CONTENTS

Facing Societal Challenges: The Need for New Paradigms in Rural Transit Service
Sandra Rosenbloom ............................................................ 1

What’s Wrong With the Railways?
Graeme McLay .................................................................... 19

The Effects of High-Speed Rail on the Reduction of Air Traffic Congestion
Andrés López-Pita and Francesc Robusté Anton .................. 37

Optimization of Bus Route Planning in Urban Commuter Networks
Steven I-Jy Chien, Branislav V. Dimitrijevic, and Lazar N. Spasovic 53

A Multiobjective Optimization Model for Flexroute Transit Service Design
Brian L. Smith, Michael J. Demetsky, and Priya K. Durvasula .... 81

Our troubled planet can no longer afford the luxury of pursuits
Confined to an ivory tower. Scholarship has to prove its worth,
Not on its own terms, but by service to the nation and the world.
— Oscar Handlin
Facing Societal Challenges: The Need for New Paradigms in Rural Transit Service

Sandra Rosenbloom
The Roy Drachman Institute
University of Arizona, Tucson

Abstract
In response to major societal trends, rural transit operators should consider different ways to organize, design, and deliver public transit services. This article summarizes societal changes affecting rural areas and identifies five “new paradigms” in rural service organization and delivery to meet the needs created in many rural areas. The article concludes that many operators in a variety of rural settings could enhance their role and better meet the transportation challenge of a changing population by adopting different service paradigms.

Introduction
A U.S. Department of Agriculture (USDA) report noted that:

Rural America has changed in many ways over the century. The rural economy in particular has changed—shifting from a dependence on farming, forestry, and mining to a striking diversity of economic activity. ...[W]hile it continues to provide most of the Nation’s food and fiber, rural America has taken on additional roles, providing labor for industry, land for urban and suburban expansion, sites for storage of waste and hazardous activities, and natural settings for recreation and enjoyment. (USDA 1997)
In response to these changes, the Transit Cooperative Research Program (TCRP) of the National Research Council has suggested that rural operators must move beyond traditional service approaches by developing “new paradigms” or different ways to organize, design, and deliver public transit services in response to the range of economic, social, and political challenges they face (TCRP 1999). This means moving beyond both direct service delivery to traditional clients and dependence on traditional sources of finance and support. This article reports on a study arising from TCRP’s emphasis on new transit paradigms. The study proposed five rural paradigms that represent a new way of thinking about the role of the transit operator and a new definition of mobility.

This article first summarizes a complex set of interrelated forces facing rural areas. The following section suggests ways in which rural operators could enhance their role to better meet the transportation challenge of a changing population by adopting five different service paradigms.

**Profound Societal Changes**

Massive societal changes may have complex implications because they create within rural areas new, different, and varied:

- housing and residential concentrations,
- community economic bases, and
- public and private service delivery systems.

Not all societal trends have the same impact in all rural areas nor do they always have major transportation implications. But most do, both directly and indirectly, because they affect the relationship of home to work, the origins, schedules, and destinations of a range of trips, the trade-offs between travel and other activities, the ability of people to give or receive transport or other services in their communities, and the capacity of communities themselves to meet rural mobility needs.

**Changes in Housing and Residential Concentrations**

Before 1990 most rural areas lost population due to out-migration of younger workers and their families (Johnson and Beale 1999). Although families dependent on agriculture or mining have continued to leave, rural areas in most parts of the United States experienced population gains beginning in 1990 (USDA 1997) due entirely to in-migration—both from metro regions and from abroad (Beale 1999).
These changes had six major components:

1. An influx of young commuters and their families (Nord and Cromartie 1999)
2. Increasing immigration of retirees (Rogers 1999a and b; Fagan and Reeder 1996; Stallman and Siegel 1995; Snyder 1994)
3. Increasing concentrations of older people aging in place (Rogers 1999a; USDA 1997)
4. Growing rural concentrations of minority populations (Cromartie 1999; Swanson 1999; USDA 1997)
5. Changing family structures and living patterns, particularly the growth of households headed by single women and increasing labor force involvement of women with children (Rogers 1999b)

These population changes have created a far more complex set of rural travel patterns and mobility needs. Many rural residents are commuting long distances to suburban or central city jobs in adjacent metro areas or rural jobs in different counties. Internal travel trends are also changing as the influx of both younger and older people—with different needs, abilities, and resources—plays out in changing travel patterns. In-migrant retirees, older people aging in place, ethnic minorities, poor families, and the increasing participation of women in the labor force have created growing mobility needs, even as the ability of family, friends, and community resources to meet those needs may be declining.

**Changing Community Economic Bases**

Rural areas have traditionally based their economies on farming, fishing, forestry, and/or mining—industries that have been declining nationally for more than three decades. At the same time there has been a sometimes startling increase in non-traditional rural employment.

Analyses show five major national economic changes in rural areas:

1. Declining role of agricultural and other land-based industries (USDA 1997; Nelson and Beyers 1998)
2. Expanding manufacturing base (McGranahan 1998; Gordon and Richardson 1998; Roth 2000)


5. Growth of other service sector industries (Beale 1996; Bell and Everett 1997; U.S. General Accounting Office 1997)

These economic trends have also created new and different worktrip patterns. Growing rural manufacturing and large-scale tourism and casinos have created employment concentrations within rural areas that not only provide a rural worktrip focus, but also draw workers from adjacent rural and metro areas. Some rural residents who once worked on farms or in mines now have jobs in local manufacturing plants, tourist destinations, or prisons.

The larger shift to a service economy creates greater variability in the timing and scheduling of work and other trips. Only a minority of service sector workers commute during traditional morning and afternoon peaks; many work different hours on different days. And multijob holding, which has helped some rural families rise above poverty level, creates even more complex and complicated travel patterns.

These trends also affect the ability of rural communities to respond to transportation needs. Many of the jobs created by retirement- or amenities-driven industries are low-skill and low-wage jobs, which partially explains why so many rural working families still live at or below the poverty level. Although these jobs do bring additional income into the community, local residents, including those aging in place, may have even higher demands for governmental or social services. Rural workers who commute out of the area for employment may spend most of their money near their metro area jobs. But at the same time, new rural industries or casinos may be willing to pay for transit services for employees or develop child care facilities for working mothers.

Changes in Public and Private Service Delivery Systems

In the last decade there have been almost unprecedented changes in transportation, communications, and service delivery systems in both rural and urban areas. In some instances these changes have been accompanied by major shifts in government programs and policies. From welfare reform to deregulation in the communications sector, rural areas are being challenged by new situations. Research suggests four major trends in these areas:


3. Changes in rural transit service delivery (Midwest Transportation Center 1996; North Carolina State 1999; Black 1999)


New commuter patterns may arise as some rural areas attract high-tech and “new-tech” firms that can locate outside major metropolitan areas because of new communications and other technology. Competition and deregulation in the communications industry can help narrow the social and economic gaps between rural and urban areas, making rural areas more attractive places to live or visit, changing a variety of travel patterns, and even creating rural congestion. Advances in communications technologies may affect both industry and the ways in which community transportation providers can respond to changing needs in low-density areas. The most obvious example is the growing use by rural transit providers of computer dispatching based on satellite communications.

Public policies can have profound rural impacts. For example, even as rural hospitals close in response to Medicare cost-containment policies, Heath Maintenance Organizations (HMOs) and other managed care programs are moving into rural areas in unprecedented numbers. This may provide more convenient medical options for some rural residents, changing medical travel and even creating new work commutes within rural areas (Ricketts and Slifkin 1995; Frenzen 1997; Kohrs 1997).
New Service Paradigms in Rural Transportation

In 1997 the Transportation Research Board established a New Paradigm Project which asserted:

Local public transit organizations and the services they currently provide are being marginalized at every turn. More specifically, traditional transit organizations:

• have been slow to adapt to fundamental changes throughout society,
• are facing circumstances that threaten their continued relevance in the future, and
• must act out of a renewed sense of urgency to reinvent themselves as agile, responsive, and responsible “managers of mobility” (TCRP 53, 1999, p. ES 1).

The study reported on here suggests that transit operators in rural areas are in a position to take advantage of opportunities by adopting five alternative approaches to meeting rural mobility needs:

1. Serving as community change agents
2. Optimizing community transportation resources
3. Becoming early adopters of technology and innovation
4. Acting as public entrepreneurs
5. Providing state-of-the-art service

The specific services, partnerships, or strategies that can support these paradigms are identified in Table 1 on the following pages.

Community Change Agents

Arguably the most important role for a rural transit system is to become an active participant in all decisions about how and where communities grow and develop. Land-use and development patterns profoundly impact the competitiveness, cost, and efficiency of transit services. Transit staff should sit on all local, county, or regional governmental committees involved in economic development, land use, or housing policies to become more aware of the potential for new and different services. By doing so, transit operators can learn about emerging population and industrial trends and influence land use and development to benefit transit. This interaction also gives the transit system a timely window of opportunity to sug-
gest site-specific exactions (e.g., requiring a firm to construct covered transit stops adjacent to the front entrance), urge firms to offer employees transit passes, or provide funding for specially designed commuter (or other) services.

At the same time rural transit officials can actively work to create denser patterns that make public transit more attractive and efficient. While rural areas will never develop high-density patterns, some commercial, industrial, and public activities can be concentrated in these regions, making public transit more attractive to potential users and service provision more efficient. When transit agencies locate facilities (e.g., transfer points or terminals), they should carefully chose sites that bring transit closer to active land uses.

**Optimizing Community Transportation Resources**

Automobiles are the dominant transportation mode in rural areas—and the greatest resource available in those communities. An equally important resource is the large number of capable drivers. By viewing their mandate as providing mobility rather than just transit service, rural transit operators could facilitate the optimal use of these vehicles and drivers by implementing a variety of ways to share both private and public vehicles. Rural transit operators could facilitate the better utilization of empty seats in privately driven cars through ordinary carpooling and matching programs or by developing more inventive programs, using the power of new communications technology to offer real-time car sharing.

Rural systems could also facilitate one driver using the private vehicle owned by another when that driver is not using it. A rural operator could also implement a car-sharing variation based on schemes tried in a few urban areas. Large residential complexes, like trailer parks or naturally occurring retirement neighborhoods, could cooperatively buy and operate a small fleet of vehicles with a system that allowed individual residents to reserve and drive them, with payment and other rules known in advance. Rural transit systems could encourage residential areas to set up their own programs or actually purchase and maintain the vehicles. Rural operators could also become car rental agencies, maintaining a fleet of vehicles for the sole purpose of rental to rural residents.

Rural operators could also provide a mechanism through which other public or nonprofit agencies could effectively sell (or barter) the underutilized capacity of their vehicles. The transit system role might include developing a matching process, providing regional maintenance facilities, supplying group insurance or umbrella policies to facilitate sharing, and/or offering driver training programs to the
### Table 1. New Rural Service Paradigms

#### OPTIMIZING COMMUNITY TRANSPORTATION RESOURCES

**Facilitate car sharing programs**
- utilize extra capacity in SOV’s
- effectively share POV down time
- facilitate residential car sharing
- use public vehicles more effectively

**Coordinate shared vehicle use**
- design mechanism to sell excess capacity
- operate regional maintenance centers
- supply group insurance or pool
- offer driver training for agency staff

**Organize vehicle purchase schemes**
- trade local match for used vehicles
- facilitate joint vehicle purchase

**Facilitate innovative volunteer programs**
- pay volunteers when efficient
- pay friends/family for cost-effective trips
- pay people to provide feeder services

#### ADOPTING TECHNOLOGY AND INNOVATION

**Implement SOA communications and dispatch technology**
- implement overall improvements
- develop statewide computer and GIS applications
- coordinate with State emergency systems

**Bank transportation credits**
- develop transportation savings accounts
- implement pre-paid options
- develop service-for-service accounts

**Provide transportation insurance**
- develop long-term relationship with appropriate agencies

#### BECOMING PUBLIC ENTREPRENEURS

**Expand contract delivery of transport services**
- contract with urban and suburban operators
- provide service to urban social service agencies

**Sell delivery services**
- contract with local stores for rural deliveries
- contract with freight companies
- provide urban delivery service for rural residents

**Provide alternatives to travel**
- deliver goods instead of rides
- organize volunteers to provide services in lieu of travel
### SERVING AS COMMUNITY CHANGE AGENTS

<table>
<thead>
<tr>
<th>Become active in community decision making</th>
<th>Help create rural trip concentrations</th>
<th>Locate transit facilities near major trip attractors</th>
</tr>
</thead>
<tbody>
<tr>
<td>- serve on all economic development and land use policy committees</td>
<td>- facilitate the co-location of rural services</td>
<td>- provide transfer and other facilities near major commercial sites</td>
</tr>
<tr>
<td>- seek transit concessions from new industry</td>
<td>- encourage concentrated developments</td>
<td>- ensure appropriate transit services</td>
</tr>
<tr>
<td>- ensure meaningful transit access in new facilities/plants</td>
<td>- ensure appropriate transit services</td>
<td></td>
</tr>
</tbody>
</table>

### PROVIDING STATE-OF-THE-ART SERVICE

<table>
<thead>
<tr>
<th>Expand services to match user needs</th>
<th>Use most cost-effective providers</th>
<th>Differentiate services by fare</th>
</tr>
</thead>
<tbody>
<tr>
<td>- restructure routes, schedules, services</td>
<td>- contract for low productivity services</td>
<td>- offer special services at premium prices to attract new users</td>
</tr>
<tr>
<td>- consider services for urban commuters</td>
<td>- pay volunteers for feeder services</td>
<td>- set fares to shift travel to off-peak</td>
</tr>
<tr>
<td>- encourage vanpooling options</td>
<td>- create and support local transportation providers</td>
<td>- use fares to achieve group travel</td>
</tr>
<tr>
<td>- provide services to children</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- provide special scheduled services</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
personnel of cooperating agencies. A rural transit operator could also act as a facilitator in the joint grant purchase of one vehicle by two (or more) agency providers.

Rural operators could also expand the volunteer-based services they have long used by paying volunteers to provide services too expensive for the system to directly provide. For example, rural operators could pay local drivers more than simple mileage charges or directly or indirectly (through user-side subsidies) pay family and friends to provide transport services for people living in areas or traveling at times when it was not efficient for the operator to respond.

**Becoming an Early Adopter of Technology and Innovation**

In industry, early adopters of technology usually surge to the head of their field. Rural transit operators can focus on both improvements in communications and dispatching technology and innovative institutional ways to capture and bank transportation resources. Potential technological strategies include overall improvements in communications and dispatching, statewide application and real-time support of technology, and coordinating improvements with state advances in rural emergency systems. Two recent TCRP projects (B-17 and A-21) concluded that there was wide scope for rural systems to implement a variety of basic or advanced technology and the use of Geographic Information Systems (GIS) databases to improve system efficiency.

In addition, many states are gradually developing statewide emergency systems, based on a variety of communications technologies. But emergencies, while extremely serious in low-density rural areas, are fairly rare and the system and equipment are substantially underutilized on a daily basis. The transit system could use the emergency system to deliver real-time information to riders waiting for service, to assist in real-time dispatching, and to improve data collection through automatic message systems.

There are also institutional innovations open to rural operators: two were first suggested by the Independent Transportation Network (ITN) system in Portland, Maine. Rural systems can develop a way for residents to “bank” transportation rides for future transport needs. Residents or family members can save for services they or their relatives need now or in the future. Residents could also offer services as drivers, escorts, or dispatchers and have rides credited to their own “transportation bank account” or that of a family member for immediate or future use. The transit operator could also offer transportation “insurance” not unlike medical
insurance, where people pay in over the years for guaranteed transportation services sometime in the future.

**Multifunction Public Entrepreneurs**

Public agencies can adopt a more entrepreneurial focus without abandoning interest in nonmonetary factors such as customer comfort or the needs of low-income people. Instead, as entrepreneurs, rural operators would focus more clearly on the customers, exploit every opportunity to increase whatever they see as the bottom line, and consider carefully how to maximize their output given their scarce resources. Acting as public entrepreneurs, rural operators could expand contract transport services to nontraditional clients, provide alternatives to travel, and sell delivery or other nonclient services to the private sector.

Although many rural providers currently offer some contract services, they could more aggressively pursue contract arrangements with both suburban and urban operators. A rural operator with downtime in urban areas (created, for example, by transporting rural workers to metro area jobs) could seek contracts with an urban transit operator or the Area Agency on Aging (AAA), etc. The rural operator could then provide congregate meal service, grocery shopping, or American with Disabilities Act (ADA) services to urban residents during the middle of the day—making money on what would otherwise be costly “dead” time.

Rural operators could also expand some of the services in which they already engage (e.g., meal deliveries) by contracting with private entities needing home deliveries. By coordinating the delivery of other goods—ranging from prescriptions to dog food—with either meal delivery or regular transportation services, operators could substantially increase their income. And, they might even reduce the need for travel by rural residents; the system could occasionally offer the delivery of groceries in lieu of a trip to the grocery store when capacity is limited or the grocery trip is difficult to serve during specific time periods. Delivery services could even be provided by a coordinated system of volunteer drivers (and perhaps shoppers).

Rural systems could become even more entrepreneurial by contracting with private delivery firms to deliver packages. In many smaller areas, national or regional freight delivery firms have only contract providers, rather than directly making delivery themselves. The rural operator could even offer a fee-based delivery service to rural residents by taking packages to an urban post office on passenger runs into urban areas. In addition, the transit operator could make urban goods deliv-
eries for rural residents, such as delivering specialty agricultural items to urban restaurants or farmers markets.

**State-of-the-Art Service Providers**

Local operators must move beyond the services they have traditionally provided, offering alternatives to new markets and new trips, expanding institutional arrangements, and seeking new partnerships. Some rural systems have been extremely innovative in this area. A recent Community Transportation Association of America (CTAA) study for the Federal Transit Administration (FTA) found that rural operators have not only expanded their vehicle fleets and the number of passengers served in the last decade but have also extended their service areas and begun providing new and different travel options (CTAA 2001).

Although some rural systems have expanded their role, not all systems have taken advantage of opportunities. Even the more innovative systems have not gone as far as they can. Overall, rural operators need to actively seek opportunities to expand service offerings to match a range of user needs, matching the most appropriate provider to each market niche (rather than viewing direct service as the first or only response) and differentiating services by fares.

Rural operators should consider route and service restructuring; park-and-ride and express services for suburban and urban commuters; vanpooling for local, regional, and metropolitan commuters; transport for child care and after-school care; and specific services geared to families and older people. Rural operators should also consider a wider range of potential contract providers, using paid “volunteers” more frequently and perhaps “growing” their own subcontractors. In a program funded by the FTA, the University of Tennessee trained welfare recipients to become transportation entrepreneurs. Three of the firms started by this program are now actively involved in a range of contract services for the transit operator itself and for other local agencies needing client services (Newsom et al. 2000).

Rural operators should also consider using fare differentials to promote some services over others or to expand services to higher-income users by charging premium prices for premium service. Operators can use lower fares to encourage people to travel in the off-peak (or whenever demand is lowest), to induce either the general public or clients of social service agencies to group trips, and to motivate users to call in advance rather than seeking real-time service (or vice versa).
Summary and Implications

Many societal trends create problems for rural operators providing traditional services. At the same time, these trends offer opportunities for rural providers to view themselves in new ways, more effectively meeting a range of mobility needs now and in the future. This article has suggested five paradigms that rural operators can adopt to take advantage of their opportunities.

To create systems that continually seek innovative solutions, plan for ongoing change, and act instead of merely react, rural operators need to do more than adopt interesting or novel ideas. Rural operators need to substantially modify how they view themselves and the strategies they employ to provide services. In fact, it is only when rural operators begin by adopting nontraditional organizational structures and new approaches to service delivery that they develop, adapt, and adopt a wide range of innovative ideas and services to better meet their changing community needs.

Acknowledgements

Some of the research reported on in this article was performed under TCRP Project B-22 by the Roy P. Drachman Institute of the University of Arizona. The author is grateful to the TCRP panel and staff for important insights and suggestions. Ghazal Farhang and Emily Steemsa Leunk, research staff at the Drachman Institute, made major contributions. All opinions, recommendations, errors, and omissions are the responsibility of the author.
References


Facing Societal Challenges


About the Author

SANDRA ROSENBLOOM (rosenblo@u.arizona.edu) is professor of planning and director of the Drachman Institute at the University of Arizona. Dr. Rosenbloom is an internationally recognized expert on the impact of societal trends on transport planning. She has written extensively on transit and paratransit planning, emerging travel patterns of women, and growing needs of older people and those with disabilities.
What’s Wrong with the Railways?

Graeme McLay
Transportation Research Group, University of Southampton
United Kingdom

Abstract
Transportation research often focuses on the problem of persuading travelers to switch from private to public transport. Lists are drawn up that note the differences between private and public transport and emphasize the environmental benefits and efficiencies that can be gained from a modal shift. New technologies or even new modes are studied to combat the ever-growing popularity of private transport. Yet, car ownership and travel continue to increase. Instead of asking what would persuade car drivers to travel by public transport, this article focuses on the question: What’s wrong with the railways that make people prefer to drive?

This research follows an earlier paper in the Journal of Public Transportation (Lyons and McLay 2000) and presents some observations on the state of passenger railway in the United Kingdom. Complaints about passenger rail continue to rise (see Office of Passenger Rail Franchising [OPRAF] 1999 and Office of the Rail Regulator [ORR] 1999). While there may be a number of reasons for this rise, including increased press coverage and improved complaint procedures, it is clear that a lack of investment in rail infrastructure led to increased delays and reliability (Department of the Environment Transport and the Regions [DETR] 1998). It is not clear from these statistics as to what people who continued to use the railway actually thought about the service. The Association of Train Operating Companies (ATOC) granted access to its collection of complaints to make further inquiries, which showed that those who do use the railways are generally satisfied with the service.
Complaints

Written complaints filed with ATOC between the beginning of January 1998 and the end of January 1999 were examined and sorted by type of complaint, according to a scenario structure created beforehand. (The process of complaints arriving at ATOC is described in Lyons and McLay 2000). A database was created in Microsoft Access to record information contained in the letters. Any information that gave details of recovery solutions, presentation of information, costs and time spent on any problem was entered into the database. Many letters gave journey details such as date, time, origin, and destination, allowing a comparison to be made between weekday, holiday, and weekend journeys. Journey time highlights differences between the actual times the passenger made on the journey and the times predicted by advance planning agencies such as the National Rail Enquiries Service (NRES).

Also detailed were journey planning information, including verification through repeat searching, showing the efforts people make to acquire information, what source they use, and how much they trust it.

Additional information discussed solutions that were presented to passengers in recovery situations (i.e., trying to continue a journey that did not progress as planned). Passengers’ opinions on the suitability of this information were recorded.

These complaint letters are an extremely biased sample, mostly covering situations that are the fault of the rail service and not the fault of the traveler. For example, they mention arriving late at the departure station and missing the train. The letters are a record of people who are unhappy enough with their rail experience that they have complained. They are also all written complaints and biased by being only the complaints of people who would write to complain. There would always be bias retrospectively treating data as a set of survey responses. An overview of the information items identified from the complaint letters and the “responses” collected is given in Table 1.

The frequency chart shows that the days most complained about are not commuting days, but weekends (and holidays) when passengers are not on their usual journeys. Table 2 shows that the current system of telephoning for information and getting a quote for the fare is not working properly. Between the NRES operator, prospective passenger, and ticket office, there is confusion over journey prices and schedules. It may be that the passenger is not clear with the phone operator about the desired trip (14 complaints concerned the wrong schedule) or that the
Table 1. Overview of the Survey Information

<table>
<thead>
<tr>
<th>Total complaints examined: 105:</th>
</tr>
</thead>
<tbody>
<tr>
<td>About actual journeys: 80</td>
</tr>
<tr>
<td>About information only: 25</td>
</tr>
</tbody>
</table>

Incorrect details about:
- Departure time: 16
- Price: 22
- Other: 14

Not told about a service: 8
Told about a nonexistent service: 21
Remaining complaints concerned late changes and delays.

Of the actual journeys made, the causes of complaints were:
- Engineering works: 7
- Natural phenomenon: 3
- Given wrong schedule: 14
- Given wrong price: 15
- Dissatisfied with information: 20
- Train either broke down or was delayed: 11
- Cause not given: 8

Of the Monday to Fridays, 7 were during holidays (Bank, Easter, Christmas)

En route information sources:
- Staff: 34
- Monitor: 1
- Timetable: 2
- PA.: 3
- N RES: 20
- Other phone: 8
- Notice board: 3
- Arrival/departure notice: 5
- Other: 4 (Internet uncited)

Passengers given another option:
- Train: 44; Suitable: 16
- Mode: 20; Suitable: 6

Additional expenses: 40; Time costs: 16
Passengers who thought recovery information was not enough: 17

ATOC refunds to passengers (from around 80 completed complaints): £1313
operator used an out-of-date timetable (21 complaints concerned a nonexistent service). In any event, the passenger is not getting the desired information. If the passenger cannot get this information before making the journey, this has implications for customer satisfaction.

### Table 2. Frequency of Occurrence

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price is different from quote</td>
<td>26</td>
</tr>
<tr>
<td>Train is different from that expected</td>
<td>20</td>
</tr>
<tr>
<td>Train cancelled</td>
<td>14</td>
</tr>
<tr>
<td>Train stops at intermediate station</td>
<td>8</td>
</tr>
<tr>
<td>Train runs behind schedule</td>
<td>6</td>
</tr>
<tr>
<td>Get to station, train gone</td>
<td>4</td>
</tr>
<tr>
<td>Get to train, does not depart</td>
<td>4</td>
</tr>
<tr>
<td>Train departs late</td>
<td>3</td>
</tr>
<tr>
<td>Get to station, can not find train</td>
<td>2</td>
</tr>
<tr>
<td>Train arrives late</td>
<td>2</td>
</tr>
<tr>
<td>Train does not stop at expected station</td>
<td>2</td>
</tr>
<tr>
<td>Train stops outside station</td>
<td>2</td>
</tr>
<tr>
<td>Get to station, imminent departure</td>
<td>1</td>
</tr>
</tbody>
</table>

While these passenger complaints represent an extremely biased sample, they show that certain situations generate dissatisfaction. A train in some way different from that expected, because the schedule or the price is different, causes the most complaints. While these situations would benefit by improving communication between the passenger and the operator, the remainder of this article describes the satisfaction of regular train passengers with the service they receive. Complaints were bundled according to the information needed to resolve them and particular attention was given to those that might affect actual rail journeys as opposed to those that would benefit from more reliable information at the outset. For this reason, and because the research was already investigating recovery situations, passengers were asked their opinions of the service they received when something went wrong with their journey to find out if there is something wrong with the railways that is encouraging people to drive.
Survey

Scenarios were developed from the complaints that matched the broad groupings in Table 2. Passengers were asked what they would do if faced with these situations for their journey (for more on this process, see McLay and Lyons 2000). Although there are many problems with survey data (Oppenheim 1996), matching scenarios to the journeys actually taken by passengers, and having them fill out the journey survey where possible, would provide more reliable information. Survey biases and the measures used to control them are noted in Table 3.

<table>
<thead>
<tr>
<th>Bias</th>
<th>Nature of Bias</th>
<th>Controlling Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>Age/gender/experience</td>
<td>Obtain demographic details of South West Trains passenger population</td>
</tr>
<tr>
<td>Experimenter</td>
<td>Participant selection, effects on participants</td>
<td>Approach all passengers with prepared speech, depart during participation</td>
</tr>
<tr>
<td>Questionnaire</td>
<td>Stated preference/memory/desire</td>
<td>Context</td>
</tr>
<tr>
<td>Train</td>
<td>Rolling stock, express/stopping</td>
<td>Sample different stock and journey types</td>
</tr>
<tr>
<td>Time</td>
<td>Temporal effects, peculiarities of the day</td>
<td>Careful selection as outlined above</td>
</tr>
</tbody>
</table>

In fact, post survey analysis showed that there were no significant differences in responses between people who returned the surveys on the train and those who sent them back later. The survey route chosen in consultation with South West Trains was from Bournemouth to Woking (as shown in Figure 1). The route was long enough to allow some modal competition, variations in the event of a breakdown, varied origins and destinations (either internal or external to the route) allowing connections, onward journeys, and have commuter and leisure users.
The route is popular with commuters and people traveling from the west into London. There are direct and stopping services and competing coach services. Passengers can disembark at Southampton or at Woking for the coach to Heathrow. Most travelers remain on board for London Waterloo. There were also a variety of ways to get to the station, including the underground in London and the ferry in Southampton.

After piloting the survey in March 2000, three dates were chosen around Easter for disseminating it. The timing of the survey around Easter would allow more people who were unfamiliar with their journey, and who, therefore, might be in more need of information, to be sampled. Wednesday, April 19, was chosen as a regular mid-week day. April 20 was included because there happened to be a strike by South West Train drivers, causing even regular travelers to seek more information than usual, and which allowed a comparison to be made with a follow-up survey one week later.

**Analysis**

Of the 1,200 questionnaires handed out, 550 were useable; 298 were collected on board the train and 252 were mailed back. (This 46% response rate compares very favorably with an industry standard of 33%.) Testing on differences of proportions shows a match within 95 percent confidence levels across all demographic categories between the survey results and South West Trains’s demographic profile, although there is a higher number of the 35–54 age group in the survey (53% v. 37%). Across almost all questions there was little statistically significant difference in responses between those mailed back and those collected by hand. There were
more 25- to 34-year-olds who mailed back their responses \( z = 2.583 \) and more 35- to 54-year-olds who returned them by hand \( z = 2.245 \). This may be a consequence of young professionals working on the train and older people using the questionnaire to pass some time on the journey. These statistics mean that the results of the questionnaire can reasonably be assumed to apply to the passenger rail-using population. They also mean that mail-back questionnaires can be used to obtain data from the rail-using population, rather than collecting surveys by hand. Unless otherwise stated, there were no significant differences in responses by gender, age, or number of passengers traveling.

Passengers were asked how often they traveled by rail to test any effect of experience of incidents on actions. Of those who traveled by rail once a week or less, 52 percent would take a car if the train were not available, while 34 percent would cancel the trip. Of those who travel most days in the week, 38 percent would use a car, but 53 percent would cancel. Perhaps people who make irregular journeys place a higher value on them, while it may be that people who commute feel more able to use the rail company as an excuse to skip work.

Nearly 60 percent of the respondents drove or were driven to the station; 48 percent would have driven if the train were not available for that day’s journey. These numbers indicate that half the sample could have driven if they wanted, but they prefer to take the train. Removing the effect of the London Underground from the sample (those traveling toward London cannot use the tube to get to the station), a chi-square test showed no significant difference at the 0.05 level between days 1 and 3 for methods of getting to the station. However, on day 2 (strike) more people drove to the station. Most people seem to have reasonably fixed methods of getting to the train station despite changing conditions (workday, strike, or holiday) as illustrated by the float times people allowed to catch a train. Analysis of variance showed no significant difference across days for leaving for the station earlier than necessary, arriving at the station early, or catching an earlier train.
Clearly, people who use the rail service, use it repeatedly. They may not feel like they have any real alternative, but they do continue to travel by rail. Table 4 shows that more than half the sample traveled at least once a week by train, with a further 17 percent traveling once a month. This experience of the rail service will effect information requirements.

Table 4. How Often Passengers Travel by Rail

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most days in the week</td>
<td>43</td>
</tr>
<tr>
<td>About once a week</td>
<td>17</td>
</tr>
<tr>
<td>Around once a month</td>
<td>17</td>
</tr>
<tr>
<td>A few times a year</td>
<td>20</td>
</tr>
<tr>
<td>Less than once a year</td>
<td>1</td>
</tr>
<tr>
<td>Unspecified</td>
<td>1</td>
</tr>
</tbody>
</table>

Part of people’s rail experience is how they have found the service to be in the past. Of the 550 respondents, 146 had previously made a written complaint about the rail service, covering 180 “offenses.” In 1998–99, there were 875,878,252 journeys nationally and 737,331 written complaints (ORR 1999). Based on this rate, 146 complaints represent 173,448 journeys, which explains why a researcher could not cover all the situations that resulted in complaints.

Table 5 demonstrates that it is the actual provision of rail services that generates the most complaints. It also shows that a substantial number of complaints concern information (or its lack) and occasions when there have been disruptions to the rail service and passengers felt they could have been better informed.

When there is a problem with a rail journey, a person’s prior experience of rail incidents can affect how they get information to recover the journey. Of those who travel by rail only a few times a year, 6 percent wanted to ask rail staff how to continue a journey that had been delayed. Less than 1 percent of more frequent travelers wanted this information. Commuter travelers seemed to know what their options were when they were on their daily trip; 49 percent of those who travel weekly wanted to know what their options were to recover a journey, but this fell to 19 percent for those traveling a few times a year. This group was more interested in getting clear information about the delay and in being told how to proceed (rather than choosing from options) than any other group.
People can only seek information when they know what is available. To understand this, passengers were asked how they got information to make the current journey. More than one information source was used, perhaps because people could not find the information they wanted from the initial source or because they wanted to double-check. The 550 survey respondents totaled 632 sources as shown in Table 6. The largest group relied on their own prior knowledge. Of the 27 percent who used a paper timetable, half did not carry one on the current journey, so 16 percent of travelers were using a paper timetable purely for pretrip information. From the survey, 69 percent could not give a reasonable approximation of the NRES number. A number of people indicated that they did not know about the NRES.

Table 6. How Passengers Received Information

<table>
<thead>
<tr>
<th>Information Source for This Journey</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior knowledge</td>
<td>36</td>
</tr>
<tr>
<td>Paper timetable</td>
<td>27</td>
</tr>
<tr>
<td>Telephone (NRES)</td>
<td>12</td>
</tr>
<tr>
<td>Didn’t use any</td>
<td>11</td>
</tr>
<tr>
<td>Internet</td>
<td>6</td>
</tr>
<tr>
<td>Information kiosk</td>
<td>3</td>
</tr>
<tr>
<td>Telephone (other)</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 5. Passenger Complaints

<table>
<thead>
<tr>
<th>Complaint Categories</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punctuality/Delay</td>
<td>27</td>
</tr>
<tr>
<td>Late train</td>
<td>22</td>
</tr>
<tr>
<td>Information</td>
<td>13</td>
</tr>
<tr>
<td>Overcrowding</td>
<td>8</td>
</tr>
<tr>
<td>Cost</td>
<td>7</td>
</tr>
<tr>
<td>Train cancelled</td>
<td>5</td>
</tr>
<tr>
<td>Train broke</td>
<td>4</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>14</td>
</tr>
</tbody>
</table>
Business travelers used the Internet more than any other group. Of 25 passengers who had an itinerary/Internet printout, 18 were on business. Fifty-three percent of all business travelers had a laptop, representing 68 percent of all laptops on the train.

Figures vary, but upwards of 25 percent of people in the United Kingdom have accessed the Internet (for Internet usage, see: British Broadcasting Corporation 2000; Richardson 2000; Allegra 2000). This number can only rise. The number who will use the Internet for pretrip travel information remains open to question. Of the 37 Internet users in the survey, all but 6 said they would use a computerized traveler information system. It is unclear why 6 people who already use the Internet would not use a computerized system, though one reason may be that their current experience has shown that while the Internet may provide reliable pretrip information, it is not capable of providing incident information. Real-time data is now available and could solve this problem. Another problem may be cost. If Internet users get free access at work, they may be less inclined to pay for Internet access themselves. Most households now have a telephone, half the U.K. population has a mobile telephone, yet only 13 percent of the sample used it to get information. Just because one has the equipment does not mean they will seek the information.

More than 70 percent of the respondents had either a mobile phone or a portable computer (14% carried a computer). Of these, 66 percent did not know the NRES number. Nearly half the sample carried a timetable booklet with them, although 31 percent had not used it to get information for the current journey. Presumably they know where they are going and carry the timetable out of habit or for emergencies.

Even under optimum conditions, useful information systems must present items wanted by passengers. Asked to pick the most important item from a list, passenger order of main preference for the current journey was first available train (30%), followed by shortest journey time (29%), lowest price (14%), and getting a seat (12%). For general rail travel, passengers prioritized their requirements as shortest journey time (41%), fewest connections (21%), lowest price (19%), and first available train (13%). These results show how important travelers view quickness of journey. However, enough respondents also requested getting a seat to have made this as desirable as catching the first available train. During an incident, 24 percent wanted updated arrival or departure information, 26 percent wanted to know the length of the delay, 13 percent wanted to know the cause, 3 percent wanted to
know how it was going to be fixed, while 21 percent did not want any information at all. They were content to wait and be informed.

Passengers were asked attitudinal questions about journey recovery situations. These questions were based on South West Trains internal Satisfaction Information System (Figure 2), which found staff was not seen as helpful, although the response of staff was thought to be clear.

Figure 2. SWT Satisfaction Information System
Figure 3 shows that survey respondents found staff as more helpful than clear. Travelers also felt that they had enough information to proceed with their journey, but they were less impressed with information available to overcome problems during a journey.

**Figure 3. Attitude to Problems**

When queried as to what they want in terms of improvements to the rail service, Figure 4 shows that the passengers’ highest requests were for more trains or seats, or using existing trains more effectively (e.g., by allowing passengers without seats to use unoccupied spaces in first class). Better quality information and advice, especially improved public announcement clarity, was the next mentioned item.
There were 12 mentions of real-time information. Some suggested it was not currently feasible, but most said it was desirable. These mentions of real-time information are valuable because they were spontaneous. People who do not know about the possibilities from information technology cannot suggest improvements, placing a higher value on unprompted recommendations.

Prompt arrival time was important for 62 percent of the respondents. Adverse consequences include missing events and connections, making a bad impression, having to work late, incurring a financial cost, stress, and having to wait. Despite the importance attached to these 183 references to arriving on time, people were not prepared to pay to obtain information that could overcome a problem with their journey. Figure 5 shows that there can be quite a delay before passengers feel the need to replan a journey.
Conclusions

The above analysis indicates that there is considerable good will toward the passenger rail industry. Passengers are habitual (nearly two thirds travel every week, half could drive). They prefer the human touch when it comes to information (79% would ask staff what to do if a train was cancelled), but they will use technology as well. The Train Operating Company (TOC) should ensure that its staff has the most accurate information possible.

When there is a problem, passengers will wait for delayed trains. They want to be kept informed, but most will remain with the rail network rather than seek an alternative mode. Passengers want to know what is going on and what the TOC is doing about it. They want to know the quickest way of getting to their destination, but in the event of an incident they are prepared to wait for the TOC to find a solution for them. TOCs should provide passengers with information that will enable them to continue with their journey by rail. This loyalty/apathy is particularly manifested in the way passengers think about punctuality. TOCs are obliged by the regulator to consider a train that arrives five minutes after the scheduled time as late. This is in contrast to the way car drivers think of punctuality or the way passengers think of lateness, which only occurs when they miss a connection or the start of a meeting. Compare this situation with a sports event in which the start may be delayed to allow spectators to get to their seats. This is considered desirable. Rail travel has not been allowed this sort of flexibility, even though passengers may consider it desirable.
During the fuel crisis of September 2000, while the rail regulator was praising the efforts of rail staff to get to work, the media reported trains failing to cope with the influx of people who were unfamiliar with the rail system (Strategic Rail Authority 2000). TOCs need to get better press. Advanced information promotes a feeling that the rail company is looking to make improvements and will also deliver network benefits in improved operation of the network, both for the TOC and the passengers. The introduction of pocket PCs will enable front-line staff to provide accurate timetable information. Some TOCs have begun loading this information onto the Internet. Furthermore, this information must be integrated. Passengers expressed an interest in a better modal interchange (see JourneyWeb 1999; PTI 2000).

Rail companies need to advertise information sources; 70 percent of passengers could not give an approximation of the NRES number. If the TOCs display the NRES number in every carriage, this would enable those phone carriers who do not recall the NRES number to use the service. It would also inform passengers who do not know about NRES that the service is available. More than a third of those surveyed by the Association of Transport Co-ordinating Officers, when prompted about the availability of information agreed that “if it were easier to get information about public transport services, I would use public transport more” (Association of Transport Co-ordinating Officers 2000). In the same survey, just 21 percent of respondents who would consider making a long-distance trip by train mentioned NRES as a source of information. None of those whose train was cancelled at the outset used NRES for recovery information despite nearly two-thirds of those who used NRES describing it as completely accurate (Harris Research Centre 1998).

It may seem obvious that passengers want modern carriages. Tables 1 and 6 show that passengers complain about the condition of the railway service (actual journeys, overcrowding, delays) more than they complain about information. Rail companies know this and are already expending most of their efforts in this area (Stagecoach has announced a £1.5 billion order for new trains that should generate an extra 70,000 peak-time seats per day. An additional £137 million will be spent refurbishing existing rolling stock.). However, TOCs may not appreciate the depth of feeling in this matter. Forty percent of the survey sample wanted to ensure they had a seat for their journey. Passengers want empty first-class carriages to be made available when there are no seats left in standard class.
There is a popular perception that rail (in the United Kingdom at least) is old, dirty, and inefficient (see Automobile Association 1997; delayed.net 1999). This view appears to be held most strongly by people who do not regularly travel by rail or who tried rail during the fuel crisis when the services were overstretched. It appears that those who regularly travel by rail are largely satisfied with the service. The upgraded rolling stock should at last be some good news to persuade car drivers that there is nothing wrong with the railways.

Acknowledgements

Many thanks are owed to all rail staff who allowed access to the railway for collecting surveys, in particular Elizabeth de Jong and her staff at South West Trains.
What's Wrong with the Railways?

References


**About the Author**

**GRAEME McLAY** (G.Mclay@soton.ac.uk) is a researcher in the Transportation Research Group at the University of Southampton. His doctoral thesis (2001) examined the way rail passengers use information to recover rail journeys. His previous academic work was in arts (philosophy and psychology) and information technology at the University of Glasgow.
The Effects of High-Speed Rail on the Reduction of Air Traffic Congestion

Andrés López-Pita and Francesc Robusté Anton
Center for Transportation Innovation
Technical University of Catalonia

Abstract

Commercial air services in Europe have experienced a spectacular growth in the last 15 years. From 1985 – 2000, the main European airports doubled their operations. Moreover, in the last 20 years, the number of regional airlines grew from 32 in 1980 to 78 in 1999.

This growth has resulted in a continuous increment of delays in air services. In 1986, 12.5 percent of air flights were delayed 15 or more minutes; the figure nearly doubled to 23 percent in 1998. In summer 1999, the average delay was more than 26 minutes. It is, therefore, not surprising that the congestion costs borne by European airlines have increased from $2,600 million in 1991 to $4,900 million in 1999.

This article presents results of studies undertaken for the Ministry of Public Works in Spain on the effect of new railway investments in reducing slot number needs at Madrid Airport. A total of 54,000 slots are compared to other studies carried out in France (40,000 slots in Paris-Charles de Gaulle) and Germany (20,000 slots in Frankfurt).
Introduction

The concept of competitiveness between different means of transport has, until very recently, been a constant factor in each mode’s historical development. So much so that requests for economic resources by the most relevant companies in each transport mode were usually made on the basis of the need to provide a higher level of service to enable them to compete with rival modes.

This way of thinking has declined in recent years because the substantial increases in movement require contributions from every mode if demand requirements are to be met effectively in both economical and environmental contexts.

Rail–air interaction probably best represents this change of approach. The introduction of the first high-speed, commercially operated service between Paris and Lyon was seen as railway’s response to the development of airline services. Today, however, the existence of the Roissy TGV station at Charles de Gaulle Airport has, together with the development of new rail infrastructure networks in France, enabled some airlines to establish collaboration agreements with French railways (Pavaux et al. 1991).

In a parallel fashion European aviation experienced a profound transformation during the last decade with the development of third-level air services into true regional services. The rapid progress experienced, with propeller planes quickly replaced by jet planes of similar capacity and undoubtedly more attractive to passengers, introduces a new variable to complementariness between the railway and airplane.

Political leaders are increasingly conscious of the need to encourage complementariness between transport modes. This is the framework for the trend observed toward converting major European airports into real centers for directing traffic onto the railways (Robusté et al. 1999). Strengthening this mode of transport not only contributes to making the transport system more fluid, but also reinforces environmental protection.

This article reviews the trends in passenger travel by both modes in the last few years. In addition, it explores possibilities for complementing air and railway services in the next few decades.
European Travel in the Last Three Decades

From the 1970s through the late 1990s, passenger traffic on all modes experienced an average annual increase of 2.8 percent. The distribution of this increase among different modes, however, was not uniform. Thus, while rail grew by 29 percent during this period, road traffic grew by 2.4 and air by 7.5.

It is not surprising, that, with respect to European medium- and long-distance passenger travel, the railway industry has a market share of 14 percent compared to 81 percent for the highway mode.

While it would take a great deal of time to analyse the causes of this distribution in detail, some data, like that outlined in Table 1, can help, at least partly, to explain the past experiences.

<table>
<thead>
<tr>
<th>Mode of Transport</th>
<th>Beginning of 1970s</th>
<th>End of 1990s</th>
<th>Rate of Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road</td>
<td>16,000 km of motorways</td>
<td>46,000 km of motorways</td>
<td>X 2.91</td>
</tr>
<tr>
<td>Plane</td>
<td>Beginning of jet planes DC-9 1966 B-737 1967 Mercure 1971 Airbus 1974</td>
<td>78 regional service companies (ERA) compared to 32 companies (ERA) in 1979</td>
<td>X 2.43</td>
</tr>
<tr>
<td>Railway</td>
<td>No new high-speed lines</td>
<td>2,000 km of new lines</td>
<td>—</td>
</tr>
</tbody>
</table>

In the case of the highway mode, motorways were first built during the beginning of the 20th century. In the early 1970s, the 15 European countries that currently form part of the European Union already had more than 16,000 km of roadways. This figure increased to 46,000 km by the end of the 1990s—a near threefold increase in the network for high-capacity, high-performance roads.

With respect to air transport, the first modern jet planes for medium distances appeared at the end of the 1960s, bringing improvements in air safety and comfort.
The French railway system, the most advanced in Europe, on the other hand, had scarcely more than 600 km of line capable of supporting maximum commercial operating speeds of 200 km/h at the beginning of the 1980s.

The comparative situation with motorway-type road infrastructure in 1995 for the four European countries where new high-quality railway infrastructure is being built (except for Belgium on account of its size) is shown in Table 2.

Table 2. High-Quality Road and Rail Infrastructure in Some European Countries (1995)

Source: Independently produced with EUROSTART data.

<table>
<thead>
<tr>
<th>Country</th>
<th>Length (km)</th>
<th>Motorways</th>
<th>High-speed Rail Lines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dual Carriageways</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>11,190</td>
<td>427</td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>6,962</td>
<td>471</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>8,275</td>
<td>1,185</td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>6,401</td>
<td>248</td>
<td></td>
</tr>
</tbody>
</table>

In the European transport system, railway distances for a given route are normally 30 to 60 km more than road, and as much as 176 km more when compared with air.

From the 1970s to late 1990s, there is no doubt that the differences between the transport modes increased significantly with respect to available resources for providing quality service. Advances made in the road network infrastructure and in air travel were unquestionable. As shown in Table 1, a major development occurred in regional air transport. This is further emphasized by the data in Table 3.

Table 3. European Regional Air Trasnport Evolution

<table>
<thead>
<tr>
<th>Variable</th>
<th>Time Period</th>
<th>Rate of Change 1988 to 1998</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air traffic (in millions)</td>
<td>1988 31.6</td>
<td>(x 2)</td>
</tr>
<tr>
<td>Average seat capacity</td>
<td>1988 35</td>
<td>(x 1.6)</td>
</tr>
<tr>
<td>Average distance covered (km)</td>
<td>1988 371 500</td>
<td>(x 1.34)</td>
</tr>
</tbody>
</table>

Source: Independently produced with Air Cosmos data.
Regional air passenger traffic actually doubled in six years, representing an average annual increase of 12 percent. The rapid use of jet planes for regional air transportation service in recent years has led to greater passenger comfort through flying at higher altitudes and reduced traveling time (Table 4).

**Table 4. Turboprops Versus Turbo Jets**

<table>
<thead>
<tr>
<th>Plane type</th>
<th>Cruising Speed</th>
<th>Altitude Operating Limit</th>
<th>Travel Time *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turboprop</td>
<td>550 km/h</td>
<td>5,200 m</td>
<td>1h 45</td>
</tr>
<tr>
<td>Turbojet</td>
<td>810 km/h</td>
<td>9,000 m</td>
<td>1h 15</td>
</tr>
</tbody>
</table>

*For same trip (Barcelona-Lyon)*

Source: Manufacturers’ figures.

Spain has been one of the countries in which regional air traffic has developed most rapidly. The Air Nostrum Company increased its fleet of planes from 6 in 1995 to 41 in 1999 (Figure 1). During this period the number of passengers carried increased from 260,000 to 1,800,000.

**Figure 1. Evolution of Air Nostrum’s Air Traffic Operations**

Source: Air Nostrum.
The major development experienced in road and air transport, added to the relative stagnation of the railway, gave rise to considerable saturation of the European transport system.

The need to increase the role of the railway arose as a natural way to face increased transport demand. This development, which constitutes one of the central pillars of the European Union's transport policy, is based on three factors:

1. Congestion problems now experienced by road and air transport
2. Reduced efficiency of both modes when they are used excessively
3. Environmental problems that would arise in the event of a generalized increase in transport capacity

Rail’s real possibilities of providing quality service with sufficient attraction for potential passengers have been confirmed with the 1981 introduction of the Paris-Lyon line.

**European Travel from 2000-2010 and Aviation Possibilities**

The trend observed over the last decade with respect to passenger movements in Europe will not change at the beginning of the new century. This is confirmed by predictions from the World Tourism Organisation, which, for the 2000–2010 period, show that the number of tourist trips will increase from 372 million to 476 million, an annual average growth of 5 percent.

The latest estimates carried out by Airbus (2000) for 1999–2019 indicate that world passenger traffic will increase by an annual accumulative average of 4.88 percent, being broken down according to corridors as indicated in Table 5. An annual average increase of 5.3 percent is forecasted for trips affecting European airspace.

**Table 5. Air Traffic Growth in Some Corridors (1999-2019)**

<table>
<thead>
<tr>
<th>Corridor</th>
<th>Annual Average Increase in Passenger Traffic</th>
<th>Traffic Growth in 2019 with respect to 1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe–United States</td>
<td>5.0 percent</td>
<td>X 2.65</td>
</tr>
<tr>
<td>Intra-European</td>
<td>5.3 percent</td>
<td>X 2.81</td>
</tr>
<tr>
<td>Europe–Asia</td>
<td>6.5 percent</td>
<td>X 3.52</td>
</tr>
<tr>
<td>Domestic Traffic</td>
<td>4.8 percent</td>
<td>X 2.55</td>
</tr>
<tr>
<td>Europe–Africa</td>
<td>4.2 percent</td>
<td>X 2.27</td>
</tr>
</tbody>
</table>

Source: Airbus. 2000.
In view of this duplication of traffic, what are the real possibilities for air travel during the next two decades? In spite of the efforts made by the airline industry, this mode faces obvious difficulties in meeting the increased passenger demand with the necessary levels of service—a demand that grew on some of the main European routes from 15 million to 34 million in the period 1986–1999.

Figure 2 presents the evolution of punctuality levels in European airspace from 1986–2000. The percentage of flights delayed between 1986 and 1989 is comparable to those between 1993–1998, with an improvement in the situation between 1989 and 1993.

**Figure 2. Evolution of Punctuality on European Airlines**

Source: Independently produced with Association of European Airlines (AEA) data.

The summer period undoubtedly represents the greatest time of difficulty for air management. With all things considered, though, the reality of the last three years, not affected by exogenous factors (1997, 1998, and 2000), is still worrying.

While traffic, expressed in terms of number of flights, increased by 12 percent, the number of delayed flights rose by 49 percent—increasing from 624,000 to 930,000. Finally, the average delay per flight also rose by 30 percent, from 20 to 26 minutes.
In addition to the negative influence the delays have had on actual passengers, they have resulted in unfavorable economic effects on the airlines. Studies carried out by the Association of European Airlines show that increased costs from airport infrastructure and air route (airways) congestion increased from 3000 MEuros in 1991 to 5700 MEuros in 1999 (Figure 3)—an annual average increase of 8 percent. For comparison sake, 5700 MEuros are equivalent to the cost of constructing the new high-speed line between Madrid and Barcelona (625 km approximately).

**Figure 3. Additional Costs Due to Air Congestion**

Source: Independently produced with Association of European Airlines (AEA) data.

Based on this information, it is not surprising that airlines are requesting the collaboration of railways to replace air services with railway services over distances in which land transport modes can ensure quality service. Against this backdrop, in 1996 Lufthansa requested that the railway replace the airplane for trips which could be made by the former in a two- to three-hour time interval.

**Medium- and Long-Distance Intra-European Travel**

Figure 4 shows existing air traffic between some of the major European cities (López-Pita 2000). A comparison of these routes with their anticipated travel time objectives by rail (Figure 5) shows how effective rail can be in decongesting air traffic (López-Pita 2000).
Figure 4. Passenger Air Traffic in Main European Corridors (1998) (in millions of passengers)

Figure 5. Estimated Travel Time in the European High-Speed Rail Network
For inland routes, rail offers travel times from city center to city center between 2.5 and 4 hours. This duration should lead, in accordance with previous experience, to a share market with respect for air of between 35 and 90 percent of the traffic total for both modes.

With respect to international routes, current available experience with high-speed trains and, in particular, with complete, newly built infrastructure, is limited to the Paris–Brussels line. There are, however, international services which partly run on newly built lines, as in the cases of the Paris–Amsterdam, Brussels–London, and Paris–London routes. Figure 7 summarizes some of the available information regarding rail–air modal distribution on these types of international routes. Market share for the origins–destinations considered is more than 45 percent, enabling us to deduce rail’s actual chances of capturing markets on certain international routes.

As indicated for the first high-speed line, even when airfare levels vary significantly from one European country to another, the comparison between rail fares and airfares (first and second class) is still very favorable for rail (Figure 6) (López-Pita 2000). This fact adds to the latter mode’s attractiveness.

**Figure 6. Comparison of Rail and Airfare Levels on Some European Routes (1998)**

*Source: A. López-Pita. 2000.*
It is reasonable to conclude that in the next few years high-speed rail could contribute very effectively to reducing congestion problems. In fact, preference for using high-speed rail on routes in which this mode offers a journey time of two to three hours, as opposed to air, has already been demonstrated. Consequently, certain airlines, such as Lufthansa, SABENA, KLM, Air France, and Iberia, have considered the possibility of withdrawing flights from these routes. This action would free a number of slots which could be used for introducing new long-distance services on routes in which air transport is irreplaceable.

Figure 8 shows the results obtained from various analyses carried out in this field. By way of illustration, if Lufthansa withdrew its current existing flights between Frankfurt, Dusseldorf, and Stuttgart, as well as other nearby destinations, 20,000 slots per year would become available. In the Benelux region (Amsterdam and Brussels), the withdrawal of short flights (d \( \geq 400 \) km) would free 20,000 slots per year. Paris-Roissy Airport estimates that 40,000 slots per year would open up by
replacing domestic flight services in France with high-speed railway services from this airport. Finally, Barajas Airport in Madrid estimates that the development of the high-speed railway network in Spain (which will guarantee journey times of three to four hours between all major Spanish cities and Madrid) will give rise to 54,000 slots per year.

**Figure 8. Estimated Number of Slots that Could Become Available at Certain Airports by Replacing Short, Middle-Distance Flights with High-Speed Railway Services**

*Source: Independently produced from different references.*

With the European network progressively extended and equipped with new infrastructure, rail’s international passenger service sector has started to take on a new dimension. Yet, this is undoubtedly only a prelude to what could happen in the near future (5 to 10 years) when the new lines programmed are physically implemented.
High-Speed Line Airport Connections

This section examines the current state and perspectives for high-speed rail services at airports as well as passenger flow connections with air services.

The Current State and Perspectives

Complementing air and high-speed rail services at airports began about 15 years ago in Europe. In October 1987, the French government built the Interconnection high-speed line, linking the TGV South-East, TGV Atlantic, and TGV North high-speed lines in the Paris area and serving Charles de Gaulle Airport. This rail–air link entered commercial service in November 1994.

The second airport connection for high-speed trains was established in July 1989 by the protocol signed by SNCF, the Rhône-Alpes region, and the Lyon Chamber of Commerce. This agreement established the financing system for a new TGV station at Lyon–Satolas Airport, which was to be opened in the second quarter of 1994.

A quick look at the physical location of the Roissy and Satolas (now Saint-Exupery) airports with respect to the high-speed routes, the Interconnection, and the Lyon–Valence line, leads one to think that the construction of the railway stations there was influenced by their relative proximity.

The novelty and importance of the subject caused the first high-speed congress, held in Brussels in 1992, to devote a section to railway complementariness with other modes of transport, particularly air travel. Executives from Paris and Frankfurt airports participated in the congress and emphasised the need for airport complementariness between both modes.

T. Norweg (1995) from Frankfurt Airport noted that many short-distance flights from the latter could not be justified beyond the pre- or post-channeling mission they carried out with respect to intercontinental flights because they were very uneconomical. He demonstrated very clearly that, in the case of the Cologne–Frankfurt flight, a Boeing 737 with approximately 100 seats would in theory require an occupation rate of 130 percent, based on existing fares, to cover costs.

From this perspective, and in collaboration with the German railways, the construction of a new high-speed railway station was planned. Land had been reserved and secured to carry out the arrangement.

New initiatives have taken hold since then. For example, the new Cologne–Frankfurt high-speed line has been in commercial service since August 2002. Two new
stations are planned at the Cologne–Bonn and Frankfurt airports. Over a longer period of time, similar development will take place at Stuttgart and Leipzig airports in Germany, as well as at Orly Airport in France. In Holland, the new high-speed line that will link Brussels with Amsterdam in 2005 will also pass through the airport in Schiphol.

Based on these examples, it appears that of all the European countries, Germany best reflects the desire to turn its main airports into true intermodal distribution centers (Grumbeier et al. 1998).

**High-Speed Passenger Flow Connections with Air Services**

Due to reasons linked to the temporary development of new railway infrastructure, available experience with respect to passenger traffic using airport railway stations as a complement between high-speed services and air services is limited to the French sector.

According to Aéroports de Paris (APD), 3.5 percent of CDG passengers arrive at Roissy Airport on high-speed trains. Given that air traffic at this airport amounted to almost 44 million passengers in 1999, it can be deduced that airport intermodality affected approximately 1.5 million passengers. By 2005, the APD predicts that this figure will rise between 5 and 10 percent. With an annual average growth rate of 6 percent in passenger traffic at Roissy Airport, passenger traffic gained by the train from the plane could be about 3 to 6 million passengers a year (Lebouef 2001).

APD executives believe a 5 percent increase in the number of passengers using high-speed air–rail services will be reached without great difficulty due to market development and, in particular, to agreements between French railways and certain airlines. Additional factors contributing to this growth include the new Thalys services and commencement of TGV Mediterranean operations to Nimes and Marseille. In regards to the Thalys services, in March 2001 Air France replaced five of its daily services between Paris and Brussels (approximately 150,000 passengers in 1999) with 1.25-hour Thalys services from Roissy Airport.

To achieve a 10 percent transfer in passengers, APD believes a suitable solution to baggage handling, enabling passengers to check in luggage as soon as possible, is necessary.

Rail–air intermodal passenger figures at Lyon–Saint Exupery Airport (approximately 150,000 passengers) are significantly less, primarily because of both lower air traffic and the lower number of TGV services (approximately 11a day) which stop at this
station. In June 2001 these services were increased to 19 a day with the commercial introduction of the TGV Mediterranean.

German transportation planners predict a 6 percent increase in passenger rail–air passenger traffic in Frankfurt. Since the introduction of new timetables in European air services (March 2000–October 2001), Lufthansa has improved its offerings between Stuttgart and Frankfurt with the introduction of ICE services. Passengers can leave their luggage in Stuttgart prior to catching the train, and pick it up at their final air destination.

KLM and Sabena will shortly replace their flights between Amsterdam and Brussels with rail services.

Conclusions
This article has illustrated the need for the European transport system to consider both rail and air travel from a complementariness perspective.

Increased travel in the coming years will merely support the fact that it is not enough to rely exclusively on road and air transport in responding to demand requirements. Evidence already exists to prove that:

1. Both modes are already experiencing saturation problems
2. Excessive use of these modes makes them less efficient
3. Their growth will not be realized in economic terms or, in particular, from an environmental perspective

Experience has also demonstrated high-speed rail’s great potential in easing saturation levels in the European transport system. The fact that high-speed rail’s market share in Europe on those interurban routes where it offers a quality service ranges from between 35 percent and 50 percent only serves to confirm this statement.
References


About the Authors

Andrés López-Pita (andres.lopez-pita@upc.es) is a lecturer in railway engineering and director of the Center for Transportation Innovation at the Technical University of Catalonia at Barcelona. He has several years of research experience in transportation planning, has been a member of the board of directors of Spanish railways. Mr. Lopez-Pita is the author of 220 publications.

Francesc Robusté Anton (r.robuste@upc.es) received his Ph.D. in engineering at the Technical University of Catalonia in Barcelona, and a master of science in operations research and master of engineering in transportation at the University of California at Berkeley. He is a lecturer in economic transportation and academic secretary of the Civil Engineering School at Barcelona. Dr. Robusté is the author of 146 publications and a member of the editorial board of Transportation Research.
Optimization of Bus Route Planning in Urban Commuter Networks

Steven I-Jy Chien, Branislav V. Dimitrijevic, and Lazar N. Spasovic
New Jersey Institute of Technology

Abstract

Bus routing is one of the most important elements of public transit system planning. This article presents a model for optimizing service headway and a bus route serving an area with a commuter (many-to-one) travel pattern. The bus route is optimized by minimizing the total system cost, including operator and user costs, while considering diagonal links in the study network. A method is developed for transforming this network into a pure grid, which enables construction of pure grid network models applicable to irregular grid networks. A case is presented to demonstrate the application of the model. Results show that the optimal bus route is sensitive to demand distribution over the service area. The developed model is particularly useful for planning a new bus service and evaluating an existing one in many cities embedded with general grid networks.

Introduction

Commuter bus routes are generally located on main thoroughfares of urban areas. However, considering realistic distributions of passenger travel demand over space and time, many route locations may not be cost-effective from either the operator or user standpoint. Therefore, relocating bus routes and redesigning headways may reduce operating costs as well as improve passenger accessibility.

Both transit operators and passengers prefer short and fast routes to reduce the operating cost and travel time, respectively. However, passengers also prefer bus routes that can be easily accessed from their origins and destinations. To reduce
access impedance, tortuous routes are often constructed. This, in turn, is likely to increase both the in-vehicle portion of user travel time as well as the bus operating cost. Transit operators are well aware of this trade-off when planning a new bus route or extending an existing service.

In the past 30 years, many researchers have analyzed the problems of optimal transit service design with many-to-one travel patterns by using analytical methods (Byrne and Vuchic 1971; Chang and Schonfeld 1991; Hurdle 1973; Spasovic and Schonfeld 1993; Spasovic et al. 1994; Wirasinghe et al. 1977). They dealt with selecting zones, route/line spacings, headways, and route lengths designed to carry people between distributed origins and a single destination (e.g., central business district [CBD], transfer station, etc.). By assuming demand homogeneity of the service area, the researchers optimized the characteristics of bus systems consisting of a set of parallel routes feeding a major transfer station of a trunk line or a single terminal point, such as the CBD.

A recent method for analyzing fixed-route bus systems is the out-of-direction (OOD) technique (Welch et al. 1991). This method improves the accessibility of a bus system by improving passenger accessibility along certain route segments. Chien and Schonfeld (1997) optimize a grid transit system in an urban area without oversimplifying the spatial and demand characteristics. They extended the model to jointly optimize the characteristics of a rail transit route and the associated feeder bus routes in an urban corridor (Chien and Schonfeld 1998).

Chien and Yang (2000) developed an algorithm to search for the best bus route feeding a major intermodal station while considering the intersection delays and realistic street network. The model optimized the bus route location and operating headway by minimizing the sum of operator and user costs. It considered irregular and discrete demand realistically distributed over the service area. The route and headway were optimized analytically.

In marked contrast to the above research, this article deals with the irregular grid street network, including diagonal streets and heterogeneous demand over the service area. To formulate bus routing problems, diagonal streets are transformed into horizontal and vertical links so the grid structure of the network could be preserved to facilitate the computational process. Actual lengths of diagonal links are taken into account when calculating the route length and travel times. The bus route, headway, and fleet size are optimized by minimizing the total cost (the sum
of operator and user costs). A computer program is developed to search for the optimal solution.

**Assumptions**

The commuter network discussed in this article is a general grid network with some diagonal links. To formulate the mathematical optimization model for such a network, the following assumptions are made:

- The irregularly shaped service area can be divided into many zones according to the street spacing and demand distribution.
- A feeder bus route provides service between the suburban area and the CBD (Figure 1a). Thus, the travel demand pattern of the area is many-to-one in the morning peak period and one-to-many during the afternoon peak period.
- A line-haul distance \( J \), connecting the CBD (or a major transfer station) and the service area at an entry point, is assumed to be constant.
- The demand is not sensitive to bus service quality or fare and is uniformly distributed within each zone. The zones may have different associated cost, demand, land-use, and traffic characteristics.
- Buses can stop anywhere along the route whenever a boarding or an alighting is requested by a passenger. Thus, the bus stop location can be ignored.
- Passengers access the route randomly, while the headway is short enough to assume that average wait time is half of the headway.
- The value of time is assumed to be additive. This assumption can be relaxed as long as the function of time value can be developed.
- The intersection (or node) delay incurred by bus is constant regardless of the bus size but may vary at different intersections.
- Vehicle layover time is negligible.
There are two types of links classified in the study network: real links and dummy links. Real links represent actual streets with the network. Dummy links do not exist in reality but are included to normalize the grid network structure. The length of dummy links is assumed to be infinite. Thus, in a minimization problem it will generate an infinite penalty for vehicles traveling through these links.

A diagonal link makes “triangular areas” with adjacent streets as shown in Figure 1a (shaded triangle A-C-D). The diagonal links are converted into horizontal and vertical links, as shown in Figure 1b. Diagonal links (AD and DG) have been replaced by horizontal links (XY and MN), while the link length remains the same. When calculating the total route length, the distance between the previous node (e.g., A) and the incident node of the horizontal link (e.g., X) is equal to zero (AX = 0). The same situation holds for link YD (YD = 0). However, the route that includes link XY must also contain links AX and YD. The length of link XC is assumed to be equal to the actual length of the vertical link originally connecting nodes A and C. The same situation holds for link BY (which has the length of the vertical link originally connecting nodes B and D).

As part of the transformation, new links marked as “dummy links” are introduced as extensions of links XY and MN (e.g., L1, L2, L3, …, L10, etc.) to preserve the grid network structure. Since new nodes are introduced in this transformation, intersection delay times for these nodes must also be defined. The intersection delay for node X is equal to the intersection delay of node A. In the same manner, the
intersection delays of nodes Y and M are equal to the intersection delay of node D, while the intersection delay of node N is equal to that of node G. To avoid double counting the intersection delay (e.g., Y, D, and M), only intersection delay at the immediate downstream node of a link is considered when the route travel time is calculated. Delays at intersections of dummy links are set to be zero.

Model Formulation

The mathematical notation and the definition of variables and parameters used in this section are summarized in Table 1. The network, shown in Figure 1a, is divided into \( m \) rows and \( n \) columns, containing \( m \times n \) zones (Figure 1b). The location of each zone is defined by indices of rows and columns. Demand of a zone, denoted by \( q_{ij} \) \((1 \leq i \leq m, 1 \leq j \leq n)\), is defined by the number of passengers of a particular zone.
Table 1. Variable and Parameter Definition

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{ij}$</td>
<td>Total access time for passengers originating from zone $(i,j)$</td>
<td>hours</td>
</tr>
<tr>
<td>$A^X_{ ij}$</td>
<td>Network horizontal links matrix</td>
<td>—</td>
</tr>
<tr>
<td>$A^Y_{ ij}$</td>
<td>Network vertical links matrix</td>
<td>—</td>
</tr>
<tr>
<td>$B_U$</td>
<td>Network node matrix</td>
<td>—</td>
</tr>
<tr>
<td>$C_A$</td>
<td>Total passenger access cost</td>
<td>$$/hour</td>
</tr>
<tr>
<td>$C_S$</td>
<td>Total supplier cost</td>
<td>$$/hour</td>
</tr>
<tr>
<td>$C_T$</td>
<td>Total system cost</td>
<td>$$/hour</td>
</tr>
<tr>
<td>$C_U$</td>
<td>Total user cost</td>
<td>$$/hour</td>
</tr>
<tr>
<td>$C_V$</td>
<td>Total in-vehicle cost</td>
<td>$$/hour</td>
</tr>
<tr>
<td>$C_W$</td>
<td>Total wait cost</td>
<td>$$/hour</td>
</tr>
<tr>
<td>$D_{ij}$</td>
<td>Minimum access distance for passengers from zone $(i,j)$</td>
<td>km</td>
</tr>
<tr>
<td>$F$</td>
<td>Fleet size</td>
<td>buses</td>
</tr>
<tr>
<td>$g$</td>
<td>Passenger walking speed</td>
<td>km/hour</td>
</tr>
<tr>
<td>$H_B$</td>
<td>Headway</td>
<td>hours</td>
</tr>
<tr>
<td>$L_J$</td>
<td>Line haul distance</td>
<td>km</td>
</tr>
<tr>
<td>$d$</td>
<td>Length of the horizontal link that replaced the diagonal link</td>
<td>km</td>
</tr>
<tr>
<td>$\ell$</td>
<td>Length of the vertical link modified after diagonal link is replaced</td>
<td>km</td>
</tr>
<tr>
<td>$m$</td>
<td>Number of rows in the network</td>
<td>—</td>
</tr>
<tr>
<td>$n$</td>
<td>Number of columns in the network</td>
<td>—</td>
</tr>
<tr>
<td>$P^x_{ij}$</td>
<td>Penalty matrix for horizontal links</td>
<td>—</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
<td>Unit</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>------</td>
</tr>
<tr>
<td>( P_y )</td>
<td>Penalty matrix for vertical links</td>
<td>—</td>
</tr>
<tr>
<td>( Q )</td>
<td>Total passenger demand</td>
<td>pass/hour</td>
</tr>
<tr>
<td>( q_{ij} )</td>
<td>Passenger demand of the zone ((i,j))</td>
<td>pass/hour</td>
</tr>
<tr>
<td>( S_D )</td>
<td>Total intersection delay incurred by passengers</td>
<td>hours</td>
</tr>
<tr>
<td>( S_J )</td>
<td>Total line haul in-vehicle time</td>
<td>hours</td>
</tr>
<tr>
<td>( S_L )</td>
<td>Total route link in-vehicle time</td>
<td>hours</td>
</tr>
<tr>
<td>( S_T )</td>
<td>Total in-vehicle time</td>
<td>hours</td>
</tr>
<tr>
<td>( T_D )</td>
<td>Total route intersection delay</td>
<td>hours</td>
</tr>
<tr>
<td>( T_{ij} )</td>
<td>Intersection delay matrix</td>
<td>hours</td>
</tr>
<tr>
<td>( T_I )</td>
<td>Line haul travel time</td>
<td>hours</td>
</tr>
<tr>
<td>( T_R )</td>
<td>Total route link travel time</td>
<td>hours</td>
</tr>
<tr>
<td>( T_B )</td>
<td>Round trip travel time</td>
<td>hours</td>
</tr>
<tr>
<td>( u_A )</td>
<td>Value of passenger access time</td>
<td>$/hour</td>
</tr>
<tr>
<td>( u_B )</td>
<td>Average bus operating cost</td>
<td>$/hour</td>
</tr>
<tr>
<td>( u_I )</td>
<td>Value of passenger in-vehicle time</td>
<td>$/hour</td>
</tr>
<tr>
<td>( u_W )</td>
<td>Value of passenger wait time</td>
<td>$/hour</td>
</tr>
<tr>
<td>( V_B )</td>
<td>Average bus operating speed in the service area</td>
<td>km/hour</td>
</tr>
<tr>
<td>( V_I )</td>
<td>Line haul speed</td>
<td>km/hour</td>
</tr>
<tr>
<td>( X_y )</td>
<td>Matrix of lengths of horizontal links in the network</td>
<td>km</td>
</tr>
<tr>
<td>( X_N )</td>
<td>Width of the row (j) of the network</td>
<td>km</td>
</tr>
<tr>
<td>( Y_y )</td>
<td>Matrix of lengths of vertical links in the network</td>
<td>km</td>
</tr>
<tr>
<td>( Y_N )</td>
<td>Width of the column (i) of the network</td>
<td>km</td>
</tr>
</tbody>
</table>
The following matrices define horizontal and vertical links, and network nodes.

**Horizontal Links**

(a) \[ A^X_{ij} \text{ for } 1 \leq i \leq (m + 1) \text{ and } 1 \leq j \leq n \text{ represents matrix of horizontal links} \]

\[ A^X_{ij} = \begin{cases} 1, & \text{if the link connecting nodes } (i, j) \text{ and } (i, j+1) \text{ is part of the bus route;} \\ 0, & \text{otherwise.} \end{cases} \]

(b) \[ X^X_{ij} \text{ for } 1 \leq i \leq (m + 1) \text{ and } 1 \leq j \leq n \text{ represents matrix of horizontal link lengths} \]

\[ X^X_{ij} = \begin{cases} X^N_j, & \text{if horizontal link } (i, j) \text{ is not transformed diagonal link; length of the diagonal link, if horizontal link } (i, j) \text{ is transformed diagonal link.} \end{cases} \]

where:

\[ P^N_{ij} \text{ takes the value of } 1 \text{ or } +\infty \text{ depending on whether the horizontal link } (i + j) \text{ is a real or dummy link.} \]

\[ X^N_j \text{ represents the width of the column } j. \]

(c) \[ P^X_{ij} \text{ for } 1 \leq i \leq (m + 1) \text{ and } 1 \leq j \leq n \text{ represents penalty matrix for horizontal links} \]

\[ P^X_{ij} = \begin{cases} +\infty, & \text{if the link connecting nodes } (i, j) \text{ and } (i, j+1) \text{ is a dummy link;} \\ 0, & \text{otherwise} \end{cases} \]
**Vertical Links**

(a) $A_y^v$ for $1 \leq i \leq m$ and $1 \leq j \leq (n + 1)$ represents matrix of vertical links of the network

$$A_y^v = \begin{cases} 
1, & \text{if the vertical link connecting nodes } (i, j) \text{ and } (i + 1, j) \text{ is a part of the bus route;} \\
0, & \text{otherwise.}
\end{cases}$$

(b) $Y_{ij}^v$ for $1 \leq i \leq m$ and $1 \leq j \leq (n + 1)$ represents matrix of lengths of vertical links

$$Y_{ij}^v = \begin{cases} 
Y_i^N \cdot P_y^v, & \text{if row } i \text{ of network is not transformed after transforming diagonal link into horizontal link;} \\
\ell, & \text{otherwise.}
\end{cases}$$

where:

- $P_y^v$ takes the value of 1 or $+\infty$ depending on whether the vertical link $(i, j)$ is a real or dummy link.
- $Y_i^N$ represents height of the row $i$.
- $\ell$ is the length of a vertical link that was introduced after transformation of the diagonal link.

As discussed in the previous section, each diagonal link is transformed into a horizontal link, and the row where the diagonal link was is split into two geometrically equal rows. There are three types of vertical links in the two new rows that replaced the original row:

i. A vertical link that represents a point where a diagonal link is beginning or ending (e.g., links AX and YD in transformed network shown in Figure 1b represent points A and D in original network shown in Figure 1a). Value of $\ell$ for these links is equal to 0.
ii. A new vertical link that represents the original vertical link leading to/from diagonal link (e.g., links XC and BY in Figure 1b represent links AC and BD in the original network in Figure 1a). Value of \( e \) for these links will be equal to original height of the initial row containing the diagonal link.

iii. A vertical link that is not connected with the diagonal link; but is in the transformed row. Value of \( e \) for such vertical links is equal to half of the original row height if the vertical link was a real link, or equal to \( +\infty \) if the vertical link was a dummy link (e.g., link PQ in the Figure 1b is equal to one half of the link PR in Figure 1a).

\[
P^V_{ij} \text{ for } 1 \leq i \leq m \text{ and } 1 \leq j \leq (n + 1) \text{ represents penalty matrix for vertical links}
\]

\[
P^V_{ij} = \begin{cases} +\infty, & \text{if the vertical link connecting nodes } (i, j) \text{ and } (i + 1, j) \\ 1, & \text{is a dummy link;} \\ 0, & \text{otherwise.} \end{cases}
\]

**Nodes**

\[
B_{ij} \text{ for } 1 \leq i \leq (m + 1) \text{ and } 1 \leq j \leq (n + 1) \text{ represents matrix of network nodes}
\]

\[
B_{ij} = \begin{cases} 1, & \text{if the node } (i, j) \text{ is in the bus route;} \\ 0, & \text{otherwise.} \end{cases}
\]

\[
T_{ij} \text{ for } 1 \leq i \leq (m + 1) \text{ and } 1 \leq j \leq (n + 1) \text{ represents matrix of intersection delay times for node } (i, j)
\]
Since the passenger demand of zone \((i, j)\), denoted by \(q_{ij}\), to the CBD is known, the total passenger demand \((Q)\) can be calculated as:

\[
Q = \sum_{i=1}^{m} \sum_{j=1}^{n} q_{ij} \quad \text{for } m \geq 1 \text{ and } n \geq 1
\]  

(1)

The objective function of the analyzed bus routing problem is the total system cost \((C_r)\), including operator cost \((C_s)\) and user cost \((C_u)\). Thus,

\[
C_r = C_s + C_u
\]  

(2)

The operator cost, in dollars per hour, is equal to the fleet size \((F)\) multiplied by the bus operating cost \((u_B)\):

\[
C_s = F \cdot u_B\]

(3)

The bus operating cost can be estimated from the average wage rate labor, insurance, and maintenance expenses. Required fleet size can be estimated from the vehicle round-trip time \((T_R)\) divided by the headway \((H_B)\). Thus,

\[
F = \frac{T_R}{H_B}
\]  

(4)

In Equation (4) the round trip travel time \((T_R)\) is double the sum of bus route travel time \((T_L)\), total route intersection delay \((T_D)\), and line-haul travel time \((T_J)\), while the vehicle layover time is negligible here. Therefore,

\[
T_R = 2(T_L + T_D + T_J)
\]  

(5)
The total local route travel time \( T_L \) is defined as:

\[
T_L = \frac{1}{V_B} \left[ \sum_{j=1}^{m} \sum_{j=1}^{n} A^Y_{ij} Y_{ij} + \sum_{j=1}^{m} \sum_{j=1}^{n} A^X_{ij} X_{ij} \right], \quad \text{for } m \geq 1 \text{ and } n \geq 1
\]  

where:

\[ V_B \] is the average bus operating speed on local streets.

The average intersection delay time \( T_{ij} \) incurred by buses can be estimated from the field data. Thus, the total intersection delay \( T_D \) per bus trip can be formulated as:

\[
T_D = \sum_{j=1}^{m} \sum_{j=1}^{n} B_{ij} \cdot T_{ij}
\]  

The bus line-haul travel time, denoted by \( T_j \), is equal to the line-haul distance \( L_j \) divided by the line-haul speed \( V_j \):

\[
T_j = \frac{L_j}{V_j}
\]  

The user cost considered in this study consists of three elements: user access cost \( C_A \), user wait cost \( C_W \), and user in-vehicle cost \( C_V \):

\[
C_U = C_A + C_W + C_V
\]
User access cost $C_A$, incurred by passengers walking to the bus route, is defined as the product of user access time for each zone $(i, j)$, denoted by $a_{ij}$, and user access cost $u_A$ (i.e., the value of access time):

$$C_A = u_A \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \quad \text{for } m \geq 1 \text{ and } n \geq 1$$

It is assumed that the passengers from zone—always walk the shortest distance—to access the bus route. The minimum distance between zone—and the access point can be estimated from the sum of horizontal and vertical distances between the gravity point of zone—and the access point. Thus, the total access time for the passengers from the zone—is:

$$a_{ij} = \frac{D_{ij} q_{ij}}{g}$$

where:

$$g \quad \text{denotes average passenger walking speed.}$$

To estimate the user in-vehicle time, denoted by $C_V$, it is necessary to determine where passengers access the bus and how much time they spend on the bus. The first step is to identify the access point for passengers from each zone. This can be done using the shortest distance rule for calculating passenger access time. Passenger in-vehicle time is defined as the average time each passenger spends in the bus. This time is equal to total travel time including link travel times and intersection delays along the bus route. Therefore, in-vehicle time for each passenger is equal to travel time between the access point and the destination. All passengers from a particular zone will incur the same average in-vehicle time since they have the same access and destination points. Thus, total in-vehicle time for passengers from one zone is equal to the product of passenger demand for that zone and average in-vehicle time for passengers from that zone. Total passenger in-vehicle time, denoted by $S_V$, is simply the sum of in-vehicle times for all the zones in the
network. And the user in-vehicle cost is equal to the product of total in-vehicle time $S_V$ and value of user in-vehicle time $u_I$:

$$C_V = S_V \ u_I$$

(12)

User wait cost, the last component of total user cost, is subject to the assumption that average wait time is half of the headway. The wait cost can be formulated as:

$$C_W = \frac{1}{2} H_B u_W \sum_{i=1}^{m} \sum_{j=1}^{n} q_{ij}$$

for $m \geq 1$ and $n \geq 1$

(13)

where:

$$\frac{1}{2} H_B$$

is the average user wait time (half of headway).

$$\sum_{i=1}^{m} \sum_{j=1}^{n} q_{ij}$$

equals total passenger boarding.

$$u_W$$

the value of passenger wait time.

**Total System Cost - $C_T$**

The objective of this study is to minimize the total system cost. The objective total cost function is:

$$C_T = C_S + C_A + C_V + C_W$$

(14)

where:

$C_S$, $C_A$, $C_V$, and $C_W$ can be obtained from Equations (3), (10), (12) and (13), respectively.
Since the decision variables in $C_T$ include $A^T_{ij}$, $A^X_{ij}$, $B_{ij}$, and $H_B$, the total cost function can be expressed as:

$$C_T(A^T_{ij}, A^X_{ij}, B_{ij}, H_B)$$

(15)

for $1 \leq i \leq m$, $1 \leq j \leq n$, $1 \leq I \leq (m + 1)$, $1 \leq J \leq (n + 1)$.

The optimal bus headway can be optimized if $A^T_{ij}$, $A^X_{ij}$, and $B_{ij}$ (elements of bus route location) are treated as exogenous variables. The optimal headway is found by setting the first derivative of the total cost function with respect to headway ($H_B$) equal to zero, as formulated in Equation (16), and solving it.

$$\frac{\partial C_T}{\partial H_B} = 0$$

(16)

Thus, the optimal bus headway $H_B$ is derived as:

$$H_B = \frac{2T_R u_B}{\sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} q_{ij} u_w}}$$

for $m \geq 1$ and $n \geq 1$

(17)

Since all variables are nonnegative, the second derivative of the total cost function with respect to $H_B$ is always positive. The objective function $C_T(A^T_{ij}, A^X_{ij}, B_{ij}, H_B)$ is thus convex, and a unique optimal headway exists for any given matrices $A^T_{ij}$, $A^X_{ij}$, and $B_{ij}$. Therefore, the minimum total system cost can be obtained by substituting the optimal headway into Equation 15.

The optimal headway must meet the route capacity constraint that states that the total route capacity should satisfy (i.e., be at least equal to) the peak-passenger demand. Thus,

$$\frac{1}{H_B} \cdot C \geq Q,$$

(18)
where:

\[ C \] represents vehicle capacity.

The value of \( H_B \) must satisfy the following condition:

\[ H_B \leq \frac{C}{Q} \]  \hspace{1cm} (19)

Therefore, the headway will decrease if the demand \( Q \) increases. Since the reduced headway may increase the fleet size, as well as the operator cost, another approach to total cost optimization would be to change the bus size. This, in turn, may cause the change in the \( u_B \), while the headway \( H_B \) and the bus route configuration need to be reoptimized.

**Solution Method**

The first step in finding the optimal solution is to identify all candidate bus routes. The route location depends on the shape and size of the street network, but even for relatively small networks it is necessary to develop an algorithm and computer program to compute the optimal solution. The *Exhaustive Search* (ES) algorithm [11] is applied to determine the optimal bus route location. The algorithm defines all possible bus routes in the network by altering matrices \( A_I^*, A_B^*, B_{IJ} \); optimizes headway; calculates the total cost of each candidate route; and then finally selects the optimal solution with the minimum total cost. Routes containing at least one dummy link are not considered candidate routes. A computer program is developed to process the network geometry and demand data, apply the ES algorithm, and generate the optimal solution. The program used network geometry data that can be extracted from the existing Geographic Information System (GIS) database. The program can be modified to directly retrieve input data about streets (links), intersections (nodes), and zones (street blocks) from the GIS database. This would automate the computational process, and at the same time, existing geographic data would be used as an input of the model. Another improvement of the code could, after the optimization is completed, enable the optimal solution to be uploaded into the GIS database, and then the optimal bus route can be displayed graphically.
An Example
To illustrate the application of the model developed in this study, a numerical example is designed based on the street network with given geometric and demographic data over the service region. The analyzed network is similar to many urban commuter networks (e.g., Washington, D.C.; Chicago; Houston; and Manhattan Island in New York City, etc., as shown in Figure 2) where this model can be applied if real-world data are available. These networks usually consist of rectangular zones with diagonal thoroughfares. As mentioned before, the developed model can become even more efficient if it is integrated with the GIS database.

Figure 2. Examples of the Grid Street Networks with Diagonal Links: (a) Chicago, IL; (b) Washington, DC; (c) Manhattan Island, New York City
To demonstrate the performance of the developed model, a hypothetical example is designed. The analyzed service area with an underlying street network is shown in Figure 3a. The width of the service area is 1.57 km, while the length is 2.24 km. The network consists of \( m \) (8) rows and \( n \) (7) columns making 56 zones. Passenger demand for each zone is given and varies from one zone to another. Total demand for the service region is 133 passengers per hour. Buses enter the network from the left (west) and move eastward on the way to the CBD. The buses operate as a local service over the service area, and express service (no stops) is assumed from the end of the service area to the CBD. Total length of the express leg is 6.43 km.

![Figure 3a: Case Study Network (in millions)](image)

There are three diagonal links in the network. These links are transformed into horizontal and vertical links to preserve the pure grid structure (Figure 3b; the gray-colored lines showing diagonal links are not part of the modified network). The lengths of the diagonal links going from the northwest to the southeast corner are 0.43 km, 0.47 km, and 0.43 km, respectively. The delays at intersections are given and vary between 30 and 45 seconds per vehicle. All the other parameters of the network are given in Table 2.
After the ES algorithm is performed, the objective total cost function is minimized. The optimal bus route is denoted as Route A and shown in Figure 4a. The total route length is found to be 2.5 km long with the optimal headway of 14.2 minutes. The one-way travel time per trip including the express line-haul portion is 15.5 minutes and the minimum total cost is $486/hour.
The fleet size necessary to operate on the optimal route with the optimal headway is 2.2 buses. Since in reality the number of buses must be an integer, the fleet size is arbitrarily rounded to two and three buses. The final headway is recalculated based on the new fleet sizes (2 or 3 buses). The fleet with three buses can be operated at the headway of 10.6 minutes achieving the total cost of $499/hour. On the other hand, the fleet with two buses can be operated at the headway of 15.9 minutes and requires $488/hour. The optimal fleet size of two buses has been chosen.

**Sensitivity Analysis**

A sensitivity analysis is performed to investigate how the model reacts to variations in the values of different parameters. Three parameters are analyzed:

1. **Bus size:** Three bus sizes are considered: 35, 50, and 70 passengers per vehicle, respectively. The average hourly bus operating costs vary with bus size.

2. **Demand:** For each zone in the network, the demand ranges from 70 percent to 150 percent of its original value.

3. **Value of passenger time:** Passenger wait time and access time are assumed identical and their values vary from $10 to $15/passenger-hour, while the passenger in-vehicle time ranges from $4 to $10/passenger-hour. For this calculation, the value of in-vehicle time is $5 less than the value of wait time.
The bus route location does not change with a variation in the value of passenger time. The same holds when the demand is increased by 50 percent. However, if the demand is increased by more than 50 percent, the configuration of the bus route changes. The new bus route is shown in Figure 4b and labeled as Route B. Route B is 2.93 km long, operates at the optimal headway of 10.2 minutes and has a one-way trip time, including the line haul, of 17.4 minutes. Total cost is $837/hour. This change of route configuration indicates that the model is sensitive to variations in demand. Route B is longer than Route A, and this increase resulted in an increase in in-vehicle cost. However, the reduction in access cost caused by the route relocation offsets this increase in in-vehicle cost.

Figure 4b: Optimal Bus Route B

The impact of the change in headway on user, operator, and total costs is also analyzed (Figure 5). Short headway resulting in high operator costs (due to large fleet size required) reduces user costs because of less waiting time. The optimal headway is reached in point B (14.2-minute headway), at which the minimum total cost is achieved, while the operator and user costs are $159/hour and $327/hour, respectively.

Figure 6 shows the relationship between demand and optimal headway. For various bus sizes, the optimal headway decreases as the demand increases. Analysis results show that regardless of the variation in demand or in the value of passenger
time, the 35-passenger-per-bus vehicle size is the most preferable as it yields the minimum total cost. Figure 7 shows that even change of headway does not change the optimality of using smaller buses to serve the analyzed region. In addition, Figure 8 shows that the increase in value of passenger time results in an increase in user cost. Thus, the optimal headway decreases.
Figure 6: Optimal Headway Versus Demand

Figure 7: Total Cost Versus Headway for Various Bus Sizes
Conclusions

This article has presented a model for optimizing bus route in an urban commuter network, while considering a more realistic street pattern and demand distribution. The network transformation procedure developed in this study facilitates the models dealing with grid networks to be applied to irregular grid networks.

The model enables transit operators to optimize bus route and headway while enhancing efficient fleet management. Results derived from the model are easy to interpret and support effective decision making. Through sensitivity analysis different demand and supply conditions were evaluated by varying the optimal bus route so that it is sensitive to demand distributions, as well as value of time. Thus, the model and developed computer program are able to efficiently search for optimal solution in various conditions. The model can be easily modified to account for changes of spatial (e.g., one-way street, roadway/lane closure, reversible lane) and temporal (e.g., incidents, special events) conditions. All these features enhance transit planners’ capability to redesign bus routes in areas that may experience significant shifts in residential density, as well as geographic or physical changes of the street network.
The total cost minimization model proposed in this article may be used iteratively with a demand reestimation model to ensure that the bus system is optimized for equilibrium demand. A model that analytically integrates supply system optimization with the demand equilibration approach (e.g., as in Chang and Schonfeld [1993] or Kocur and Hendrickson [1992]) would be a desirable extension to the proposed model. It can also be further improved by introducing multiple bus routes to optimization procedure, which would make the model even more realistic and applicable.

A very important extension of this research is design of an interface with GIS databases containing real-world information on street geometry, demography, and traffic conditions in the studied service area. This information includes average vehicle operating speeds in different zones, street patterns, rights-of-way, street directionality, and time and spatial distribution of passenger demand. The GIS interface would enable data exchange between model and GIS: input data would be automatically retrieved from the GIS database, and the solution would be geocoded, loaded back into database, and displayed using graphic capabilities of the GIS software. This would improve analysis and let the user easily browse through different scenarios and corresponding solutions and to compare them.

**Acknowledgments**

This research was partially supported by a grant from the U.S. Department of Transportation, University Transportation Centers Program, through the National Center for Transportation and Industrial Productivity (NCTIP) at New Jersey Institute of Technology (NJIT). This support is gratefully acknowledged but implies no endorsement of the conclusions by these organizations.
References


About the Authors

Steven I-Jy Chien (chien@adm.njit.edu) is an associate professor of civil engineering and has a joint appointment with the interdisciplinary program in transportation at NJIT. Dr. Chien earned his master’s and Ph.D. degrees from the University of Maryland at College Park.

His teaching interests are in public transportation systems, transportation planning, traffic simulation, and urban systems engineering. His research interests are focused on intermodal transit systems, microscopic traffic simulation, transportation systems analysis, intelligent transportation systems, and application of artificial intelligence in transportation problems.

Branislav V. Dimitrijevic (bxd1947@njit.edu) is a research assistant and Ph.D. candidate in the interdisciplinary program in transportation at NJIT. He earned his master’s degree from NJIT and his bachelor’s degree from the University of Belgrade, Yugoslavia. His current research areas are network optimization, multimodal freight transportation, and transportation systems analysis.

Lazar N. Spasovic (spasovic@njit.edu) is a professor of transportation and management at NJIT and director of the National Center for Transportation and Industrial Productivity (NCTIP) and the International Intermodal Transportation Center (IITC). He is codirector of the Brownfields Economic Redevelopment initiative.

A recognized expert in intermodal freight transportation and transportation system analysis, Dr. Spasovic holds a Ph.D. degree in systems engineering from the University of Pennsylvania. He teaches multimodal freight transportation and urban transportation planning.
A Multiobjective Optimization Model for Flexroute Transit Service Design

Brian L. Smith and Michael J. Demetsky
University of Virginia

Priya K. Durvasula,
ConvaTec

Abstract

Flexroute transit (also referred to as route-deviation) is an innovative public transportation approach in which service is provided at fixed stops on a predetermined schedule, while also providing “on-demand” service to customers between the fixed stops. A significant challenge to the implementation of flexroute transit has been the complexity inherent in the design process. This research explored two of the key design parameters available to a transit planner: service zone size (the area between fixed stops where deviations are permitted), and slack time distribution (the method used to distribute among zones the total “slack” time built into the schedule to allow for deviations). It was found that the objectives of maximizing feasible deviations and minimizing unused slack time are conflicting and highly sensitive to the design parameters. To assist in the design process, the article formulates a multiobjective binary optimization model that can be used to develop an efficient frontier—guiding transit planners as they seek to explore design trade-offs.
Introduction

To address challenges resulting from such factors as low-density development and compliance with the Americans with Disabilities Act (ADA), many public transportation providers have turned to nontraditional service options such as flexroute transit (Rosenbloom 1996). In a flexroute system (also referred to as route-deviation), service is provided at fixed stops on a predetermined schedule, while also providing “on-demand” service to customers between the fixed stops. As seen in Figure 1, these characteristics suggest that flexroute transit is a hybrid of conventional fixed-route and paratransit services.

Figure 1. Comparison of Flexroute, Paratransit and Fixed-Route Services

While there is significant interest in flexroute and other forms of demand responsive transit, transit providers are impeded by a relative lack of literature on the effective design and operation of such a service (Farwell 1998). This demonstrates the need to develop a methodology to aid in the design of an effective flexroute system. The research presented in this article addresses this need by using a three-step methodology:

1. Identifying the critical design parameters and outlining the relevant performance measures to evaluate the service under different designs

2. Performing a sensitivity analysis of the design parameters using a Geographic Information System (GIS)-based support system
3. Developing a mathematical programming model to support optimal service design

In step 2, a GIS-based support system developed for scheduling and dispatching a flexroute service was used to evaluate the system performance given different design parameters. The GIS-based support system aides in routing and scheduling the on-demand aspect of the service within the bounds of the fixed-stop schedule constraints. The support system identifies the location of requested origins and destinations, determines the location within the on-demand service area, and calculates the time for the shortest deviated route to pick up and drop off the passenger. In addition, the support system keeps track of all committed rides and continually checks to ensure that any new ride will not prevent the transit vehicle from arriving at the fixed stops on schedule. A detailed description of the GIS-based support system can be found in Durvasula et al. (1998). For the purpose of this research, the GIS-based support system emulates the operation of a flexroute service, and therefore was used to conduct the sensitivity analysis.

In step 3, the results of the sensitivity analysis conducted in step 2 serve as inputs to a mathematical programming model. A multiobjective binary nonlinear programming model was developed to guide the design process that yielded the design parameter values for effective system performance. A description of the design methodology as well as the results of a case study are presented in this article. Before describing the service design methodology, a brief review of flexroute transit and a description of the study area used in this research are presented.

Core Flexroute Concepts

The following are the fundamental terms, definitions, and assumptions concerning flexroute transit considered in this research.

- A service zone is defined as the region between a pair of two consecutive fixed stops on a given route. Each zone is given a unique identification number based on the associated route and the fixed stops that bound the region. For example, Zone 1 is the service area between the fixed stops 1 and 2; Zone 2 is the area between fixed stops 2 and 3 and so on. This definition is used to locate the position of the origin and destination locations of a trip request with respect to the fixed stops.

- The transit vehicle will make mandatory stops at the fixed stops and adhere to the constraint of arriving on schedule at these stops.
• The total time needed to travel from one end of the route to the other end is fixed. This time includes the built-in slack in the schedule for deviating to pick up/drop off on-demand passengers in service zones.

**Study Area**

To provide a context for this effort, the research team worked with Hampton Roads Transit (HRT), an agency that was considering the introduction of flexroute service in the Peninsula region of Virginia (the region anchored by the cities of Hampton and Newport News). In addition to traditional fixed-route service, HRT also provides paratransit services, HandiRide, for the disabled under ADA mandates.

Two of HRT’s existing fixed routes were chosen to serve as the “routes” for flexroute service in this study. The two routes, Routes 10 and 11, were chosen in consultation with HRT management and were “conceptually” modified to enable flexroute service. Fixed stops with low ridership were eliminated for the two routes. Each retained up to a maximum of five major fixed stops. Additionally, the schedules at these stops were changed to incorporate some slack time between the stops to allow for deviation. A methodology described in Welch et al. (1991) was used to realign the fixed routes.

**Step 1: Design Parameters and Performance Measures**

The critical design parameters and relevant performance measures identified for flexroute service are described in this section.

**Flexroute Design Parameters**

The two important design parameters identified and studied in this research are: service zone size and the slack time distribution method. These are discussed below.

**Service Zone Size**

The service zone area of a flexroute service defines how far away from the standard route a vehicle may deviate to pick up or drop off a passenger. The service area defines the maximum feasible deviation distance. It does not imply that everyone within the zone is guaranteed a deviated trip. At times, previously committed rides will force a requested trip from within a service zone to be denied.

By virtue of the definition of a service zone, the total number of zones for a given route will always be one less than the total number of fixed stops. Prima-
rily due to a lack of more complete guidance, most transit systems have limited flexroute service zones to a ¾ mile distance on either side of the nominal fixed route. However, due to the myriad of factors that differ from zone to zone, such as potential ridership, street network connectivity, population density etc., a service area of ¾ mile on either side of the nominal fixed route may not be universally ideal. This warrants the need to analyze the effect of different service zone sizes on the performance of a route deviation system. Three service zones’ distances were considered in this research: 400m, 800m, and 1,200m (1/4, 1/2, and 3/4 mile).

Slack Time Distribution

Clearly, the total route running time for a flexroute system must be greater than the sum of the direct travel times between fixed stops. Otherwise, no deviations would be possible in the system. This “extra” time that is built into the schedule is referred to as slack time.

Given:

- the total nonstop travel time or running time between two consecutive fixed stops (F)
- the total running time of the entire route in minutes (T, which for this study is 60 minutes)

The total slack time for a route can be computed from the relation: \( S = (T - F) \). Clearly, there is a desire to limit slack time to prevent “idle” vehicles in the absence of flexroute requests. The slack time built into each route (in each direction, i.e. inbound/outbound) for the purposes of this study ranged from 5 to 15 minutes, with an average slack time of 8 minutes. The next challenge in flexroute service design is to determine the best way to distribute the slack time among a route’s service zones. In this investigation, two approaches to slack time distribution were considered:

1. Slack time distribution as a weighted average of the nonstop travel time between the two fixed stops of a zone (SDWR), and
2. Slack time distribution as a weighted average of the total number of origins and destinations of ADA certified trips from HandiRide logs (SDNP)
System Performance Measures

The influence of the design parameters was investigated using system-specific performance measures. The two performance measures identified and selected in this research, and the reasons for choosing them, are described below.

System Perspective

The *maximum allowable or feasible (given all schedule constraints) deviations per hour* was chosen as the performance measure from the perspective of a flexroute provider. This measure was chosen because it measures the viability of a flexroute service in terms of the number of requests/calls that can be served within the bounds of fixed-stop schedules. That is, it measures the maximum supply of service that an operator can provide within the given constraints of the system. Once the limit of the supply is known, it can be compared against the demand for flexroute service to see if the demand can be met adequately. The objective of a flexroute provider would be to maximize the value of this performance measure.

Customer Perspective

The *cumulative unused slack time remaining at all fixed stops per route* was chosen as the performance measure from the perspective of a customer using the flexroute service. Although the additional dwell time does not affect the estimated time of arrival at the customer’s destination, it measures the “perceived” inconvenience of a rider in the event of a transit vehicle arriving at the fixed stop earlier than the scheduled time. Early arrivals at scheduled stops are always discouraged in fixed-route services, and provide an important determinant in transit level of service and transit network design (Gray and Hoel 1992). The objective of a flexroute transit provider would be to minimize this value for each service zone.
The two performance measures were also chosen because they are universally applicable in measuring the level of service provided by flexroute transit from the perspectives of the \textit{provider} and \textit{customer}.

\textbf{Step 2: Sensitivity Analysis of Design Parameters}

Table 1 illustrates the sensitivity analysis scenarios and performance measures used to evaluate each scenario. A combination of each tested design parameter value yields a total of 6 possible scenarios.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
Design Parameter & Scenarios \\
\hline
Service zone size (Distance from the route) & 1. 400m (1/4 mile) \\
 & 2. 800m (1/2 mile) \\
 & 3. 1,200m (3/4 mile) \\
\hline
Slack Time (ST) distribution & 1. ST distributed by weighted average of running time between stops (SDWR) \\
 & 2. ST distributed by weighted average of number of ADA riders (SDNP) \\
\hline
\end{tabular}
\caption{Sensitivity Analysis Scenarios and Performance Measures}
\end{table}

The analysis was conducted using the GIS-based support system described earlier to emulate the system performance under each scenario. For the network analysis, the vehicle dwell time values at the fixed and deviated stops were estimated based on common methods available in the literature (see Durvasula et al., 1998 for details). For each service zone, a random number generator was used to randomize the order in which persons were served (emulating different request patterns). Requests for deviations for each zone were picked from a list of HandiRide addresses that were within that particular service zone. After randomizing the order, the requests were added one after the other from the list to the support system, each time checking whether the additional request was feasible within the flexroute constraints. If the addition of a point was not feasible, it was dropped.
from the list of feasible points. Several iterations were performed to exhaustively include all the different orders of requests.

**Results of the Sensitivity Analysis**

Tables 2 and 3 summarize the results of the sensitivity analysis. The numbers presented in the two tables are the mean values for the two performance measures obtained at the end of all iterations for each zone.

The experiments indicate that the performance measures are indeed quite sensitive to the design parameter values. Upon a close examination of the results above, one can see that for Route 11, a zone distance of 400m with a slack time distribution method of SDNP will yield the largest number of feasible deviations, whereas for the additional dwell-time performance measure a zone distance of 800m with the same slack time distribution method yields the best value. Hence, the “best” design parameter combination is not the same under the two performance mea-
sures for a given route. This is further complicated by the fact that, unlike the assumption made in the sensitivity analysis, the buffer distance design parameter is actually zone-specific while the slack time distribution method is route-specific. Thus, it is evident that the design of a flexroute service is a complex, multiobjective problem. Attempting to develop a “near”-optimal design manually, particularly as the number of routes and zones increase, is practically infeasible. Therefore, the research team explored the use of automated optimization techniques as a means to support the design of flexroute systems.

Table 3. Effect of Slack Time Distribution Methods on Additional Dwell Time (in seconds) at the Fixed Stops

<table>
<thead>
<tr>
<th></th>
<th>400m Zone Distance</th>
<th></th>
<th>800m Zone Distance</th>
<th></th>
<th>1,200m Zone Distance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SDWR</td>
<td>SDNP</td>
<td>SDWR</td>
<td>SDNP</td>
<td>SDWR</td>
<td>SDNP</td>
</tr>
<tr>
<td>Zone 1</td>
<td>52.96</td>
<td>53.46</td>
<td>52.96</td>
<td>1.46</td>
<td>46.11</td>
<td>5.39</td>
</tr>
<tr>
<td>Zone 2</td>
<td>41.43</td>
<td>27.27</td>
<td>31.38</td>
<td>55.78</td>
<td>30.21</td>
<td>54.92</td>
</tr>
<tr>
<td>Zone 3</td>
<td>263.49</td>
<td>54.87</td>
<td>110.64</td>
<td>39.79</td>
<td>76.56</td>
<td>40.35</td>
</tr>
<tr>
<td>Route 11 total</td>
<td>357.88</td>
<td>135.60</td>
<td>194.98</td>
<td>97.03</td>
<td>152.88</td>
<td>100.67</td>
</tr>
<tr>
<td>Zone 4</td>
<td>38.78</td>
<td>59.42</td>
<td>38.78</td>
<td>20.89</td>
<td>31.02</td>
<td>20.89</td>
</tr>
<tr>
<td>Zone 5</td>
<td>35.93</td>
<td>35.93</td>
<td>63.16</td>
<td>63.16</td>
<td>15.28</td>
<td>25.4</td>
</tr>
<tr>
<td>Zone 6</td>
<td>11.05</td>
<td>71.05</td>
<td>39.52</td>
<td>57.02</td>
<td>33.66</td>
<td>41.06</td>
</tr>
<tr>
<td>Zone 7</td>
<td>54.11</td>
<td>150.11</td>
<td>30.69</td>
<td>30.69</td>
<td>39.14</td>
<td>29.18</td>
</tr>
<tr>
<td>Route 10 total</td>
<td>139.87</td>
<td>316.51</td>
<td>172.16</td>
<td>171.76</td>
<td>119.1</td>
<td>116.53</td>
</tr>
<tr>
<td>Overall totals</td>
<td>497.75</td>
<td>452.11</td>
<td>367.14</td>
<td>268.79</td>
<td>271.98</td>
<td>217.2</td>
</tr>
</tbody>
</table>
**Step 3: Multiobjective Nonlinear Choice Model**

The problem of identifying the “best” values of design parameters per service zone and route in a flexroute system has been formulated as a multiobjective nonlinear integer-programming model. Furthermore, the model constrains all decision variables as binary (0-1) integer variables. As such, the model is an “assignment” or “coded” model that searches for the optimal matching or pairing of objects of two distinct types under assignment constraints (i.e., each object of each set may be paired only once). The multiobjective aspect of the model is addressed by identifying Pareto optimal or efficient points and using them to construct an “efficient frontier” (Rardin 1997).

**Binary Nonlinear Program: Model Formulation**

The variables used in formulating the objective functions and constraints of the nonlinear model are defined in Table 4. It is important to carefully consider the design parameter values that must be identified for a given route and its service zones: the slack time distribution method (e.g., SDWR and SDNP) for a route, and service zone sizes (e.g., 400m, 800m, and 1,200m) for each zone within a route. Clearly, there is a distinction in the assignment of the two design parameters. Namely, slack time distribution is a route-specific design parameter while service zone size is a zone-specific design parameter. This distinction leads to the following assignment constraint: performance measure values for zones within a given route cannot be chosen from two different slack time distribution methods as it would not be practicable.

The binary nonlinear model is presented by equations 1 through 6. The decision variables are discrete and are represented by $X_{rij}$, $S_{r1}$, and $S_{r2}$ and the model is rendered nonlinear by the two objective functions (which must be computed using the GIS-based support system). Equations 1 and 2 represent the two objective functions, where (1) minimizes the additional dwell time (or the unused slack time) and (2) maximizes the number of feasible deviations performance measure. The index ranges in the objective functions assure that each combination of service zone and design parameter is considered only once. Rather than utilize a decision variable for each combination of design parameters for each zone, an additional set of decision variables representing each slack time distribution method is used. This incorporates the route-specific nature of the slack time distribution method.
## Table 4. Variables Used in the Integer Programming Model

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{q}$</td>
<td>The nonlinear 0-1 integer decision variable for the $r^{th}$ route, $i^{th}$ zone, $j^{th}$ buffer distance</td>
</tr>
<tr>
<td>$S_{r1}$</td>
<td>The nonlinear 0-1 integer decision variable for $r^{th}$ route using Slack time distribution method 1</td>
</tr>
<tr>
<td>$S_{r2}$</td>
<td>The nonlinear 0-1 integer decision variable for the $r^{th}$ route using Slack time distribution method 2</td>
</tr>
</tbody>
</table>

### Inputs or Known Values

| $D_{r_{q_{-1}}}$ and $N_{r_{q_{-1}}}$ | For nonlinear 0-1 model—the additional dwell time and maximum number of feasible deviations for the $i^{th}$ zone, in the $r^{th}$ route, and for the $j^{th}$ buffer distance, under slack time distribution method 1 ($s1$) |
| $D_{r_{q_{-2}}}$ and $N_{r_{q_{-2}}}$ | For nonlinear 0-1 model—the additional dwell time for the $i^{th}$ zone, in the $r^{th}$ route, and for the $j^{th}$ buffer distance, under slack time distribution method 2 ($s2$) |
| $N_{z}$  | The total number of zones in the system |
| $Z$      | Number of zones in the $r^{th}$ route |
| $N_{R}$  | Total number of routes in the system |
| $N_{zR}$ | Average number of zones per route in a system |
| $N_{B}$  | Total number of buffer distances |
| $S_{N}$  | Total number of methods for distributing the slack time |
The two objective functions represented by equations (1) and (2) have been formulated for the case considering only two slack time distribution methods. The formulation could be extended to a general case where one considers more than two methods, i.e., \( S_n > 2 \), and the objective functions may be rewritten as:

\[
\begin{align*}
& \text{Minimize} \sum_{r=1}^{N_z} \sum_{i=1}^{Z_r} \sum_{j=1}^{N_x} x_{r ij} \left[ (D_{r ij, s_1} s_{r_1}) + (D_{r ij, s_2} s_{r_2}) \right] \\
& \text{Maximize} \sum_{r=1}^{N_z} \sum_{i=1}^{Z_r} \sum_{j=1}^{N_x} x_{r ij} \left[ (N_{r ij, s_1} s_{r_1}) + (N_{r ij, s_2} s_{r_2}) \right] \\
\end{align*}
\]

s.t.
\[
\sum_{j=1}^{N_x} x_{r ij} = 1 \text{ for all } r \text{ and } i \\
S_{r_1} + S_{r_2} = 1 \text{ for all } r \\
\sum_{r=1}^{N_z} \sum_{i=1}^{Z_r} \sum_{j=1}^{N_x} x_{r ij} = N_Z \\
x_{r ij} s_{r_1} s_{r_2} = 0 \text{ or } 1 \text{ for all } r, i, j
\]

The constraint modeled by equation (3) ensures that a single zone distance is assigned to each service zone in the system. Equation (4) is important in that it ensures that only one slack time distribution method is chosen for each route.
That is, this constraint accounts for the route specific nature of the slack time distribution design parameter. Constraint (5) satisfies the condition that the sum of all the decision variables’ values must equal the total number of zones in the system. This implies that each zone must be assigned a zone distance and a slack time distribution method. Finally, equation (6) incorporates the 0-1-integer constraint for the binary decision variables. The objective functions of the model have only second order terms, and all the constraints are linear, thus making it a quadratic program.

The algorithm used to solve the nonlinear integer program is the generalized reduced gradient (GRG) method developed by Lasdon et al. (1978). This method is known to be susceptible to converging to a local optimum. To improve the likelihood that the solution obtained was globally optimal, this limitation was addressed by frequently changing the initial values of the decision variables (known as the “multistart” procedure). The global optimum was determined from the range of the objective values obtained from all the trials.

Results: Two Objectives Considered Separately

The model was solved first by considering each objective independently, before considering the multiobjective formulation. Although only one objective function was optimized in each model, the resulting values of the second objective are also reported in Table 5.

Table 5. Local and Global Optima for Each Objective Function Solved Separately

<table>
<thead>
<tr>
<th>Model 1: Maximize No. of Deviations (1st Obj.)</th>
<th>Model 2: Minimize Additional Dwell Time (2nd Obj.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Obj. Value</td>
<td>1st Obj. Value</td>
</tr>
<tr>
<td>34.94</td>
<td>1st Obj. Value</td>
</tr>
<tr>
<td>426.56</td>
<td>2nd Obj. Value</td>
</tr>
<tr>
<td>30.19</td>
<td>2nd Obj. Value</td>
</tr>
<tr>
<td>156.56</td>
<td></td>
</tr>
</tbody>
</table>
The decision variable values that result in the optimal objective function values of 34.94 for the first objective, and 156.56 for the second objective, solved separately, are presented in Table 6. In the case of the first objective (i.e., maximize number of deviations), the slack time distribution method chosen for each route is SDNP. For the second objective, SDNP was chosen for Route 11 while SDWR was chosen for Route 10. The choice of SDWR for Route 10 of the second objective can be understood by comparing the totals in Tables 2 and 3 for SDWR and SDNP for Route 10 (Zones 4 through 7). Notice that the total additional dwell time for Route 10 under 1,200 and 800m zone sizes for SDWR is only slightly higher than the totals for SDNP. For a zone size of 400m, however, the total additional time under SDWR is much lower than for SDNP (139.87 v. 316.51). For this reason the minimize dwell time objective SDWR is preferred over SDNP. This illustrates the effectiveness of the optimization model in identifying the combination of design parameters that yield the best design. Additionally, the results in Table 6 indicate that zones within a given route may indeed need to have different sizes in order to improve the flexroute service level. This substantiates the assumption that there is a need for an automated optimization model to assist in the search for the best design parameters given the zone-to-zone variability within a single route.

**Results: Efficient Frontier—Multiobjective Problem**

Based on the results obtained from solving the optimization model considering the objective functions independently, it is clear that the objectives are conflicting. That is, achieving a larger number of feasible deviations results in a longer additional dwell time. This “conflict” requires a decision-maker to consider trade-offs between these objectives. The designs identified above represent extremes; thus, there is a need to carefully consider possible “trade-off” designs between the extremes. This is referred to as developing the efficient frontier.

The notation \([\text{ND}, \text{ADT}]\) is used to represent the pair of objective function values resulting from a particular flexroute design, where ND stands for number of feasible deviations (to be maximized) and ADT for additional dwell time (to be minimized). The feasible range of these values are known from solving the two objectives separately, and are \([34.94, 426.56]\) and \([30.19, 156.56]\). To construct the efficient frontier, the second objective (i.e., additional dwell time) was incorporated into the 0-1 INLP model as an inequality constraint with a new parameter a (where the values of a \(\bar{a}\) [156.56, 426.56], the limits of the range of values for the
Table 6. Actual Performance Measure Values Selected by the Model for the Global Optimum

<table>
<thead>
<tr>
<th>Zone</th>
<th>No. of Deviations</th>
<th>Additional Dwell Time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,200m</td>
<td>800m</td>
</tr>
<tr>
<td>Zone 1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zone 2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zone 3</td>
<td>-</td>
<td>5.00</td>
</tr>
<tr>
<td>Zone 4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zone 5</td>
<td>2.90</td>
<td>-</td>
</tr>
<tr>
<td>Zone 6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zone 7</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

II. Minimizing Add. Dwell Time

<table>
<thead>
<tr>
<th>Zone</th>
<th>No. of Deviations</th>
<th>Additional Dwell Time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,200m</td>
<td>800m</td>
</tr>
<tr>
<td>Zone 1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zone 2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zone 3</td>
<td>-</td>
<td>39.79</td>
</tr>
<tr>
<td>Zone 4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zone 5</td>
<td>25.40</td>
<td>-</td>
</tr>
<tr>
<td>Zone 6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zone 7</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
second objective function). The 0-1 INLP model formulation with the second objective introduced as a constraint is given as follows:

\[
\begin{align*}
\text{Maximize} & & \sum_{r=1}^{N_r} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} x_{rij} \left( (N_{rj,s_1} s_{r1}) + (N_{rj,s_2} s_{r2}) \right) \\
\text{s.t.} & & \sum_{r=1}^{N_r} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} x_{rij} \left( (D_{rj,s_1} s_{r1}) + (D_{rj,s_2} s_{r2}) \right) \leq \alpha \\
& & \sum_{j=1}^{N_j} x_{rij} = 1 \text{ for all } r \text{ and } i \\
& & s_{r1} + s_{r2} = 1 \text{ for all } r \\
& & \sum_{r=1}^{N_r} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} x_{rij} = N_Z \\
& & x_{rij} s_{r1} s_{r2} = 0 \text{ or } 1 \text{ for all } r, i, j
\end{align*}
\]

The above model was solved using different feasible values of \( a \) to construct the efficient frontier, which is displayed in Figure 2.
An important insight gained from Figure 2 is that over the efficient frontier, the values of the number of feasible deviations objective vary over a smaller range (30.19, 34.93) when compared to the range of the dwell time objective (156.56, 426.5). This indicates that the number of feasible deviations objective would not be considerably degraded (a change of only 4) if the solution for the best value of the additional dwell time objective were to be selected (i.e., 156.56 seconds). Whereas, if the best value for the number of deviations objective is selected, then the corresponding dwell time objective will be 426.5 seconds, nearly 3 times the lowest value of 156.56 seconds. Therefore, based on the results for this particular case study, it would be recommended that the design parameter values for the solution with the objective values, (30.19, 156.56), be used for the flexroute service.

**Conclusions**

While there is significant interest in the concept of flexroute transit service, a significant implementation impediment is the complexity inherent in designing the service. This research identified and explored key flexroute transit design parameters: service area and slack time distribution. Specifically, the results of this effort indicate that for a constant slack time, as the area of the service zone is increased, the number of feasible deviations that can be accommodated decreases. However, the additional dwell time, or the excess slack time left over, increases with the
reduction in service area. Thus, the objectives of increasing the number of feasible deviations and simultaneously reducing the excess slack time are conflicting. For this reason, and the large number of design options available when designing multiple route, multiple zone flexroute service, a multiobjective binary optimization model was formulated to (1) choose the “best” combination of design parameter values for each of the two objectives, and (2) determine the relative change in the values of the two objectives by constructing the efficient frontier. The model is the final step in the service design methodology and will be very useful for a system with a large number of routes and zones. Finally, additional design parameter values can be introduced in the model to determine their effect on design of route service.
Flexroute Transit Service Design

References


About the Authors

Brian L. Smith (briansmith@virginia.edu) is an assistant professor of civil engineering at the University of Virginia. He is the university director of the Smart Travel Laboratory where he conducts research in the area of intelligent transportation systems.

Michael J. Demetsky (mjd@virginia.edu) is professor and chair of civil engineering at the University of Virginia. He is the former director of the Center for Transportation Studies and conducts research in the area of transportation planning and operations.
**Priya K. Durvasula** (priya.durvasula@bms.com) obtained her Ph.D. in civil engineering from the University of Virginia. She is currently a principal analyst at ConvaTec.