Pergamon

Transpn Res.-A, Vol. 32, No. 1, pp. 19–38, 1998

0965-8564/98 $19.00+0.00

PII: S0965-8564(97)00016–5

3423

SERVICE OPTIMIZATION FOR BUS CORRIDORS WITH SHORT-TURN STRATEGIES AND VARIABLE VEHICLE SIZE

PAOLO DELLE SITE and FRANCESCO FILIPPI*
DITS Dipartimento Idraulica, Trasporti e Strade, Università di Roma La Sapienza, Via Eudossiana 18, 00184 Rome, Italy

(Received 11 June 1996; in revised form 12 February 1997)

Abstract—An optimization framework for intermediate-level planning of bus operations on a given corridor is presented. Service patterns over different operation periods, which include short-turn strategies (full-length and short-turn line) and variable vehicle size, are considered. Variables optimized are: positions where turn-backs are operated, vehicle size, frequencies (by operation period) and, where bus arrivals are regular, relative offset (by operation period) of full-length and short-turn line, and fare. The objective function is the net benefit represented by benefits that accrue to the users for lower average waiting times minus operator's costs not covered by fares. A numerical procedure to solve the problem is provided. An application of it to a case representative of radial corridor conditions shows the effects on service patterns and the trade-offs over operation periods between users' and operator's costs, stemming from different design criteria according to the weights attached to the two players. © 1998 Elsevier Science Ltd. All rights reserved.

1. INTRODUCTION

Considerable interest in service provision along corridors is prompted by the economies of scale of transit services, with respect to market size, which stem from concentration. The term corridor is here used to refer to the heavy-demand linear element that, in the approach to transit network redesign based on transit centers (Schneider and Smith, 1981), represents a link of the primary network which connects two transit centers and is fed by local bus services. Heuristics can be used (as in Filippi and Gori, 1991) to determine the layout and the technology (bus/rail) on corridors.

On bus corridors, productivity can be increased by the use of routing and scheduling strategies tailored to the demand pattern, replacing the conventional local service. These include short-turning and zonal services, of which a synthetic review is found in Furth and Day (1985) where the most favorable conditions for application are provided.

In the base configuration, the short-turn strategy consists of a system of two service patterns, in which the short-turn line is entirely overlapped by the longer full-length line. The short-turn pattern covers the more heavily used part of the corridor. The strategy can be extended to include more than two service patterns. The patterns can have the same terminus at one end of the corridor with turnbacks at different distances, or turnbacks at both ends.

In current practice the strategy is meant as scheduling adjustment to reduce the number of vehicles required. Productivity is enhanced by higher occupancy ratios and lower capital and operating costs. Passengers who are not served by the short-turn line, that is, the full-length market, increase waiting time, while those who can choose between the full-length and short-turn line, the choice market, do not change the service level. The use of variable vehicle size instead of a fixed vehicle capacity provides an additional choice dimension in designing the service.

The criterion often adopted for urban bus services, which aim for a given line capacity at minimal cost to the operator, has led to a gradual increase in bus size, mainly to offset the high incidence of personnel costs in operating costs. This approach ignores the user benefit possible from smaller buses, which can provide the same line capacity with a more frequent service. Further advantages often claimed for smaller buses are higher average speed, from which both users and operator can benefit, and lower unit capital and operating costs.

Design of short-turn strategies is addressed by Furth (1988), Vijayaraghavan (1988), Ceder (1989), Vijayaraghavan and Anantharamaiah (1995). Vijayaraghavan (1988) and Vijayaraghavan

*Author for correspondence.

and Anantharamaiah (1995) provide a methodology for reducing the fleet requirement on a bus route by inserting short-turn trips taking the headway as input.

Furth (1988) and Ceder (1989) formulate an optimal design problem where the set of decision variables includes headway and (Furth only) vehicle size. Both methodologies provide the minimum required fleet size. In the procedure proposed by Ceder, allowance for level of service improvements is made by a further step where the number of short-turn trips is minimized, provided that the minimum fleet size computed is maintained. In the approach proposed by Furth, the search for an optimal trade-off between fleet size and waiting time can be accomplished only by comparing values obtained with different combinations of feasible design parameters.

Ceder's approach considers aggregate volumes and capacity without providing a model of passenger behavior in the case of overlapping service patterns.

Furth's approach includes a model to provide loads on the two patterns (short-turn and full-length) in the case of regular bus arrivals only. Furth claims regularity as an operating requirement to balance the loads on the two lines, i.e. to avoid overcrowding. The relative offset, i.e. the time elapsed since the previous trip of either lines, determines the loads on the two lines (as well as the average waiting time), and is hence taken as decision variable of the design problem. Moreover, for easier schedule coordination, the frequency of the short-turn line is set to a multiple of that of the full-length line (the ratio of the two frequencies, e.g. 1:1, is called the scheduling mode).

In both Ceder's and Furth's optimization models, demand is constant, i.e. insensitive to service level and fare, and the analysis is restricted to a single period of operation.

Work on optimal bus size has mostly been restricted to the analysis of a single line which is treated independently from other lines of the network, as in Jansson (1980), or to the search for the average vehicle size of a network, as in Oldfield and Bly (1988). Jansson (1980) considers constant demand and two operation periods, peak and off-peak, Oldfield and Bly (1988) elastic demand and a single operation period. Both provide solutions analytically. It is shown that minimization of total (operator's plus users') costs (Jansson, 1980), or maximization of the net benefit [users' and external benefits minus operator's costs not covered by fares and external costs (Oldfield and Bly, 1988)], would result in a smaller bus size than that now used. The result holds even in the conservative hypothesis where smaller buses have the same speed as larger buses (Oldfield and Bly, 1988).

Recently Shih and Mahmassani (1994) have proposed a procedure for systemwide optimization of vehicle size where demand on each line varies, according to an assignment model, with the optimal bus sizes and associated frequencies of other lines. As in the above bus size optimization models, the total (users' plus operator's) cost is adopted as objective function. One period of operation is considered.

The optimal solution they achieve is provided by a heuristic due to the impossibility, stemming from the assignment problem, of explicitly expressing the users' costs on a line-by-line basis. Vehicle size is treated as a continuum, and the solution with a limited set of commercially available bus sizes is obtained by a further, nearest feasible integer heuristic.

The approach here extends the objective of total costs adopted in the works on optimal bus size to the design of bus service with short-turn strategies. The configuration with two lines, short-turn and full-length, and turnbacks at both ends of the corridor is considered. Vehicle size is variable for both lines. Different periods of operation are considered.

For a given capacity at peak volume point, the use of a short-turn line, the vehicle size unchanged, reduces operator's costs but increases waiting time costs due to the penalty imposed on the full-length market. When the lowest capacity consistent with demand is provided at peak volume point in each operation period, the operator minimizes his costs. Solutions where benefits accrue to all corridor users is obtained with smaller bus sizes and greater frequencies all along the corridor over the whole operation time.

Between these two extremes lies an optimal trade-off to maximize the net benefit, the difference between users' benefits and operator's costs not covered by fares (elastic corridor demand), or minimize total users' and operator's costs (constant corridor demand). This trade-off is sought here by setting up a multi-period optimization model, where positions of turn-backs, frequencies and vehicle size for both lines (full-length and short-term), and average fare (in case of elastic demand) are the design variables.

Both the cases of random (Poisson distributed) and regular bus arrivals are considered. The former is likely to occur in mixed-traffic, heavily congested corridors. The latter requires schedule coordination to be maintained. For this, frequencies of the short-turn line are set to a multiple of those of the full-length line, as in Furth (1988), and scheduling mode and relative offset are added to the set of design variables. A model of passenger behavior provides the loads on both lines and the average waiting time in both cases. Frequencies and, in the case of regular bus arrivals, scheduling mode and relative offset, differentiate by operation period, thus allowing the short-turn line to be operated in certain periods only, while the settings of the other variables are unchanged over the whole operation time.

The extension to a multi-period problem is not trivial, since interdependencies between operation periods which stem on the operator's side (variables taking the same value over the whole operation time, financial constraint binding over the whole operation time, bus capital costs non-additive over operation periods) make the optimization problem not separable into single-period subproblems. Variables representing frequency, relative offset (in the case of regular bus arrivals), and fare are treated as a continuum. The set of feasible vehicle sizes restricts to that available to the operator.

Since the problem addressed deals with two lines only, the drawbacks originating from the attempt of a systemwide optimization that have previously been highlighted (Shih and Mahmassani, 1994) are avoided here: (i) the maximization of net benefit (minimization of total costs) is explicitly sought since the assignment problem reduces to two lines only and hence the objective function is separable by line, and (ii) the optimal solution on the set of feasible vehicle sizes is achieved by complete enumeration. The optimization is carried out by a numerical method.

The paper continues as follows. Section 2 presents the optimal design problem with the solution procedure. Section 3 describes the application to a hypothetical but realistic case referable to a radial corridor. Section 4 concludes the paper, emphasizing contributions provided and highlighting directions of research.

2. THE OPTIMIZATION PROBLEM

2.1. Framework and design variables

The problem is formulated according to the notation listed in Appendix A. The design problem is set up for a bus corridor with given stops where no other routes overlap, in a time of operation which is subdivided into periods of length t_i, $i \in I$, I being the set of operation periods. Over each operation period i, demand, represented by the origin–destination (OD) matrix, and operating arrangements are constant.

The operating arrangements affect both the level of service (represented by the average generalized cost) perceived by the users and hence demand, which is supposed to be elastic, and the operator's costs. The design objective includes users' and operator's perspectives and is based on the assessment of costs which accrue to both players. The design is subject to a set of operational and financial constraints.

The service is operated with two lines: a full-length line (subscript $h = 1$), traveling the whole length of the corridor, and a short-turn line (subscript $h = 2$) with turnbacks at both ends of the corridor.

Depending on the distribution of bus arrivals, the following cases are considered:

(a) random arrivals with exponential distribution of headways, i.e. the arrivals are viewed as a Poisson process, with arrivals of one line independent of those of the other line;
(b) regular arrivals with frequency of the short-turn line set to an integer multiple m_i (called scheduling mode) of that of the full-length line in each operation period; thus, between two trips of the full-length line there will be m_i trips of the short-turn line. The offsets of multiple trips of the short-turn line are equal to approach the condition where loads are balanced (which is strictly fulfilled when passenger arrivals have a uniform random distribution).

Variables and formulas relevant to one case only will hereafter be indicated by the extensions 'a' and 'b', respectively. The decision variables of the optimization problem represent the operating arrangements and include the following:

(1) stops where turnbacks occur, which remain unchanged over the set I of operation periods; these define the sets, R_h, $h = 1, 2$, of corridor arcs served by each line (R_1 always includes all corridor arcs);

(2) bus size of each line h, also unchanged over the set I, represented by unit capacity (total number of passenger spaces, sitting and standing) z_h, $h=1,2$ measured in spaces per bus hereafter sps/bus);
(3a) bus frequency of each line h, differentiated by operation period, $f_{h,i}$, $h = 1, 2$, $i \in I$ (in bus/h);
(3b) bus frequency of the full-length line by operation period, f_i, $i \in I$, scheduling mode by operation period m_i, $i \in I$, and relative offset, i.e. time interval between a full-length trip and the preceding short-turn trip divided by headway (reciprocal of frequency) by operation period ϕ_i, $i \in I$;
(4) variables representative of the fare level: these include the base charge τ_1 (in money unit per passenger, hereafter abbreviated pr), equal to the whole fare in case of flat fare, and the rate τ_2 of increase with length traveled (in money unit/km/pr), which is greater than zero in case of distance-related fare.

Variables (2) are assumed to take discrete values only, corresponding to the finite number of bus sizes available to the operator; variables in (3a), (3b) except for scheduling modes m_i, $i \in I$, which are integer, and (4) are taken as a continuum. The relative offset ϕ_i, $i \in I$ is defined in the interval [0,1].

2.2. Users' costs and demand functions

The average generalized cost $G_{i,\mathrm{OD}}$, $i \in I$, OD$\in P$, P being the set of OD pairs, includes the cost of time components, walking to and from the service, waiting, riding and transfer (if any), and the fare cost.

Walking and riding times are taken as input independent of the optimizable operating parameters. Walking to and from the service is a constant that depends on the relative position of stops and catchment areas. Riding time depends on operating speed which, if the effects of variations of dwell times are negligible, is independent of demand. Operating speed can be assumed independent of bus size also, if, besides constant dwell times, it is assumed that differences in acceleration and maneuverability produce negligible effects. The hypothesis of equal operating speed for both lines is consistent with that of relative offset equal at every stop [case (b) of regular bus arrivals]. The operating speed differentiates by operation period and direction to take into account different traffic conditions.

Waiting time depends on the hypotheses made on passenger behavior. The full-length market, i.e. passengers whose origin and destination are connected without transfer by the full-length line only, is entirely assigned to the full-length line. The choice market, i.e. the rest of the market, is assigned to either the full-length or the short-turn line, the passengers taking the first bus, since, in the hypotheses adopted, the level of service, except for waiting time, is not differentiated by line.

Passengers are assumed to arrive randomly, according to a uniform distribution, regardless of the arrival time of the buses. This hypothesis is suitable for heavily congested and high-frequency corridors where schedule followers are not a major concern. In any case the hypothesis is consistent within a frequency design framework to avoid underestimates of users costs because of the need to account for the schedule delay, even in the off-peaks, relative to the actual time schedule followers would have wanted to depart. Under this hypothesis overcrowding is avoided when bus arrivals are regular if sufficient capacity is provided, thus allowing all passengers to board the first bus. Moreover, the hypothesis ensures equal loads on each trip of the short-turn line due to the equal offsets.

In case (a) of random bus arrivals and under the assumption that passengers can always board the first available bus, the average waiting time in operation period i is given (Chriqui and Robillard, 1975) by $1/f_{1,i}$ for the full-length market, $1/(f_{1,i}+f_{2,i})$ for the choice market.

In case (b) of regular bus arrivals, the average waiting time in operation period i is given by $1/(2 \cdot f_i)$ for the full-length market,

$$\left[\phi_i^2 + \frac{(1-\phi_i)^2}{m_i}\right]/(2 \cdot f_i)$$

for the choice market.

Transfer time is dropped for the above hypotheses on behavior of passengers of the full-length market.

Thus, assuming that time components are valued linearly, the following expressions hold for the average generalized cost:

$$G_{i,\mathrm{OD}} = c_u \cdot u + c_w \cdot \frac{1}{f_{1,i} + x_a} + c_r \cdot \frac{l_{\mathrm{OD}}}{v_{i,d}} + \tau_1 + \tau_2 \cdot l_{\mathrm{OD}} \quad i \in I, \ \ \mathrm{OD} \in P \tag{1a}$$

$$x_a = \begin{cases} 0 & \text{if OD} \in M_1 \\ f_{2,i} & \text{if OD} \in M_2 \end{cases} \qquad d = \begin{cases} 1 & \text{if OD} \in P_1 \\ 2 & \text{if OD} \in P_2 \end{cases}$$

$$G_{i,\mathrm{OD}} = c_u \cdot u + c_w \cdot x_b \cdot \frac{1}{2 \cdot f_i} + c_r \cdot \frac{l_{\mathrm{OD}}}{v_{i,d}} + \tau_1 + \tau_2 \cdot l_{\mathrm{OD}} \quad i \in I, \ \ \mathrm{OD} \in P \tag{1b}$$

$$x_b = \begin{cases} 1 & \text{if OD} \in M_1 \\ \phi_i^2 + \frac{(1-\phi_i)^2}{m_i} & \text{if OD} \in M_2 \end{cases} \qquad d = \begin{cases} 1 & \text{if OD} \in P_1 \\ 2 & \text{if OD} \in P_2 \end{cases}$$

where the new symbols introduced represent:

c_u = unit walking time value
c_w = unit waiting time value
c_r = unit riding time value
u = average walking time to and from the service
l_{OD} = OD trip length
$v_{i,d}$ = operating speed in operation period i and direction d
M_1,M_2 = set of OD pairs of, respectively, full-length and choice market
P_1,P_2 = set of OD pairs of, respectively, outward and return direction

Demand is modeled under the hypothesis that trip distribution is not affected by changes in the operating arrangements, and the new demand for each OD pair is either diverted from other modes or newly generated. Thus, demand of each OD depends on average generalized cost of the single OD only, i.e. separable demand functions are adopted.

Demand flows $D_{i,\mathrm{OD}}$, $i \in I$, OD$\in P$, are expressed according to demand functions with constant elasticity e_i, $i \in I$, with respect to average generalized cost:

$$D_{i,\mathrm{OD}} = D_{i\,\mathrm{OD}}^0 \cdot \left(\frac{G_{i,\mathrm{OD}}}{G_{i\,\mathrm{OD}}^0} \right)^{e_i} \quad e_i \leqslant 0, \quad i \in I, \quad \mathrm{OD} \in P \tag{2}$$

where quantities with superscript 0 denote values for the base (initial) operation, e.g. with one line, standard bus size and minimum frequencies consistent with demand. For $e_i = 0$, $i \in I$, demand is constant. Compared with other frequently used public transport demand functions, say the exponential function, the functional form adopted makes it possible to vary the elasticity explicitly and, hence, renders straightforward any sensitivity analysis that should be needed in the absence of data. For the hypotheses set forth on the average generalized cost, walking and riding time keep constant values, equal to those in base operation.

2.3. Operator's costs

The operator's costs can be allocated to each line h, as is common practice, as follows:

(1) fleet size, i.e. maximum number of buses required for the service N_h;
(2) total bus-km operated $\sum_{i \in I} K_{h,i}$, $K_{h,i}$ being the number of bus-km in operation period i;
(3) total bus-hours operated $\sum_{i \in I} H_{h,i}$,$H_{h,i}$ being the number of bus-hours in operation period i.

The cost categories and the corresponding items which are usually assigned to the above factors (overheads being assumed of no relevance here) are, respectively:

(1) fixed costs—bus capital costs (interest paid and vehicle depreciation), bus insurance and road tax;
(2) running costs—consumption (fuel, lubricant, tires), bus maintenance, spare parts and repairs;
(3) personnel costs—bus crew.

The sum of (2) and (3) is also referred to as operating costs.

Thus, under the assumption of absence of economies of scale (constant unit costs) and if the fleet size in each operation period is not restricted to taking integer values because of the possibility of interlining with other routes sharing one corridor's terminus, the operator's costs C for both lines are given by:

$$C = \sum_{h=1,2} \left(a_h \cdot N_h + b_h \cdot \sum_{i \in I} K_{h,i} + c \cdot \sum_{i \in I} H_{h,i} \right) \tag{3}$$

$$N_h = \max_{i \in I} \left\{ f_{h,i} \cdot \left(2 \cdot \vartheta + \sum_{d=1,2} \frac{L_{h,d}}{v_{i,d}} \right) \right\} \quad h = 1,2 \tag{3.1a}$$

$$K_{h,i} = f_{h,i} \cdot t_i \cdot \sum_{d=1,2} L_{h,d} \quad h = 1,2, \quad i \in I \tag{3.2a}$$

$$H_{h,i} = f_{h,i} \cdot t_i \cdot \left(2 \cdot \vartheta + \sum_{d=1,2} \frac{L_{h,d}}{v_{i,d}} \right) \quad h = 1,2, \quad i \in I \tag{3.3a}$$

$$N_h = \max_{i \in I} \left\{ \omega \cdot f_i \cdot \left(2 \cdot \vartheta + \sum_{d=1,2} \frac{L_{h,d}}{v_{i,d}} \right) \right\} \quad h = 1,2 \tag{3.1b}$$

$$K_{h,i} = \omega \cdot f_i \cdot t_i \cdot \sum_{d=1,2} L_{h,d} \quad h = 1,2, \quad i \in I \tag{3.2b}$$

$$H_{h,i} = \omega \cdot f_i \cdot t_i \cdot \left(2 \cdot \vartheta + \sum_{d=1,2} \frac{L_{h,d}}{v_{i,d}} \right) \quad h = 1,2, \quad i \in I \tag{3.3b}$$

$$\omega = \begin{cases} 1 & \text{if } h = 1 \\ 0 & \text{if } h = 2 \text{ and } \phi_i = 1 \\ m_i & \text{if } h = 2 \text{ and } \phi_i \neq 1 \end{cases}$$

where the new symbols introduced represent:

a_h, b_h, c = unit costs, respectively fixed, running and personnel, of which the first two differentiate by line to take into account dependence of fixed and running costs on bus size;
ϑ = layover time;
$L_{h,d}$ = length of line h in direction d,

and t_i is expressed in hours.

The variable ω takes into account that, in case (b) of regular bus arrivals, the short-turn line is operated, with a frequency $m_i \cdot f_i$, only for $\phi_i \neq 1$. In fact, for the hypotheses set forth on passenger behavior, when the relative offset takes the unit value the short-turn line is never used (the whole choice market boards the full-length line).

The model adopted, where fixed costs are represented in a separate term, takes into account the higher incidence of fixed costs on total unit cost per vehicle-km for buses with lower usage, as experienced by bus operators.

2.4. *Objective function*

The objective function of the optimization problem is the net benefit given by the users' benefits minus the operator's costs for both lines not covered by the fares (fares only represent a transfer between both players). External effects are not considered for the limited scale of intervention.

The users' benefit generated by a drop in average generalized cost from G^0 to G and by the resulting increase in demand from D^0 to D, can be measured by the variation of the Marshallian consumer surplus, given by the integral $-\int_{G^0}^{G} D(G)\cdot dG$, where $D(G)$ represents the demand function. The measure applies to one OD pair and extends straightforwardly to more than one OD pair by summation if, as supposed here [eqn (2)], demand functions are separable (i.e. OD demand depends on its own generalized cost only).

A frequently used measure of users' benefits, which will be adopted here, is given by the so-called rule-of-a-half (RH). This provides for the benefits the expression $\frac{1}{2}\cdot(D^0+D)\cdot(G^0-G)$. The RH can either be seen as approximation, which provides a good measure in case of marginal changes of costs, of the Marshallian consumer surplus variation, or can be justified on a purely intuitive basis (the issue is discussed in Jara-Díaz and Farah, 1988). In the hypotheses about demand functions set forth above, the RH extends straightforwardly by summation to more than one OD pair.

Thus, in the case here, the RH yields for the users' benefits U:

$$U=\frac{1}{2}\cdot\sum_{i\in I}\sum_{\mathrm{OD}\in P}\left(D^0_{i,\mathrm{OD}}+D_{i,\mathrm{OD}}\right)\cdot\left(G^0_{i,\mathrm{OD}}-G_{i,\mathrm{OD}}\right) \qquad (4)$$

with average generalized cost $G_{i,\mathrm{OD}}$ and demand $D_{i,\mathrm{OD}}$, $i\in I$, $\mathrm{OD}\in P$, given by, respectively, eqns (1) and (2).

The adopted expression of users' benefits reduces to total generalized cost variation when demand is inelastic, and also applies for ODs that experience increases in generalized cost.

Fare revenues F in the operation time are given by:

$$F=\sum_{i\in I}\sum_{\mathrm{OD}\in P}(\tau_1+\tau_2\cdot l_{\mathrm{OD}})\cdot D_{i,\mathrm{OD}} \qquad (5)$$

Thus, the net benefit B in the operation time, representing the design objective function to be maximized, is expressed by:

$$B=U-(C-F) \qquad (6)$$

with users' benefits U given by eqn (4), operator's costs for both lines C by eqn (3) and fare revenues F by eqn (5). When demand is constant, the objective function B reduces to the sum of users' waiting costs and operator's costs, to be minimized.

2.5. *Constraints*

The feasible settings of short-turn points and of bus sizes of both lines belong to the Cartesian product $T\times Z^2$ where T is the discrete set of stop pairs where turn-backs can be operated and Z the discrete set of unit capacities corresponding to the bus types available to the operator.

Operational constraints determine lower bounds on frequencies, which must satisfy the capacity constraint and, for the full-length line, cannot be lower, in any case, than the minimum policy value. The capacity constraint imposes that, on each arc of either line, passenger flows cannot exceed the offered capacity. This is given by frequency times unit capacity, if any maximum allowable load factor adopted to meet comfort considerations is subsumed in the values taken for the unit capacity.

Passengers of the choice market being assumed always to take the first bus, the fraction on either of the two lines of flow that comes from the choice market equals, in case (a) of random bus arrivals, the probability that the first bus belongs to the line. With uniformly distributed random passenger arrivals, the probability that the first bus belongs to line h is given, in operation period i, by (Chriqui and Robillard, 1975) $f_{h,i}/(f_{1,i}+f_{2,i})$.

In case (b) of regular bus arrivals and with uniformly distributed random passenger arrivals, the fraction of the choice market that boards, in operation period i, the full-length line is ϕ_i, that boarding the short-turn line is $(1-\phi_i)$.

Thus, operational constraints are expressed by:

$$\begin{array}{l} f_{1,i} \geqslant \max\{q_{1,i}, f_0\} \\ f_{2,i} \geqslant q_{2,i} \end{array} \quad i \in I \tag{7a}$$

$$q_{h,i} = \max_{r \in R_h} \left\{ \frac{Q_{h,r,i}}{z_h \cdot t_i} = \frac{y \cdot \sum_{OD \in M_{1,r}} D_{i,OD} + \frac{f_{h,i}}{f_{1,i}+f_{2,i}} \cdot \sum_{OD \in M_{2,r}} D_{i,OD}}{z_h \cdot t_i} \right\} \quad h = 1, 2, \quad i \in I \tag{7.1a}$$

$$y = \begin{cases} 1 & \text{if } h = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{array}{l} f_i \geqslant \max\{q_{1,i}, f_0\} \\ m_i \cdot f_i \geqslant q_{2,1} \end{array} \quad i \in I \tag{7b}$$

$$q_{h,i} = \max_{r \in R_h} \left\{ \frac{Q_{h,r,i}}{z_h \cdot t_i} = \frac{y \cdot \sum_{OD \in M_{1,r}} D_{i,OD} + y_b \cdot \sum_{OD \in M_{2,r}} D_{i,OD}}{z_h \cdot t_i} \right\} \quad h = 1, 2, \quad i \in I \tag{7.1b}$$

$$y = \begin{cases} 1 & \text{if } h = 1 \\ 0 & \text{otherwise} \end{cases} \quad y_b = \begin{cases} \phi_i & \text{if } h = 1 \\ 1 - \phi_i & \text{otherwise} \end{cases}$$

where the new symbols introduced represent:

$q_{h,i}$ = minimum frequency of line h in operation period i required by capacity constraint
f_o = minimum policy frequency
r = subscript for arc
$Q_{h,r,i}$ = flow on arc r of line h in operation period i
$M_{1,r}, M_{2,r}$ = set of OD pairs of, respectively, full-length and choice market, with origin before arc r and destination after arc r

The set of constraints is completed by the operator's financial constraint which imposes an upper bound on the operating ratio (ratio of costs C for both lines to the fare revenue F in the operation time), if subsidy is given as percentage s of costs:

$$\frac{C}{F} \leqslant \frac{1}{1-s} \tag{8}$$

or, alternatively, on the deficit, if subsidy is given as fixed sum S:

$$C - F \leqslant S \tag{9}$$

with operator's costs C and fare revenues F given by, respectively, eqns (3) and (5).

2.6. Solution procedure

The optimal solution is found by breaking the problem down into subproblems, each corresponding to the search for the optimal values of the variables taken as a continuum for a given setting of the remaining variables.

In case (a) of random bus arrivals, each subproblem consists in finding the optimal frequencies $f_{h,i}$, $h = 1, 2$, $i \in I$, and fare variables τ_1, τ_2, for one of the feasible combinations of stops where turnbacks can occur, represented by the arc sets R_h, $h = 1, 2$, and of unit capacities z_h, $h = 1, 2$.

In case (b) of regular bus arrivals, each subproblem consists in finding the optimal frequencies of the full-length line f_i, $i \in I$, the relative offset ϕ_i, $i \in I$, and the fare variables τ_1, τ_2, for one of the feasible combinations of stops where turnbacks can occur, unit capacities z_h, $h = 1, 2$, and scheduling modes m_i, $i \in I$.

The number of subproblems generated is given, in case (a) of random bus arrivals, by the cardinality $| T \times Z^2 |$ of the set of feasible combinations of turnbacks and bus sizes, in case (b) of regular arrivals, by $| T \times Z^2 |$ times the number of feasible combinations of scheduling modes $n^{|I|}$, where n is the maximum integer for the scheduling mode and $| I |$ the cardinality of the set of operation periods. In both cases the number of subproblems to be solved is limited, and hence the breaking down is here suitable for real-size problems, owing to the reduced magnitude of the number of stop pairs where turnbacks can be operated, available bus types and reasonable scheduling modes. Hence solution is given by the maximum value of the objective function over the set of subproblems generated.

For given settings of arc sets, $R_h, h = 1, 2$, unit capacities $z_h, h = 1, 2$, and, in the case (b) of regular bus arrivals, scheduling modes m_i,$i \in I$, the problems are:

$$\begin{cases} \max B[f_{h,i}(h = 1, 2, \quad i \in I), \tau_1, \tau_2] \\ \text{s.t.}: \\ g_j[f_{h,i}(h = 1, 2, \quad i \in I), \tau_1, \tau_2] \leqslant 0 \quad j \in J \\ f_{h,i} \geqslant 0 \quad h = 1, 2, \quad i \in I \\ \tau_1, \tau_2 \geqslant 0 \end{cases} \tag{10a}$$

$$\begin{cases} \max B[f_i(i \in I), \phi_i(i \in I), \tau_1, \tau_2] \\ \text{s.t.}: \\ g_j[f_i(i \in I), \phi_i(i \in I), \tau_1, \tau_2] \leqslant 0 \quad j \in J \\ f_i \geqslant 0 \quad i \in I \\ 0 \leqslant \phi_i \leqslant 1 \quad i \in I \\ \tau_1, \tau_2 \geqslant 0 \end{cases} \tag{10b}$$

where B is the net benefit given by eqn (6), g_j, $j \in J$, represent operational constraints (7) and financial constraint (8) or (9), and the arguments thereof the decision variables.

Due to the expressions taken [eqns (3.1a) and (b)] by the variables N_h, $h = 1, 2$, representing the fleet size, the optimization problems above have a nonstandard form of the min-max type. It can be proved (see Appendix B), however, that problems (10a) and (b) can be reduced to standard problems by adding N_h, $h = 1, 2$, to the set of decision variables and the conditions:

$$N_h \geqslant f_{h,i} \cdot \left(2 \cdot \vartheta + \sum_{d=1,2} \frac{L_{h,d}}{v_{i,d}} \right) \quad h = 1, 2, \geqslant i \in I \tag{11a}$$

$$N_h \geqslant \omega \cdot f_i \cdot \left(2 \cdot \vartheta + \sum_{d=1,2} \frac{L_{h,d}}{\mu_{i,d}} \right) \quad h = 1, 2, \quad i \in I \tag{11b}$$

$$\omega = \begin{cases} 1 & \text{if } h = 1 \\ 0 & \text{if } h = 2 \text{ and } \phi_i = 1 \\ m_i & \text{if } h = 2 \text{ and } \phi_i \neq 1 \end{cases}$$

to the set of constraints.

Thus, the equivalent problems to be solved are:

$$\begin{cases} \max B'\{f_{h,i}(h = 1, 2, \quad i \in I), N_h(h = 1, 2), \tau_1, \tau_2] \\ \text{s.t.}: \\ g'_j[f_{h,i}(h = 1, 2 \quad i \in I), N_h(h = 1, 2), \tau_1, \tau_2] \leq 0 \quad j \in J' \\ f_{h,i} \geqslant 0 \quad h = 1, 2, \quad i \in I \\ N_h \geqslant 0 \quad h = 1, 2 \\ \tau_1, \tau_2 \geqslant 0 \end{cases} \tag{12a}$$

$$\begin{cases} \max B'[f_i(i \in I), \phi_i(i \in I), N_h(h=1,2), \tau_1, \tau_2] \\ \text{s.t.}: \\ g'_j[f_i(i \in I), \phi_i(i \in I), N_h(h=1,2), \tau_1, \tau_2] \leqslant 0 \quad j \in J' \\ f_i \geqslant 0 \quad i \in I \\ 0 \leqslant \phi_i \leqslant 1 \quad i \in I \\ N_h \geqslant 0 \quad h = 1,2 \\ \tau_1, \tau_2 \geqslant 0 \end{cases} \tag{12b}$$

where the set of constraints J' includes now inequalities (11a) and (b) also.

The penalty function method is applied to the nonlinear constrained problems (12a) and (b) which reduce, as the penalty coefficient ε tends to zero, to the problems:

$$\begin{cases} \min \psi; \left[f_{h,i}(h=1,2, \quad i \in I), N_h(h=1,2), \tau_1, \tau_2, \varepsilon\right] = -B' + \frac{1}{\varepsilon}\sum_{j \in J'} \max\left\{0, g'_j\right\}^2 \\ \text{s.t.}: \\ f_{h,i} \geqslant 0 \quad h = 1,2 \;\; i \in I \\ N_h \geqslant 0 \quad h = 1,2 \\ \tau_1, \tau_2 \geqslant 0 \end{cases} \tag{13a}$$

$$\begin{cases} \min \Psi[f_i(i \in I), \phi_i(i \in I), N_h(h=1,2), \tau_1, \tau_2, \varepsilon] = -B' + \frac{1}{\varepsilon}\sum_{j \in J'} \max\left\{0, g'_j\right\}^2 \\ \text{s.t.}: \\ f_i \geqslant 0 \quad i \in I \\ 0 \leqslant \phi_i \leqslant 1 \quad i \in I \\ N_h \geqslant 0 \quad h = 1,2 \\ \tau_1, \tau_2 \geqslant 0 \end{cases} \tag{13b}$$

where ψ is the penalty function, the arguments of B' and $g'_j, j \in J'$, which are omitted, are the same of problems (12a) and (b), and the problems have been changed into minimization problems by changing the sign of the function B'.

Problems (13a) and (b) are solved by global optimization methods. A technique in the class of controlled random search methods is used. The algorithm implemented, written in FORTRAN and run on IBM-RISC 6000, is an improvement of the Price method (Price, 1978). The method does not require the calculation of the derivatives of the objective function. Solution is provided on a compact set. This can be defined for each problem here by the box constraints obtained by simply adding upper bounds, where needed, on each decision variable.

The Price method generates trial points in the domain of interest that tend to converge upon the region where the function takes the lowest (or highest) values. The improvement (Brachetti *et al.*, 1994) consists of a different modality in the selection of the trial points so as to exploit the information on the objective function acquired during the execution of the algorithm.

3. ILLUSTRATIVE APPLICATION

3.1. *Case study*

The model is applied to a hypothetical but realistic case. The stop density and volume profiles are constructed to represent the frequently found case of radial corridors with one end in CBD. Stops on corridors are aggregated so as to reduce to 10 stops per direction only (1..10 inbound, 10..1 outbound) with arc lengths, equal in both directions, given in Table 1.

The operation time (12 h) refers to the average weekday and is subdivided into three periods, a.m. peak (2 h), off-peak (7 h) and p.m. peak (3 h). Figure 1 shows the corresponding OD matrices (elements above the leading diagonal represent the outward direction, those below the return) and Fig. 2 the volume profiles derived from them. High flows are concentrated on a short part near one end of the corridor (stop 10 corresponding to the CBD) on the a.m. and p.m. peaks, in a direction which reverses in the two peaks. The a.m. peak shows the highest hourly peak volume. The volume profile is almost constant in both directions in the off-peak.

Table 1. Arc lengths

Arc (outward and return)	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10
Length (m)	1600	1500	1100	900	700	600	550	600	450

Values used for supply parameters are based on data provided by the bus operator in Rome. The operating speed is 8 km/h in the peak directions of the a.m. and p.m. peaks, 14 km/h in the other cases. Buses are one-person operated with unit personnel cost of 40,000 lire/h. Three bus sizes have been considered: 40, 100 and 160 sps/bus. The unit fixed and running costs, obtained by linear interpolation on data relating to different buses in operation, are, respectively, for the three sizes considered: 42,000, 78,000 and 114,000 lire/bus and 245, 350 and 455 lire/(bus-km). Fare is flat, with average value of 400 lire/pr. A value of 5 min has been assumed for the layover time.

a.m. peak

	1	2	3	4	5	6	7	8	9	10	to/ from
1	—	29	14	64	4	3	3	1	1	25	1
2	14	—	15	70	4	4	3	1	1	27	2
3	5	5	—	49	3	3	2	1	0	19	3
4	8	7	4	—	18	15	12	4	3	111	4
5	74	63	35	0	—	5	4	1	1	37	5
6	4	4	2	0	0	—	5	2	1	50	6
7	1	1	0	0	0	3	—	20	16	636	7
8	8	6	3	0	0	26	5	—	7	262	8
9	16	14	7	0	0	58	11	0	—	77	9
10	13	11	6	0	0	47	9	0	10	—	10
from /to	1	2	3	4	5	6	7	8	9	10	

off-peak

	1	2	3	4	5	6	7	8	9	10	to/ from
1	—	8	0	7	40	0	0	5	9	11	1
2	12	—	0	9	48	0	0	6	12	13	2
3	4	4	—	4	22	0	0	3	5	6	3
4	7	6	3	—	2	0	0	0	1	1	4
5	62	53	29	0	—	0	0	2	3	3	5
6	3	3	2	0	0	—	0	29	53	62	6
7	1	1	0	0	0	2	—	3	6	7	7
8	6	5	3	0	0	22	4	—	4	4	8
9	13	12	6	0	0	48	9	0	—	12	9
10	11	9	5	0	0	40	7	0	8	—	10
from /to	1	2	3	4	5	6	7	8	9	10	

p.m. peak

	1	2	3	4	5	6	7	8	9	10	to/ from
1	—	8	0	7	40	0	0	5	9	11	1
2	24	—	0	9	48	0	0	6	12	13	2
3	12	12	—	4	22	0	0	3	5	6	3
4	53	58	41	—	2	0	0	0	1	1	4
5	3	4	2	15	—	0	0	2	3	3	5
6	2	3	2	12	4	—	0	29	53	62	6
7	2	3	2	10	3	4	—	3	6	7	7
8	0	1	1	3	1	1	17	—	4	4	8
9	0	1	1	2	1	1	13	5	—	12	9
10	22	23	16	94	31	41	530	219	64	—	10
from /to	1	2	3	4	5	6	7	8	9	10	

Fig. 1. OD matrices by operation period (hourly).

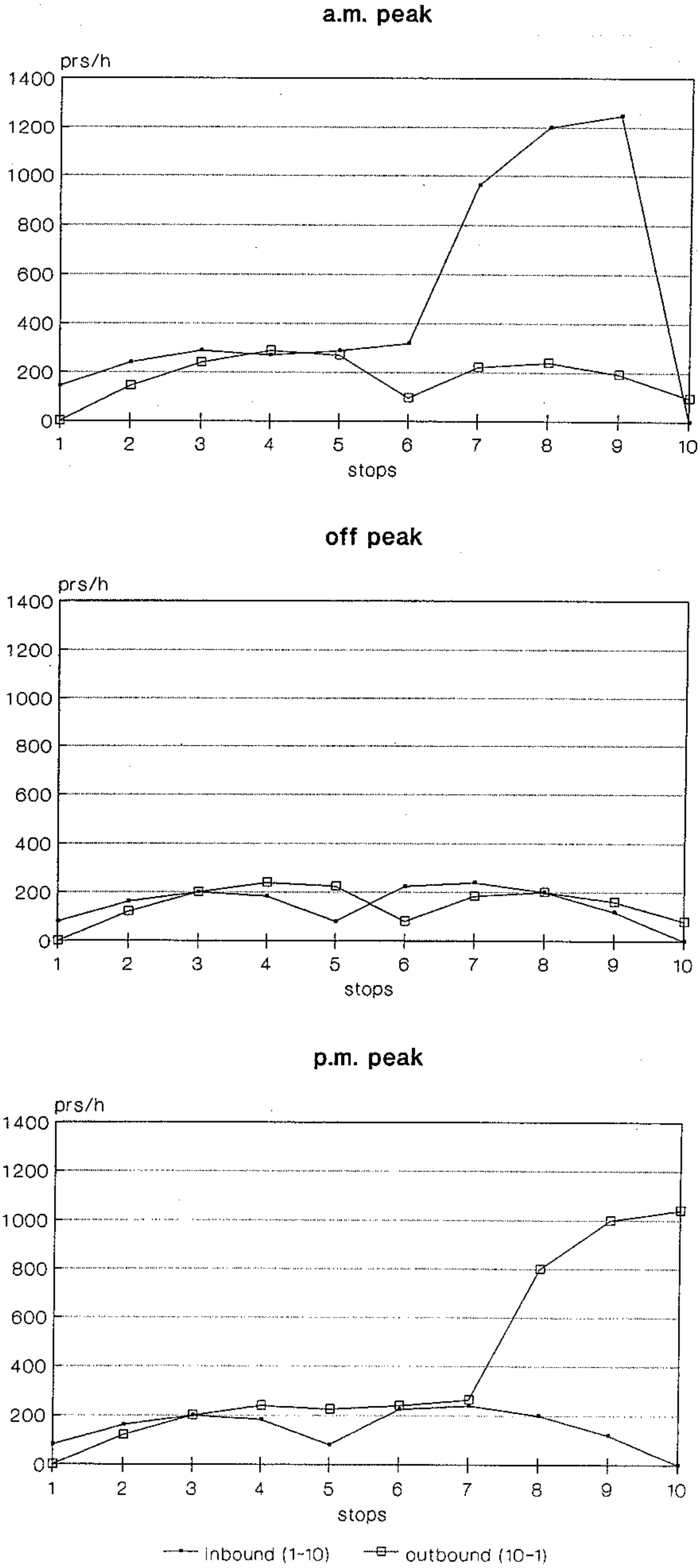

Fig. 2. Volume profiles by operation period (hourly).

The case of random bus arrivals is considered. Base operation assumed for calculation of average generalized costs is that with one line, standard bus size of 100 sps/bus and maximum headways in integer minutes consistent with demand (respectively in the three operation periods 4, 24 and 5 min). Walking time has been dropped from the generalized cost expression.

The model is run under different hypotheses of demand elasticity and unit time costs. Two different values of demand elasticity with respect to average generalized cost, equal over the three operation periods, namely 0 (constant demand) and −0.4, are investigated, each combined with two hypotheses of time values. Both hypotheses assume value of waiting time twice that of riding time: the first assumes respectively 8,000 and 4,000 lire/h, the second half of these values. In the cases with elastic demand (if demand is constant fare cannot be optimized in the framework here) the flat fare system is maintained.

The solution sought is subject to a minimum policy frequency of 3 bus/h and a constraint on the operating (cost/revenue) ratio, as is the practice in urban bus operation in Italy. The maximum of the operating ratio is set to the value calculated for the base operation (1.39), which corresponds to a subsidy equal to 0.28 of costs. Parameter values used are summarized in Tables 2 and 3, with reference to the notation introduced in Section 2.

3.2. Results

Solutions obtained with higher values of user time cost represent operating arrangements which are more favorable to the users; those obtained with lower values represent an operation which is relatively more favorable to the operator. We will refer to the former as the 'users-oriented' (US) solutions, to the latter as the 'operator-oriented' (OP) solutions. Each includes the case with constant demand and that where demand is elastic with respect to average generalized cost.

Because of the demand pattern, which shows a sharp increase of volume profile in both peaks at stop 7 (Fig. 2), the search for the optimum has been limited in all cases to the solutions with turnbacks at stops 7 and 1. Optimal values of the other design variables, bus sizes, frequencies and fare, are shown in Tables 4 (US) and 5 (OP).

Quantities useful for assessment are given in Tables 6 and 8 (US) and Tables 7 and 9 (OP), where values at optimum are compared with those in base operation. Operator costs in base operation are calculated rounding the fleet size in each operation period to the next integer.

In Tables 4 and 5 it can be seen that the short-turn line is operated in the peaks only, with smaller bus sizes (40 sps/bus) and higher frequencies than the full-length line in the US solutions, with the same bus size (100 sps/bus) and lower frequencies in the OP solutions. Fare is higher than the value in base operation in the US solutions, lower in the OP solution. Differences between constant and elastic demand consist of higher frequencies in the elastic demand case in both US and OP solutions.

Tables 6 and 7 give quantities referred to the whole operation time. The average waiting time is calculated as weighted average, with weights equal to market (full-length and short-turn) sizes in different operation periods.

Table 2. Parameter values used: quantities independent of bus size

c	Unit personnel cost	40,000	lire/(h·bus)
c_r	Unit riding time value	4000	lire/(h·pr)
		2000	lire/(h·pr)
c_u	Unit walking time value	0	lire/(h·pr)
c_w	Unit waiting time value	8000	lire/(h·pr)
		4000	lire/(h·pr)
e_1, e_2, e_3	Demand elasticity	0	
		−0.4	
f_0	Minimum policy frequency	3	bus/h
t_1	a.m. Peak length	2	h
t_2	Off-peak length	7	h
t_3	p.m. Peak length	3	h
u	Average walking time	0	h
$v_{1,1}, v_{3,2}$	Operating speed in peak directions of peak periods	8	km/h
$v_{1,2}, v_{2,1}, v_{2,2}, v_{3,1}$	Operating speed other cases	14	km/h
$1/(1-s)$	Maximum operating (costs/revenues) ratio	1.39	
ϑ	Layover time	5	min
τ_2	Rate of increase of fare with length traveled	0	lire/(pr·km)

Variations with respect to base operation show the same sign for both constant and elastic demand cases within each solution (US and OP). Absolute values show minor differences, more favorable to users in the case of elastic demand, to the operator in that of constant demand, in both solutions (US and OP).

Compared with base operation, both US and OP solutions show gains in waiting times and fixed costs, losses in running and personnel costs. In the US solutions, gains and losses result in worse total costs and deficit, while in the OP solutions (where benefits still accrue to the users), total costs and deficit are almost unchanged.

Table 3. Parameter values used: quantities dependent on bus size

z_h	Bus size (sps/bus)	40	100	160
a_h	Unit fixed cost (lire/bus)	42,000	78,000	114,000
b_h	Unit running cost [lire/(bus·km]	245	350	455

Table 4. Optimal bus sizes, frequencies and fare (users-oriented solutions)

Line	Constant demand		Elastic demand	
	Full-length	Short-turn	Full-length	Short-turn
Bus size (sps/bus)	100	40	100	40
Frequency (bus/h)				
a.m. Peak	9.9	14.5	9.7	15.5
Off-peak	10.5	0	12.8	0
p.m. Peak	9.9	10.5	9.7	11.7
Fare (lire/pr)	—		493	

Table 5. Optimal bus sizes, frequencies and fare (users-oriented solutions)

Line	Constant demand		Elastic demand	
	Full-length	Short-turn	Full-length	Short-turn
Bus size (sps/bus)	100	100	100	100
Frequency (bus/h)				
a.m. Peak	7.6	7.4	7.4	7.7
Off-peak	7.4	0	8.8	0
p.m. Peak	7.6	5.6	7.4	6.4
Fare (lire/pr)	—		360	

Table 6. Comparisons between base and optimal operation (users-oriented solutions)

	Base operation	Optimal solutions		Variations (optimal–base)	
		Constant demand	Elastic demand	Constant demand (%)	Elastic demand (%)
Average waiting time (min)	11.8	4.9	4.5	−58.5	−61.9
Passengers (prs/day)	15,100	15,100	17,250	0	+14.2
Revenues (1) (million lire/day)	6.0	—	8.5	—	+42
Fixed costs (2) (million lire/day)	2.1	1.6	1.6	−24	−24
Running costs (3) (million lire/day)	0.5	0.7	0.8	+40	+60
Personnel costs (4) (million lire/day)	5.8	8.4	9.3	+45	+60
Total operator's costs (5=2+3+4) (million lire/day)	8.4	10.7	11.7	+27	+39
Deficit (= 5−1) (million lire/day)	2.4	—	3.2	—	+33

Table 7. Comparisons between base and optimal operation (operators-oriented solutions)

	Base operation	Optimal solutions		Variations (optimal–base)	
		Constant demand	Elastic demand	Constant demand (%)	Elastic demand (%)
Average waiting time (min)	11.8	6.8	6.4	−42.5	−45.8
Passengers (prs/day)	15,100	15,100	16,700	0	+10.6
Revenues (1) (million lire/day)	6.0	—	6.0	—	0
Fixed costs (2) (million lire/day)	2.1	1.3	1.3	−38	−38
Running costs (3) (million lire/day)	0.5	0.5	0.6	0	+20
Personnel costs (4) (million lire/day)	5.8	6.0	6.5	+3	+12
Total operator's costs (5=2+3+4) (million lire/day)	8.4	7.8	8.4	−7	0
Deficit (= 5−1) (million lire/day)	2.4	—	2.4	—	0

Tables 8 and 9 show quantities referred to the single operation period and shed light on the extent to which the operation of the short-turn line contributes to the overall result. The average waiting time is the weighted average with weights equal to market sizes. The average occupancy ratio is calculated as the ratio of p-km carried to s-km offered.

Differences between constant and elastic demand cases within each solution show the same hallmarks reported above for quantities of Tables 6 and 7.

The average waiting time improves on base operation in the off-peak and worsens in the a.m. peak, which shows the highest hourly peak volume and the lowest length (2 h) in both US and OP

Table 8. Comparisons between base and optimal operation: relevant quantities by operation period (users-oriented solutions)

	Base operation	Optimal solutions		Variations (optimal-base)	
		Constant demand	Elastic demand	Constant demand (%)	Elastic demand (%)
Average waiting time* (min)					
a.m. Peak	4	6.1 2.5 4.3	6.2 2.4 4.2	+7.5	+5.0
Off-peak	24	5.7	4.7	−76.2	−80.4
p.m. Peak	5	6.1 2.9 4.5	6.2 2.8 4.4	−10.0	−12.0
Average occupancy ratio					
a.m. Peak	0.208	0.282	0.275	+35.6	+32.2
Off-peak	0.654	0.155	0.171	−76.3	−73.9
p.m. Peak	0.211	0.236	0.235	+11.8	+11.4
Fleet size† (buses)					
a.m. Peak	27	17.1 7.0 24.1	16.8 7.4 24.2	−10.7	−10.4
Off-peak	4	13.8	16.8	+245	+320
p.m. Peak	21	17.1 5.0 22.1	16.8 5.6 22.4	+5.2	+6.7
Operating costs (million lire/day)					
a.m. Peak	2.3	2.0	2.0	−13	−13
Off-peak	1.2	4.3	5.2	+258	+333
p.m. Peak	2.7	2.8	2.9	+4	+7

*in peaks for optimal solutions: full-length market, choice market, average (variations over average).
†in peaks for optimal solutions: full-length line, short-turn line, total (variations over total).

Table 9. Comparisons between base and optimal operation: relevant quantities by operation period (users-oriented solutions)

	Base operation	Optimal solutions		Variations (optimal–base)	
		Constant demand	Elastic demand	Constant demand (%)	Elastic demand (%)
Average waiting time* (min)					
a.m. Peak	4	7.9	8.1	+47.5	+47.5
		4.0	4.0		
		5.9	5.9		
Off-peak	24	8.1	6.8	−66.2	−71.7
p.m. Peak	5	7.9	8.1	+24.0	+24.0
		4.5	4.3		
		6.2	6.2		
Average occupancy ratio					
a.m. Peak	0.208	0.341	0.334	+63.9	+60.6
Off-peak	0.654	0.220	0.236	−66.4	−63.9
p.m. Peak	0.211	0.289	0.285	+37.0	+35.1
Fleet size† (buses)					
a.m. Peak	27	13.2	12.9	−37.8	−38.5
		3.6	3.7		
		16.8	16.6		
Off-peak	4	9.7	11.6	+142.5	+190
p.m. Peak	21	13.2	12.8	−24.3	−24.3
		2.7	3.1		
		15.9	15.9		
Operating costs (million lire/day)					
a.m. Peak	2.3	1.4	1.4	−39	−39
Off-peak	1.2	3.0	3.6	+150	+200
p.m. Peak	2.7	2.1	2.1	−22	−22

*in peaks for optimal solutions: full-length market, choice market, average (variations over average).
†in peaks for optimal solutions: full-length line, short-turn line, total (variations over total).

solutions. In the p.m. peak, which shows a lower hourly peak volume but a higher length (3 h), it worsens in the OP solution, improves in the US solution. This is one case where the short-turn line produces benefits to the users in the period in which it is operated. The full-length market worsens all the times the short-turn line is operated. Thus, in the case where users benefit from a reduction of average waiting time when the short-turn line is operated (p.m. peak in the US solutions), gains accrued to the short-turn market outweigh losses to the full-length.

The average occupancy decreases with respect to base operation in the off-peak, i.e. without short-turn line, but increases in both peaks where the short-turn line is operated.

Variations in fleet size and operating (i.e. running plus personnel) costs follow those of average waiting time. There is always an increase in the off-peak and a decrease in the a.m. peak. The p.m. peak shows decreases in the OP solutions where there is a loss for the users, increases in the US solutions where a gain is accrued to users.

3.3. Discussion

Results should be regarded as only indicative because the case to which the model is applied is hypothetical. They do, however, indicate the type of conclusions that it is possible to draw. The sensitivity analysis with respect to user time values is not exhaustive, even though the solutions investigated, namely those referred as 'users-oriented' and those as 'operator-oriented,' make it possible to highlight the main effects of different weights attached to users' costs in the design criteria.

The base operation, with which the above solutions are compared, can be regarded as a benchmark representing many present service conditions where no allowance is made for users' costs.

Positions of turnbacks investigated have been restricted to one setting only, since the demand pattern is such that an extension of the short-turn line beyond the point where the sharp rise in volume occurs would have provided the users with few benefits compared with the increase in fixed and operating costs that would have been involved. The main findings for a radial corridor, in the hypothesis of random bus arrivals, and with financial constraint on the operating ratio, i.e. subsidy given as a fixed proportion of costs, can be summed up as follows:

(1) the short-turn line is operated in the periods where demand shows a suitable pattern (peaks only);
(2) the more heavily used part of the corridor is served by a short-turn line operated with smaller buses when users' time is relatively more valued (users-oriented solutions), by bus of the same size as those of the full-length line when users' time is relatively less valued;
(3) frequency should be much higher in the off-peak compared with current practice (the same result is provided by Jansson, 1980);
(4) users gain an overall benefit in the range of 60% (users-oriented) and 44% (operator-oriented solutions) of average waiting time;
(5) the operator reduces fixed costs and increases operating costs, resulting in higher total costs (27–39%) in the users-oriented solutions, in total costs which do not change significantly in the operator-oriented solutions;
(6) in the case of elastic demand, both solutions (users-oriented and operator-oriented) show an increase in the overall demand level of almost the same magnitude (10–14%), but with changes in fare opposite in sign (increase when time is relatively more valued, decrease otherwise); this results, the operating ratio remaining unchanged, in higher deficit for the operator (about 33%) in the users-oriented solution, in a deficit that does not change significantly in the operator-oriented solution.

The overall result stems from a trade-off of gains and losses for both players in the different operation periods, as represented in Fig. 3. It is shown that benefits accrue to users in the off-peak, while, except for the p.m. peak in the users-oriented solutions, in the periods when the short-turn line is operated, users' average waiting time worsens. The different service patterns shown by the two peaks are a consequence of both demand patterns and lengths of operation periods, thus the results must be interpreted in light of the fact that, particularly in this regard, they hold strictly for the case here. The increase in waiting time in the peaks, which can reach high percentage values in the operator-oriented solutions (up to 48%), is in any case low in absolute values (a few minutes), owing to the relatively high frequencies. Differences between short-turn and full-length market are also shown in Fig. 3.

The results provided by the model confirm those of models proposed for short-turning design (e.g. Furth, 1988; Ceder, 1989), which can be compared with the case represented in the upper left-hand corner of Fig. 3, and yield original insights in all the other cases.

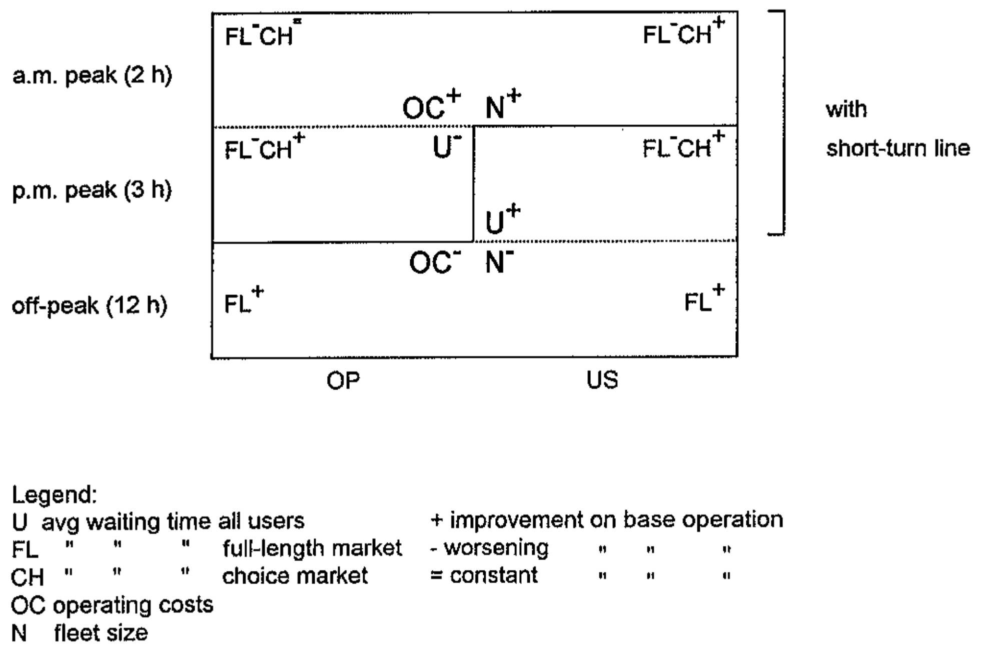

Fig. 3. Optimal trade-offs by operation period with US and OP solutions.

The results that, in the cases of elastic demand, relate to optimal fares are to be interpreted in the light of the relative values of partial elasticities with respect to time and fare, which are proportional to the ratio of the corresponding cost to the total generalized cost. The low absolute value of fare in base conditions makes demand less sensitive to fare changes than to time changes, as also resulted from a number of studies (Cervero, 1990). In both solutions (users-oriented and operator-oriented) users benefit from improvements in waiting times, while increase in demand is almost equal. The greater waiting-time improvement shown by the users-oriented solution comes with a fare increase to maintain the operating ratio at base value. Users can benefit in the operator-oriented solution from both waiting-time improvements, though of lower magnitude, and decreased fare.

The adopted hypothesis on passenger behavior, according to which transfers never occur (induced transfers, which could arise for passengers with origin served by the short-turn line and destination beyond the turn-back point, have been assumed negligible), is validated by the high frequencies resulting for the full-length line in the operation periods where the short-turn line is operated.

Improvements on the solutions could be obtained, when the short-turn line is operated with minibuses, by interlining between the short-turn and full-length line over different operation periods. From an investigation on the maximum occupancy ratio in the off-peak (always less than 0.4), it is seen that it would be possible to operate the full-length line in the off-peak using minibuses also. This would yield a saving in the running costs but not in the bus fleet which is in all cases given by the peak-period requirements.

Further savings in operator's costs could be achieved by including the possibility of deadheading in the off-peak directions. It can be argued, however, that the underestimate of operator's costs implied in the model, where fleet size is not rounded up to the next integer, can account for savings gained not only from interlining with routes having one corridor's terminus in common but also from deadheading.

Further tests have been carried out to assess the marginal effects of including bus size and fare in the set of design variables. The investigation with either bus size or fare kept fixed to the current values has led to suboptimal solutions which, compared with the corresponding optimal solutions, have the following features in common: lower deficit for the operator, in the range 3–18%, lower users' benefits in the range 2–7%, and standard size for buses of both lines. Thus, reducing the number of degrees of freedom in a design framework where both users and operator are included in the objective shifts the solution toward the operator's side.

4. CONCLUSION

The approach to optimal service design on bus corridors presented here provides a more comprehensive framework for operations including short-turning and variable vehicle size than do both common practice and existing theoretical analysis. The model developed is meant as a tool for intermediate-level planning of bus networks, where the variables optimized are then fed into the stage of scheduling design (where timetables are provided). In particular, the approach is valuable for the assessment over different operation periods of different criteria in the design objective according to the weights attached to the operator and to the users. The application to a case study referable to a radial corridor has shown the nature and the wide scope of the outcome that the model can provide.

The framework can be extended, with minor modifications, to the design of bus routes with two lines, each serving only part of the route length. Further, less straightforward, extension relates to the development of a design framework for branching corridors, i.e. tree-shaped bus networks whose root is at one corridor's terminus.

Acknowledgement—The work reported on here was partially supported by CNR (Italian National Research Council) under PFT2 (2nd Research Programme on Transportation), contract No. 93.01836.74.

REFERENCES

Brachetti, P., De Felice Ciccoli, M., Di Pillo, G. and Lucidi, S. (1994) A new version of the Price's algorithm for global optimisation. Report 26-94, Department of Computer Science and Systems, University La Sapienza, Rome.

Ceder, A. (1989) Optimal design of transit short-turn trips. *Transportation Research Record* **1221**, 8–22.

Cervero, R. (1990) Transit pricing research. *Transportation* **17**, 117–139.

Chriqui, C. and Robillard, P. (1975) Common bus lines. *Transportation Science* **9**, 115–121.

Filippi, F. and Gori, S. (1991) Un sistema integrato per la progettazione di un sistema di trasporto pubblico. Atti del III Convegno *SIDT* [An integrated system for urban transit network design, *Proceedings of the 3rd SIDT Conference*], Reggio Calabria, Italy, 10–11 October.

Furth, P. G. (1988) Short turning on transit routes. *Transportation Research Record* **1108**, 42–52.

Furth, P. G. and Day, F. B. (1985) Transit routing and scheduling strateies for heavy demand corridors. *Transportation Research Record* **1011**, 23–26.

Jansson, J. H. (1980) A simple bus line model for optimisation of service frequency and bus size. *Journal of Transport Economic Policy* **14**, 53–80.

Jara-Díaz, S. R. and Farah, M. (1988) Valuation of users' benefits in transport systems. *Transport Review* **8**, 197–218.

Oldfield, R. H. and Bly, P. H. (1988) An analytic investigation of optimal bus size. *Transportation Research* **22**, 319–337.

Price, W. L. (1978) A controlled random search procedure for global optimization. In *Towards Global Optimization 2*, eds L. C. W. Dixon and G. P. Szego, North-Holland, Amsterdam.

Schneider, J. B. and Smith, S. P. (1981) Redesigning urban transit systems: a transit-center-based approach. *Transportation Research Record* **798**, 56–65.

Shih, M. and Mahmassani, H. S. (1994) Vehicle sizing model for bus transit networks. *Transportation Research Record* **1452**, 35–41.

Vijayaraghavan, T. A. S. (1988) Vehicle scheduling in urban transportation with quick and cut trip insertions for fleet size reduction. *Transportation Planning and Technology* **12**, 105–120.

Vijayaraghavan, T. A. S. and Anantharamaiah, K. M. (1995) Fleet assignment strategies in urban transportation using express and partial services. *Transportation Research-A* **29**, 157–171.

APPENDIX A

Notation used

The dimensions and quantities represented are given for each of the symbols used. The following quantities have been assumed as fundamental: bus [B], length [L], passenger [P], time [T] and monetary value [V].

a_h[VB^{-1}]	fixed cost per bus of line h
b_h[VB^{-1}L^{-1}]	running cost per bus-km of line h
$B, B'[V]$	net benefit
$c[VT^{-1}B^{-1}]$	crew cost per bus-h
$C[V]$	operator's costs for the lines
c_u[VT^{-1}P^{-1}]	unit walking time cost
c_r[VT^{-1}P^{-1}]	unit riding time cost
c_w[VT^{-1}P^{-1}]	unit waiting time cost
d	subscript for direction
$D_{i,\mathrm{OD}}$[P]	passengers in operation period i of the OD pair
$D_{i,\mathrm{OD}}$[P]	as above in base operation
e_i	demand elasticity in operation period i
f_o[BT^{-1}]	minimum policy frequency
f_i[BT^{-1}]	frequency of the full length line in operation period i
$f_{h,i}$[BT^{-1}]	frequency of line h in operation period i
$F[V]$	fare revenues
g, g'	optimization problem constraint
$G_{i,\mathrm{OD}}$[VP^{-1}]	average generalized cost in operation period i of the OD pair
$G_{i,\mathrm{OD}}$[VP^{-1}]	as above in base operation
h	subscript for line
$H_{h,i}$[BT]	bus-hours of line h in operation period i
i	operation period subscript
I	set of operation periods
l_{OD}[L]	OD trip length
$L_{h,d}$[L]	length of line h in direction d
$K_{h,i}$[BL]	bus-km of line h in operation period i
m_i	scheduling mode in operation period i
M_1, M_2	OD pairs of, respectively, full-length and choice markets
$M_{1,r}, M_{2,r}$	as above with origin before arc r and destination after arc r
n	maximum value for scheduling mode
N_h[B]	number of buses of line h required for service
OD	subscript for OD pair
P	set of OD pairs
P_1, P_2	OD pairs of, respectively, outward and return direction
$q_{h,i}$[BT^{-1}]	minimum frequency of line h in operation period i required by capacity constraint
$Q_{h,r,i}$[P]	flow on arc r of line h in operation period i
r	subscript for arc

R_h	set of arcs served by line h
s	subsidy as fraction of operator's costs
S[V]	subsidy as fixed sum
T	set of stop pairs where turnbacks can be operated
t_i[T]	length of operation period i
u[T]	walking time to and from the service
U[V]	users' benefits
$v_{i,d}[LT^{-1}]$	operating speed in operation period i and direction d
$z_h[PB^{-1}]$	bus unit capacity of line h
Z	set of available unit capacities
$\in$	penalty coefficient
ϑ[T]	lay-over time
$\tau_1[VP^{-1}]$	base charge
$\tau_2[VL^{-1}P^{-1}]$	rate of increase of fare with length travelled
ϕ_i	relative offset in operation period i
ψ	penalty function

APPENDIX B

Proof of equivalence of problems (10) and (12)

Optimization problems (10a) and (b) differ from problems (12) and (b) only by constraints (11a) and (b), which in problems (12a) and (b) replace the expression (3.1a) and (b) giving the fleet size of each line. Constraints (11a) and (b) reduce to conditions (3.1a) and (b) if there is one operation period for each line such that the corresponding constraint (11a) or (b) is satisfied as an equality. This occurs due to the formulation of the objective function which implies minimization of fixed costs of each line a_hN_h and hence of variables N_h, $h = 1, 2$.