Solving school bus routing problems through integer programming

T Bektaş[1,2]* and Seda Elmastaş[2,3]

[1]*Université de Montréal, HEC Montréal, Quebec, Canada;* [2]*Başkent University, Ankara, Turkey; and* [3]*Bilkent University, Ankara, Turkey*

In this paper, an exact solution approach is described for solving a real-life school bus routing problem (SBRP) for transporting the students of an elementary school throughout central Ankara, Turkey. The problem is modelled as a capacitated and distance constrained open vehicle routing problem and an associated integer linear program is presented. The integer program borrows some well-known inequalities from the vehicle routing problem, which are also shown to be valid for the SBRP under consideration. The optimal solution of the problem is computed using the proposed formulation, resulting in a saving of up to 28.6% in total travelling cost as compared to the current implementation.

1. Introduction

The school bus routing problem (SBRP) is simply concerned with transporting students to schools via public transportation systems (such as buses). This is a problem which has received attention from the scientific community for over 30 years. The generality of the problem makes it very important and scientific ways should be employed to deal with it. However, it is often the case that the routings are planned rather intuitively in real life, which may result in excessive cost for the transportation.

In some ways, the SBRP resembles the vehicle routing problem (VRP), which is extensively studied in the operations research literature (see, eg, the book by Toth and Vigo (2002)). However, as also pointed out by Mandl (1979), there are some specific characteristics of the problem and these are outlined below:

1. The buses in general do not have to return to the school after completing their tours. In specific, they may end their tours at any point other than the depot. Consequently, the routes of the buses in an SBRP will be *paths* as opposed to the *tours* of the VRP.
2. The total number of students each bus carries cannot exceed the capacity of the bus.
3. The length (or time) of each tour is restricted by a certain amount, since the students must be transported to the school before a specific time.

Correspondence: T. Bektaş, Centre for Research on Transportation, Université de Montréal, HEC Montréal, 3000 chemin de la Côte-Sainte-Catherine, Montréal, Canada H3T 2A7.
E-mail: tolga@crt.umontreal.ca

In fact, the problem described above falls into another version of the VRP, namely the open vehicle routing problem (OVRP), which has recently attracted attention of several researchers (see, eg, Fu *et al* (2005), Tarantilis *et al* (2005), Li *et al* (2006), and Repoussis *et al* (2006)).

In this study, we are concerned with an SBRP that arises in transporting the students of an elementary school that is located in central Ankara, Turkey. The school has a contractor firm, which takes care of transporting the students. It is of interest for the contractor firm to minimize the total cost associated with transporting the students to and from the school to their homes. The firm also has to obey the capacity and distance constraints stated above.

The problem considered here corresponds to a capacitated and distance constrained OVRP. In this paper, we present an integer linear programming formulation for this problem which is solved to optimality using a commercial integer programming optimizer.

The paper is organized as follows: We briefly review the related work on the SBRP and the OVRP in the next section. Section 3 formally defines the problem considered in this paper and presents the associated integer linear programming formulation, along with some valid inequalities. Input data and the solution of the integer linear programming formulation are discussed in Section 4. Finally, the last section presents our conclusions.

2. Related work

In this section, we present some previous work related to school bus routing. Our attempt is by no means to provide an

exhaustive review on the subject, but rather to focus on the type of problem and the solution approach considered in each study (For a detailed literature review, the reader is referred to the relevant sections of these studies).

One of the earliest studies on the subject is due to Angel *et al* (1972). In this study, the authors presented an algorithm for a school bus scheduling problem, which minimizes the number of buses and the total distance covered, such that capacity and time constraints are respected. The case study for which the algorithm is developed consists of transporting 1500 students located on a region of about 150 square miles, in Indiana, USA.

Braca *et al* (1997) offered a computerized approach for the transportation of students to multiple schools located in New York, USA. Their problem includes capacity, distance and time window constraints. In addition, they require that a lower bound on the number of students that form a route should also be respected. The problem consists of 4619 students to be picked up from 838 bus stops and transported to 73 schools. The authors proposed a routing algorithm based on the location-based heuristic for the capacitated VRP. One interesting aspect of this study is the estimation of distances and travel times, which are performed via a geographic information systems-based program (MapInfo) and a regression analysis, respectively. The authors also presented two integer programming formulations, namely a set partition model and an assignment-based model, although these do not explicitly include the capacity and distance constraints and are not utilized in the proposed routing algorithm.

A recent study related to the subject is due to Li and Fu (2002). These authors provided planning techniques for a single SBRP in Hong Kong, China, which consists of transporting 86 students located at 54 pick-up points. The problem is of a multi-objective nature, including the minimization of the total number of buses used, the total travel time of all the students, the total bus travel time and balancing the loads and travel times between buses. A heuristic algorithm is proposed to solve the problem. The authors also presented a three-index flow-based integer programming formulation, which, however, is not utilized in the solution algorithm.

The OVRP has recently attracted attention from the operations research community. One of the first studies on the OVRP is due to Sariklis and Powell (2000), who proposed a heuristic to solve the capacity constrained version. Later on, Tarantilis *et al* (2005) described a single parameter metaheuristic algorithm for the problem. To the best of our knowledge, the only exact method to solve the capacitated OVRP is due to Letchford *et al* (2006).

The OVRP may have additional constraints besides the capacity restrictions. In a recent study, Repoussis *et al* (2006) studied the OVRP with capacity and time window constraints and described a greedy look-ahead route construction heuristic algorithm to solve the model that is proposed for the problem. For the OVRP with capacity and distance constraints, Brandão (2004) and Fu *et al* (2005) presented tabu search algorithms, and Li *et al* (2006) described a record-to-record travel heuristic. To our knowledge, there is no exact solution algorithm offered for the solution of the OVRP with distance and/or time window constraints.

In what follows, we break away from the previous studies' problem-specific solution approaches for the SBRP. Our approach and contribution lies in presenting an integer linear program for the problem under consideration, which also has a potential to be used for other similar problems. As the case study is of moderate size, our focus is on well-model building rather than devising a specialized solution procedure. The formal definition of the problem and the corresponding integer programming formulation are presented in the next section.

3. Problem definition and integer programming formulation

The SBRP we consider in this paper is defined as follows: A number of buses are to be used to transport the students between an elementary school and their houses. Each bus is dedicated to a single path and this bus utilizes the same path for both picking-up or dropping-off the students. In addition, each bus has a limited capacity with an upper bound on the total amount of distance it may traverse. Each pick-up or drop-off point is visited only once by a single bus, that is, no partial pick-ups are allowed. The problem lies in finding the minimum number of buses required for the transportation of all the students and their corresponding routes, so as to minimize the total cost of transportation.

We formulate the problem on a complete graph $G = (V, A)$ with $|V|$ nodes and $|A|$ arcs. The distances between each node pair is characterized by a symmetric distance matrix $D := [d_{ij}]$, where d_{ij} represents the distance required to traverse from node i to node j (and also from node j to node i). The node set is partitioned as $V = \{0\} \cup I$, where node 0 is the depot (the school) and I is the set of *intermediate nodes*. Each intermediate node $i \in I$ has a number of students (denoted by q_i) to be picked up. Then, the SBRP consists of determining k node disjoint paths connected to the depot such that the total capacity on each path does not exceed a pre-determined capacity limit (denoted by Q) and the total length of each path does not exceed some amount (denoted by T). A feasible solution to an SBRP resembles a star-like topology on the graph G, where there are k paths all starting from the depot, that is, node 0. Hence, we have a number of paths originating from a single origin point rather than a collection of tours (as in the case of the VRP). Note that this problem corresponds to the OVRP with capacity and distance constraints. Therefore, the following formulation is also valid for this problem.

In order to model the SBRP, we introduce a 'dummy' node d, to which all the last nodes of each path will be connected to. In this case, the problem reduces to finding k node disjoint paths between two points on an expanded graph $G' = (V', A')$,

such that all intermediate nodes are visited exactly once. In the expanded graph, $V' = V \cup \{d\}$ and the new distances in A' are defined as follows:

$$d'_{ij} = \begin{cases} 0 & \text{if } i \in I \text{ and } j = d \\ M & \text{if } i \in 0 \text{ and } j = d \\ d_{ij} & \text{otherwise} \end{cases}$$

Before presenting the integer linear program, we define the following binary variable:

$$x_{ij} = \begin{cases} 1 & \text{if arc } (i, j) \text{ is traversed in the solution} \\ 0 & \text{otherwise} \end{cases}$$

Then, the integer program is constructed as follows:

$$\text{minimize} \quad \sum_{i \in V'} \sum_{j \in V'} c_{ij} x_{ij} + f.k \tag{1}$$

subject to

$$\sum_{i \in I} x_{0i} \leqslant k \tag{2}$$

$$\sum_{i \in I} x_{id} \leqslant k \tag{3}$$

$$\sum_{j \in I \cup \{d\}} x_{ij} = 1 \qquad \forall i \in I \tag{4}$$

$$\sum_{i \in I \cup \{0\}} x_{ij} = 1 \qquad \forall j \in I \tag{5}$$

$$+ \text{ capacity constraints} \tag{6}$$

$$+ \text{ distance constraints} \tag{7}$$

$$x_{ij} \in \{0, 1\} \qquad \forall i, j \in V' \tag{8}$$

In this formulation, the objective function represents the total cost of the travel and the fixed cost of the number of buses used. Here, c_{ij} is the cost of traversing arc (i, j) and is typically a function of the distance as $c_{ij} = \alpha d_{ij}$, with α being the unit distance cost. In the second term, f is the fixed unit cost of dispatching a bus. Constraints (2) and (3) allow at most k vehicles to depart from node 0 and arrive to node d. Constraints (4) and (5) are the degree constraints, which force each intermediate node to be visited exactly once. Capacity and distance restrictions imposed on each bus are denoted by (6) and (7). In what follows, we will present the associated capacity and distance constraints that are polynomial in size and which also ensure that valid paths will be formed.

3.1. Capacity restrictions

By definition, the SBRP requires that the capacity of each bus is respected, that is, no bus can carry more than its capacity. To impose such a restriction, we use the Miller, Tucker, and Zemlin (MTZ)-based constraints (Miller *et al*, 1960) for the

capacitated vehicle routing problem. These constraints are given as follows:

$$u_i - u_j + Q x_{ij} + (Q - q_i - q_j) x_{ji} \leqslant Q - q_j$$
$$\forall i \neq j \in I \tag{9}$$

$$u_i \geqslant q_i \qquad \forall i \in I \tag{10}$$

$$u_i - q_i x_{0i} + Q x_{0i} \leqslant Q \quad \forall i \in I \tag{11}$$

Note that under these constraints, variables x_{ij} and x_{ji} are only defined if $q_i + q_j \leqslant Q$. Constraints (9) and (10) are those presented by Kara *et al* (2004). Constraint (11) is derived specially for the SBRP. In these constraints, the variable u_i represents the total amount of students picked up by the vehicle just after leaving node i. The following propositions show the validity of these constraints.

Proposition 1 *The constraints* (9), (10) *and* (11) *are valid capacity constraints for the SBRP.*

Proof We first observe that constraints (10) and (11) imply that if $i \in I$ is the first node on a path (ie, $x_{0i} = 1$), then $u_i = q_i$. Also observe that if $x_{ij} = 1$ in a solution to SBRP, writing constraints (9) for pairs (i, j) and (j, i) results in $u_j = u_i + q_j$. Now, consider a path $\mathscr{P} := \{i_1, i_2, \ldots, i_l, i_k\}$ in a solution to the SBRP. Writing constraints (9) for each pair of arc in path $\mathscr{P}$ results in $u_{i_k} = q_{i_k} + q_{i_l} + \cdots + q_{i_2} + q_{i_1}$, that is, u_{i_k} represents the total capacity on path $\mathscr{P}$. Since i_k is the last node on the path, constraints (10) implying $u_{i_k} \leqslant Q$ restrict the total capacity of the path by Q. $\square$

We now show that the capacity constraints presented above also prevent the formation of illegal tours that are not connected to the depot, which are named as *subtours* in the routing literature.

Proposition 2 *Constraints* (9) *prevent the formation of subtours within the intermediate nodes.*

Proof It is shown in the proof of Proposition 1 that constraints (9) 'link' all the nodes in a single path via the u_i variables. Now, assume an illegal subtour as (k, l, m, k) where $k, l, m \in I$. Then, constraints (9) imply $u_k = u_m + q_k = u_l + q_m + q_k = u_k + q_l + q_m + q_k$, which is impossible since $q_i > 0$, $\forall i \in I$ (any i with $q_i = 0$ can be removed from the graph without loss of generality). By raising a similar argument for all the subtours disconnected from the depot, we can conclude that no subtour will be formed among the intermediate nodes. $\square$

3.2. Distance restrictions

Similar to the capacity restrictions, we now present distance constraints that restrict the total length of each path to some pre-determined amount T. The constraints are given in the

following proposition:

Proposition 3 *Constraints*

$$v_i - v_j + (T - d_{id} - d_{0j} + d_{ij})x_{ij} + (T - d_{id} - d_{0j} - d_{ji})x_{ji}$$
$$\leqslant T - d_{id} - d_{0j} \quad \forall i \neq j \in I \qquad (12)$$

$$v_i - d_{0i}x_{0i} \geqslant 0 \quad \forall i \in I \qquad (13)$$

$$v_i - d_{0i}x_{0i} + Tx_{0i} \leqslant T \quad \forall i \in I \qquad (14)$$

are valid for the SBRP for $d_{0i} + d_{ij} + d_{j0} \leqslant T$ for every pair (i, j), where the variable v_i denotes the total length travelled from the depot to node i.

Proof   Similar to that of Proposition 1.   □

In these constraints, variables v_i denote the distance that the vehicle travelled up until point i. Constraints (12) are liftings of those proposed by Naddef (1994) (see Desrochers and Laporte (1991) and Kara and Bektaş (2005) for the lifting results), whereas constraints (13) and (14) are specifically derived for the SBRP. The former constraint is used to 'connect' the nodes in each tour and the latter two are used to set initialize the value of v_i to d_{0i} if i is the first node on the tour. We would like to note that these constraints can also be used to restrict the total travelling time of each bus in a similar fashion, since time is typically a function of the travelled amount of distance.

As a result of the preceding discussion, the integer programming formulation of the SBRP may now be given in full as Minimize $\sum_{i \in V'} \sum_{j \in V'} c_{ij}x_{ij}$: s.t. (2)–(5), (8), (9)–(14). In the next section, we will describe the solution of the problem under consideration using the proposed formulation.

4. Solution and proposed implementation

This section describes the current implementation of the case study, explains how the input data was processed and presents the results obtained by solving the integer program.

4.1. Current implementation

In the current implementation, the transportation of the students is handled by a contractor firm, which has $k = 26$ identical vehicles with a common capacity of $Q = 33$. There are 519 students to be picked up from different locations in Ankara. The routing is planned rather intuitively, and the one currently implemented is presented in Figure 1. As demonstrated in the figure, most of the buses (23 out of 26) are dedicated to a single destination only. Total distance travelled by all buses in the current implementation is calculated to be 246.736 km. The school requires that each student should not travel more than a total of $T = 25$ km by bus.

4.2. Input data processing

Since the locations of the students are points scattered throughout Ankara, we have grouped all of these points
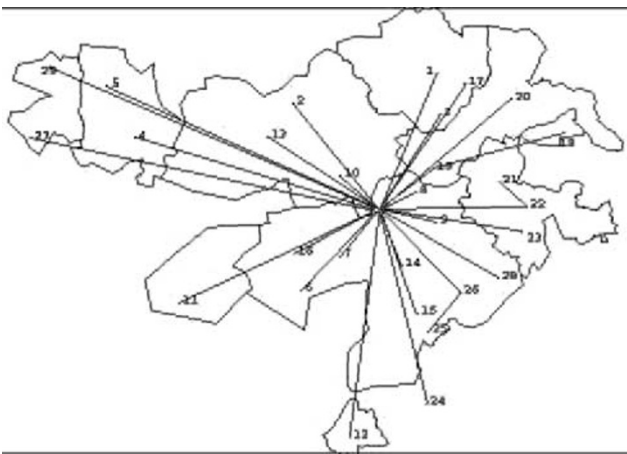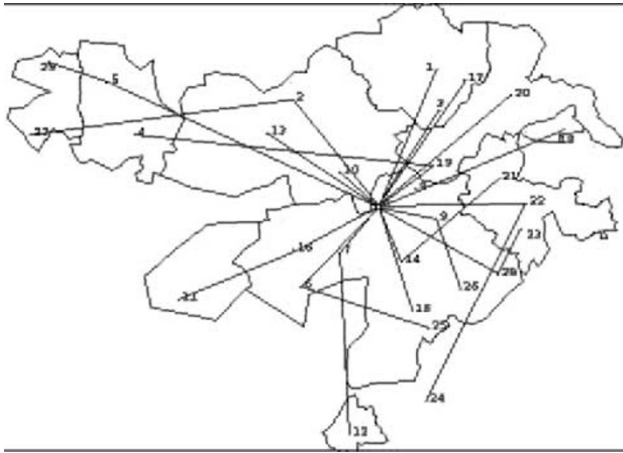

Figure 1   The current routing plan.

Table 1   The number of students located in each subregion

Subregion	Number of students
1	25
2	24
3	26
4	19
5	18
6	22
7	24
8	19
9	22
10	22
11	9
12	8
13	24
14	19
15	24
16	23
17	32
18	12
19	13
20	24
21	12
22	12
23	22
24	21
25	11
26	11
27	5
28	9
29	7

into approximately equal-sized clusters, which resulted in 29 different subregions. These sub-regions along with the number of students located in each are given in Table 1. The centroid of each subregion is considered to be the pick-up point of all the students located in this subregion. Since the size of each sub-region is quite small as compared to the entire routing area, the inter-travel walking distances from the homes of each student to the central point in each region can be neglected.

Figure 2 The proposed routing plan.

The distances between each pair of centres are calculated via MapInfo, by taking into account the most often used paths in the current implementation and the paths on which the buses are allowed to travel according to the traffic regulations. The resulting distance matrix is symmetric but clearly non-Euclidean.

The unit distance cost for each bus is calculated to be $\alpha = 300$ Turkish Liras (TLs) per metre and the fixed cost of each bus is calculated to be $f = 18\,568\,181$ TLs. The total cost of transporting the students to school in the current implementation is calculated as $556\,793\,506$ TLs, using 26 buses in total.

4.3. Solution of the model

Using CPLEX 9.0 as the commercial optimizer, the integer linear program presented was solved to optimality in 202.31 CPU seconds on a Sun UltraSPARC $12 \times 400\,\text{MHz}$ with 3 GB RAM. The optimal solution came up with a total cost of $397\,151\,058$ TLs, using 18 buses in total. Compared to the current implementation, the reduction in the total cost is 28.6%. The corresponding routing plan of the optimal solution is given in Figure 2.

We also present some data in Tables 2 and 3, relevant to the capacity utilization of the 26 buses used in the current implementation and the 18 buses used in the optimal solution, respectively.

In Tables 2 and 3, the first column denotes the bus number, the second column indicates how many students are carried by this bus, the third column shows the respective capacity utilization (in percent) and the last column indicates the amount of distance traversed by this bus. These tables also present the maximum, minimum and average capacity utilization rates of all the buses. Table 2 indicates that the capacity utilization in the current implementation varies highly (about 82% between the maximum and minimum values) and the average utilization rate of 60.49%, is rather low. On the other hand, the optimal solution has an average capacity utilization

Table 2 Capacity utilization and distance figures for the buses in the current implementation

Bus No.	Used Cap.	Cap (%)	Distance
1	25	75.76	7450
2	24	72.73	6920
3	26	78.79	6512
4	19	57.58	13 820
5	18	54.55	18 020
6	22	66.67	3290
7	24	72.73	5420
8	19	57.58	850
9	22	66.67	4430
10	22	66.67	3550
11	9	27.27	13 220
12	8	24.24	16 980
13	24	72.73	7960
14	19	57.58	3810
15	24	72.73	5814
16	23	69.7	4310
17	32	96.97	8620
18	25	75.76	10 940
19	24	72.73	7990
20	24	72.73	9850
21	22	66.67	8350
22	21	63.64	13 980
23	22	66.67	8950
24	5	15.15	26 580
25	9	27.27	8100
26	7	21.21	21 020
Avg		60.49	9489.84
Min		15.15	850
Max		96.97	26 580

Table 3 Capacity utilization and distance figures for the buses in the optimal solution

Bus No.	Used Cap.	Cap (%)	Distance
1	25	75.76	7450
2	29	87.88	24 500
3	26	78.79	6512
4	25	75.76	21 020
5	33	100	10 355
6	32	96.97	17 600
7	31	93.94	10 890
8	33	100	6770
9	22	66.67	3550
10	24	72.73	7960
11	31	93.94	12 960
12	24	72.73	5814
13	32	96.97	13 220
14	32	96.97	8620
15	32	96.97	15 590
16	24	72.73	7990
17	33	100	16 975
18	31	93.94	11 970
Avg		87.37	11 652.55
Min		66.67	3550
Max		100	24 500

of 87.37%, as presented in Table 3, with the variation in the capacity usage being decreased (to about 33%).

Comparing the current implementation with the optimal solution terms of distance, the total distance that the buses travel has decreased from 246 736 m to 209 746 m. However, the average distance that each bus travels has increased from 9489.84 m to 11 652.55 m.

5. Conclusion and further remarks

In this paper, we have presented an integer linear programming formulation to solve a real-life SBRP. The integer program was solved to optimality easily, due to the moderate size of the case study. The optimal solution obtained for the SBRP considered here resulted in a 28.6% savings in total cost as compared to that of the current routing scheme. As the school bus routing is most often intuitively planned in real life, we see that one can surely benefit from a better plan offered through integer programming.

It goes without argument that large-size problems do need specialized algorithms, as is the case in other studies mentioned previously, since such problems cannot in general be directly solved using commercial packages. However, as the computer hardware and software technology is rapidly improving, we believe that the focus should be on developing better formulations for problems of moderate size, rather than devising solution algorithms that are problem-specific. This is exactly the approach taken in this paper. In fact, it is quite interesting to note that the integer program presented here could not be solved by CPLEX 8.0 on a Pentium III 1400 Mhz PC running Linux (it was stopped after 900 CPU seconds without reaching an optimal solution), whereas the same integer program was easily solved on a faster computer using CPLEX 9.0.

We finally note that the formulation presented here is also capable of accommodating several additional constraints, such as the upper and lower bounds on the number of students each bus carries or time-window constraints. In specific, the time window constraints presented by Desrochers and Laporte (1991) can be directly included in the formulation to restrict each node to be visited in certain time intervals. Further research might therefore consider testing the applicability of the proposed formulation in solving OVRP with capacity, distance and time window constraints using test problems taken from the literature.

References

Angel RD, Caudle WL, Noonan R and Whinston A (1972). Computer-assisted school bus scheduling. *Mngt Sci* **18**: B279–B288.

Braca J, Bramel J, Posner B and Simchi-Levi D (1997). A computerized approach to the New York City school bus routing problem. *IIE Trans* **29**: 693–702.

Brandão J (2004). A tabu search algorithm for the open vehicle routing problem. *Eur J Opl Res* **157**: 552–564.

Desrochers M and Laporte G (1991). Improvements and extensions to the Miller–Tucker–Zemlin subtour elimination constraints. *Oper Res Lett* **10**: 27–36.

Fu Z, Eglese R and Li LYO (2005). A new tabu search heuristic for the open vehicle routing problem. *J Opl Res Soc* **56**: 267–274.

Kara I and Bektaş T (2005). *Integer linear programming formulations of distance constrained vehicle routing problems*. Research Report, Department of Industrial Engineering, Başkent University (available on request from the authors).

Kara I, Laporte G and Bektaş A (2004). A note on the lifted Miller–Tucker–Zemlin subtour elimination constraints for the capacitated vehicle routing problem. *Eur J Opl Res* **158**: 793–795.

Letchford AN, Lysgaard J and Eglese RW (2006). *A branch-and-cut algorithm for the capacitated open vehicle routing problem*. Unpublished research report available at http://www.lancs.ac.uk/staff/letchfoa/ovrp.pdf.

Li F, Golden B and Wasil E (2006). The open vehicle routing problem: Algorithms, large-scale test problems, and computational results. *Comput Oper Res*, forthcoming.

Li L and Fu Z (2002). The school bus routing problem: A case study. *J Opl Res Soc* **53**: 552–558.

Mandl C (1979). *Applied Network Optimization*. Academic Press: London.

Miller CE, Tucker AW and Zemlin RA (1960). Integer programming formulations and traveling salesman problems. *J Assoc Comput Mach* **7**: 326–329.

Naddef D (1994). A remark on 'Integer linear programming formulation for a vehicle routing problem' by N.R. Achutan and L. Caccetta, or how to use the Clark & Wright savings to write such integer linear programming formulations. *Eur J Opl Res* **75**: 238–241.

Repoussis PP, Tarantilis CD and Ioannou G (2006). The open vehicle routing problem with time windows. *J Opl Res Soc*, doi: 10.1057/palgrave.jors.2602143.

Sariklis D and Powell S (2000). A heuristic method for the open vehicle routing problem. *J Opl Res Soc* **51**: 564–573.

Tarantilis CD, Ioannou G, Kiranoudis CT and Prastacos GP (2005). Solving the open vehicle routing problem via a single parameter metaheuristic algorithm. *J Opl Res Soc* **56**: 588–596.

Toth P and Vigo D (2002). *The Vehicle Routing Problem, SIAM Monographs on Discrete Mathematics and Applications*. SIAM: Philadelphia.