Transportation Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Modeling Train Delays in Urban Networks

A. Higgins, E. Kozan,

Modeling Train Delays in Urban Networks[1]

A. HIGGINS

CSIRO Tropical Agriculture, 306 Carmody Road, St. Lucia, QLD 4067, Australia

E. KOZAN

School of Mathematical Sciences, Queensland University of Technology, PO Box 2434, QLD 4001, Australia

The reliability of urban passenger trains is a critical performance measure for passenger satisfaction and ultimately market share. A delay to one train in a peak period can have a severe effect on the schedule adherence of other trains. This paper presents an analytically based model to quantify the expected positive delay for individual passenger trains and track links in an urban rail network. The model specifically addresses direct delay to trains, knock-on delays to other trains, and delays at scheduled connections. A solution to the resultant system of equations is found using an iterative refinement algorithm. Model validation, which is carried out using a real-life suburban train network consisting of 157 trains, shows the model estimates to be on average within 8% of those obtained from a large scale simulation. Also discussed, is the application of the model to assess the consequences of increased scheduled slack time as well as investment strategies designed to reduce delay.

When passenger trains are scheduled over an urban network, the objective is to produce a timetable to satisfy the peak period demand while keeping the number of scheduled train services, operating costs, and expected train delay as small as possible. A schedule that is constructed under ideal conditions to satisfy passenger demand, while minimizing operating costs, may not be the best schedule when applied in practice. Two distinct timetables, but with identical operating costs and efficiency, will most likely be subject to different expected delays when applied. Given that a train is delayed, considerable knock on delay may occur to other trains in the timetable. This occurs particularly at bottleneck parts of the track network, peak periods, or when scheduled connections are to take place. A schedule with a small amount of expected delay can be operated close to its theoretical capacity with little instability.

The model described in this paper estimates the expected positive delay to individual trains, track links, and the schedule as a whole in an urban train network. Such a network can contain multiple unidirectional (plus some bidirectional) track links, intersections, and sidings. A schedule can include trains of varying upper velocities, scheduled stops at any station (or run express), crossing and passing of other trains (providing sidings are available), and nonuniform scheduled departure times. Although urban passenger trains are the primary focus, long distance and nonpassenger trains are also permitted to use the network. Freight trains are usually scheduled away from peak periods so as not to conflict with the passenger services. The specific delay types addressed by the model are as follows.

- **Direct delay to trains.** This is the direct consequences to a train because of minor delays that are not knock-on (referred to as source delay).
- **Knock-on delay to other trains.** When a train is delayed as a consequence of a source delay, it may prevent other trains from passing it (or crossing it in the case of bidirectional track).
- **Delay due to late connections.** Train connections include:
  1. one train waiting at a station for another to arrive for the transfer of passengers;

[1] Accepted by Gilbert Laporte.

Transportation Science
Vol. 32, No. 4, November 1998

0041-1655/98/3204-0346 $01.25

2. a set departure order (i.e., one train must depart before the other at a station). An example of this situation is when two tracks (same direction) merge into one after a train station from which the faster train departs first;
3. Commencement of a new service after arrival at the destination using the same physical train. This type of connection is important to rail planners because it is desirable to keep the initial expected delay to the new train service to a minimum, while at the same time satisfying the constraints of demand and other connections.

Minor source delays include long dwell times at stations because of passengers, a late commencement of a train service, waiting for signals, interferences from maintenance activities, minor train faults, and poor weather conditions. The consequences of major or extremely frequent source delays are more difficult to estimate accurately because these can cause excessively long term knock-on delays that are subject to a large amount of variation in practice. This is particularly so if the network consists largely of bidirectional track. In Queensland Rail (QR) Australia, major source delays such as train breakdowns (which lead to cancellations) are infrequent, which makes it harder to model using probability distributions. The model of this paper does not address these types of delays or train cancellations, which usually only occur in QR if a train breaks down (i.e., major source delay) and can no longer be used to provide the service.

The model, which provides the train controller with knowledge of the expected delay to individual trains and track links, can be used to assess possible investment strategies designed to improve reliability or reduce average travel time of the overall system. Such investment strategies include:

1. new track or sidings, which reduces the large delay due to bottlenecks;
2. altering track alignments to increase train speeds;
3. a more advanced train control system to decrease minimum headway and thus increase track capacity; and
4. new systems to allow more efficient or safer entry and exit of passengers from trains.

Although the consequences of delay have been modeled for freight services over a single line regime (HIGGINS, FERREIRA, and KOZEN 1995), there are major differences to a suburban network. In terms of train priority, load, type, and customer agreements are some of the key determinants for freight services, whereas peak period passenger trains are generally of a higher priority than those outside of peak period. Freight services do not stop as frequently from origin to destination, thus the final arrival time is usually most important. The delay to freight services is heavily characterized by the length and position of sidings. Suburban networks are usually at least double track, thus a train traveling in one direction does not normally disrupt those traveling in the opposite direction. Although delay is usually shorter than for freight services (HIGGINS, 1996), passengers are more sensitive to delay. The economic consequences of a schedule with a large amount of expected delay is lost opportunity cost because of passengers choosing another mode of transport.

1. PAST WORK

ALTHOUGH LIMITED LITERATURE exists for analytically modeling the delay of a urban network as a whole, there has been an abundance of research aimed at freight services on a single line track with multiple track sidings (CHEN and HARKER, 1990; HALLOWELL, 1993; and Higgins et al., 1995). These papers focused on the expected delay caused by a combination of train conflicts and random delay, given that trains can only cross and overtake at sidings. The single equation analytical models (PETERSEN, 1974; and KRAFT, 1983) related expected delay to the number of train conflicts and the distance between sidings. The latter model also considered train priorities as well as acceleration and deceleration characteristics.

The literature aimed at passenger services only considered the knock-on delays of scheduled trains (sometimes uniform) along a single track. The delay to a train due to another traveling in the same direction, was analyzed by CAREY and KWIECINSKI (1994). In that paper, the knock-on delay to any train was dependent on the random delay (source delay) as well as the headway between two trains. Estimates of the expected trip times given these delays, were calculated by nonlinear regression and heuristic methods. OZEKICI and SENGOR (1994) modeled the knock-on delays and average passenger waiting time using Markov chain techniques. The scheduled departure times were assumed uniform and a service was canceled if the delay exceeded a predefined level. Given a travel time probability density function for a train on a track link, a departure time transition matrix was constructed for the calculation of the expected departure delay.

The use of computer simulation (PETERSEN and

TAYLOR, 1987; LIDEN, 1993; and CHENG, 1996) as an alternative to analytical techniques has the advantage of allowing increased detail in track layout and more realistic rules for when two trains compete for the same track link. The simulation model proposed by Cheng (1996) was able to adopt different rescheduling strategies for when delays occur. The drawbacks for simulation include a much longer computation time than analytical methods and a lesser ability to be applied for schedule optimization.

Queuing theory (TAKAGI, 1991; and GROSS and HARRIS 1985) is an efficient alternative to simulation for estimating delay but cannot be accurately applied to an entire rail network because of the following reasons.

- A train network is complex in that it includes many intersections, uni- and bidirectional track links of various lengths, sidings, and track capacity. Train services vary with different upper velocities, slack time, scheduled stops, nonuniform departure times, and include train connections as described in the introduction of the paper. In the case of train connections and intersections, a train can suffer a delay from another that is scheduled much earlier and from a different part of the network.
- As well, the distribution of arrival times for each train at any station or intersection depends on the distribution of current delay, which can be different for each train service. Hence, delay to both the trains and at stations (or intersections) are interdependent. Therefore, the calculation of expected delay requires a solution of equations.

Rail systems are interested in determining which individual trains or track links have (and cause) the greatest delay, so as to modify the schedule or infrastructure. Given the above reasons, queuing theory, unlike the model of this paper, cannot provide delays to individual track links or trains, nor would it give an accurate average estimate of delay for a group of trains (e.g., peak period).

2. MODEL SPECIFICATION

Definitions

Delay: For the purposes of this paper, only positive delay (i.e., tardiness) is considered. QR does not allow early departures for passenger trains, so there is no advantage of negative delay.

Source Delay: The delay to a train that is not caused by another. This type of delay can be a result of lengthy dwell time at stations due to embarking and disembarking of passengers, maintenance activities, signal problems, minor train failures, and poor weather conditions.

Minimum Headway: Minimum length of time (or distance) separating two trains traveling on the same line. This is determined by the length of the track link.

Link: A section of track on which only one train is permitted at any time. Track links are usually separated by signals. If a train station is located on a link, passengers can embark and disembark at that link.

Siding: A track link that can be used for the crossing and passing of trains.

Assumptions

The following assumptions are made with regard to the model.

- The scheduled arrival and departure times include any slack time. If a train arrives early at a station, it will wait until the scheduled departure time.
- Trains are permitted different upper average velocities. As an example, express passenger trains will traverse a link faster than those that must stop. Dwell time at a station is included in the scheduled travel time along the link.
- A train delayed by another can pass it if a siding or parallel track is available and not used by other trains at the time of delay.
- Conflicts in the event of delay are resolved on a first come first serve basis. This priority rule is used most of the time with QR (WARDROP, SWIFT, and MATHESON 1993).
- Capacity at a train station is determined by the number of sidings or unidirectional track links at the station.

Model Parameters

The following model parameters are supplied by the user:

I = Set of train services for one cycle of the schedule (usually daily).

J = Set of track links for the whole suburban network.

Q_j^i = Set of remaining links from link $j \in J$ along which train $i \in I$ is scheduled to travel.

Q^i = Set of links along which train $i \in I$ is scheduled to travel from origin to destination.

R_j^i = Set of links prior to link $j \in J$ along which $i \in I$ is scheduled to travel. $Q^i = \{R_j^i, j, Q_j^i\}$

V_j^i = The link immediately prior to link $j \in J$ on which train $i \in I$ is scheduled to travel. $V_j^i \in R_j^i$.

Y_j^i = Scheduled departure time of train $i \in I$ from link $j \in Q^i$.

T = Duration of a source delay.

$PRO(T, i, j)$ = Probability that train $i \in I$ will have a source delay of duration T on link $j \in Q^i$.

$E_{k,j}^i$ = Amount of time train $i \in I$ is able to recover between links $k \in Q^i$ and $j \in Q^i$ (i.e., scheduled slack time). The shortest possible travel time for train $i \in I$ between these two links is $Y_k^i - Y_j^i - E_{k,j}^i$ (i.e., scheduled travel time minus slack time).

$CONN$ = Set of train connections. A train connection takes place between train $i \in I$ at link $j \in J$ and train $l \in I$ at link $h \in J$. Links j, $h \in J$ are located at the same station. $(i, j, l, h) \in CONN$.

The following delay information is supplied as output from the model.

$PRS(TL, i, j)$ = Probability that train $i \in I$ is TL minutes late at the departure of link $j \in Q^i$.

${}^1TC_j^i$ = Direct delay to train $i \in I$ at link $j \in Q^i$ as a result of source delays.

${}^2TC_j^i$ = Delay to train $i \in I$ at link $j \in Q^i$ as a result of knock on delays.

${}^3TC_j^i$ = Delay to train $i \in I$ at link $j \in Q^i$ as a result of late connections.

TC_j^i = Delay to train $i \in I$ at link $j \in Q^i (= {}^1TC_j^i + {}^2TC_j^i + {}^3TC_j^i)$.

Direct Delay to Trains

In this section, the expected delay to a train due to source delays is addressed. Consider the situation in which a train $i \in I$ (with a current delay of TL minutes) suffers a source delay of T minutes on link $k \in R_j^i$. If the scheduled slack time, $E_{j,k}^i$, from link $k \in R_j^i$ to $j \in Q_k^i$ is less than the current delay, TL, the delay to train $i \in I$ due to the source delay is T minutes. If the slack time is greater than the current delay, some or all of the source delay will be recovered by the time train $i \in I$ departs link $j \in Q_k^i$. In this situation, the delay to train $i \in I$ at link $j \in Q_k^i$ due to the source delay is $\max(TL + T - E_{j,k}^i, 0)$. If the train is capable of recovering more time than the source delay and current delay combined, it will suffer no delay and is why the $\max(\cdot, 0)$ is included. Given both situations of the scheduled slack time being greater than and less than the current delay, the delay at link $j \in Q_k^i$ is $\max(TL + T - \max(TL, E_{j,k}^i), 0)$.

Taking into account all possible source delay and current delay durations, T and TL respectively, as well as all track links $k \in R_j^i$ (before link $j \in Q_k^i$), the expected delay to train $i \in I$ at link $j \in Q^i$ due to source delays is

$$\mathrm{E}({}^1TC_j^i) = \sum_{k \in R_j^i} \sum_{T} \sum_{TL} \mathrm{PRS}(TL, i, V_k^i) * \mathrm{PRO}(T, i, k) * g_1(i, k, j, T, TL), \quad (1)$$

where $g_1(i, k, j, T, TL) = \max(TL + T - \max(TL, E_{j,k}^i), 0)$.

Knock-on Delay to Other Trains

A train can suffer delay on both a unidirectional and bidirectional track. For a unidirectional track, only trains traveling in the same direction can be affected. Both cases are considered in this subsection.

Unidirectional Track

Given that train $l \in I$, with a current delay of TL, suffers a source delay of T minutes on link $k \in Q^l$, let train $i \in {}_{TL,TM}^{\ \ T}S_k^l$ (where train $i \in I$ has a current delay of TM) if it suffers some knock-on delay due to train $l \in I$. This is illustrated by the time distance graph in Figure 1 for which $i \in {}_{TL,TM}^{\ \ T}S_k^l$ if

$$Y_k^i + TM > Y_k^l + TL \quad (2a)$$

and

$$Y_k^i + TM < Y_k^l + T + TL + {}_{TL,TM}^{\ \ T}P_k^{l,i} * M_k \quad (2b)$$

where ${}_{TL,TM}^{\ \ T}P_k^{l,j}$ is the position of train $i \in I$ in ${}_{TL,TM}^{\ \ T}S_k^l$ (in ascending order of Y_k^i) and M_j is the minimum average time for a train to travel along link $j \in J$.

A train will not belong to ${}_{TL,TM}^{\ \ T}S_k^l$ if it can bypass train $l \in I$ by taking a different track or siding for which there is no other scheduled traffic at the time of delay. In Eq. 2a, train $i \in I$ (with current delay TM) will not suffer knock-on delay by train $l \in I$ (with current delay TL) if it arrives at link $k \in Q^l$ before train $l \in I$. Eq. 2b states that train $i \in I$ will not suffer knock-on delay from train $l \in I$ if it arrives at link $k \in Q^l$ after train $l \in I$ plus source delay T plus expected time required for delayed trains between $i \in I$ and $l \in I$ to clear first $({}_{TL,TM}^{\ \ T}P_k^{l,j} * M_k)$. For the example shown in Figure 1, ${}_{TL,TM}^{\ \ T}P_k^{l,i}$ is equal to 1.

Rearranging Eq. 2b, the knock-on delay to train

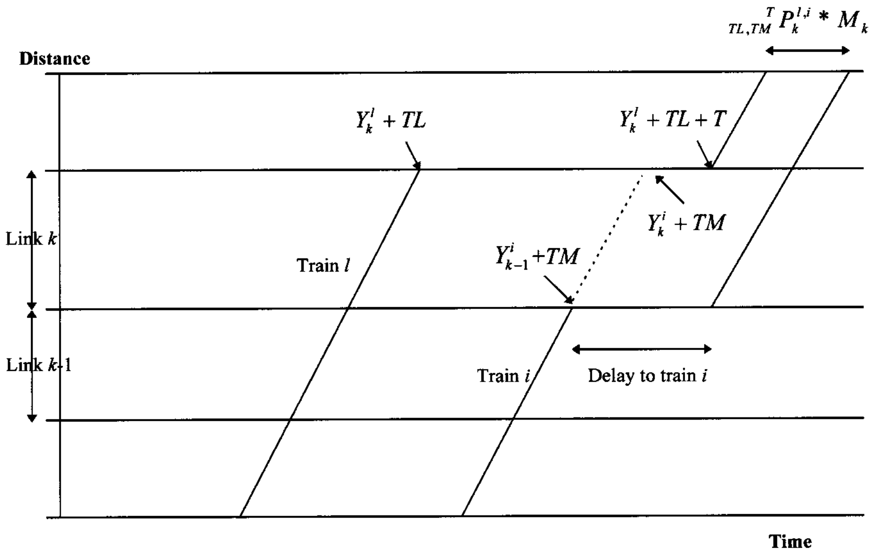

Fig. 1. Illustration of delay to train $i \in I$.

$i \in I$ due to train $l \in I$ is

$$Y_k^l + T + TL + {}_{Tl,TM}^{\ \ T}P_k^{l,i} * M_k - TM - Y_k^i.$$

Taking into account schedule slack time and no negative delay, the knock-on delay to train $i \in I$ at a link $j \in Q^i$ due to train $l \in I$ (delayed for duration T on link $k \in Q^l$) is

$$\begin{aligned} g_2(i, k, j, l, T, TL, TM) &= \max(Y_k^l + T + TL + {}_{TL,TM}^{\ \ T}P_k^{l,i} * M_k - TM \\ &\qquad - Y_k^i - E_{k,j}^i, 0) \\ &\qquad \text{if} \quad i \in {}_{TL,TM}^{\ \ T}S_k^l, \quad 0 \quad \text{otherwise.} \end{aligned}$$

When considering all possible source delays, T, and current delays of trains i, $l \in I$ at each link $k \in Q^l$ (TM, TL), the expected knock-on delay to train $i \in I$ at link $j \in Q^i$ due to all other trains $l \in I$ is

$$\begin{aligned} \mathrm{E}({}^2TC_j^i) = \sum_k \sum_T \sum_l \sum_{TL} \sum_{TM} & * \mathrm{PRO}(T, l, k) \\ & * \mathrm{PRS}(TM, i, k) * \mathrm{PRS}(TL, l, k) \\ & * g_2(i, k, j, l, T, TL, TM) \end{aligned} \tag{3}$$

Bidirectional Track

Bidirectional track usually contains many siding links to allow trains to cross and overtake. Although this regime is common in rural areas, it is sometimes used in urban areas when traffic is not sufficient to justify two unidirectional tracks (one in each direction).

If a train suffers a source delay on a bidirectional track where a siding link is not located (position A of Figure 2) trains traveling in the same direction will tend to bunch because of the setup of signals. To allow the least average knock on delay and safe operation, delayed trains traveling in the same directions are to have priority over opposing trains until all have cleared. Hence, the knock-on delay to trains on bidirectional track, where no siding is located, are per unidirectional track except with ${}_{TL,TM}^{\ \ T}S_k^l$ containing all same-direction trains first (in ascending order of Y_k^i), followed by all opposing trains.

If a train suffers a source delay where a siding is located (position B of Figure 2), trains traveling in the same direction suffer the same knock-on delay as per unidirectional track, unless there are no opposing trains at the time of the source delay. Opposing trains are assumed to be not affected.

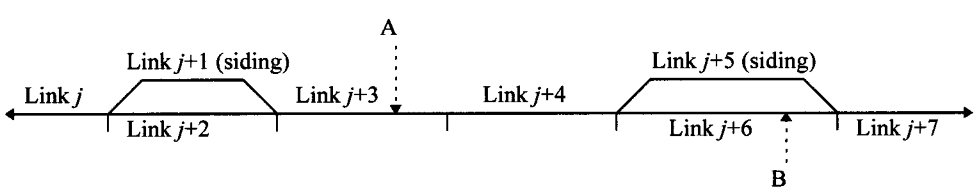

Fig. 2. Bidirectional track with siding links.

Delay Because of Late Connections

For a scheduled connection, trains $i \in I$, $l \in I$ are to depart from links $k \in Q^i$ and $h \in Q^l$, respectively, at times Y^i_k and Y^l_h, respectively. Links $k \in Q^i$ and $h \in Q^l$ are located at the same station (different platforms) and train $i \in I$ cannot depart until $l \in I$ arrives (i.e., $Y^i_k \geqslant Y^l_h$).

If train $i \in I$ has a current delay of TL minutes on arrival at link $k \in Q^i$, and train $l \in I$ has a current delay of TM minutes on arrival at link $h \in Q^l$, the delay to train $i \in I$ due to the connection is $\max(Y^l_h + TM - Y^i_k - TL, 0)$. The $\max(\cdot, 0)$ is included to prevent negative delay because train $i \in I$ is not permitted to depart until it is scheduled to. Given the scheduled slack time from departure of link $k \in Q^i$ to link $j \in Q^i$, the delay to train $i \in I$ at link $j \in Q^i$ due to the connection is

$$g_3(i, k, l, h, j, TL, TM) = \max(Y^l_h + TM - Y^i_k - TL - E^i_{j,k}, 0).$$

Given all scheduled connections for train $i \in I$ and all possible current delays for train $i \in I$ and its connecting train $l \in I$, the expected delay to train $i \in I$ at link $j \in Q^i$ due to connections is

$$\mathrm{E}({}^3TC^i_j) = \sum_{\substack{k,l,h \\ \text{s.t.}(i,k,l,h)\in \text{CONN}}} \sum_{TL} \mathrm{PRS}(TL, i, V^i_k) \sum_{TM} \mathrm{PRS}(TM, l, V^l_h) * g_3(i, k, l, h, j, TL, TM). \tag{4}$$

3. SOLUTION OF THE EQUATIONS

SIMPLE EVALUATION of Eqs. 1, 3, and 4 is not possible because of the unknown delay probabilities $\mathrm{PRS}(\cdot, \cdot, \cdot)$. The solution algorithm used here to solve this system of equations is based on iterative refinement (WOODFORD, 1992). For the algorithm, $\mathrm{PRS}(\cdot, \cdot, \cdot)^r$ and $\mathrm{E}(TC^i_j)_r$ are, respectively, the delay probability distribution and the expected delay at iteration r of the iterative refinement algorithm, with

$$\mathrm{E}(TC^i_j)_r = \mathrm{E}({}^1TC^i_j)_r + \mathrm{E}({}^2TC^i_j)_r + \mathrm{E}({}^3TC^i_j)_r \tag{6}$$

and variance

$$\mathrm{V}(TC^i_j)_r = \sum_{c=1,2,3} \mathrm{V}({}^cTC^i_j)_r + \sum_{\substack{c,d \\ c\neq d}} \mathrm{cov}({}^cTC^i_j, {}^dTC^i_j)_r, \tag{7}$$

where

$$\mathrm{V}({}^cTC^i_j)_r = \mathrm{E}({}^cTC^{i^2}_j)_r - (\mathrm{E}({}^cTC^i_j)^2_r)$$

$$\mathrm{cov}({}^cTC^i_j, {}^dTC^i_j)_r = \mathrm{E}({}^cTC^i_j * {}^dTC^i_j)_r - \mathrm{E}({}^cTC^i_j)_r\, \mathrm{E}({}^dTC^i_j)_r.$$

The requirement of the variance (Eq. 7) in the solution algorithm depends on the distribution of $\mathrm{PRS}(\cdot, \cdot, \cdot)^r$. The distributions found in practice (Sections 4 and 5) are of an Erlang shape and do not require a variance. If, for other applications, $\mathrm{PRS}(\cdot, \cdot, \cdot)^r$ or $\mathrm{PRO}(\cdot, \cdot, \cdot)^r$ are continuous distributions such as negative exponential, the distributions are discretized when applied to Eqs. 1, 3, and 4. The solution to these equations is determined by the following iterative refinement algorithm where an initial solution is obtained by setting the delay probabilities, $\mathrm{PRS}(\cdot, \cdot, \cdot)^0$, to zero.

Step 0. **Let** $r = 0$. Set $\mathrm{PRS}(TL, i, j)^r = 0$, $\mathrm{E}(TC^i_j)_r = 0$ and $\mathrm{V}(TC^j_i)_r = 0\ \forall i \in I, j \in Q^i, TL$

Step 1. **Let** $r = r + 1$. Determine $\mathrm{E}(TC^i_j)_r$, $\mathrm{V}(TC^i_j)_r$ and $\mathrm{PRS}(TL, i, j)^r\ \forall i \in I, j \in Q^i, TL$

Step 2. **IF** $r > 0$ and $|\mathrm{E}(TC^i_j)_r - \mathrm{E}(TC^i_j)_{r-1}| < \varepsilon_1$, $\forall i \in I, j \in Q^i$ (where ε_1 is small)
THEN
Go to **Step 3.**
ELSE
Go to **Step 1.**
END {IF}

Step 3. Evaluate Eqs. 1, 3, and 4 using $\mathrm{PRS}(TL, i, j)^r\ \forall i \in I, j \in J, TL$

In Step 1 of the above algorithm, $\mathrm{PRS}(TL, i, j)^r$ is constructed from $\mathrm{E}(TC^i_j)^{r-1}$ and $\mathrm{V}(TC^i_j)_{r-1}$ (if the distribution is characterized by a variance). The convergence criteria in Step 2 states that the difference between all expected delay estimates, $\mathrm{E}(TC^i_j)_r$, from one iteration to the next is to fall below the user-specified tolerance ε_1. This criteria was chosen because the expected delay estimates were found to converge at different rates. Although the algorithm is not guaranteed to converge (ORTEGA and RHEINBOLDT, 1970), it has always converged to a unique solution when applied to the real life problem in Sections 4 and 5.

4. MODEL APPLICATION

IN THIS SECTION, the model described earlier is applied and validated using the Brisbane city train network of QR. This network (Figure 3), which covers a region with a population of about 1.5 million people, consists of over 400 km of track and is split into 557 links as separated by signal posts. Illus-

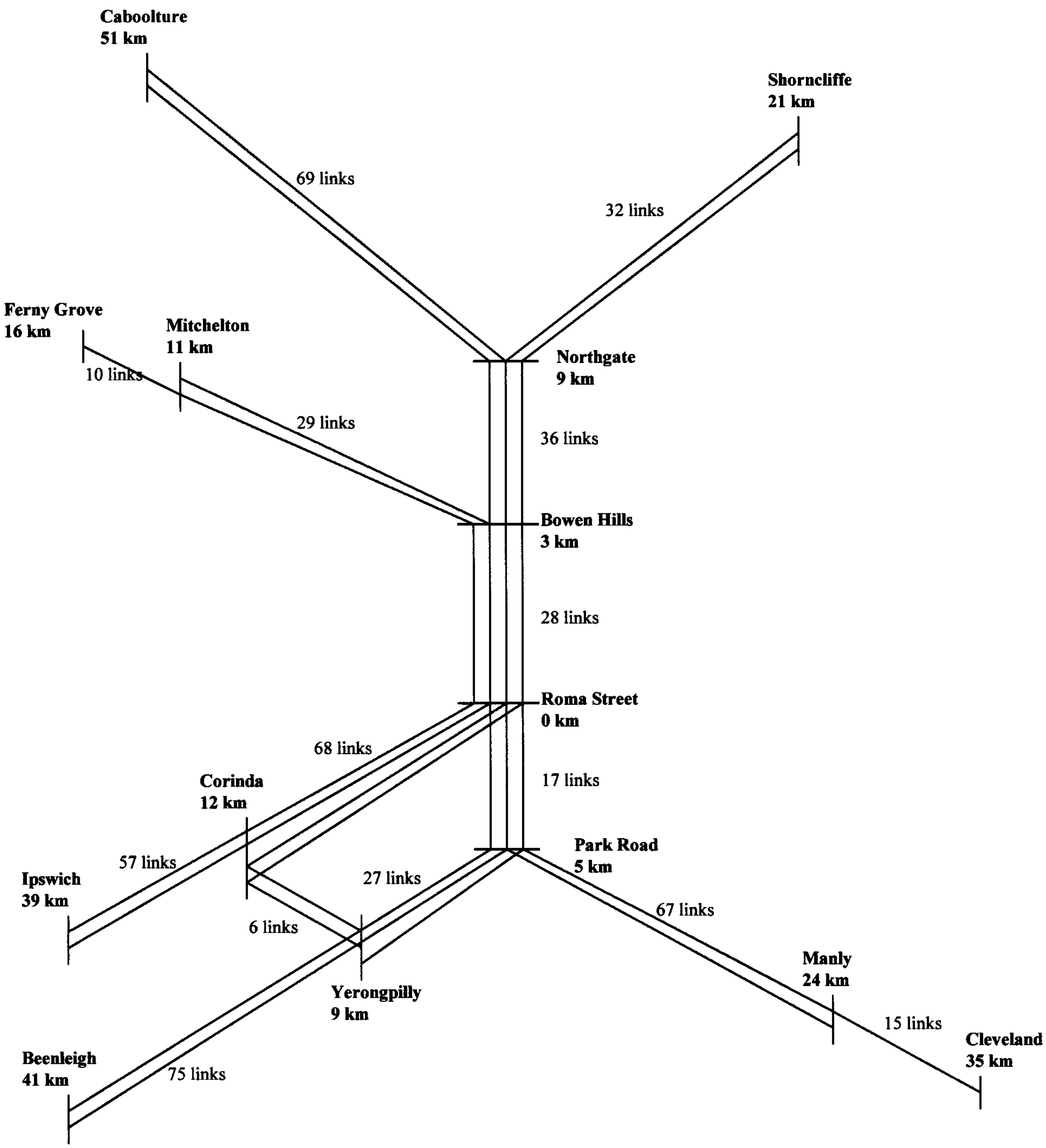

Fig. 3. Queensland Rail city train network.

trated in Figure 3 is the number of links between each key station and the distance from the central station (Roma Street). Individual links (including those that allow trains to transfer from one main line to another), sidings, and train stations were not shown because of the size of the network and level of detail required. All tracks are unidirectional, except for one of the tracks between Bowen Hills and Northgate, and between Roma Street and Yerongpilly. The single line tracks (such as between Ferny Grove and Mitchelton) are also bidirectional.

On each weekday, 320 trains are scheduled to use this network. The trains' services are split into six main types, namely Caboolture, Shorncliffe, Ferny Grove, Ipswich, Beenleigh, and Cleveland lines. On each weekday, there is a morning and afternoon peak period. For the application here, only the morning trains are considered. The morning period con-

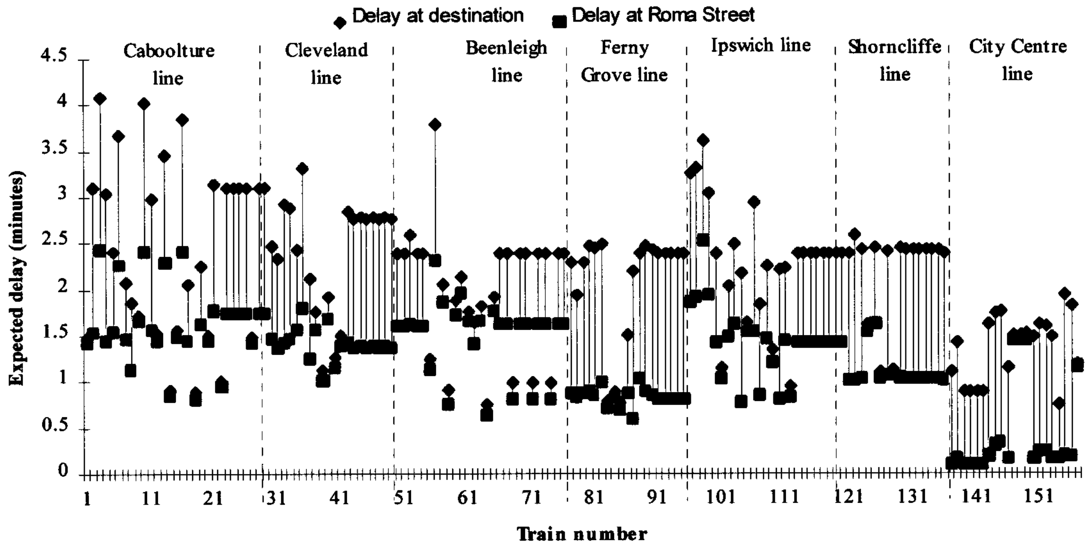

Fig. 4. Expected delay for each train using the analytical model.

sists of 157 trains. A train usually commences its journey from one of the six stations mentioned above. It then passes through Roma Street and terminates at one of these six stations, except at the other side of the city. For example, a train may commence its journey at Caboolture, and terminate at Cleveland. In a peak period, many trains commence or terminate their journeys in the city or at intermediate stations. Freight trains may also use the city train network. These trains do not pass through the Bowen Hills to Roma Street corridor and are usually scheduled away from peak periods.

Model validation is performed by comparing the delay estimates derived from the model with those obtained using stochastic simulation techniques (PIDD, 1984). The delay estimates obtained using simulation (described in Appendix A) are assumed to reflect accurately actual conditions. The availability of historical train delay data was limited for passenger services because there was no data for individual train delays on each link or delay distributions. QR did have available the percentage of passenger trains that are less than three and five minutes late for the morning and afternoon peak periods. The historical distribution of source delays for QR freight services was found (Higgins 1996) to follow a negative exponential distribution, whereas overall delays were found to follow an Erlang distribution with a shape parameter of 3. From Higgins (1996), the probability of a source delay occurring on a track link with no train station was estimated to 0.0001. For track links with busy train stations (e.g., between Bowen Hills and Roma Street or stations adjacent to intersections), the probability of delay was estimated at 0.08. The probability of a source delay was estimated at 0.002 for links with nonbusy stations. Given that a source delay does occur, it was estimated (Higgins 1996) to have a mean duration of 3.03 minutes regardless of track link.

The city train schedule only contains scheduled slack time (between 2 and 4 minutes) at Roma Street station. This is the central station in the Brisbane network, and far more passengers embark (and disembark) at this station than any other. There are 148 scheduled connections most of which occur at the destination station of the train services (i.e., a train commences a new service after terminating at the destination). Other scheduled connections occur at Bowen Hills, Roma Street, and Park Road.

The model, which was allowed to converge to an accuracy of 0.01 minutes for expected delay, took 1 minute and 35 seconds to run on a Pentium PRO-180 PC. Illustrated in Figure 4 is the expected delay for each train at its destination station and Roma Street. Trains within each service type are in order of arrival time at Roma Street. For example, each train in the Caboolture line service type commences its journey at Caboolture or some other station on that line. Trains in the City Centre line commence their journey from either Roma Street or Bowen Hills. In Figure 4, the delay estimates are generally characterized by the distance traveled by each train upon arrival at Roma Street and its destination, as well as whether the train is in peak period or not. In Figure 4, peak period is about the middle of each service type. Trains that commence their journey on the Ipswich or Caboolture lines have large expected

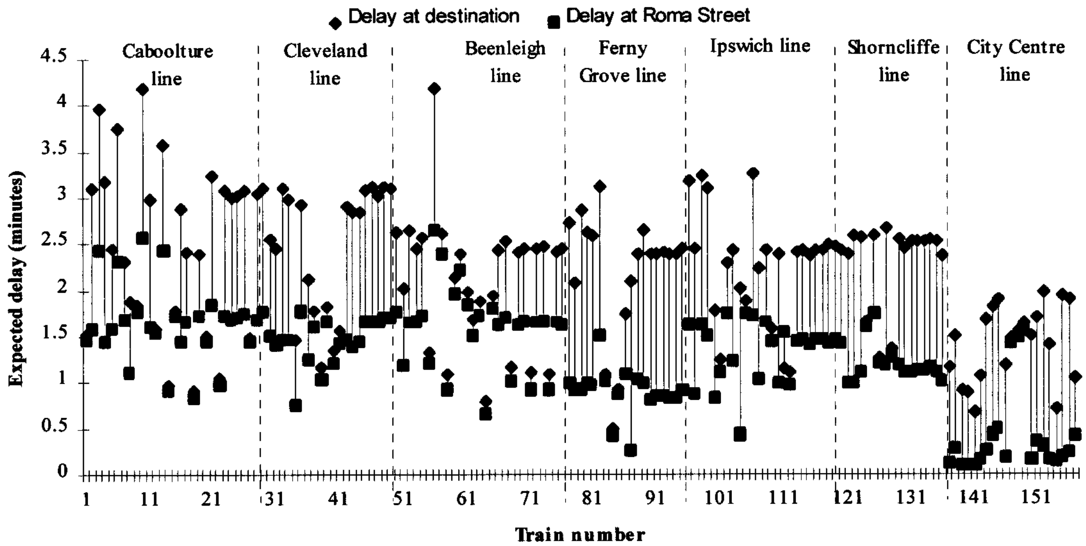

Fig. 5. Expected delay for each train using simulation.

delays in peak periods because these two lines are the most congested and convey the largest number of passengers. The low delays for trains in the City Centre line are due to the distances traveled being about half that of other trains.

The simulation model required 6000 iterations (5.5 hours on a Pentium Pro PC) for the standard deviation of expected train delay estimates to be sufficiently small (i.e., standard deviation for 95% of the train delay estimates being less than 0.1 minutes). Visual comparison of analytical (Figure 4) and simulation results (Figure 5) show that the analytical model did give accurate estimates and was able to pick out trains with large expected delays.

Illustrated in Figure 6 is a plot of the analytical model versus the simulation model expected delay. In this figure, the model tends to slightly underestimate the simulation results because most of the points fall slightly below the line slope = 1. This is also indicated in Table I, which contains (for each scheduled service type) the average absolute relative error (AARE), the percentage of trains with an absolute relative error (ARE) of less than 10 and 20%, and average ratio of model to simulation results (M/S). From Table I, the model estimates of delay are, on average, about 2% less than the simulation, with only the Ipswich line giving a slight overestimate. This slight model inaccuracy may be due to the distribution of current delay (Erlang with shape parameter 3), which is sensitive to long term

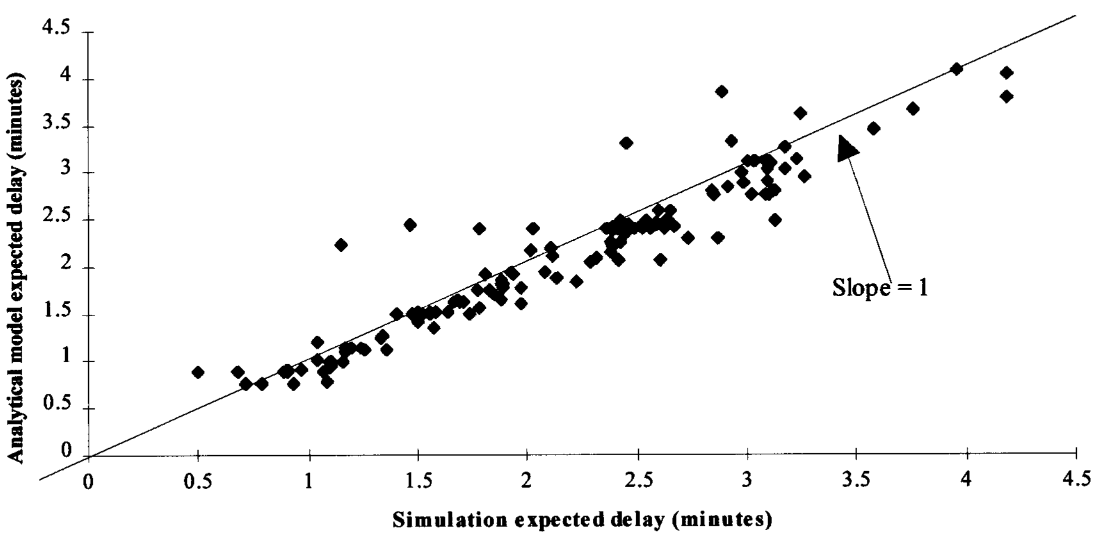

Fig. 6. Analytical model versus simulation expected delay.

TABLE I

Comparison of Analytical Model and Simulation by Service Type

Service Type	M/S	AARE	Percentage of Trains with ARE >10%	Percentage of Trains with ARE >20%
Caboolture	0.99	0.05	10	3
Cleveland	1.00	0.09	15	5
Beenleigh	0.94	0.07	24	4
Ferny Grove	0.97	0.11	30	20
Ipswich	1.03	0.12	28	12
Shorncliffe	0.95	0.05	13	0
City centre	0.99	0.07	17	7
Average	0.98	0.08	20	7

M/S, average of analytical model divided by simulation delay; AARE, average absolute relative error; ARE, absolute relative error.

knock on delays such as those that occur in peak periods, being slightly different from the actual for some trains. The low AARE in Table I (8% on average) further indicates there is a strong relationship between model and simulation estimates. As well, most scheduled service types have less than 7% of trains with an ARE of greater than 20%.

5. CONSEQUENCES OF MODIFYING INPUT PARAMETERS

USING THE QR base problem of Section 4, the first options analysis assesses the consequences of modifying scheduled slack time for each link in the train network. In Table II, slack time is varied between 0 and 16% of the scheduled travel time and expected delay for each train is measured at the central station. Results in Table II show that, although large reductions in average expected delay per train are achievable with small amounts of slack time, there is little improvement when slack time is increased further (e.g., 8 to 16%). This result is also evident when considering the number of trains with a delay of greater than 1 and 2 minutes (Table II). When planning a timetable, such a result would be applied for assessing the ideal schedule slack time in terms of overall train travel time and reliability. Passenger

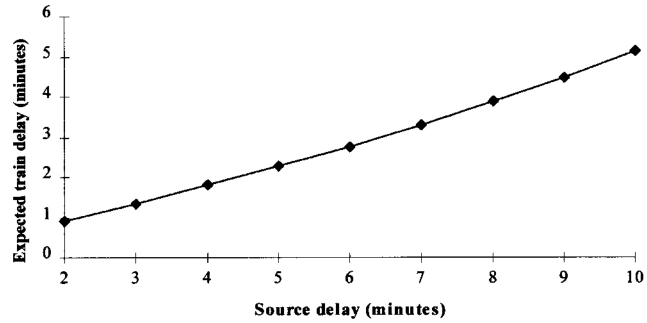

Fig. 7. Consequences of modifying average source delay.

satisfaction and competition with other modes of transport is a trade-off between both.

The second options analysis assesses the impacts of modifying the mean of the source delay distribution (negative exponential). In Figure 7, the mean of the distribution, given that a delay occurs, is varied from 2.0 to 10 minutes, where the base case is 3.03 minutes. The results in Figure 5 show the increase in expected delay being slightly greater for larger values of source delay. This is due to delayed trains causing larger knock-on delays and to an increased number of trains.

6. CONCLUSIONS

THIS PAPER HAS PRESENTED a model to estimate the expected delay to individual trains and track links on a complex multitrack rail network. The delay types addressed were direct delay to trains, knock-on delay to other trains, and delay due to late connections. The resultant system of delay equations were solved using an iterative refinement algorithm.

The model was verified using an actual suburban rail network as supplied by Queensland Rail. The network consisted of 557 links and 157 scheduled trains in the morning period. Model expected delay estimates were compared to those obtained from stochastic simulation, results of which were used to reflect actual conditions. For 93% of scheduled trains, the model delay estimates were within 20% of simulated delays. The slight model inaccuracy

TABLE II

Consequences of Modifying Scheduled Slack Time

	Slack Time as a Percentage of Scheduled Travel Time: 0	2	4	6	8	10	12	14	16
Expected delay/train (minutes)	1.32	1.12	0.94	0.80	0.69	0.61	0.54	0.48	0.44
Percent of trains >1 minute delay	64	53	23	15	8	5	4	3	3
Percent of trains >2 minutes delay	6	4	2	2	2	2	2	1	1

may have been due to the assessed distribution of current delay, which is sensitive to long term knock-on delays, being slightly different from the actual distribution for some trains. Sensitivity tests were performed to assess the consequences of modifying schedule slack time and expected source delay. These results, which provide a quantified means of assessing various options in terms of change in reliability and average train travel time, can be applied as an aid to a cost-benefit analysis for future schedule or track upgrading. The model has been applied as part of a Queensland Rail project to investigate the impact of new proposed timetable changes, which are as a result of expanded infrastructure.

APPENDIX A

THE SIMULATION OF SECTION 4 was conducted according to the following algorithm.

X_j^i = actual departure time of train $i \in I$ from link $j \in J$.

$CDEL_j^i$ = cumulative delay of train $i \in I$ at link $j \in J$

$EDEL_j^i$ = expected delay of train $i \in I$ at link $j \in J$

DO H1 times
  FOR each train $i \in I$ on each link $j \in J$
    Generate a random number and use it to determine the duration of source delay T.
    Increased the travel time of train $i \in I$ on link $j \in J$ by T minutes.
  END {FOR}
  Generate an actual schedule for each of the trains. In the event of two trains conflicting for a track link, a first come first served priority rule is applied. Once the actual schedule, given delays to individual trains, is established, determine X_j^i for each train $i \in I$ on each link $j \in J$.
  FOR each train $i \in I$ on each link $j \in J$
    $\mathrm{CDEL}_j^i = \mathrm{CDEL}_j^i + X_j^i - Y_j^i$
  END {FOR}
END {DO}
FOR each train $i \in I$ on each link $j \in J$
  $\mathrm{EDEL}_j^i = \mathrm{CDEL}_j^i/\mathrm{H1}$
END {FOR}

At each repetition of the above simulation, the timetable suffers random source delays to individual trains on various links according to the input source delay distributions, and the actual timetable is calculated given these source delays. The delay to each train at each link is calculated by subtracting the scheduled departure time at each link from the actual departure time. The expected delay for each train on each link is calculated by dividing the cumulated delays for all repetitions of the simulation by the total number of repetitions H1.

ACKNOWLEDGMENTS

THE AUTHORS WOULD LIKE to thank Rohan Schippan of Queensland Rail for supplying the necessary knowledge and data to conduct the real life application and the anonymous reviewers for their helpful comments to earlier versions of the paper.

REFERENCES

CAREY, M., AND A. KWIECINSKI, "Stochastic Approximation to the Effects of Headways on Knock-on Delays of Trains," *Transportation Research B* **28,** 251–267 (1994).

CHEN, B., AND P. HARKER, "Two Moments Estimation of Delay on Single-Track Rail Lines with Scheduled Traffic," *Transportation Science* **24,** 261–275 (1990).

CHENG, Y., "A Network Based Train Traffic Simulation with Changeable Rescheduling Strategies," in *Proceedings of the Second International Symposium on Parallel Architectures, Algorithms, and Networks*, IEEE Computer Society Press, Los Alamitos, California, 214–220 (1996).

GROSS, D., AND C. M. HARRIS, *Fundamentals of Queuing Theory*, Wiley, New York, 1985.

HALLOWELL, S., "Optimal Dispatching under Uncertainty: With Applications to Railroad Scheduling," PhD dissertation at the Operations and Information Management Department, Wharton School, University of Pennsylvania, 1993.

HIGGINS, A., L. FERREIRA, AND E. KOZAN, "Modelling Delay Risks Associated with Train Schedules," *Transportation Planning and Technology* **19,** 89–108 (1995).

HIGGINS, A., "Optimisation of Train Schedules to Minimise Transit Time and Maximise Reliability," PhD dissertation at the Queensland University of Technology, Brisbane, Australia, 1996.

KRAFT, E. R., "Analytical Models for Rail Line Capacity Analysis," *Transportation Research Forum Proceedings* **24,** 153–162 (1983).

LIDEN, T., "The New Train Traffic Simulation Program," *ASU Newsletter* **21,** 2–15 (1993).

ORTEGA, J. M., AND W. C. RHEINBOLDT, *Iterative Solution of Non-linear Equations in Several Variables*, Academic Press, New York, 1970.

OZEKICI, S., AND S. SENGOR, "On a Rail Transportation Model with Scheduled Services," *Transportation Science* **28,** 246–255 (1994).

PETERSEN, E. R., "Over the Road Transit Time for a Single Track Railway," *Transportation Science* **8,** 65–74 (1974).

PETERSEN, E. R., AND A. J. TAYLOR, "Design of a Single-Track Rail Line For High Speed Trains," *Transportation Research A* **21,** 47–57 (1987).

PIDD, M., *Computer Simulation in Management Science*, Wiley, New York, 1984.

TAKAGI, H., *Queuing Analysis: A Foundation of Performance Evaluation*, North Holland, Amsterdam, 1991.

WARDROP, A. W., A. G. SWIFT, AND P. G. MATHESON, "Timetable Capture and Train Flow Modeling," Transport Management Group International Internal Report, Australia, 1993.

WOODFORD, C., *Solving Linear and Non-linear Equations*, Ellis Horwood, London, 1992.

(Received: October 1996; revisions received: October 1997; accepted: October 1997)