A Nested Location–Routing Heuristic Using Route Length Estimation

Gábor Nagy and Saïd Salhi

Management Mathematics Group,
School of Mathematics and Statistics,
University of Birmingham,
Edgbaston, Birmingham, B15 2TT, U.K.

Accepted by *S.O. Krumke and N. Mladenović*

Abstract

A nested heuristic approach that uses route length approximation is proposed to solve the location-routing problem. A new estimation formula for route length approximation is also developed. The heuristic is evaluated empirically against the sequential method and a recently developed nested method for location-routing problems. This testing is carried out on a set of problems of 400 customers and around 15 to 25 depots with good results.

Keywords: location-routing, route length estimation, heuristics, tabu search.

1 Introduction

While distribution management is traditionally divided into the long-term sub-problem of location and the short-term sub-problem of routing, it has been shown in the last decades, that an integrated approach to distribution management is both more logical and more efficient. This is due to the fact that the two sub-problems referred to above are inter-related in practice. This integrated approach is named location-routing. The location-routing problem (LRP) arises in several applications of locational problems where the customers' demand is less than a full truck-load.

The LRP can be defined as follows. Given are the following:

- the set of customers J,
- the set of potential depot locations I,
- the set of vehicles K,
- the distance d_{ij} between locations i and j $(i, j \in I \cup J)$,
- the fixed cost f_i of establishing depot i $(i \in I)$,
- the demand q_j of customer j $(j \in J)$,
- the drop time δ,
- the maximum capacity of the vehicles Q, and
- the maximum distance allowed for the vehicles D.

We wish to determine:

- the number and *location* of depots,

- the *allocation* of customers to depots, and
- the number and *routing* of delivery vehicles,

such that:

- the total system cost (the sum of depot and delivery costs) is *minimised*,
- the demand of every customers is *satisfied*, and
- no vehicle load or route length *exceeds* the given constraints.

The LRP is a complex combinatorial problem and it can be formulated as a 0-1 ILP. We present one possible formulation below. This formulation is a slightly modified version of problem (LRP1) of Laporte (1989). Let us begin by presenting some additional notation.

$$x_{ijk} = \begin{cases} 1 \text{ if customer } j \text{ follows customer } i \text{ on route } k,\ i,j \in I \cup J, k \in K \\ 0 \text{ otherwise} \end{cases}$$

$$y_i = \begin{cases} 1 \text{ if depot } i \in I \text{ is open} \\ 0 \text{ if depot } i \in I \text{ is closed} \end{cases}$$

$$z_{ij} = \begin{cases} 1 \text{ if customer } j \in J \text{ is served from depot } i \in i \\ 0 \text{ otherwise} \end{cases}$$

Using the above notation, the problem formulation is as follows:

Minimise $\sum_{i\in I} f_i y_i + \sum_{k\in K} \sum_{i\in I\cup J} \sum_{j\in I\cup J} d_{ij} x_{ijk}$
subject to

$$\sum_{j\in J} q_j \sum_{i\in I\cup J} x_{ijk} \le Q \qquad (k \in K) \qquad (1)$$

$$\sum_{i\in I\cup J} \sum_{j\in I\cup J} (d_{ij} + \delta) x_{ijk} \le D + \delta \qquad (k \in K) \qquad (2)$$

$$\sum_{k\in K} \sum_{i\in I\cup J} x_{ijk} = 1 \qquad (j \in J) \qquad (3)$$

$$\sum_{k\in K} \sum_{i\in S} \sum_{j\in((I\cup J)\setminus S)} x_{ijk} \ge 1 \qquad (2 \le |S| \le |I \cup J|; S \subseteq I \cup J; S \cap J \ne 0) \qquad (4)$$

$$\sum_{j\in I\cup J} x_{ijk} = \sum_{j\in I\cup J} x_{jik} \qquad (k \in K; i \in I \cup J) \qquad (5)$$

$$\sum_{i\in I} \sum_{j\in J} x_{ijk} \le 1 \qquad (k \in K) \qquad (6)$$

$$\sum_{s\in I\cup J} (x_{isk} + x_{sjk}) \le z_{ij} + 1 \qquad (i \in I; j \in J; k \in K) \qquad (7)$$

$$x_{ijk} = 0, 1 \qquad (i \in I; j \in J; k \in K) \qquad (8)$$

$$y_i = 0, 1 \qquad (i \in I) \qquad (9)$$

$$z_{ij} = 0, 1 \qquad (i \in I; j \in J) \qquad (10)$$

In the above formulation the objective function is the sum of depot fixed costs and vehicle routing costs. Constraints (1) and (2) are the maximum capacity and the maximum distance constraints respectively. Constraints (3) to (7) are standard route logic and flow conservation constraints. Constraints (8), (9) and (10) stipulate that x_{ijk}, y_i and z_{ij} are binary variables.

The LRP is NP-hard and can be solved optimally only for small sized problems. Thus, heuristics are widespread in solving the LRP. Most of them are based on iterating between solving the two sub-problems, and are thus called *iterative* methods. If the interdependence of the sub-problems is ignored then we have a *sequential* approach: this consists of solving the locational problem first, and the resulting routing problem second.

Despite the fact that it has only recently been investigated, a number of papers have been published on this topic. It is not possible for us to provide a comprehensive review of LRP papers here: instead we refer the reader to the reviews of Laporte (1989), and of Salhi and Fraser (1995).

The need for route length estimation (RLE) formulae arises from the fact, that computing *the sum of radial distances* is much easier to perform than determining *the total routing cost.* The problem of RLE can be stated as follows. Given the radial distances between the depot and the customers assigned to it, find the total length of the routes originating from that depot. This means finding a *formula* which describes the relationship between the above two quantities. There is a distinct relationship between radial distances and route length and a linear relationship was assumed by Webb (1968). However, more complex formulae were proposed by Christofides and Eilon (1969) and later extended by Daganzo (1984). Stokx and Tilanus (1991) used linear regression to approach the problem of RLE, using trips created by a vehicle routing algorithm as the basis for regression.

While it is possible to devise exact solution methods for many hard combinatorial problems, they may be too slow and inappropriate for practical purposes. This prompts us to use heuristics, which seem to be the best way forward to tackle hard combinatorial problems. In addition, these approaches have the advantage of producing more than one single solution and they are also easy to understand, to modify and to implement. However, heuristics have a drawback when it comes to evaluation: they usually do not yield easily to analytic approaches and so the mathematically less well founded empirical evaluation has to be used.

The aim of this study is to:

(i) demonstrate how important it is to integrate location and routing,
(ii) propose some heuristic approaches that use the idea of nested methods and route length approximation to solve the LRP,
(iii) develop a new RLE formula which can incorporate constraints on tour length and on vehicle capacity.

The rest of the paper is organised as follows. We first review the structure of nested methods in section 2. We present the new nested methods that use route length approximation in section 3. Computational results are provided in section 4. Conclusions and future research directions are given in section 5.

2 General Methodology of Nested Methods

Most LRP heuristics fall into the categories of *sequential* or of *iterative* methods. Sequential methods solve the locational problem first and solve the resulting routing problem second. This approach is clearly inadequate, as routing information is not taken into account when solving the locational problem. It has been shown, see Salhi and Rand (1989), that sequential methods provide inferior solutions to combined heuristics, such as iterative methods.

Iterative methods iterate between the location phase and the routing phase until a suitable stopping criterion is met (see Salhi and Fraser (1995) for further references). Although these algorithms have been shown to produce better quality solutions than the sequential methods, they still have some drawbacks.

These methods let the location algorithm run until the end, and then restart it using some information from the routing stage. Thus, if the locational algorithm leads to a solution far from the "true" optimum, than the routing information may come "too late" to lead the method back to a good search direction. We can also object to iterative methods from the *modelling* point of view. These methods treat the two constituent components of the location-routing problem as if they were on the same footing. Observe, that a location-routing problem is essentially a location problem, with the routing factor taken into consideration. So, instead of treating the two sub-problems as equal we can observe a *hierarchical structure*, with location as the main problem and routing as a subordinate one. This concept of hierarchy is also strongly emphasised by Balakrishnan, Ward and Wong (1987).

This observation suggests an appropriate method: an algorithm, where location is the master problem, and routing is a subproblem. We refer to this approach "nested" because the routing stage is *embedded* into the location phase. More details on this concept are given in Nagy and Salhi (1995).

2.1 The Location Phase

Many good heuristics exist for locational problems, such as Körkel (1989) and Beasley (1993), but they are not search-based methods which could be easily embedded into nested methods. The need to use methods which rely on visiting feasible solutions at each step is essential for our purpose. The location algorithm adopted here is a modification of the add/drop/shift heuristic, initially introduced by Kuehn and Hamburger (1963).

In our location phase the neighbourhood of a current solution is defined by the three moves of dropping an existing depot, adding a new depot to the existing set of open depots, and finally simultaneously closing a given depot and opening a new depot from the set of customers served by that given depot. Our locational algorithm starts with an initial feasible solution: this is usually the outcome of a sequential solution method applied to the problem. Then, in the *investigation* stage, the neighbourhood of the current solution is investigated: all the possible moves are evaluated. In the *implementation* stage, the best move is implemented and the cost of the new solution is calculated. One run of these two phases is called an iteration. These iterations are repeated until some stopping criterion is met: we shall explain the criteria used later.

Finally, we wish to note that this neighbourhood search algorithm has been enhanced through the use of a simple version of *tabu search.* After implementing a move, the city or cities involved in the move will be made tabu, i.e. they cannot be considered for another move for a given number of steps. We also introduced a form of diversification: if the search does not produce new better solutions after a given number of iterations, we re-initialise the system starting from the best solution and make some of the cities involved in the previous moves tabu. This concept is also used in establishing our termination criteria: a maximum of ten re-initialisations are allowed. Furthermore, a maximum time limit of two hours was also imposed. For more details on tabu search in general, we refer the reader to Reeves (1993).

2.2 The Routing Phase

Many different routing algorithms could be used in our method. However, the algorithm chosen has to be a multi-depot vehicle routing algorithm which can produce good solutions quickly and consistently. We have hence chosen the MDVFM heuristic developed by Salhi and Sari (1995). This composite heuristic is based on the concept of "borderline" customers. It relies on a route construction method, an insertion-based heuristic and some improvement routines. This method caters for the vehicle fleet mix problem, a generalisation of the vehicle routing problem, where the vehicles do not necessarily have the same capacity.

2.3 How Routing Information is Used in the Location Stage

Routing information is needed in our locational algorithm in two distinct stages. In each iteration, the cost of the current solution has to be found. As this information needs to be determined only once per iteration, we can use a *slow* but *accurate* method to find routing costs. This is done by applying the MDVFM routine for the entire set of cities. We also need to find the cost improvement associated with each move. As there are several moves within each iteration to be evaluated, we use a *fast* but *approximate* algorithm to do this. This involves ideas drawn from computational geometry and route length estimation formulae. Geometric ideas in location-routing have been investigated in Nagy and Salhi (1995) and implemented successfully in our previously developed method REGIONAL. Finally, we wish to note that we will use the routing algorithm MDVFM to provide us with information other than just route configuration or routing costs.

2.4 How Do Changes in Location Influence Routing?

We wish to demonstrate that the area of influence of a change in location is a connected area incorporating the depot concerned, its customers and also some "nearby" depots with their associated customers. Let us begin with an example. Suppose that during one of the iterations of our algorithm we wish to evaluate the move of shifting a depot from location a to location b (within the catchment area of a). We make three observations about the changes in routing caused by such a move.

Firstly, a change in location influences routing locally: moving the depot from a to b implies that all the routes previously assigned to a will now start from and terminate at b.

Secondly, it is incorrect to assume that there will be no change outside the catchment area of the depot. In the above example, it is likely that some customers of a will now be re-allocated to a neighbouring depot c rather than to b, thus influencing the routing in the catchment area of c.

Thirdly, we observe that the area where the changes take place is connected. Generalising our second observation, a change in the location of a may influence c, this can in turn influence a neighbour of depot c, and so on. Thus the area of influence of the move is the union of the above catchment areas, and is therefore connected.

This area which may be affected by the change is referred to from now on as a *region*. The depot concerned in the change is called the *central depot*. As we do not know the exact area of influence of a locational move, we have to approximate it, using some proper subset of the set of all cities as our region. Finally, we wish to note that regions are specific to their central depot.

2.5 Approximating Costs Using Regions

Suppose we have created a region around the central depot and that we also know the full routing solution to the current situation. Then, to calculate the cost improvement of a move, we only calculate the cost improvement *within* the region, thus making the implicit assumption that route configuration and thus routing cost remains unchanged outside the region. While this assumption may not hold in practice, it enables us to rely on calculations pertaining to only a subset of all cities as opposed to the whole system for routing cost calculations. This considerably speeds up our algorithm. It is clear that a good choice of region would avoid both excessive computational time associated with too large regions and loss of accuracy associated with too small regions. Finally, we wish to note that by employing MDVFM to evaluate cost improvements within the region we get a location-routing methodology. This method is called REGIONAL and is described in Nagy and Salhi (1995).

2.6 How Do We Construct These Regions?

Although the original idea for the structure of the regions was inspired by computational geometry, the current construction is based on our routing algorithm

MDVFM. Note that this is one example of using the routing algorithm for a purpose widely different from what it was originally designed to do. This routing algorithm finds at some stage the nearest and the second nearest depot to each customer. We define an incidence relation $\mathcal{N}$ on the set of depots as follows. If two depots are nearest and second nearest respectively for any customer, let these two depots be incident, or neighbours, to each other. The set of all the neighbours of a depot d is denoted by $\mathcal{N}(d)$. The region pertaining to a depot d is defined to consist of the depot d, its customers $E(d)$, its neighbouring depots $\mathcal{N}(d)$ and their customers $E(\mathcal{N}(d))$. While other constructions are also possible, see Nagy and Salhi (1995) for further details, the above structure was found to be the most logical and it also produced the best results.

3 Nested Methods Based on Route Length Estimation

Balakrishnan, Ward and Wong (1987) suggested the idea of incorporating RLE formulae into a locational method. This suggestion was not taken up by many researchers; we know of two works, where RLE was used in LRP: namely Salhi (1987) and Laporte and Dejax (1989). However, Salhi (1987) only calculated bounds for the number of vehicles and did not use a RLE formula; while Laporte and Dejax did not incorporate the RLE formula into a LRP algorithm and relied on evaluating all possible combinations for depot locations.

ESTIMATE incorporates the ideas presented earlier. It fits into the general framework of nested methods, and it relies on the concept of regions. However, unlike REGIONAL, it does not employ a routing algorithm when investigating the possible moves within an iteration. Instead, it calculates the sum of direct distances for the relevant region (a much simpler task than routing) and employs a RLE formula to provide us with an approximation of the routing cost per region. Thus the change in *estimated* cost is evaluated for each move. The move with the largest estimated cost improvement is chosen for execution; only after the move has been made is the route configuration and total routing cost determined.

3.1 The Structure of ESTIMATE

To facilitate understanding of the concepts involved in ESTIMATE, and of the different versions thereof, a pseudo-code of the locational part of the method is given below, followed by some comments about the steps of the algorithm.

1*. update the *estimation formula* $\mathcal{E}$ for the *whole system* M

2. for all $d_i \in M$ do
3. construct the region R_i
4*. update the *estimation formula* $\mathcal{E}$ within the region R_i
5. calculate $\mathcal{E}(R_i) - \mathcal{E}(R_i \backslash \{d_i\})$
6. for all $c_{ij} \in E(d_i)$ do

7. calculate $\mathcal{E}(R_i) - \mathcal{E}(R_i \cup \{c_{ij}\})$
8. calculate $\mathcal{E}(R_i) - \mathcal{E}(R_i \cup \{c_{ij}\} \backslash \{d_i\})$
9. end
10. end
11. implement the move with the largest *estimated* cost improvement
12. calculate $C(M_{new})$
13. calculate $\mathcal{E}(M_{new})$
14. repeat steps 1 to 13 until termination criteria reached

where $M = \{d_1, d_2, \ldots, d_k\}$ is the set of open depots,
$E(d_i) = \{c_{i1}, c_{i2}, \ldots, c_{iw}\}$ is the set of customers served by depot d_i,
R_i is the region associated with the depot d_i,
$\mathcal{E}(S)$ is the estimated cost of the system S, and
starred (*) steps are optional.

Steps 1 and 4. These steps provide the *flexibility* of the algorithm. The five different versions of ESTIMATE differ from each other in steps 1 and 4, and in the formula $\mathcal{E}$ they employ.

Step 3. We wish to note that as regions are specific to their central depot, they only need to be created once per iteration: all moves involving the central depot are investigated before the algorithm moves on to another depot.

Steps 5, 7 and 8. These steps are associated with the *trial moves* "close depot d_i", "open depot d_i" and "shift depot d_i to customer location c_{ij}" respectively.

Step 11. We would like to remark that the consequences of a move are not fully known until step 12, thus the move made in step 11 may well not result in the best improvement. Thus, it is important to have a good choice of region and of estimation formula: a wrong choice of either may result in misleading moves being implemented.

3.2 Local Estimation

An easy way of estimating route length is to assume that it is a simple function of the sum of direct distances. Usually a *linear* relationship is assumed (see e.g. Webb (1968)). The total routing cost per depot T is approximated by rD, where D is the sum of direct distances from the depot to all its customers and r is a constant. We refer to this constant r as the *ratio*. Although this is a very crude measure, it is better than approximating T by D, which is what the sequential method does.

We have created a version of ESTIMATE called ESTIMATE1 based on the formula $T \approx rD$. Referring to the pseudo-code given above, we note that step 1 is omitted, while step 4 consists of determining the ratio r for each region. This is done by dividing $\sum T$ by $\sum D$, where summation refers to the whole of the region. Cost improvements are then calculated as improvements in the value of rD.

This measure is called a *local estimate*, as the ratio r is calculated for each region, and thus it represents the *local* characteristics of the system. It is a flexible measure, depicting the spatial characteristics of the cities within each region. As in each iteration the r values are re-calculated, this measure is flexible not only in *space*, but also in *time*. However, this RLE formula is a very crude measure. Webb (1968) noted several deficiencies of this formula. Although many better models exist in the literature, they cannot cater for our requirements. So, we decided to develop a new RLE formula.

3.3 Drawbacks of Current RLE Formulae

A standard regression formula for route lengths was given by Christofides and Eilon (1969), and later extended by Daganzo (1984). Recently, Stokx and Tilanus (1991) extended this formula by introducing regression coefficients. The formula we present here is an aggregation of their work. The total routing cost per depot is approximated in equation (11) below:

$$T \approx aD/c + bD/\sqrt{n} \tag{11}$$

where n is the total number of customers belonging to the depot, c is the average number of stops on a feasible vehicle tour, and a and b are regression coefficients. Note that for the case of a single travelling salesman tour, the above formula reduces to:

$$T = aD/n + bD/\sqrt{n} \tag{12}$$

Following Daganzo (1984) we call the terms $aD/\sqrt{n}$ and $bD/\sqrt{n}$ in (2) above *linehaul* and *detour* distances respectively.

While this formula can be developed further in many different directions, we shall concentrate on one particular defect of the above formula. We would need to know the average number of stops per tour it c to use it. This particular measure is, however, not available to us during the run of our locational algorithm. Thus we need to develop the formula such that the use of the variable c is avoided. We can, however, make use of other data which we do know, namely the maximum capacity and distance constraints, the demand levels of the customers and the drop time.

3.4 Developing a New Route Length Estimation Formula

In this study, a new route length estimation formula is derived, using the existing models and the knowledge of the routing algorithm used in this work. T he formula is constructed by mimicking the *multiple giant tour* algorithm, which forms a part of the MDVFM heuristic of Salhi and Sari (1995). A similar exercise is presented in Daganzo (1984), but the author used a cluster-first, routing-second method, while we use a routing-first, cluster-second algorithm. We first determine the routing cost per depot with t vehicles, following the partitioning of the giant tour. Then, we determine t , the number of vehicles, using the above result and the constraints on both vehicle capacity and tour length. Finally, the above two results will be put together to yield a new RLE formula. The mathematical derivation is relegated to the appendix and here we only quote the result as given in equation (A14):

$$T \approx \left(\frac{aD}{n} - \frac{bD}{\sqrt{n}(n-1)}\right) \times \mathrm{Int}\left(\alpha + \beta \mathrm{Max}\left(\frac{2nQ}{2nP-Q}; \frac{2bn\sqrt{n}D + 2n^2(n-1)\delta}{(n-1)(2Zn-2aD-\delta n)+b\sqrt{n}D}\right)\right) + \frac{bD\sqrt{n}}{n-1} \quad (13)$$

where Q is the sum of customer demands, P is the maximum capacity of a vehicle, δ is the drop time and Z is the maximum distance. a, b, α and β are regression coefficients, with default values of 1.8, 1.8, 1.0 and 1.0 respectively. The importance of this result stems from the fact, that both c and t are eliminated from the RLE formula, and all the quantities involved in this formula are known to us at each iteration. From now on, we refer to the above formula as $T \approx f(D)$.

3.5 Global Estimation

We developed two versions of ESTIMATE under this heading, namely ESTIMATE2 and ESTIMATE3. In these methods, the total routing cost T is approximated by $f(D)$. Referring to our pseudo-code, we note that step 4 is omitted in both versions. In ESTIMATE2, step 1 is also omitted, and the formula used relies on fixed coefficients (default values). ESTIMATE3 uses regression in step 1 to determine the regression coefficients a, b, α and β, and thus is able use them in the RLE formula. Cost improvements in both versions are calculated as improvements in the value of *f(D)*.

We wish to note that the regression procedure we used was created specifically for this purpose, as no general regression codes were available which could have tackled regression for formulae containing both "integer part" and "maximum". We wrote a small code relying on the NAG library which uses the *Simplex Method of Function Minimisation* of Nelder and Mead (1965). Using regression is yet another example of the subtle use of routing information.

This measure is called a *global estimate*, as the estimation formula *f(D)* relates to the entire set of cities, so it is a *globally* valid formula. Although the powerful formula used in these methods is certainly an advantage, these methods are rigid in *space* and do not appreciate the local characteristics of the system. ESTIMATE2 is also rigid in *time*, as regression is not used. ESTIMATE3, on the other hand, offers flexibility in time. A further drawback of global estimation is that it relies on the RLE formula provided. While this formula could always be developed further, there is always some difference between the actual and the estimated costs, and it would be desirable to account for these differences.

3.6 Combined Estimation

We have seen that the local estimation measure is flexible but not very accurate, while the global measures are good estimates but do not offer sufficient flexibility. Thus, the idea of *combining* the two comes to one naturally. We make the assumption that routing costs T and their estimation $f(D)$ are linearly related, $T \approx rf(D)$. Two further versions of ESTIMATE using the formula $rf(D)$ were created, namely ESTIMATE4 and ESTIMATE5. Referring to the pseudo-code given earlier, we note that step 4 consists of determining the *ratio* of routing cost and its approximation. In ESTIMATE4, step 1 is omitted, and the formula used relies on fixed coefficients (default values). ESTIMATE3 uses regression in step 1 to determine the regression coefficients a, b, α and β, and thus is able to use them in the RLE formula. Cost improvements in both versions are calculated as improvements in the value of $rf(D)$.

The advantage in using combined estimation is that we retain the desirable attributes of both local and global methods. Flexibility in space is provided, as we use a ratio which differs from region to region. Flexibility in time is also achieved by updating the ratios in each iteration; furthermore, ESTIMATE5 also updates the coefficients in its RLE formula. A sophisticated RLE formula is incorporated into both of these methods, which makes them very accurate. The quality of the formula is further enhanced by the fact, that discrepancies between real costs and their estimations are noted and corrected by the method.

4 Computational Results

The heuristics proposed above were written in VAX Fortran and executed on a VAX 4000-500 computer at the University of Birmingham. They were evaluated using empirical testing.

4.1 Data Generation

Five sets of coordinates and demands, each consisting of 400 customers, are tested with three levels of fixed depot costs to provide solutions around 25, 20 and 15 depots respectively: this gives 15 problems in total. These data sets are generated using some of the problems given by Christofides, Mingozzi and Toth (1979), and enlarging them to produce problems with 400 customers. We have chosen 400 for the size of our test problems as a compromise. On one hand too small a problem size would not have enabled us to appreciate the structure of regions, on the other hand too many customers would have slowed down the underlying vehicle routing algorithm. This generation is explained in more detail in Nagy and Salhi (1995). The problem sets are numbered as $11, 12, 13, 21, \ldots, 53$: problem 11 is the first problem set with low depot costs, etc. We also wish to note that customers are clustered for problems 21 to 33. Furthermore, maximum capacity and maximum distance constraints were chosen so that they are tight for all problems except 41 to 43.

4.2 Plan for Comparison

The combined location-routing methods are compared to the sequential location-routing method. This way of comparing the methods was suggested by Balakrishnan, Ward and Wong (1987) and used previously by Salhi and Rand (1989) and by Salhi and Fraser (1995). In order to do this type of comparison, we ran the sequential method first. A number of trial runs with different levels of depot costs were executed in order to give the required number of depots. A "strong" local optimum was found for all the problems, i.e. a local optimum, from which the method could not "climb out" after about a hundred iterations. Then, all the other methods used this solution as their initial solution. We had two termination criteria: a cutoff time of two hours, and a maximum of ten restarts.

Two main comparison criteria are used: solution quality and solution time. Solution quality of the combined methods is measured as the percentage improvement they offer over the solution obtained by our benchmark sequential algorithm. Solution time is best judged as the time the different methods take to execute one iteration. Furthermore, a number of subsidiary observations are also made. The results for our test runs are tabulated in Tables 1 and 2.

4.3 Analysis of Results

Looking at Table 1, we can see that all versions of ESTIMATE offer improvements to the solution found by the sequential method. Despite the differences in their structure, all of them produce about the same solution quality, and in many cases the same solution. The solution quality is better than that of REGIONAL, with cost improvements ranging, on average, between 7-8% as opposed to 4-6% in the case of REGIONAL. As the five versions of ESTIMATE each use a more sophisticated method of estimation than the previous one, we can expect them to produce progressively better results. This is indeed the case, if we look at Table 1, we can clearly see the cost improvement figures rising from 7.21% to 7.69%. However, these differences are minor. It seems to us that this method is not too sensitive to the type of estimation used. Observing the change in cost improvement across the different problems, we see, that for problems 41 to 43 a large improvement is found. This is probably due to the fact that the constraints are loose for these problems: the same phenomenon has been observed for REGIONAL. Furthermore we wish to note that while REGIONAL only offered small improvements for some of the problems, in ESTIMATE5 the minimum improvement was 1.73%. This makes ESTIMATE a more consistent method, giving good solutions across the board.

We can also observe, using Table 2, that the solution speed of ESTIMATE has significantly improved compared to the speed of REGIONAL. We can also observe that as this method has a speed comparable to that of CLASSIC, the algorithm is usually stopped by the termination criterion based on the number of restarts. Furthermore we note that the radial cost has always increased during the run of our combined methods. This reinforces our view that minima of the radial cost and of the total cost do not coincide. We can see, looking at Table 2,

	R3	E1	E2	E3	E4	E5
11	2.60	2.60	3.89	3.89	3.89	**4.78**
12	1.95	1.48	2.49	**3.58**	**3.58**	**3.58**
13	3.83	**5.25**	**5.25**	**5.25**	**5.25**	**5.25**
21	2.45	**5.32**	**5.32**	**5.32**	**5.32**	**5.32**
22	2.31	2.31	2.45	2.45	**2.51**	**2.51**
23	1.90	1.90	**2.01**	**2.01**	**2.01**	**2.01**
31	0.49	0.49	0.58	0.58	0.58	2.38
32	**1.21**	0.75	0.88	**1.21**	**1.21**	**1.21**
33	**1.73**	**1.73**	**1.73**	**1.73**	**1.73**	**1.73**
41	11.5	**23.9**	**23.9**	**23.9**	**23.9**	**23.9**
42	**25.7**	**25.7**	**25.7**	**25.7**	**25.7**	**25.7**
43	**26.0**	**26.0**	**26.0**	**26.0**	**26.0**	**26.0**
51	**3.09**	**3.09**	**3.09**	**3.09**	**3.09**	**3.09**
52	4.21	4.01	4.21	4.21	**4.23**	**4.23**
53	**3.75**	**3.75**	**3.75**	**3.75**	**3.75**	**3.75**
P1	2.79	3.11	3.88	4.24	4.24	**4.54**
P2	2.22	3.18	3.26	3.26	**3.28**	**3.28**
P3	1.14	0.99	1.06	1.17	1.17	**1.77**
P4	21.1	**25.2**	**25.2**	**25.2**	**25.2**	**25.2**
P5	3.68	3.62	3.68	3.68	**3.69**	**3.69**
A1	4.02	7.08	7.36	7.36	7.36	**7.89**
A2	7.07	6.84	7.14	7.43	**7.44**	**7.44**
A3	7.44	7.72	**7.74**	**7.74**	**7.74**	**7.74**
AV	6.18	7.21	7.41	7.51	7.51	**7.69**

Table 1. A cost somparison of the different methods.

Entries are percentage improvements over the sequential method. The best solution is highlighted in **bold**. R3 refers to REGIONAL3, the best method in the REGIONAL group of methods, described in Nagy and Salhi (1995). E1 to E5 stand for ESTIMATE1 to ESTIMATE5. P1 is the average of problems 11, 12 and 13; P2 to P3 likewise. A1 is the average of problems 11, 21 and 31; similarly for A2 and A3. AV stands for the average of all the 15 test problems.

that the number of depots and of vehicles has always decreased. This points to the fact that the sequential method usually overestimates the necessary number of depots and vehicles. Our methods on average resulted in having 4 to 6 depots and 2 to 3 vehicles less in the final solution than the outcome of the sequential method.

	COST	D.D	ND	NV	IT	TIME	RST	TPI
C1	0.00	0.00	0	0	117/-	8:26/-	10	4/-
R3	6.18	17.69	5	2	53/36	121:32/82:28	5	147/148
E1	7.21	20.22	6	3	117/76	28:26/18:32	10	14/14
E2	7.41	20.56	6	3	120/91	27:59/21:31	10	13/14
E3	7.51	20.15	6	3	119/92	41:40/32:27	10	20/20
E4	7.51	20.23	6	3	119/93	27:58/21:57	10	14/14
E5	7.69	20.56	6	3	119/93	42:29/33:29	10	20/20

Table 2. A summary of computational results.

C1 refers to the sequential method. COST is the percentage cost improvement, while D.D. is the percentage increase in the sum of direct distances w.r.t. C1. ND and NV stand for the decrease in the number of depots and of vehicles respectively. IT is the number of iterations executed. TIME is the running time (in minutes and seconds), while TPI is the time per iteration (in seconds). RST denotes the number of restarts. In columns IT, TIME and TPI there are two entries per row: the first refers to the whole run, the second refers to the period until the best solution is found.

5 Conclusions and Future Research Directions

Heuristic approaches using the concept of nested methods are developed for the combined location-routing problem. These techniques use ideas borrowed from computational geometry to define the concept of proximity and route length estimation to approximate the routing cost. To our knowledge, this is the first algorithm which fully integrates RLE techniques into an LRP procedure.

Encouraging results were obtained when compared against a recently developed nested method. These estimation-based models consistently produce good results without resorting to severe computational effort. In this study, we have demonstrated how inefficient it is to refer to the commonly applied approach of location-routing which uses location first and routing second. Our findings re-enforce even further the previous conclusions obtained by Salhi and Rand (1989) and also by Salhi and Fraser (1995) beside providing a clearer insight to this strategic distribution problem. All nested location-routing methods gave encouraging results, producing between 4% to 7% average improvement over the sequential method. In addition, the computational effort for these methods was reasonably good.

A number of issues may be worth investigating in connection with the research presented in this paper. These include:

(i) carrying out a computationally intensive simulation in order to assess the consistency and the robustness of nested methods. This field of investigation may serve as a justification for combining the long-term problem of location and the short-term problem of routing into a common framework. The authors obtained encouraging preliminary results.

(ii) introducing capacity limits on depots. This would place a limit on the number of customers which can be assigned to a depot and thus the allocation phase of our routing heuristic would need to be modified.

(iii) extending the vehicle routing part of the location-routing problem to the case of vehicle fleet mix. This extension would be facilitated by the fact that we used a fleet mix heuristic in the routing phase of our algorithm. However, the RLE formula would need to be modified to allow for the case of several alternative constraints on vehicle capacity and on maximum distance.

Acknowledgements

The authors would like to acknowledge the financial support of the Committee of Vice-Chancellors and Principals of the United Kingdom (ORS Awards) and of the School of Mathematics and Statistics, The University of Birmingham. We thank both referees for their useful comments which improved the presentation of the paper.

References

[1987] Balakrishnan,A., Ward,J.E., and Wong,R.T.: Integrated facility location and vehicle routing models: recent work and future prospects, *American Journal of Mathematics and Management Science.* **7** (1987) 35-61.

[1993] Beasley,J.E.: Lagrangean heuristics for location problems, *European Journal of Operational Research.* **65** (1993) 383-399.

[1969] Christofides,N., and Eilon,S.: Expected distances in distribution problems, *Operational Research Quarterly.* **20** (1969) 437-443.

[1979] Christofides,N., Mingozzi,A., and Toth,P.: The vehicle routing problem, *in:* N. Christofides, A. Mingozzi, P. Toth and C. Sandi: *Combinatorial Optimization* (1979) Wiley, Chichester.

[1984] Daganzo,C.F.: The distance traveled to visit n points with a maximum of c points per vehicle: an analytic model and an application, *Transportation Science.* **18** (1984) 331-350.

[1989] Körkel,M.: On the exact solution to large-scale simple plant location problems, *European Journal of Operational Research.* **39** (1989) 157-173.

[1963] Kuehn,A.A., and Hamburger,M.J.: A heuristic program for locating warehouses, *Management Science.* **9** (1963) 643-666.

[1989] Laporte,G.: A survey of algorithms for location-routing problems, *Investigácion Operativa.* **1** (1989) 93-123.

[1989] Laporte,G., and Dejax,P.: Dynamic location-routing problems, *Journal of the Operational Research Society.* **40** (1989) 471-482.

[1996] Nagy,G., and Salhi,S.: Nested heuristic methods for the location routing problem, *Journal of the Operational Research Society.* (to appear) (1996).

[1965] Nelder,J.A., and Mead,R.: A simplex method for function minimization, *The Computer Journal.* **7** (1965) 308-313.

[1993] Reeves,C.: *Modern Heuristic Techniques for Combinatorial Problems*, Blackwell, Oxford (1993).

[1987] Salhi,S.: *The Integration of Routing into the Location-Allocation and Vehicle Fleet Composition Problems*, Ph.D. thesis, University of Lancaster (1987).

[1995] Salhi,S., and Fraser,M.: An integrated heuristic approach for the combined location-vehicle fleet mix problem, *Studies in Locational Analysis.* 8 (1996) 3-22.
[1989] Salhi,S., and Rand,G.K.: The effect of ignoring routes when locating depots, *European Journal of Operational Research.* **39** (1989) 150-156.
[1995] Salhi,S., and Sari,M.: A p-level composite heuristic for the the multi-depot vehicle fleet mix problem, *Preprint 95/13* (1995), School of Mathematics and Statistics, The University of Birmingham.
[1991] Stokx,C.F.M., and Tilanus,C.B.: Deriving route lengths from radial distances: empirical evidence, *European Journal of Operational Research.* **50** (1991) 22-26.
[1968] Webb,M.J.: Cost functions in the location of depots for multiple-delivery journeys, *Operational Research Quarterly.* **19** (1968) 311-320.

Appendix. Developing a new route length estimation formula

In this appendix we construct a new RLE formula, following the strategy outlined in section 3. First, we find routing costs per depot for a given number of vehicles t. Then, we use this information and the constraints to find t. Substituting t into the formula for routing costs yields the desired RLE formula.

A.1 Routing Costs for Single and Multiple Tours

Let us now see what happens when the giant tour (travelling salesman tour) is partitioned into feasible vehicle routes. Using equation (12) we can see that the average line-haul and detour distances for a travelling salesman tour are ad/n and $bD/\sqrt{n}$, respectively.

Suppose that we have t feasible routes, each serving, on average, $c = n/t$ customers. We can observe that the average line-haul distance for each tour remains aD/n. It is somewhat more cumbersome to calculate the average detour distance. In the travelling salesman tour, the detour distance is divided into $n-1$ route segments, connecting two neighbouring customers. Thus, the average distance between neighbouring customers on a tour is $\frac{bD}{\sqrt{n}(n-1)}$. For a feasible route serving c customers, the detour distance will be $c-1$ times this average distance, i.e. $\frac{bD(c-1)}{\sqrt{n}(n-1)}$. This gives us the length of an average feasible route, $\frac{aD}{n}+\frac{bD(c-1)}{\sqrt{n}(n-1)}$. To find the total routing cost, we simply multiply this by t. This gives us the routing cost as

$$T \approx \frac{at}{n} + \frac{bD(n-t)(n-1)}{\sqrt{n}} \qquad \text{(A1)}$$

A.2 The Effect of the Constraints on the Number of Routes:

We need to find t to complete the above formula. To do this, we first calculate c, then use the relation $n = ct$ to get t. There are two constraints in our problem, the *maximum capacity constraint* P and the *maximum distance constraint* Z. Let us look again at the routing algorithm. When partitioning the giant tour, we go along the tour as long as customers can be added without violating either of the above constraints. If we have on average c stops on a route, this means that c customers per route do not violate any of the constraints, but $c+1$ would. This allows us to establish both lower and upper bounds for c.

We begin with the maximum capacity constraint. The average number of stops per route times the average customer demand Q/n should not exceed the maximum load P. On the other hand, this constraint should be violated by allowing $c+1$ stops:

$$c\tfrac{Q}{n} \leq P < (c+1)(\tfrac{Q}{n}) \qquad \text{(A2)}$$

Rearranging this gives:

$$(nP/Q) - 1 < c \leq nP/Q \qquad \text{(A3)}$$

This gives us an estimate for c as:

$$c \approx (2nP - Q)/2Q \qquad \text{(A4)}$$

The above method of estimation is more accurate than estimating c by nP/Q, since it accounts for the fact that the vehicles are usually not fully utilised. We can easily transform equation (A4) to find t:

$$t \approx 2nQ/(2nP - Q) \qquad \text{(A5)}$$

This formula needs one modification. We have to make allowances for the fact that an integer number of tours is required. This usually results in the "last" tour constructed being under-utilised, although different algorithms may spread the slack over all the routes. This modification is especially important if we only deal with one or a few routes per depot. Thus, we take t to be the upper integer of the right-hand side of equation (A5), to obtain

$$t \approx \text{Int}(2nQ/2nP - Q)) + 1 \qquad \text{(A6)}$$

Now we proceed to find a formula for t in terms of the maximum distance constraint Z. Unfortunately, this is more complicated as a number of other variables exist in the expression. One of them will be the *drop time* δ, as the maximum distance requirement includes this term. Otherwise, we proceed as in the previous paragraph. The average distance per vehicle route plus the total drop time should not exceed the maximum distance Z. We look back to equation (A3) to find the average length of a vehicle route. Taking the total drop time per route $c\delta$ into account, the maximum distance requirement gives:

$$a\frac{D}{n} + \frac{bD(c-1)(n-1)}{\sqrt{n}} + c\delta \leq Z < aD/n + \frac{bDc(n-1)}{\sqrt{n}} + (c+1)\delta \tag{A7}$$

using c and $c+1$ stops per route respectively. This can be re-formulated to give upper and lower bounds for c:

$$\frac{Zn(n-1)-aD(n-1)-\delta n(n-1)}{b\sqrt{n}D+n(n-1)\delta} < c \leq \frac{Zn(n-1)-aD(n-1)+b\sqrt{n}D}{b\sqrt{n}D+n(n-1)\delta} \tag{A8}$$

We estimate c in the same manner as in the previous section:

$$c \approx \frac{2Zn(n-1)-2aD(n-1)-\delta n(n-1)+b\sqrt{n}D}{2b\sqrt{n}D+2n(n-1)\delta} \tag{A9}$$

We transform this into an estimate for t:

$$t \approx \frac{2bn\sqrt{n}D+2n^2(n-1)\delta}{2Zn(n-1)-2aD(n-1)-\delta n(n-1)+b\sqrt{n}D} \tag{A10}$$

We note again that this estimation for t accounts for the slack between tour lengths and maximum distance requirements.

The same modification as in the case of the maximum capacity constraint should be made, this gives us an integer estimate for t, namely:

$$t \approx Int\left(\frac{2bn\sqrt{n}D+2n^2(n-1)\delta}{2Zn(n-1)-2aD(n-1)-\delta n(n-1)+b\sqrt{n}D}\right) + 1 \tag{A11}$$

Unfortunately, this formula is much more complicated than equation (A6), and it also includes regression coefficients. The reason for this is, that while the maximum capacity requirement produces an estimate of t in terms of the total demand, which is a known quantity, the maximum distance requirement, if used in a similar fashion, would produce t as a function of the total tour length, which is an unknown quantity. In fact, it is the very quantity we wish to estimate. Thus, we have to use "estimation within estimation" to get our result. This is the reason for this formula being much more complicated than the previous one. Furthermore, the drop time has to be included, giving extra twist to the formula.

Now we combine the two formulae in the previous paragraphs to allow for the case of either of the constraints being the binding one. The binding constraint is the one which necessitates a larger number of tours, so t should be taken to be the maximum of the two expressions given in equations (A6) and (A11). This approach was also adopted by Salhi (1987), but in his approach the tour length was estimated using an approximation procedure, as opposed to our algebraic manipulations. This gives us:

$$t \approx \text{Int}\left(\text{Max}\left(\frac{2nQ}{2nP-Q}; \frac{2bn\sqrt{n}D+2n^2(n-1)\delta}{2Zn(n-1)-2aD(n-1)-\delta n(n-1)+b\sqrt{n}D}\right) + 1\right) \tag{A12}$$

This formula can be further improved. We note, that it may well be probable that for some of the routes of a depot the distance constraint, and for others the capacity constraint, is binding. This is most likely to occur when the two expressions for t in the previous paragraphs are very near. Thus, we have decided to introduce further regression coefficients, to give us better flexibility. The modified formula is as follows:

$$t \approx \text{Int}\left(\alpha + \beta \text{Max}\left(\frac{2nQ}{2nP-Q}; \frac{2bn\sqrt{n}D + 2n^2(n-1)\delta}{2Zn(n-1) - 2aD(n-1) - \delta n(n-1) + b\sqrt{n}D}\right)\right) \tag{A13}$$

where a, b, α and β are regression coefficients. Should we not wish to use regression, or if we need starting values for the regression process, the following values may be used: $a = 1.8$, $b = 1.8$, $\alpha = 1.0$ and $\beta = 1.0$. The values for a and b are as suggested by Daganzo (1984), while α and β are taken from equation (A12).

A.3 Completing the Estimation Formula

Now we insert the formula for t given by equation (A13) into our formula for routing cost per depot provided by equation (A1). This gives

$$\begin{aligned} T \approx & \left(a\frac{D}{n} - b\frac{D}{\sqrt{n}(n-1)}\right) \\ & \times \text{Int}\left(\alpha + \beta \text{Max}\left(\frac{2nQ}{2nP-Q}; \frac{2bn\sqrt{n}D + 2n^2(n-1)\delta}{(n-1)(2Zn - 2aD - \delta n) + b\sqrt{n}D}\right)\right) + b\frac{D\sqrt{n}}{n-1} \end{aligned} \tag{A14}$$

Thus, we have found a formula for the total routing cost per depot T in terms of D, n, Q, P, Z and δ, all known quantities during our solution process.

ONSTANTINE
ORPHYROGENETUS
TERNATIONAL
SSOCIATION

JOURNAL OF THE INSTITUTE OF INTERDISCIPLINARY
STUDIES IN ENVIRONMENTAL MANAGEMENT,
LOCATIONAL DECISIONS AND REGIONAL PLANNING

IS EN
LO RE

emper
xcellens

Studies in LOCATIONAL ANALYSIS

S

Recent Advances in Locational Analysis

Guest Editors:

Horst W. Hamacher
Kathrin Klamroth
Stefan Nickel

UE 10

OBER 1996

I 1105-5162

Studies in
LOCATIONAL ANALYSIS

ISSN 1105-5162

Aims

The Studies in Locational Analysis is a journal each of whose issues is devoted to a particular topic or theme, or is a report on a workshop or the proceedings of a conference. Specifically, the aims of The Studies include:
* the production of interesting and academically valuable issues;
* each issue devoted to a special theme or topic, or a meeting report;
* issues to appear periodically with, on average, two issues per year;
* not to be in direct competition with regular journals;
* guest editor(s) expert on the subject of the specific issue;
* to be relatively inexpensive to encourage individual subscription.

Subscription rates

While the aims above dictate a relatively low subscription rate, one which will largely cover running costs is required. Current rates ***per issue*** are:
20$USA (5000 Greek drachmas) for individual subscribers
40$USA (10000 Greek drachmas) for institutions

INSTITUTE OF INTERDISCIPLINARY STUDIES IN ENVIRONMENTAL MANAGEMENT, LOCATIONAL DECISIONS AND REGIONAL PLANNING (ISENLORE)

Correspondence: J. Karkazis, Smirnis 1, 15772 Zografou, Greece (tel/fax: 30-1-8227977, 30-1-6123631)

CONSTANTINE
PORPHYROGENETUS
INTERNATIONAL
ASSOCIATION

Studies in
LOCATIONAL ANALYSIS

Semper Excellens

CONTENTS

ISSUE 10

OCTOBER 1996

ISSN 1105-5162