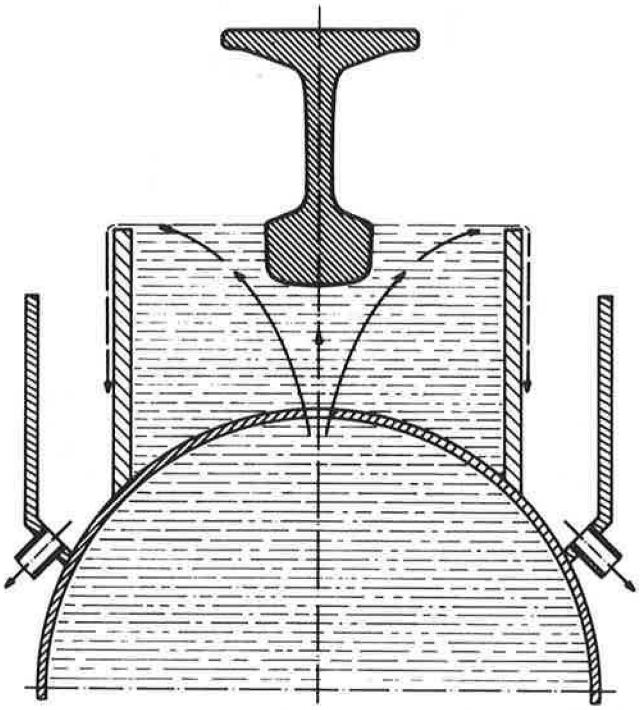


FIGURE 9 Position of the rail in bath.

with a high-performance saw equipped with hard alloy saw teeth.

After straightening, each rail is tested for internal defects in the head and web in a testing center with automatic ultrasonic equipment. The ultrasonic equipment is currently furnished with 6 probes that ensure that all critical areas of the rail cross section are scrutinized.

The measurement of the running surface straightness is executed without contact by 4 laser measuring probes. The results from these probes are evaluated according to the traveling fiber principle, and the values obtained for the total length of the rails are printed in tabular as well as graphic form in test certificates.

SUMMARY

Today head-hardened rails are used in tracks subjected to maximum stress. These rails are generally manufactured by induction heating of the rail head with subsequent accelerated cooling. In the opinion of the authors, this process has technical and, above all, economical disadvantages. Therefore, the authors developed a process by which rails of Grade 900A according to UIC 860-V are cooled down immediately from the rolling heat in such a way that the fine pearlitic structure essential for optimum wear properties is achieved in practically the whole area of the head. The heads of these rails have a minimum hardness of 341 BHN to a minimum depth of 20 mm. The web and base have a pearlitic structure typical of Grade 900A after normal cooling. For mechanical properties the rails fully correspond in these areas to the requirements of UIC 860-V for Grade 900A. Because they also comply with this grade in its chemical composition, the rails head hardened according to this technique can be welded without problems. To achieve an optimum structural shape, only the welded area must be cooled down with compressed air to approximately 500°C immediately after welding.

Results of comprehensive laboratory tests revealed that these rails have a wear behavior that is distinctly better than that of the common highly wear-resistant alloy Grade 1100. Field tests confirmed this favorable behavior of rails heat treated according to the procedure discussed here. This is why an automated plant was built and put into operation early in 1990 that permits heat treatment of rails up to 400 ft long. Together with the new flow line finishing shop, which was activated at the same time, head-hardened rails 200 ft long can be supplied. Because these rails are heat treated according to a special process and have especially favorable service properties, they are called HSH rails.

Publication of this paper sponsored by Committee on Railroad Track Structure System Design.

Turnout Rehabilitation with Bituminous Concrete Underlayments

JERRY G. ROSE

The importance of providing adequate trackbed support to maximize turnout life is emphasized. A relatively new procedure for construction and rehabilitation turnout trackbeds is described. The technique involves placing a subballast layer of bituminous (hot-mix asphalt) concrete below the ballast to provide proper support, impermeability, and positive drainage for the track structure. Design practices, installation procedures, 10-year performance evaluations, and economic studies are presented.

Turnouts represent one of the special trackwork components of the total railroad track system. Turnouts and other special trackwork, such as crossing diamonds, crossovers, interlockings, and highway crossings, typically require higher initial capital costs and higher maintenance costs to operate effectively than comparable regular two-rail trackbeds.

A recent study by a Roadmasters and Maintenance of Way Association of America Committee researching the economics of turnouts concluded the following (*1*):

> [T]he real economic benefits rest in the proper and timely maintenance of the turnout. Ignoring proper turnout maintenance contributes to accelerated wear to the turnout components and a significant reduction in normal serviceable life. It also leads to an increase in the premature replacement of various turnout components.

Obviously turnout maintenance involves not only maintaining the metal wearing surfaces and flangeways to the proper contour, smoothness, and adjustment, and the ties to proper spacing and condition, but also providing proper support for the metal and ties. This in turn provides a means to maintain proper geometric features for the turnout by reducing impact stresses and rapid deterioration and wear of the turnout components.

The selection of the support system depends on several factors. Lines having high-speed traffic and heavy tonnages and wheel loads require the highest quality geometric features for safe operations. Providing proper support is most important. Also, the poorer the quality of the underlying native materials and the more difficulty in obtaining adequate drainage, the more likely maintenance costs will accelerate unless proper roadway stabilization and drainage improvements are provided. The choice of an adequate quality ballast that does not deteriorate and degrade under traffic and weathering is important, particularly on lines that traverse poor quality subgrade materials combined with difficult drainage conditions.

The conventional new trackbed support system consists of an open-graded crushed granular ballast material—typically crushed granite, basalt, or other hard, crushed aggregate. Underlying the ballast is a more dense-graded granular aggregate called a subballast. One or more layers of a geosynthetic may be used above or below the subballast.

Existing trackbeds typically do not have a well-defined layer of ballast. It generally transitions into a subballast. Thus, the relative and total thicknesses of the ballast and subballast vary from site to site, resulting in varying levels of support. Areas exhibiting poor quality subgrade materials and difficult drainage conditions provide less support for the track. Increased deflections of the track and increased relative movements between the rail, tie, and ballast interfaces cause the track components to deteriorate faster. Track geometry is adversely affected, and speeds must be reduced for safe operations.

A report presented at the 1991 American Railway Engineering Association (AREA) Technical Conference on the Economic Assessment of Increased Axle Loads Based on Heavy Axle Load Tests at the Association of American Railroads Transportation Test Center (*2*) indicated that when axle loads were increased from 33 to 39 tons, turnouts showed a much higher deterioration rate, which was reflected in higher routine maintenance requirements and in a shorter life for the major components of the turnout, and furthermore, turnout maintenance costs were substantially increased by heavier axle loads.

An ideal track support system provides a consistent support that is reasonably stiff, but has sufficient resiliency to absorb shocks and deflect slightly under loadings. This is particularly important at special trackworks where impact loadings are typically greater than along regular track. Because high and variable moisture contents of many subgrade and granular subballast materials will adversely affect their ability to provide a proper level of support, it is important that the track system provides adequate drainage to minimize the occurrence of high moisture levels in the subgrade and granular subballast.

HOT-MIX ASPHALT UNDERLAYMENTS

Recorded investigations involving the use of a layer of bituminous concrete, commonly known in the paving industry as hot-mix asphalt (HMA), within trackbeds began during the late 1960s (*3,4*) and continued sporadically during the 1970s. Concerted efforts began during the early 1980s, and today HMA underlayments are considered as an optional trackbed maintenance and construction technique. The AREA *Manual For Railway Engineering* (*5*) recently included a statement

---

Anderson Hall, University of Kentucky, Lexington, Ky. 40506.

that "hot mix asphalt concretes have been used with success as a flexible stabilized roadbed."

During the 1980s significant research was funded by the National Asphalt Pavement Association and the Asphalt Institute to (a) evaluate the applicability of HMA trackbeds, (b) develop design criteria, (c) optimize field installation procedures, and (d) evaluate long-term maintenance costs and operating efficiency concerns based on the performance of in-service HMA trackbeds. Most of this effort has been carried out through the University of Kentucky with support from several railroad companies.

HMA has been applied in hundreds of trackbeds during the past 11 years. The incidence appears to be increasing each year. It has been used in new construction installations such as passing siding extensions, new alignments, yards, terminals, and loading facilities. Wider use has been made at specific locations as a maintenance solution to specific trackbed instability problems in existing tracks. Typically the HMA serves to compensate for inadequate subgrade support and drainage conditions at sites where conventional maintenance and rehabilitation procedures have failed. Specific projects have included short sections of regular track, special trackworks (turnouts, crossings, hump tracks, and highway crossings), tunnel floors, bridge approaches, and loading facilities.

With the exception of a few initial test sites on low tonnage lines, almost all recent installations have been on high tonnage lines where attainment of high-quality support and adequate drainage are more critical. Exceptions are high-volume highway crossings intersecting lightly traveled branch or spur lines where the quality of the highway crossing is considered paramount. A paper presented at the 1991 TRB Annual Meeting (*6*) documents in detail the performance of several test trackbeds incorporating HMA.

TURNOUT DESIGN PRACTICES

A typical cross-sectional view of HMA underlayment trackbed is shown in Figure 1. The HMA mat is placed directly on new subgrade or on an existing roadbed. Most, if not all, turnout projects have involved rehabilitation of a turnout in an existing line. It may be desirable to undercut a particularly troublesome roadbed, but generally this has not been necessary. The increased level of support and enhanced drainage provisions of the HMA layer reduces the stresses to within acceptable limits for the underlying material.

A layer of ballast is placed between the top of the HMA mat and the ties. The HMA mat basically serves as a waterproofing subballast, in place of an open granular subballast, and does not require precise grade control because the layer of ballast serves as a leveling course.

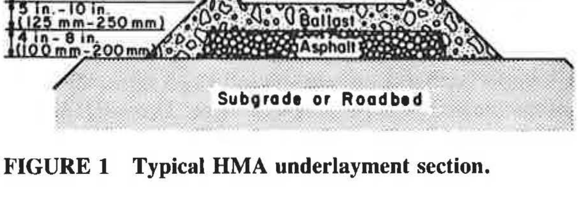


FIGURE 1 Typical HMA underlayment section.

The basic asphalt mix is fairly similar to that used for a conventional 37.5 mm (1.5 in.) maximum size dense-graded highway base. Highway base mixes have been used and are performing satisfactorily. However, it is possible to use a more plastic (low modulus), denser mix in trackbed applications to optimize certain mix properties. This is achieved by using a more dense aggregate grading and a higher than normal asphalt content. The low-void, high-asphalt-content mix is even more impermeable, compacts easier, and is more durable than a typical highway base mix. Details of the design are given elsewhere (*6*).

A rational and practical structural design procedure, KENTRACK, was developed during the early 1980s and later confirmed to be a reasonable, although conservative, procedure for HMA trackbeds. The design is based on two failure criteria: limiting the horizontal tensile strain (fatigue) at the bottom of the HMA and limiting the vertical compressive stress (permanent deformation) on the top of the subgrade. Charts were developed from the computer program to determine thicknesses of ballast and HMA as functions of subgrade support and the amount of train traffic. Discussions on the development and application of the KENTRACK charts for the thickness design of HMA trackbeds are presented elsewhere (*7,8*). Typical thicknesses are shown in Figure 1.

The recommended minimum ballast thickness is 125 mm (5 in.), so conventional roadbed maintenance equipment can be used when required for routine track adjustments. The required ballast thickness increases as the traffic level increases and as the subgrade support level decreases. Typical ballast thicknesses range from 125 to 250 mm (5 to 10 in.), although greater thicknesses have been used.

HMA mat thicknesses range from 100 to 200 mm (4 to 8 in.). These thicknesses are sufficient to provide an impermeable mat. Increased load carrying capability can be more economically achieved by increasing ballast thickness.

The HMA mat should extend about 0.5 to 0.6 m (1.5 to 2.0 ft) beyond the ends of the ties, which requires a mat 3.4 to 3.7 m (11 to 12 ft) wide on single track installations. The mat will extend proportionally wider on turnouts or other special trackworks having longer ties. This will provide adequate width to achieve the desired waterproofing, support, and confinement properties of the HMA mat.

TURNOUT REHABILITATION PROCEDURES

The primary use of HMA underlayments in turnouts is to rehabilitate turnouts that have exhibited poor performance when conventional procedures have been used, which has resulted in high maintenance costs and poor operating efficiency. Pumping of the fines (mud) from the underlying subgrade or old roadbed is the common condition (see Figure 2). Degradation of the ballast and wear of the ties are contributing factors. Geometry is adversely affected. Conventional in-place jacking and raking, plowing the mud from the shoulders, or adding a geotextile have not been successful in prohibiting the re-occurrence of mud pumping and contamination of the track.

The turnout must first be removed from the line. The contaminated mixture of mud, ballast, and old roadbed are excavated, typically dozed, to the desired grade, which ranges from 0.6 to 0.9 m (2 to 3 ft) below top-of-rail elevation, depending on the specified thicknesses of HMA and ballast.

FIGURE 2 Typical pumping turnout before placement of HMA underlayment.

The HMA is hauled by dump truck from a hot-mix plant and is either spread with a standard highway asphalt paver or dumped on the grade and spread with a dozer blade. Precise grade and thickness control of the HMA is more difficult when the mix is spread with a blade; however, time and costs are reduced if a paver and paving crew are not required. The blade spreading technique has been used for the majority of turnout rehabilitation projects. The railroad and contractor forces already on site handle the spreading and placing activities. It is desirable to slightly slope or crown the HMA mat to prevent ponding of water and facilitate drainage away from the roadbed. Installation of drainage pipe is normally not necessary. Figure 3 depicts a typical HMA placement.

The mix can generally be placed in a single lift up to 150 mm (6 in.) thick. Compaction should follow soon before the mix cools and stiffens, which would adversely affect the compaction effectiveness. A standard roller, preferably a steel-wheel vibratory type, is commonly used to obtain a well-compacted mat with minimum air voids.

The turnout can be immediately dragged back on or lifted on the hot HMA mat. Rubber-tired equipment will minimize scuffing of the hot mat. Sometimes it is preferred to place a layer of ballast on the HMA mat before repositioning the turnout. Ballast will slightly indent a hot HMA mat. This is not detrimental and may even be desired to increase the frictional resistance between the ballast and HMA mat. After the rails are joined, all or the remainder of the ballast is distributed. The turnout is pulled and surfaced to provide the specified ballast thickness below the ties and the required line and surface. Either Number 24 or Number 4 ballast is generally specified. It should provide a shoulder 0.30 to 0.45 m (1 to 1.5 ft) wide.

One hr or less is typically required to spread and compact the HMA mat for turnout rehabilitation. The time will be slightly longer if a paver is used to spread the mix.

PERFORMANCE

Four longer sections of HMA underlayment trackbeds were subjected to periodic instrumented tests and measurements during several years of weathering in the trackbed environment. Mat temperatures were monitored throughout the seasons, and HMA cores were taken for analysis of the recovered asphalt and properties of the mix. A detailed analysis, a summary of which follows, of the performance findings is presented elsewhere (*6*).

The range of temperature extremes in the HMA mat in the insulated trackbed environment was considerably less than is typical for HMA highway applications. Also less weathering and hardening of the HMA mixes and more consistent modulus values were noted for the HMA trackbed mixes than are

FIGURE 3 Placement of HMA with paver (*top*) and backdumping and spreading with dozer (*bottom*).

typical for HMA highway mixes. No cracking or other distress was noted in any of the HMA mats.

Subgrade (roadbed) moisture contents under HMA mats were found to have stabilized at or near optimum values. This waterproofing effect provides uniformly strong subgrade support for the service life of the trackbed.

HMA trackbeds have maintained an optimum track stiffness with correspondingly less settlement and smaller deflections than normally obtained on conventional trackbeds. No significant changes in track geometry have been detected on the HMA underlayment trackbeds.

These performance evaluations were conducted on longer sections of track than is common for turnouts 30 to 90 m (100 to 300 ft) long. However, it is logical to assume similar test results would have been obtained for turnouts.

TURNOUT INSTALLATIONS

During the past several years numerous turnouts have been underlain with HMA and periodically observed. Possibly an even greater number of HMA crossing diamonds and road crossings are in service. Design and application considerations are similar.

The cost-effectiveness (or overall economics) of several HMA underlayment installations relative to costs associated with maintenance and operating efficiency before and after rehabilitation were investigated. The findings follow. Performance has been excellent for every HMA installation.

Ravenna, Kentucky

One of the first turnouts underlain with HMA was in 1981 on a Number 10 turnout on the L&N (now CSX) mainline through this East-Central Kentucky yard. An estimated 15 MGT of mostly 0.91-t (100-ton) coal trains use the mainline through the yard. The old turnout and contaminated ballast were removed as part of a general rehabilitation of the yard. An HMA mat 200 mm (8 in.) thick, 35.7 m (117 ft) long, and 4.3 to 7.3 m (14 to 24 ft) wide was placed on the old roadbed. After the preassembled new turnout was positioned on the HMA mat and joined to the mainline, 150 mm (6 in.) of ballast was unloaded and the track was pulled, aligned, and surfaced with conventional maintenance equipment.

Precise maintenance records were not available before 1981; however, the local maintenance crew indicated that the particular turnout had been a muddy spot for many years and had required a considerable number of raises, ballast, and tamping to keep the track in service. In the 10 years since it was rehabilitated, no maintenance has been required, and the turnout has remained high and dry with no fouling or settlement (Figure 4). The abutting track has performed poorly compared with the turnout. A skin lift was applied to the mainline during 1991.

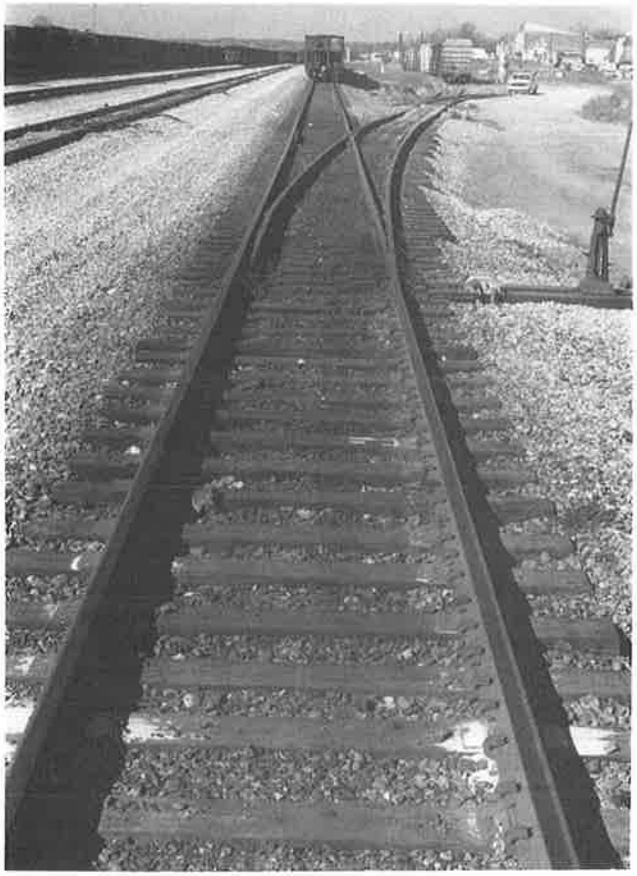


FIGURE 4 CSX Ravenna Yard mainline HMA turnout after 10 years of service.

Flynn Yard Hump Track

An HMA underlayment was used by the Santa Fe in 1985 to rehabilitate the hump track and scale lead track (including four turnouts) at their 20 MGT Flynn Yard in Oklahoma City. This decision followed the favorable 3-year performance of a similar HMA underlayment on the trimlead track at the other end of the bowl. The original hump trackbed, built in 1981 and consisting of lime stabilized soil and ballast, had exhibited periodic major pumping and associated track irregularities. Cranes were used to remove the 183+ m (600+ ft) of line track and four turnouts. Next, the fouled ballast and lime-treated soil were excavated. An HMA mat 150 mm (6 in.) thick was placed in one lift (Figure 5) by means of a highway paving machine. The width and length of the mat varied from 3.7 to 16.5 m (12 to 54 ft) and 191 m (625 ft), respectively. Repositioning the original trackwork and applying 200 mm (8 in.) of ballast completed the installation. A recent view is shown in Figure 6.

Unfortunately, the track maintenance costs incurred on the hump between 1981 and 1985, before the HMA application, are not readily available. However, the magnitude was sufficient for the Santa Fe to essentially reconstruct the roadbed at tremendous cost. Two large mobile cranes, several pan scrapers, dozers, loaders, and the like were used during the 3-day main portion of the project.

The cost of the HMA underlayment was minimal compared with the total project costs. Since 1985 the hump and scale lead trackage, including the turnouts, have required only one skin lift and have exhibited no settlement or pumping. Detailed data on long-term HMA mixture characterization and

FIGURE 5 Santa Fe Flynn Yard hump track with HMA in place.

trackbed performance testing for the Flynn Yard HMA projects are presented elsewhere (*6*).

Red River Army Depot

During 1985, HMA was utilized at this Texarkana, Texas, facility to demonstrate the applicability of the underlayment technique as a rehabilitation alternative on Army railroad track. Three tracks and three turnouts totaling 370 m (1,215 ft) at the locomotive maintenance shop yard were removed. A layer 100 mm (4 in.) thick was placed before the addition of 200 mm (8 in.) of ballast and replacement of the trackage. Particular attention was given to arranging the slope of the HMA mat to minimize leakage of water into the locomotive service pits. A contiguous area of 1,580 m^2 (17,000 ft^2) was paved to accommodate the multiple track configuration. The native soils in the area possess low-quality engineering characteristics.

An extensive evaluation of the design, construction, and initial performance of the installation was made by the U.S. Army Corps of Engineers Waterways Experiment Station, which culminated in a detailed report (*9*). Assuming no trackbed maintenance was required on the HMA section, the additional cost of using the HMA was estimated to yield a break-even point of 7.5 years, when the maintenance savings would exceed the higher initial cost. After 6.5 years, the HMA section has required no maintenance, and there is no indication if or when any maintenance will be required. The problem with leakage into the locomotive pit was also solved. An additional long-term benefit cited in the report was that the effects of the better drainage of the track structure would reduce tie decay, reduce track deterioration, and improve operating conditions. These conditions have been observed and further support the long-term economic benefits of HMA underlayments. A view of the yard is shown in Figure 7.

Falmouth, Kentucky

Two Number 10 turnouts at this town on the CSX Transportation single-track 40+ MGT mainline south of Cincinnati were rehabilitated in 1986 with underlayments to demonstrate the applicability of the technique under traffic. The turnouts had historically required periodic trackbed maintenance with accompanying slow orders. Two dozers and a loader were used to remove the old trackwork and excavate the fouled ballast and several layers of fabric. The HMA was back-dumped and spread with a dozer to a 125-mm (5-in.) thickness, 3.7- to 6.1-m (12- to 20-ft) width, and 46-m (150-ft) length. The dumping, spreading, and compacting of the HMA required less than 1.5 hr. The new trackwork was immediately dragged on the compacted HMA mat and joined to the existing track. Ballast totaling 125 mm (5 in.) thickness was dumped, and the turnouts were tamped and aligned.

The turnouts were rehabilitated on two separate days. A ten-hr curfew was in effect each day; however, only 8 hr was needed. The required track time could have been further reduced, if necessary, by employing larger equipment and

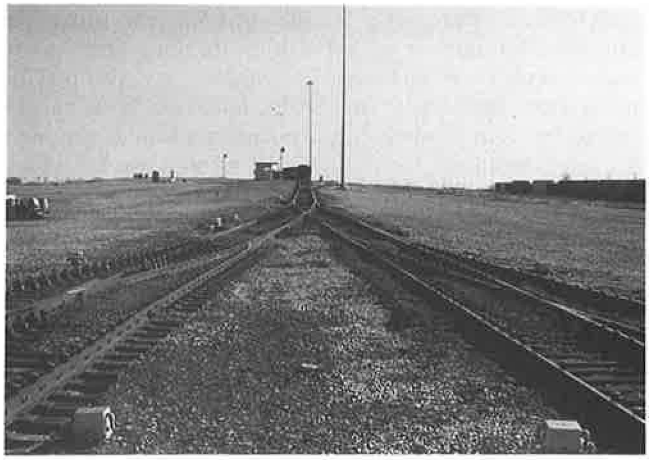


FIGURE 6 Santa Fe Flynn Yard HMA hump track after 7 years of service.

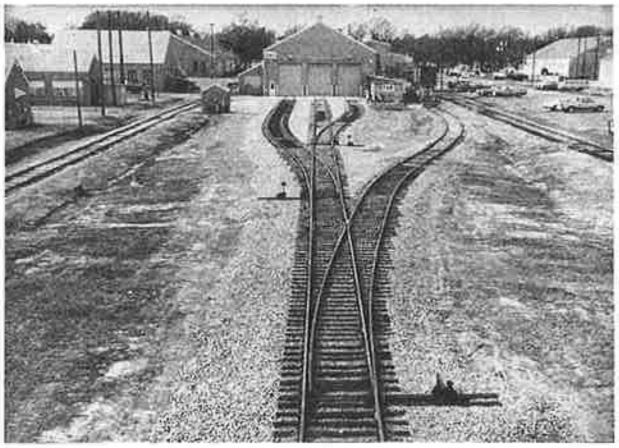


FIGURE 7 Red River Army Depot HMA locomotive maintenance shop yard.

TABLE 1 FALMOUTH REHABILITATION COST PER TURNOUT

Item	Cost
New turnout (metal and ties)	$13,000
Remove old turnout, excavate and place new turnout	
Section gang, 6 workers for 1 day	1,118
Welding gang, 2 workers for 1 day	412
Equipment rental	520
Welds (4 @ $250)	1,000
Surfacing gang for 10 hr	1,174
Tamper	727
Regulator	1,269
Ballast, 200 mm (8 in.)	750
HMA, 125 mm (5 in.), 65 t (72 tons)	2,350
Total	$22,320

NOTE: HMA cost = $\frac{2,350}{22,320} \times 100 = 10.5$ percent.

performing certain preparatory work before cutting the rail.

Detailed costs were maintained for each turnout rehabilitation project. The average cost for a turnout is presented in Table 1. The HMA represented 10.5 percent of the project costs. However, had HMA not been used, a granular subballast layer and geotextile would have been used in its place. Thus, the effective increase in the cost of the HMA was reduced to $1,450, or 6.9 percent, as noted in Table 2.

The affected area was undercut and a geotextile installed in 1982. This proved ineffective. Information gathered from two former roadmasters and the present roadmaster revealed that during the 12 years preceding the HMA applications, the turnouts had required frequent maintenance for adequate geometry for safe operations. On average, portions of the turnouts were raised and tamped at 3-month intervals. Ballast was dropped with mechanized raising and surfacing once a year. The average annual cost for these routine maintenance operations, excluding the 1982 undercutting and geotextile installations, was $2,650 per turnout, as noted here.

Hand raising and tamping four times a year:
5-person crew × ½ day × $175/person/day × 4 times/year = $1,750

Mechanized surfacing once a year:
(½ car of ballast × $800/car × 1 time/year = 400) + (1 tamper/regulator/operator × $1,000/day × ½ day × 1 time/year = 500) = $2,650 total per turnout per year

During the 5.5 years since the turnouts were underlain with HMA no trackbed maintenance has been needed or applied. The effective HMA cost of $1,450 was recovered during the first 6 months. Total savings for the following 5 years is $2,650 × 5, or $13,250 per turnout. In addition, traffic interruptions due to maintenance curfews and slow orders have been eliminated, which has provided additional savings. The excellent performance of the turnouts is depicted in Figure 8.

TABLE 2 FALMOUTH EFFECTIVE HMA COST PER TURNOUT

Item	Cost
Actual HMA cost	$2,350
Minus subballast cost (est.)	−450
Minus geotextile cost (est.)	−450
Effective HMA cost	$1,450

NOTE: $\frac{1,450}{20,870} \times 100 = 6.9$ percent, where 20,870 is the estimated rehabilitation cost without HMA.

Livingston, Kentucky

This represents a more recent (1988) HMA turnout rehabilitation project on which the costs and performance have been closely monitored. Included in the 90-m (300-ft) -long section on the CSX 50 MGT single-track mainline in southeast Kentucky were a Number 10 turnout for the Livingston sidetrack and the approach to north end of the Rockcastle River bridge. The turnout and bridge approach had required periodic trackbed maintenance for several years. A geotextile had been placed previously under the turnout, but had not rectified the chronic pumping and instability problems. It was decided to remove the badly deteriorated turnout and underlying materials and replace them with an HMA underlayment and new ballast and turnout.

The track was out of service for 8 hr. Dozers were used to remove the track and excavate 330 mm (13 in.) below bottom of ties. The 109 t (120 tons) of HMA was backdumped on the grade, spread with the dozers, and compacted to a 125-mm (5-in.) thickness in 45 min. A 200-mm (8-in.) thickness of granite ballast and a new wood tie turnout were installed along with the existing track panels (Figure 9).

Table 3 provides a comparison of the actual costs for the HMA underlayment on 90 m (300 ft) of track and turnout and estimated costs for conventional rehabilitation using additional ballast and geotextile. The HMA basically replaced 100 mm (4 in.) of ballast and a geotextile. No additional labor or equipment costs were assigned directly for placing the HMA because the labor and equipment were on-site and it would have taken as long, if not longer, to place the extra ballast and geotextile.

The HMA was delivered and dumped for $28/ton, for a total cost of $3,360. This represented 7.1 percent of the total project cost of $46,935. However, the increased cost of using HMA over that of conventional materials and design was only $1,160, or 2.5 percent, as indicated in Table 3.

The turnout and bridge approach track sections had been undercut and a geotextile installed in 1983. This proved ineffective. Information gathered from the local roadmaster, maintenance crew, and retired roadmaster revealed that for many years the turnout and bridge approach areas had required frequent maintenance to maintain adequate geometry for safe operations. On average, the turnout was raised and tamped at 2-month intervals. Ballast was dropped with mechanized raising and surfacing at 6-month intervals. The average annual cost for these routine maintenance operations was $4,425, as noted here.

Hand raising and tamping every 2 months:
5-person crew × ½ day × $175/person/day × 6 times/year = $2,625

Mechanized surfacing every 6 months:
(½ car of ballast × $800/car × 2 times/year = 800) + (1 tamper/regulator/operator × $1,000/day × ½ day × 2 times/year = 1,000) = $4,425 total per turnout per year

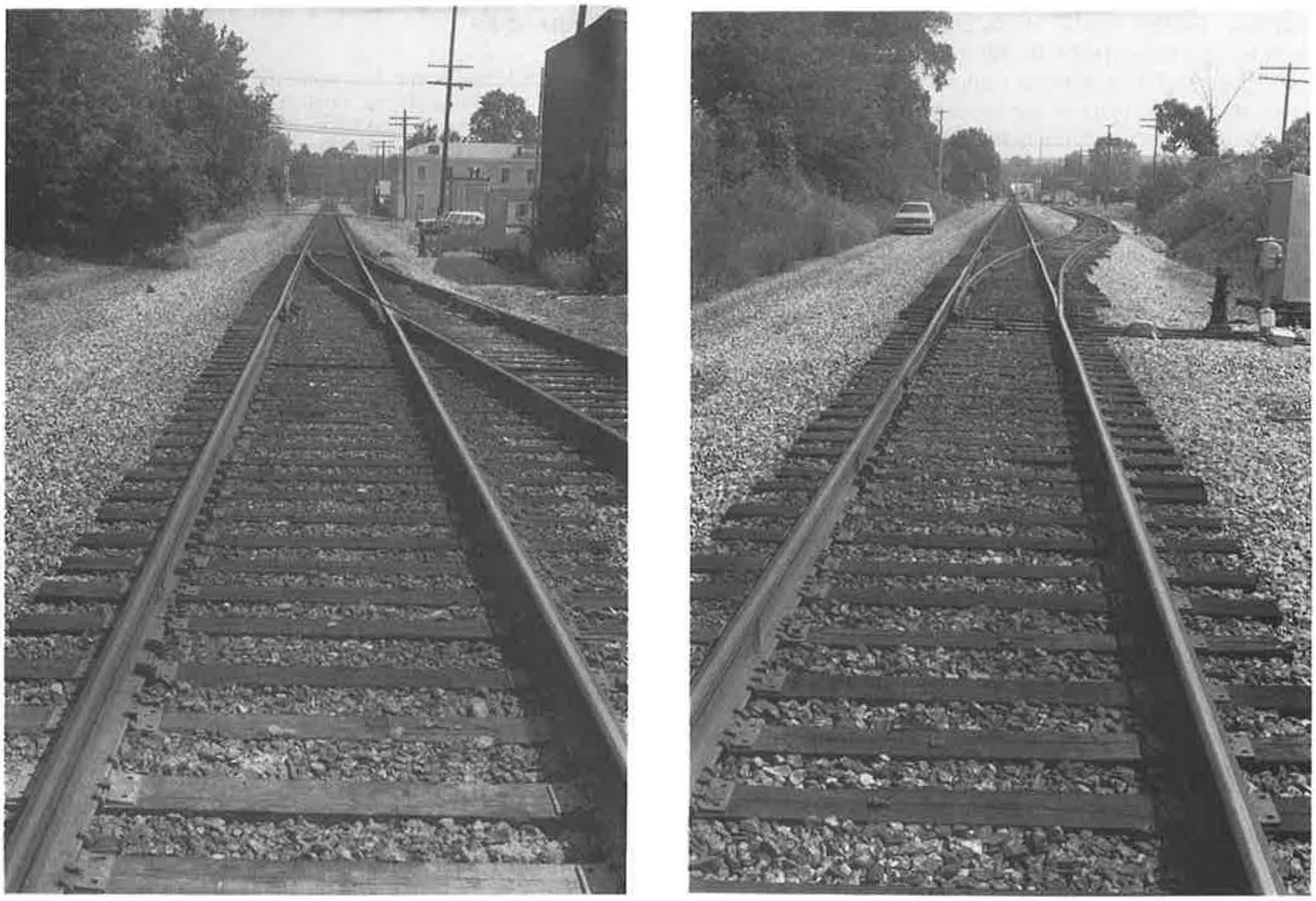


FIGURE 8 CSX Falmouth, Kentucky, mainline HMA turnouts north (*left*) and south (*right*).

The increased cost for the HMA, $1,160, was recovered during the first 3 months by maintenance savings on the turnout alone. During the following 1.5 years no maintenance was required, resulting in a savings of $6,637. During 1990 the wood tie trackage in the area, including the turnout, was replaced with concrete ties. Obviously, the section was surfaced following the concrete tie installation. No maintenance has been required since, and the geometry has remained essentially perfect.

FIGURE 9 CSX Livingston, Kentucky, mainline HMA turnout and bridge approach installation.

CLOSURE

HMA has been successfully used to rehabilitate numerous special trackworks, including turnouts, where conventional maintenance and rehabilitation procedures had not been successful. Performances of all monitored HMA turnout projects have been excellent. No ballast fouling, settlement, or other trackbed deterioration has been observed.

Typically the old turnout and fouled ballast and subballast are removed before placing the layer of HMA and new ballast. The additional time required to spread and compact the HMA is minimal, provided reasonable access for truck delivery is

TABLE 3 NUMBER 10 TURNOUT AND 90 m (300 ft) OF TRACK REHABILITATION AND RENEWAL COSTS, LIVINGSTON, KENTUCKY, 1988

Items		Conventional (Estimated)		HMA Underlayment (Actual)
New Turnout (Metal & Ties)		$16,775		$16,775
Remove Old Turnout, Track & Excavation		11,310		11,310
Replace New Turnout & Track		5,220		5,220
Welds		570		570
Surface & Align		1,000		1,000
18-oz Geotextile		900		–
Ballast & Unloading	(12 in.)	10,000	(8-in.)	8,700
5-in. HMA 120 tons @ $28/ton		--		3,360
Total		$45,775		$46,935
$\frac{3,360}{46,935} \times 100 = 7.1\%$				$\frac{1,160}{45,775} \times 100 = 2.5\%$

Note: 1 in. = 25.4 mm, 1 oz/sq yd = 34 g/m², 1 ton = 0.91 t available. Further modifications and optimization of equipment for delivering HMA by rail to remote locations and for spreading the HMA without removing the turnout will enhance the acceptability of the procedure. The increased cost of using HMA is small and is typically recovered in less than a year through reductions in trackbed maintenance costs and improved train operating efficiency.

The primary benefits of the HMA layer are to improve load distributions to the subgrade, waterproof and stabilize the moisture content and strength of the subgrade, confine and thereby improve the load-carrying capacity of the ballast, and provide an impermeable layer to separate ballast from intermingling with subgrade. The resilient HMA mat eliminates subgrade pumping and does not substantially increase the stiffness of the trackbed.

To date the studies have mainly concentrated on evaluating ballast, HMA, and subgrade performance for HMA underlayment installations. The consistent high-quality support features of the HMA underlayment system should also enhance the performance of the turnout components (metal, ties, and fastenings) by reducing deflections and impact forces that adversely affect the fatigue, wear, and ultimately the life of the components. These factors will become more significant as wheel loads, tonnages, and speeds increase on the mainline routes. Reduced maintenance expenditures and increased operating efficiency are important goals for a profitable, competitive railroad system.

ACKNOWLEDGMENTS

The research reported herein was funded by the National Asphalt Pavement Association and the Asphalt Institute.

REFERENCES

1. Material-Methods and Economics of Extending Turnout Component Life. *Proc., 102nd Annual Conference of the Roadmasters and Maintenance of Way Assoc. of America*, 1990, pp. 106–108.
2. M. B. Hargrove. Economic Assessment of Increased Axle Loads Based on Heavy Axle Load Tests at the AAR Transportation Test Center—Pueblo. *Proceedings Volume 92*, Bulletin No. 732, American Railway Engineering Association, Washington, D.C., 1991, pp. 227–244.
3. A Railroad "First" is an Asphalt "First." *ASPHALT*, The Asphalt Institute, College Park, Md., Oct. 1968, p. 13.
4. Smoother, Stronger Roadbeds for Railway Cars. *ASPHALT*, The Asphalt Institute, College Park, Md., Oct. 1970, pp. 10–11.
5. American Railway Engineering Association. *Manual for Railway Engineering*. Chapter 1, Articles 1.4.1.5 and 1.4.1.6, Washington, D.C., 1988.
6. J. G. Rose and M. J. Hensley. Performance of Hot-Mix Asphalt Railway Trackbeds. In *Transportation Research Record 1300*, TRB, National Research Council, Washington, D.C., 1991, pp. 35–44.
7. Y. H. Huang, J. G. Rose, and C. J. Khoury. Thickness Design for Hot-Mix Asphalt Railroad Trackbeds. *Proc., Association of Asphalt Paving Technologists,* Vol. 56, 1987, pp. 427–453.
8. J. G. Rose and Y. H. Huang. Hot-Mix Railroad Trackbed Systems. *Proc., ASME/IEEE Joint Railroad Conference,* Chicago, Ill., 1990, pp. 85–90.
9. D. M. Coleman. Structural Enhancement of Railroad Track Structures Using Asphalt Underlayment. Miscellaneous Paper GL-86-21. U.S. Army Waterways Experiment Station, Vicksburg, Miss., 1986, 108 pp.

The contents of this paper reflect the views of the author and not necessarily those of the sponsoring agencies or cooperating railroads.

Publication of this paper sponsored by Committee on Railway Maintenance.