ITLS

WORKING PAPER
ITLS-WP-07-03

Planning Approximations to the Length of TSP and VRP Problems

By

Miguel Andres Figliozzi

March 2007

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-07-03

TITLE: Planning Approximations to the Length of TSP and VRP Problems

ABSTRACT: This paper studies parsimonious, intuitive, and effective formulas to approximate the length of Traveling Salesman Problems (TSP) and Vehicle Routing Problems (VRP). Using intuition derived from continuous models and graph theory, a formula to approximate the length of vehicle routes is proposed. In instances with different patterns of customer spatial distribution, time windows, customer demands, and depot locations are used to test the proposed approximation. Regression results show that the approximation can reasonably predict the length of TSP and VRP problems in randomly generated problems and real urban networks. Expressions for the incremental cost of serving an additional customer or increasing the number of routes are derived and estimated. The main contribution of this paper is to develop and test intuitive approximations to TSP and VRP problem in general settings. The approximations are valuable for strategic and planning analysis of transportation and logistics problems.

KEYWORDS: ***Traveling salesman problem, vehicle routing problem, distance estimation, marginal cost, simulated experiments, case study.***

AUTHOR: Miguel Andres Figliozzi

CONTACT: Institute of Transport and Logistics Studies (C37)
An Australian Key Centre
The University of Sydney NSW 2006 Australia

Telephone: +61 9351 0071
Facsimile: +61 9351 0088
E-mail: itlsinfo@itls.usyd.edu.au
Internet: http://www.itls.usyd.edu.au

DATE: March 2007

1. INTRODUCTION

In many logistics problems it is necessary to estimate the distance that a fleet of vehicles travel to meet a set of costumer demands. Traveled distance is not only an important element of carriers' variable costs but it is also a key input in tactical and strategic models to solve problems such as facility location, fleet sizing, and network design.

Previous research has either focused on approximating distances in Traveling Salesman Problem (TSP) or Capacitated Vehicle Routing Problem (CVRP). Although the CVRP can be considered a TSP with capacity constraints, there is no unifying framework to approximate distances in both problems. In addition, existing models to estimate distance have not considered Vehicle Routing Problems with time windows[1] (VRP). This is a critical gap in the literature as time windows are becoming increasingly important with the wider implementation of customer-responsive and made-to-order supply chains in both the manufacturing and service sectors.

The unique contributions of this work are threefold: (a) it provides an intuitive and unifying mathematical framework to estimate distances in TSP and VRP problems, (b) it considers routing problems with time windows, and (c) it tests the approximation in not only randomly generated scenarios but also in real world urban networks. Although the approximations can be used for operational purposes, the approximations are intended for strategic and planning analysis of transportation and logistics problems.

The paper is organized as follows: Section 2 provides a literature review; Section 3 derives approximation formulas based on intuition and insights borrowed from graph theory and continuous approximation models; Section 4 describes the experimental design, including how the test problems were generated. The experimental results are analyzed and discussed in Section 5; the description and analysis of a real world application is presented in Section 6, followed by concluding comments in the final section.

2. LITERATURE REVIEW

There exists an extensive body of TSP and VRP related literature in operations research and transportation journals. The goal of this section is not to present a review of TSP and VRP solution methods but to focus on the literature that deals with the estimation of distances in TSP and VRP problems. Comprehensive reviews of solution methods for TSP and VRP problems are found in Gutin and Punnen (2002) and Toth and Vigo (2001) respectively.

A seminal contribution to estimate the length of a shortest closed path or tour through a set of points was established by Beardwood et al. (1959) . These authors demonstrated that for a set of n points distributed in a *s*-dimensional space $\Re$, with a probability of one, the following result holds:

[1] For the sake of brevity, VRP is used to indicate vehicle routing problems with both capacity constraints (CVRP) and time windows (VRPTW).

$$\lim_{n\to\infty}\; n^{-(s-1)/s}\, TSP(\boldsymbol{V}^n) = \beta_s d^{1/2} \int_{\aleph} p^{(s-1)/s} du \tag{1}$$

Where a sequence of points or vertices in $\Re$ is $\boldsymbol{V} = v_1, v_2, ...$ and the first n points of $\boldsymbol{V}$ are denoted $\boldsymbol{V}^n = \{v_1, ..., v_n\}$, the distance between two points is the ordinary Euclidian distance and $TSP(\boldsymbol{V}^n)$ is the distance of the shortest closed path tour through $\boldsymbol{V}^n$. The tour through $\boldsymbol{V}^n$ is confined to the bounded set $\aleph$ and its points are independently distributed over $\aleph$ with a common probability function p, whose density is taken with respect to a Lebesgue measure. For a two dimensional space and uniform distribution of points in a circle of area one, Bearwood et al. (1959) established that $\beta_s s^{1/2} \approx 0.53\sqrt{2}$.

Although the results from Bearwood et al. are valid for any density function, posterior contributions have applied expression (1) to problems where the points are randomly and uniformly distributed in an area A with a constant density $\delta = n / A$. For two dimensions or $s = 2$, Euclidian metric, and uniform distributions, the constant term $k = \beta_s s^{1/2}$ has been estimated at $k = 0.765$ (Stein, 1978). For reasonably compact and convex areas, the limit provided by expression (1) converges rapidly (Larson and Odoni, 1981). In these types of areas, the following approximation formula can be used:

$$E[TSP(\boldsymbol{V}^n)] \approx 0.765\sqrt{nA} \tag{2}$$

Where A denotes the area of the set $\aleph$. As long as feasibility is satisfied, economies of density are achieved in expressions (1) and (2) because distance grows slower than the number of customer requests. Formula (2) requires a Euclidean travel metric or L_1 metric. Jaillet (1988) estimated the constant $k \approx 0.97$ for Manhattan travel metric or L_2 metric.

Approximations to the length of capacitated vehicle routing problems were first published in the late 1960's and early 1970's (Webb, 1968, Christofides and Eilon, 1969, Eilon et al., 1971). Webb studies the correlation between route distance and customer-depot distances. Eilon et al. (1971) propose several approximations to the length of CVRP based on the shape and area of the delivery area, the average distance between customers and the depot, the capacity of the vehicle in terms of the number of customers that can be served per vehicle, and the area of a squared delivery region.
Daganzo (1984) proposed a simple and intuitive formula for the CVRP when the depot is not necessarily located in the area that contains the customers.

$$E[CVRP(\boldsymbol{V}^n)] \approx 2\bar{r}\, n / C + 0.57\sqrt{nA}) \tag{3}$$

Where $CVRP(\boldsymbol{V}^n)$ is the total distance of the CVRP problem serving n customers, the average distance between the customers and the depot is $\bar{r}$, and the maximum number of customers that can be served per vehicle is C. Expression (3) can be interpreted as having: (a) a term related to the distance between the depot and customers and (b) a term related to the distance between customers. The coefficients of expression (3) were derived assuming $C > 6$ and $N > 4C^2$. Daganzo's approximation works better in elongated areas as the routes were formed following the "strip" strategy. Robuste et al.

(2004) use simulations to analyze elliptical areas and propose adjustments based on area shape, C, and n.
Chien (1992) carried out simulations and linear regressions to test the accuracy of different models to estimate the length of TSP. Chien tested rectangular areas with 8 different length/width ratios ranging from 1 to 8 and circular sectors with 8 different central angles ranging from 45 to 360 degrees. Exact solutions to solve the TSP problems were used and the size of the problems is 5 to 30 customers. The depot was always located at the origin, the left-lower corner of the rectangular areas. Chien randomly generated test problems and using liner regressions found the best fitting parameters. The mean absolute percentage error (*MAPE*) was the benchmark to compare specifications.

The fitted models were:

$$E[TSP(\boldsymbol{V}^n)] \approx 0.26\bar{r} \qquad R^2 = 0.80 \qquad MAPE = 40.1 \tag{4}$$

$$E[TSP(\boldsymbol{V}^n)] \approx 0.88\sqrt{(n+1)A} \qquad R^2 = 0.97 \qquad MAPE = 18.6 \tag{5}$$

$$E[TSP(\boldsymbol{V}^n)] \approx 1.95\bar{r} + 0.62\sqrt{nA}) \qquad R^2 = 0.99 \qquad MAPE = 8.3 \tag{6}$$

The R^2 is obtained from the linear regression. *MAPE* is a measure of the average accuracy of the estimator. Chien finds that the lowest *MAPE* equals 6.9% using the area of the smallest rectangle that covers only the customers; this area is denoted R.

$$E[TSP(\boldsymbol{V}^n)] \approx 2.1\bar{r} + 0.67\sqrt{nR}) \qquad R^2 = 0.99 \qquad MAPE = 6.9 \tag{7}$$

However, expression (7) is not convenient for planning purposes when there may be many possible subsets of costumers; expression (7) requires the estimation of R for each subset of customers. Using the analytical approximations derived for compact areas, the resulting *MAPE*s are:

$$E[TSP(\boldsymbol{V}^n)] \approx \bar{r}n \qquad MAPE = 202.1 \tag{8}$$

$$E[TSP(\boldsymbol{V}^n)] \approx 0.75\sqrt{(n+1)A} \qquad MAPE = 23.0 \tag{9}$$

$$E[TSP(\boldsymbol{V}^n)] \approx 2\bar{r} + 0.57\sqrt{nA}) \qquad MAPE = 8.8 \tag{10}$$

The previous models were also estimated for each of the 16 different regions; R^2 and *MAPE* are reported for each type of region and model. The estimated parameters change according to the shape of the region.

Kwon et al. (1995) also carried out simulations and linear regressions but in addition they also used neural networks to find better approximations. To test the accuracy of different models they tested TSP problems in rectangular areas with 8 length/width ratios ranging from 1 to 8. Models were estimated with the depot being located at the origin and at the middle of the rectangle. The sizes of the problems range from 10 to 80 customers. Kwon et al. (1995) compare models (6) and (7) with two additional models that make use of the geometric information proportioned by the ratio length/width of the rectangle (length and width defined in such a way that the ratio is always larger or equal to 1). The results obtained for the depot located at the origin are as follows:

$$E[TSP(\boldsymbol{V}^n)] \approx 2.02\bar{r} + 0.57\sqrt{(n+1)A} \qquad R^2 = 0.99 \qquad MAPE = 4.68 \qquad (11)$$

$$E[TSP(\boldsymbol{V}^n)] \approx 2.28\bar{r} + 0.55\sqrt{nR} \qquad R^2 = 0.99 \qquad MAPE = 5.00 \qquad (12)$$

$$E[TSP(\boldsymbol{V}^n)] \approx [0.83 - 0.0011(n+1) + 1.11S/(n+1)]\sqrt{nA} \qquad R^2 = 0.99 \qquad MAPE = 3.71 \qquad (13)$$

$$E[TSP(\boldsymbol{V}^n)] \approx 0.41\bar{r} + [0.77 - 0.0008(n+1) + 0.90S/(n+1)]\sqrt{nA} \qquad R^2 = 0.99 \qquad MAPE = 3.61 \qquad (14)$$

Accounting for the shape of the area improves accuracy, although this is at the expense of adding one and two extra terms in the last two expressions. With the depot located at the center of the rectangle the results obtained are as follows:

$$E[TSP(\boldsymbol{V}^n)] \approx 3.38\bar{r} + 0.59\sqrt{(n+1)A} \qquad R^2 = 0.99 \qquad MAPE = 4.72 \qquad (15)$$

$$E[TSP(\boldsymbol{V}^n)] \approx 3.98\bar{r} + 0.58\sqrt{nR} \qquad R^2 = 0.99 \qquad MAPE = 4.45 \qquad (16)$$

$$E[TSP(\boldsymbol{V}^n)] \approx [0.87 - 0.0016(n+1) + 1.34S/(n+1)]\sqrt{nR^1} \qquad R^2 = 0.99 \qquad MAPE = 3.88 \qquad (17)$$

$$E[TSP(\boldsymbol{V}^n)] \approx 1.15\bar{r} + [0.79 - 0.0012(n+1) + 0.97S/(n+1)]\sqrt{nR^1} \qquad R^2 = 0.99 \qquad MAPE = 3.70 \qquad (18)$$

Where R^1 is defined as the area of the smallest rectangle that covers the customer and the depot. Kwon et al. (1995) also used neural networks to find a model that better predicts TSP length. They conclude that the capability of neural networks to find "hidden" relationships provides a slight edge against regression models.

This section has thoroughly reviewed approximations and simulation results for TSP. There are strong theoretical and intuitive reasons to include both $\sqrt{nA}$ and $\bar{r}$ terms in the models. More accuracy can be obtained if additional terms related to the shape of the region and customers are added, as in Kwon et al. Although Daganzo (1984) and Robuste et al. (2004) propose distance formulas for the CVRP, to the best of the author's knowledge there is no published research that reports $MAPE$ and simulation results for the CVRP. Distance-wise, problems with time windows have not been analyzed in the literature. In the next section a simple and intuitive approximation to the distance required for TSP and VRP problems is proposed.

3. ALTERNATIVE FORMULA TO APPROXIMATE THE DISTANCE OF TSP AND VRP PROBLEMS

In this paper a VRP is associated with a set of points in $\Re^2$ that is denoted $\boldsymbol{V}^n = \{v_0, v_1, ..., v_n\}$, where the graph associated with the set of vertices $\boldsymbol{V}^n$ is denoted $G = (V, L)$. The set of links or arcs is denoted $L = \{(v_i, v_j) : v_i, v_j \in V, i \neq j\}$. The distance between two points is the ordinary Euclidian distance. Vertex v_0 represents the depot and $\boldsymbol{C}^n = \boldsymbol{V}^n \setminus v_0$ represents the set of customers' requests. There is a set of vehicles denoted by $\boldsymbol{T}$. Each vehicle route must start and finish at the depot; the routes must satisfy a set of time windows and capacity constraints denoted by $\boldsymbol{Z}$. The first objective is to build the smallest number of m routes such that $m \leq |\boldsymbol{T}|$, $\boldsymbol{Z}$ is satisfied. The secondary objective is to minimize total distance. Each route j serves a set of $\boldsymbol{V}_j^n \subseteq \boldsymbol{V}^n$ customers and $\boldsymbol{V}_j^n \cap \boldsymbol{V}_k^n = \emptyset$ for all routes where $j \neq k$. Using this notation, a special case of a VRP with $|\boldsymbol{T}| = 1$ and $\boldsymbol{Z} = \emptyset$ is equivalent to a TSP.

For the sake of simplicity, let $\boldsymbol{V}_j^n = \{v_{j,1}, v_{j,2}, ..., v_{j,n}\}$ also denote the customer service sequence or order, e.g., vehicle j first visits $v_{j,1}$, then $v_{j,2}$, and so on. The distance of the path connecting $\{v_{j,1}, v_{j,2}, ..., v_{j,n}\}$ is denoted *local* or intra-customer distance $l\{v_{j,1}, v_{j,2}, ..., v_{j,n}\}$ or simply $l\{\boldsymbol{V}_j^n\}$. The distance of the remaining links, that is $\{(v_0, v_1); (v_n, v_0)\}$ is denoted by the *connecting* distance $c\{\boldsymbol{V}_j^n\}$. For any set of customers $\boldsymbol{V}_j^n$ it follows that $TSP\{\boldsymbol{V}_j^n\} = l\{\boldsymbol{V}_j^n\} + c\{\boldsymbol{V}_j^n\}$. The definition of these distances can be extended to each of the m routes of a VRP problem with:

$$VRP\{\boldsymbol{V}^n\} = \sum_m TSP\{\boldsymbol{V}_m^n\} = \sum_m l\{\boldsymbol{V}_m^n\} + \sum_m c\{\boldsymbol{V}_m^n\} = VRP_l\{\boldsymbol{V}^n\} + VRP_c\{\boldsymbol{V}^n\} \tag{19}$$

For one route and n customers, any solution to the TSP uses only $n+1$ links. In a TSP, $n-1$ links are *local* and 2 links are *connecting*. If capacity and/or window constraints are added to the TSP, the resulting VRP has a number of routes $m \geq 1$. In general, for m routes and n customers any solution to the VRP uses $n+m$ links. The number of *connecting* links is $2m$ and the number of *local* links is $n-m$. The following observations can be made assuming a uniform distribution of customers:

- For a TSP, as the number of customers increases the relative importance of the local and connecting distance increases and decreases respectively – Vice versa as the number of customers decreases

- As more constraints are added to a TSP and the number of routes m grows, the number of connecting links increases and the number of local links decreases.

- A TSP for the set of customers $\boldsymbol{C}^n$ has exactly n links.

Based on these observations, the following formula is proposed to estimate the length of TSP and VRP:

$$VRP\{\boldsymbol{V}^n\} \approx k_l \frac{n-m}{n}\sqrt{An} + k_b\sqrt{A/n} + k_m m \tag{20}$$

Where k_l, k_b, and k_m are parameters to be estimated by linear regression.

The first term of expression (20) tries to approximate the *local* tour distance. The first term has the desirable property that when $n = m$, the *local* distance is zero. Alternatively, when $n >> m$ the first term approximates the local tour distance as suggested by the Bearwood et al. The third term approximates the connecting distance. Observing Daganzo's expression (4), this term is related to the average distance from the depot to the customers. In expression (4) the term n/C is equivalent to the number of routes necessary due to capacity constraints. The second term is readily associated with the connecting distance when the depot is located within or close to A. Intuitively, for a depot located within or sufficiently close to A, the distance from the depot to the first and last customer of the tour tends to decrease as the number of customers increases. The second term is the *bridging* component between the local tour and the average distance from the region to the customer area. The value of parameters k_l, k_b, and k_m are to be determined by linear regression and capture the influence of remaining factors such as spatial customer distributions, depot location, and time windows.

The incremental distance of an additional route or customer is denoted $\Delta(m)$ and $\Delta(n)$ respectively and can be obtained taking the derivative of the distance with respect to m and n. These values can also be thought of as the average marginal cost of adding a new customer to the existing routes and the marginal cost of tightening the constraints $\boldsymbol{Z}$ such that a new route is needed while keeping the number of customers served constant (see derivation in the appendix).

$$\Delta(m) = k_m - \frac{k_l\sqrt{A}}{\sqrt{n}} \qquad (21)$$

$$\Delta(n) = \frac{k_l\sqrt{A}}{\sqrt{n}}\left(\frac{1}{2} + \frac{m}{2n} - \frac{k_b}{2nk_l}\right) \qquad (22)$$

The value of $\Delta(m)$ is readily interpreted as the difference between the added connecting links and the replaced inter-customer link. The value of $\Delta(n)$ represents the distance of the additional inter-customer link added. This distance decreases when the number of customers increases and when the number of routes decreases. The value of $\Delta(n)$ is always less than the average inter-customer distance as expressed by $k_l\sqrt{A}/\sqrt{n}$ when k_b is positive or when $n-m$ is larger than the absolute value of the ratio k_b/k_l. Expression (20) is a linear function of m and the derivative $\Delta(m)$ is the real average rate of change. However, expression (20) is a convex non-linear function of n and the derivative $\Delta(n)$ slightly overestimates the average distance increase when a new customer is added.

The next section describes the experimental design used to test expression (20) in settings with diverse customer demands, time windows, depot location, and geographic distribution of customers. To evaluate the prediction accuracy the $MAPE$ and the MPE (Mean Percentage Error) are used, which are calculated as follows:

$$MPE = \frac{1}{p}\sum_{i=1}^{p}\frac{D_i - E_i}{D_i}*100\%$$

$$MAPE = \frac{1}{p}\sum_{i=1}^{p}\frac{|D_i - E_i|}{D_i}*100\%$$

Where the actual distance for instance i is denoted D_i and the estimated distance is denoted E_i. For a given set of instances it is always the case that $MPE \le MAPE$. The MPE indicates whether the estimation on average overestimates or underestimates the actual distance; the $MAPE$ provides the average deviation between actual and estimated distance as a percentage of the actual distance.

It remains to be seen how well the derived expression works, more specifically:

1. How does expression (20) compare with the expression derived for the CVRP?
2. Does the adjusting factor $(n-m)/n$ improve the accuracy of the predictions?
3. What is the impact of time windows on the accuracy of the approximations and estimated coefficients?
4. What is the impact of depot location on the accuracy of the approximations and estimated coefficients?

To answer the first two questions two alternative models are used: (a) the “CVRP” model, expression (23), which is similar to previously derived formulas for the CVRP; and (b) the “base” model, expression (24), that is equivalent to the “adjusted” model (20) but without the correcting factor.

$$VRP\{\mathbf{V}^n\} \approx k_1\sqrt{An} + k_3 m \tag{23}$$

$$VRP\{\mathbf{V}^n\} \approx k_1\sqrt{An} + k_2\sqrt{A/n} + k_3 m \tag{24}$$

Expression (20), herein is referred to as the “adjusted” model since it includes the factor $(n-m)/n$. The next section describes the experimental setting.

4. EXPERIMENTAL SETTING

The classical instances of the VRP with time windows proposed by Solomon (1987) are used in this research. The Solomon instances are chosen because they incorporate key factors that affect the behavior of routing algorithms such as geographical distribution of customers, the number of customers serviced by a vehicle, time-constrained customers, vehicles capacities, and customer demands. These problems have not only been widely studied in the operations research literature and but they are readily available[2].

In the Solomon problems there are 100 randomly generated customers per instance. The distances and travel times are Euclidean. There are six different classes of problems depending on the geographic location of customers (R: random; C: clustered; RC: mixed random and clustered) and time windows length (1: short time windows; 2: long time windows). The customer coordinates are identical for all problems within one type (i.e., R, C and RC). The sets R1, C1 and RC1 have vehicle capacity C of 200 units, allowing fewer customers per route than the remaining sets. In contrast, problem sets R2, C2, and RC2 have vehicle capacity C equal to 1000, 700 and 1000 units, respectively, allowing a larger number of customers per route.

Due to the short time windows, problem sets R1, C1 and RC1 allow only a few customers per route (approximately 5 to 10). Problem sets R2, C2 and RC2 have longer

[2] Many websites maintain downloadable datasets of the instances including Solomon’s own website: http://web.cba.neu.edu/~msolomon/problems.htm

time windows and route sizes are in the order of 30 customers serviced by the same vehicle. Figures 1, 2, and 3 illustrate the distribution of customer locations in instances of problems R, C, and RC respectively. Table 1 provides statistical information regarding customers coordinates, demands, time windows (beginning, end, length), and service time spent at the customer.

The approximations (20), (23), and (24) are proposed and tested for planning purposes. Out of n possible customers in the service area A, a problem or instance is formed by any non-empty subset of $\boldsymbol{C}^n$ and the depot v_0. Instances can differ not only in the subset of customers served but also in their demand levels. Hence, using the first instance of the six problem types proposed by Solomon, 15 subsets of customers of size 70, 60, 50, 40, 30, 20, and 10 were randomly selected from the original 100 customers. To incorporate different levels of customer demand, new instances were created applying the demand factors presented in Table 2 to each subset of customers. Applying the factors in the first row of table 2 (row of ones) the customers have similar demands as in the original Solomon problems. The resulting problems using the highest demand multipliers (last row of table 2) are such that some customers are truckload (TL) or almost TL customers. This was done to test the approximations when problems are highly constrained and have a large number of routes. In the Solomon problems the depot has a central location with respect to the customers. To test the approximation when the depot is located in the periphery all the created instances where also solved with the depot located at the origin, i.e. coordinates (0,0). To study the approximation quality and parameter values without time windows all the problem instances were also solved without time windows. The combination of settings renders 15*7*6*2*2 = 2,520 instances.

The instances were solved with a VRP heuristic that have reported solutions within the 3% of the best solutions results found by any algorithm or solution approach (Figliozzi and Bain, 2007). It is worth mentioning that out of the many solutions approaches for the VRPTW, there is no method that can provide optimal or even the best known solution for all instances as indicated by Braysy and Gendreau (2005a, 2005b). Even for CVRPs there are only relatively few instances that can be solved optimally. For example, no method can consistently solve instances where the CVRP has more than 50 customers (Cordeau et al., 2002). Ideally, optimal solutions will be used to compute routes and distances. However, the high fitting of the approximations strongly suggests that trends and results presented in this research will be valid when other solution algorithms are used.

5. ANALYSIS AND DISCUSSION OF EXPERIMENTAL RESULTS

To illustrate and provide intuition about the behavior of the different terms of expression (20) several graphs have been prepared. The relationship between tour distance and connecting distance in TSP problems is shown in Figure 4. Sets of 2, 5, 10, 20, 30, 40, and 50 customers were randomly extracted from the original R101 Solomon problem with 100 customers and the depot has a central location. As shown in figure 4 the connecting distance quickly decreases and the number of customers increases. As expected, the rate of decrease tapers down as $k\sqrt{A/n}$ anticipates. The tour distance increases but at a rate that tapers down as the number of customers grows; this change is well captured by $k\sqrt{An}$. These results show that when the number of customers per tour is less than 10 the importance of the connecting distance is evident.

The relationship between tour distance and connecting distance in VRP problems is shown in figures 5, 6, and 7. Sets of 10, 20, 30, 40, 50, 60, and 70 customers were randomly extracted from the original R101 problem. To produce problems with different number of routes, the customer demand was multiplied by the demand factors shown in table 2. Sets with low customer demand have few routes while sets with high customer demand have many routes. Time windows were not considered, therefore, the results correspond to distances of CVRP instances. Figure 5 shows that tour distance follows the trend already seen in figure 4. The connecting distance shows a different pattern (see figure 6). Clearly, the number of customers alone is not a good predictor of the connecting distance. As shown in figure 7, the number of routes is a good predictor of the connecting distance and a poor predictor of the tour distance.

Results for CRVP instances, i.e. no time windows, and the depot located at the center are shown in tables 3, 4, and 5 respectively. Model fit (R^2, MAPE, and MPE), estimated coefficients, and the probability of coefficients being equal to zero are displayed for the CVRP, base, and adjusted model in tables 3, 4, and 5 respectively. Herein, all the regression results were obtained forcing the intercept or constant term to be zero, which is consistent with previous studies by Chien (1992) and Kwon et al. (1995). In the regression models, the average distance per sample size is the dependent variable.

For the sake of clarity, only 3 decimals are displayed in the tables. All three models have such a good R^2 that no significant differences can be observed; in all three models the estimated coefficients are significant, and the probability at a 95% confidence level is shown with a 3 decimal level of detail. The *MAP* shows that the base and adjusted model do not significantly overestimate or underestimate the distances. All the models present good *MAPE* values but the adjusted model is superior to the base model. The CVRP model, despite its good R^2, does not predict distances as well as the other models.

The effect of time windows is shown in tables 6, 7, and 8. These results are obtained using the same instances used previously in obtaining tables 3, 4, and 5 but considering all the customer time windows as originally intended in the Solomon problems. A slight decrease in the R^2 values is observed; imposing time windows decreases the predicting ability of the three models. However, the CVRP model is the most affected; its *MPE* and *MAPE* increase almost 7 and 3 times respectively. The base and adjusted models fare better with a *MPE* and *MAPE* increase of 50% with the adjusted model still showing the best prediction power with an average *MAPE* of 3.7%.

Time windows also affect the value of the estimated parameters, although the impact is different for tours and connecting distances. The adjusted model shows the most consistent behavior. All local tour parameters k_l show an increase that is statistically significant at the 95% confidence level. This is intuitively correct since time window constraints do not allow the formation of close and tied routes. All connecting parameters k_b show a significant decrease. The parameters k_m are not significantly different at the 95% confidence level. This is intuitively correct as the k_m parameter is related to the number of routes and the distance between the depot and customers, and therefore, it is not influenced by changes in time windows. With the depot at the center the average increase in the distance associated with the parameters $k_l(n-m)\sqrt{An}/n + k_b\sqrt{A/n}$ is 50% for 10 customers and over 120% for 70 customers. There are also significant increases in the distance associated with the term $k_m m$ as the number of routes increase up to 30% on average for type 1 problems. The increase in the number of routes is more modest in type 2 problems, on average less than a 5% increase of the number of routes.

The same models were also estimated with the depot located at the corner, i.e., coordinates (0,0). Moving the depot to the corner increases the average distance between the depot and the customers considerably. The increase in average distance is indicated in table 9. Despite the increase, the same trends are still observed: a) the adjusted model is the best performer followed by the base model, b) time windows decreases the accuracy of the models, and c) with time windows the parameter k_l increases, the parameter k_b decreases, and the parameter k_m does not change (with a 95% confidence level). With the corner depot, all three models perform better in terms of *MPE* and *MAPE*.

Comparing the parameters obtained with a central an corner depot, it is observed that coefficient k_m increases on average at almost the same rate as the average depot-customer distance (roughly two times larger). However, the average of coefficients k_l and k_b show a slight change. This indicates that the $k_m m$ term becomes the dominant term of the distance as the depot moves further away from the customers. The ratio between $k_m/2$ and the average distance between depot and customers is within the interval (1.03-1.25) with a central depot and within the interval (0.97-1.10) with a corner depot. Intuitively, for large enough distances between the depot and a bounded customer area, the route distance can be estimated simply using the number of routes and the average distance between the depot and the customers.

Despite the change in sign for some k_b parameters – it becomes negative, with the depot at the corner the average increase in the distance associated with the parameters $k_l(n-m)\sqrt{An}/n + k_b\sqrt{A/n}$ is 70% for 10 customers and over 135% for 70 customers. There are also significant increases in the distance associated with the term $k_m m$, however, the percentage increase associated with type 1 and 2 instances is similar to the increase observed when the depot is at the center.

The estimated distance increases when a customer or route is added as shown in tables 16 and 17. Table 16 shows the values of $\Delta(n)$ and $\Delta(m)$ when the depot is located at the center and type 2 problems. The value of $\Delta(n)$ decreases as n increases which is an indication of economies of scope. The value of $\Delta(n)$ increases as m increases which is intuitive and expected from the formula. When time windows are present (TW column) the increase per customer added is in all cases significantly higher than without time windows (no TW column). The value of $\Delta(m)$ is the difference between the added connecting links and the replaced inter-customer link. The value of $\Delta(m)$ decreases with time windows because: (a) the average depot-customer distance remains constant and (b) the inter-customer distance is larger with time window constraints. Table 17 shows the values of $\Delta(n)$ and $\Delta(m)$ when the depot is located at the corner. When comparing tables 16 and 17 it is clear that the values of $\Delta(n)$ are similar in both tables. However, the values of $\Delta(m)$ increase reflecting the increase in the average depot-customer distance.

If the adjusted model is used for operational purposes, e.g. without averaging the 15 instances with the same number of customers, the parameters do not experience a significant change but the *MAPE* is increased. Roughly, the *MAPE* doubles without time windows and triples with time windows. For illustrative purposes, the operational results with a central depot and no time windows are presented in table 18. If table 5 and 18 are compared, it can be observed that the parameters have slightly changed but the *MAPE* have increased from a 2.6% to a 6.0% average. Therefore, if a higher predictive accuracy is required for operational purposes more parameters have to be added.

The experimental results have confirmed the better performance of the adjusted model, expression (20), for vehicle routing problems with different levels of customer demand, customer geographic distribution, time windows, and depot location. In addition, the behavior of the approximation can be intuitively understood.

6. REAL WORLD APPLICATION

Previous literature has solely tested TSP or CVRP distance approximations on simulated environments with Euclidian distances. Although approximation formulas have theoretical applications in transport and logistics planning, they can also be used to estimate distance, costs, and times in practical planning applications. The original motivation for this research came from the study of distribution routes for a freight forwarding company based in Sydney, Australia. Distribution tours originated at a depot located close to the port of Sydney; the customers were mostly located in different industrial suburbs. The pattern of customer distribution resembles the mix of random and clustered customers as in the RC Solomon problems. The company customers are in the hundreds but they are not visited every day. The freight forwarding company consolidates LTC (less than container) shipments and customers are visited if a consignment has arrived before the distribution cutoff time. Further details about the tour characteristics can be found in Figliozzi et al. (2007).

The adjusted model was tested with customers located in the industrial suburb of Bankstown with thirty costumers distributed in an irregular area of 39.5 squared kilometers (see map in figure 8). The delivery area is bordered by the Bankstown local airport in the west, a freeway in the south, and secondary highways in the east and north. The average distance between the depot and the industrial suburb is approximately 22 kilometers in the connecting freeway. To test the adjusted model five sets of 2, 4, 6, 8, 10, 15, and 20 customers were randomly chosen among the existing customers in the suburb to simulate the daily demand. Selecting random subsets of customers from the pool of existing customers in the area is a fair representation of the real demand. The number of customers visited per day varies widely; it may be as low as 1 or 2 or exceptionally close to thirty. In the results presented hereafter all customers have the same probability of a visit. Although this is not the case in reality, it simplifies the exposition and introduces greater variation in the customer subsets.

Due to contract and labor policies, the main distribution cost is associated with the number of driver hours needed. Therefore, the objective is to minimize total route durations avoiding expensive overtime (overtime pay rate is 50% higher). An important consideration when working with travel times in an urban area is that speeds are strongly influenced by road characteristics and speed limits. In this application the speeds used are: 65 km/hour in freeways, 35 km/hour in main connecting streets – four lanes or more and traffic lights, and 25 km/hour in local streets. With this speed information a matrix of shortest travel times between customers and depot was constructed using the real network and geographic information system (GIS) software. Figure 8 displays the relationship between the Euclidian distance and the distance based on the shortest time path – for all customers and the depot. The high concentration of short distance points close to the origin correspond to the distances between customers in the suburb while the longer distances are mostly depot-customer. The R^2 =0.93 indicates that despite the irregular shape of the distribution area and the mix of travel speeds the Euclidian distance is a fairly good predictor of the actual distance traveled between customer pairs or customer-depot pairs. From existing customer data, an average service time of 45 minutes is used.

Three different routing scenarios were constructed: (a) no constraints or TSP case, (b) with tour duration constraint of 8 hours, and (c) adding 4 hour time windows per customer. The number of routes varied from 1 route in the TSP instances to 5 routes in the instances with time windows. The model was estimated with all the data provided by pooling all three scenarios and instances together. The results are shown in table 19. The network distance traveled is well approximated with a *MAPE* of 4.2%. The prediction of travel time in hours has a *MAPE* of 7.9%. The good *MAPE* is not surprising giving the good correlation between distance traveled and time driven (see Figure 10). These results are encouraging and show that the proposed approximations may have useful applications in urban networks. While results are promising, from this example it is impossible to generalize these results. Further research efforts are necessary to study how effective the proposed formula may be in different cities and applications.

7. CONCLUSIONS

A new expression to approximate the distance traveled by vehicles in TSP and VRP problems was derived. The approximation is intended for strategic and planning analysis of transportation and logistics problems. The proposed approximation was tested in instances with different patterns of customer spatial distribution, time windows, customer demands, and depot locations. The experimental results indicate that the approximation outperforms other approximations in randomly generated instances. The approximation is parsimonious, effective, and intuitive. Expressions for the average marginal cost of adding a new customer or route are derived and estimated. It was found that time windows negatively affect the accuracy of the approximations. Time windows not only increase traveled distance because the number of routes is increased but also because the separation between customers per route is increased. As the distance between the depot and delivery region increases the accuracy of the approximation increases. The approximation was also tested in a real urban network with encouraging results. The proposed approximation may be also useful to estimate travel times in urban networks, though further research is necessary to validate and generalize this claim.

APPENDIX

To obtain the values of $\Delta(n)$ and $\Delta(m)$ it is easier to write expression (20) into the equivalent expression (25):

$$k_l \frac{n-m}{n}\sqrt{An} + k_b\sqrt{A/n} + k_m m \qquad (20)$$

$$= k_l (1-\frac{m}{n})\sqrt{A}n^{1/2} + k_b\sqrt{A}n^{-1/2} + k_m m =$$

$$= k_l\sqrt{A}[n^{1/2} - mn^{-1/2} + k_b / k_l n^{-1/2}] + k_m m \qquad (25)$$

Taking the derivative of expression (25) with respect to n:

$$\Delta(n) = k_l\sqrt{A}[\frac{1}{2}n^{-1/2} + \frac{1}{2}mn^{-3/2} - \frac{k_b}{2k_l}n^{-3/2}]$$

$$\Delta(n) = \frac{k_l\sqrt{A}}{\sqrt{n}}[\frac{1}{2} + \frac{m}{2n} - \frac{k_b}{2k_l n}]$$

Taking the derivative of expression (25) with respect to m:

$$\Delta(m) = k_m - \frac{k_l\sqrt{A}}{\sqrt{n}}$$

REFERENCES

Bearwood, J., Halton, H. & Hammersley, J. (1959) The Shortest Path Through Many Points. *Proceedings of the Cambridge Philosophical Society,* 55**,** 299-327.

Braysy, I. & Gendreau, M. (2005a) Vehicle routing problem with time windows, part 1: Route construction and local search algorithms. *Transportation Science,* 39**,** 104-118.

Braysy, I. & Gendreau, M. (2005b) Vehicle routing problem with time windows, part II: Metaheuristics. *Transportation Science,* 39**,** 119-139.

Chien, T. W. (1992) Operational Estimators For The Length Of A Traveling Salesman Tour. *Computers & Operations Research,* 19**,** 469-478.

Christofides, N. & Eilon, S. (1969) Expected Distances In Distribution Problems. *Operational Research Quarterly,* 20**,** 437-443.

Cordeau, J. F., Gendreau, M., Laporte, G., Potvin, J. Y. & Semet, F. (2002) A guide to vehicle routing heuristics. *Journal of The Operational Research Society,* 53**,** 512-522.

Daganzo, C. F. (1984) The Distance Traveled to Visit N-Points with a Maximum of C-Stops per Vehicle - An Analytic Model and an Application. *Transportation Science,* 18**,** 331-350.

Eilon, S., Watson-Gandy, D. & Christofides, N. (1971) *Distribution Management: Mathematical Modelling and Practical Analysis,* New York, Hafner.

Figliozzi, M. & Bain, S. (2007) An Iterative Route Construction Heuristic for the Vehicle Routing Problem with Time Windows. *Working Paper, Institute of Transport and Logistics Studies, University of Sydney*.

Figliozzi, M., Kingdon, L. & Wilkitzki, A. (2007) Commercial Vehicle Tours and Road Congestion in Urban Areas: Implications for Carriers' Operations and Public Data Collection and Dissemination Efforts. *Proceeding of the 86th Transportation Research Board Annual Meeting CD rom- January 2007 - Washington DC. USA.*

Gutin, G. & Punnen, A. (Eds.) (2002) *The Traveling Salesman Problem and its Variations,* Boston, MA, Kluwer.

Jaillet, P. (1988) Apriori Solution Of A Traveling Salesman Problem In Which A Random Subset Of The Customers Are Visited. *Operations Research,* 36**,** 929-936.

Kwon, O., Golden, B. & Wasil, E. (1995) Estimating The Length Of The Optimal Tsp Tour - An Empirical-Study Using Regression And Neural Networks. *Computers & Operations Research,* 22**,** 1039-1046.

Larson, R. C. & Odoni, A. R. (1981) *Urban Operations Research,* Prentice-Hall, Inc.

Robuste, F., Estrada, M. & Lopez-Pita, A. (2004) Formulas for Estimating Average Distance Traveled in Vehicle Routing Problems in Elliptic Zones. *Transportation Research Record,* 1873**,** 64-69.

Solomon, M. M. (1987) Algorithms For The Vehicle-Routing And Scheduling Problems With Time Window Constraints. *Operations Research,* 35**,** 254-265.

Stein, D. (1978) An asymptotic probabilistic analysis of a routing problem. *Mathematics Of Operations Research,* 3**,** 89-101.

Toth, P. & Vigo, D. (Eds.) (2001) *The Vehicle Routing Problem*, Society for Industrial and Applied Mathematics.

Webb, M. (1968) Cost functions in the location of depots for multiple delivery journeys. *Operational Research Quarterly,* 19**,** 311-315.

LIST OF TABLES

Instance	Statistic	X COORD.	Y COORD.	DEMAND	TW BEGIN	TW END	TW LENGTH	SERVICE TIME
C1	MIN	0	5	10	10	67	37	90
	MEDIAN	40	48	10	418	480	61	90
	AVERAGE	42	49	18	427	488	61	90
	MAX	95	85	50	1,054	1,127	89	90
R1	MIN	2	3	1	18	28	10	10
	MEDIAN	31	35	13	93	103	10	10
	AVERAGE	34	36	15	96	106	10	10
	MAX	67	77	41	200	210	10	10
R1	MIN	0	3	2	11	41	30	10
	MEDIAN	40	41	16	87	117	30	10
	AVERAGE	40	44	17	92	122	30	10
	MAX	95	85	40	192	222	30	10
C2	MIN	0	5	10	8	168	160	90
	MEDIAN	41	45	10	1,449	1,609	160	90
	AVERAGE	42	48	18	1,470	1,630	160	90
	MAX	95	85	50	3,119	3,279	160	90
R2	MIN	2	3	1	17	172	27	10
	MEDIAN	31	35	13	386	459	117	10
	AVERAGE	34	36	15	391	507	116	10
	MAX	67	77	41	849	980	212	10
RC2	MIN	0	3	2	11	131	120	10
	MEDIAN	40	41	16	351	471	120	10
	AVERAGE	40	44	17	371	491	120	10
	MAX	95	85	40	822	942	120	10

Table 1: Statistics of customer distributions by problem type

Instance	C1	R1	CR1	C1	R1	CR1
Capacity	200	200	200	700	1000	1000
Max. Demand	50	41	40	41	41	40
Demand Factors	1	1	1	1	1	1
	1.6	1.78	1.8	3.6	5.68	5.8
	2.2	2.56	2.6	6.2	10.36	10.6
	2.8	3.34	3.4	8.8	15.04	15.4
	3.4	4.12	4.2	11.4	19.72	20.2
	4	4.9	5	14	24.4	25

Table 2: Truck capacity and customer demand data by problem type

Instance	R2	MAPE	MAP		k1	k3
C1	0.999	3.1%	-0.9%	Coeff.	0.47	58.47
				Prob.	0.000	0.000
R1	0.999	3.9%	2.4%	Coeff.	0.58	54.67
				Prob.	0.000	0.000
RC1	1.000	2.2%	0.3%	Coeff.	0.48	72.07
				Prob.	0.000	0.000
C2	0.999	4.3%	0.2%	Coeff.	0.55	57.09
				Prob.	0.000	0.000
R2	0.998	6.4%	3.5%	Coeff.	0.74	49.27
				Prob.	0.000	0.000
RC2	0.999	3.9%	0.4%	Coeff.	0.63	66.84
				Prob.	0.000	0.000
AVERAGE	0.999	4.0%	1.0%	AVERAGE	0.58	59.74

Table 3: CVRP model – Depot at the center- Variable demand and NO time Windows

Instance	R2	MAPE	MAP		k1	K2	K3
C1	1.000	3.0%	-0.3%	Coeff.	0.45	0.66	58.84
				Prob.	0.000	0.051	0.000
R1	1.000	2.0%	-0.1%	Coeff.	0.67	-2.18	53.34
				Prob.	0.000	0.000	0.000
RC1	1.000	2.4%	-0.1%	Coeff.	0.50	-0.58	71.80
				Prob.	0.000	0.175	0.000
C2	0.999	4.3%	-0.5%	Coeff.	0.57	-0.68	56.83
				Prob.	0.000	0.097	0.000
R2	0.999	3.1%	-0.5%	Coeff.	0.85	-3.04	48.24
				Prob.	0.000	0.000	0.000
RC2	0.999	3.5%	-0.7%	Coeff.	0.66	-1.26	66.51
				Prob.	0.000	0.008	0.000
AVERAGE	0.999	3.1%	-0.4%	AVERAGE	0.62	-1.18	59.26

Table 4 Base Model – Depot at the center- Variable demand and Time Windows

Instance	R2	MAPE	MAP		kl	Kb	km
C1	1.000	2.6%	-0.1%	Coeff.	0.43	2.08	64.02
				Prob.	0.000	0.000	0.000
				St. Dev.	0.013	0.253	0.386
R1	1.000	1.5%	0.1%	Coeff.	0.65	-0.48	59.49
				Probl.	0.000	0.016	0.000
				St. Dev.	0.013	0.193	0.334
RC1	1.000	2.0%	0.1%	Coeff.	0.48	1.20	77.53
				Prob.	0.000	0.001	0.000
				St. Dev.	0.017	0.343	0.446
C2	0.999	3.4%	-0.3%	Coeff.	0.55	0.94	63.50
				Prob.	0.000	0.004	0.000
				St. Dev.	0.013	0.310	0.448
R2	1.000	2.2%	-0.1%	Coeff.	0.82	-1.13	56.27
				Prob.	0.000	0.000	0.000
				St. Dev.	0.011	0.215	0.304
RC2	1.000	2.6%	-0.3%	Coeff.	0.63	0.86	74.23
				Prob.	0.000	0.012	0.000
				St. Dev.	0.013	0.328	0.475
AVERAGE	1.000	2.4%	-0.1%	AVERAGE	0.60	0.58	65.84

Table 5. Adjusted Model – Depot at the center- Variable demand and NO Time Windows

Instance	R2	MAPE	MAP		k1	k3
C101	0.996	9.4%	5.5%	Coeff.	0.98	56.54
				Prob.	0.000	0.000
R101	0.996	8.4%	5.1%	Coeff.	0.45	67.97
				Prob.	0.000	0.000
RC101	0.998	5.8%	3.5%	Coeff.	0.67	69.15
				Prob.	0.000	0.000
C201	0.994	10.8%	5.7%	Coeff.	1.20	48.23
				Prob.	0.000	0.000
R201	0.987	17.8%	12.3%	Coeff.	1.69	38.89
				Prob.	0.000	0.000
RC201	0.992	13.7%	8.9%	Coeff.	1.56	52.22
				Prob.	0.000	0.000
AVERAGE	0.994	11.0%	6.8%	AVERAGE	1.09	55.50

Table 6 CVRP model – Depot at the center- Variable demand and Time Windows

Instance	R2	MAPE	MAP		k1	k2	k3
C101	0.998	3.4%	0.0%	Coeff.	1.25	-7.08	51.44
				Prob.	0.000	0.000	0.000
R101	0.999	3.8%	-0.4%	Coeff.	1.01	-7.25	55.48
				Prob.	0.000	0.000	0.000
RC101	0.999	2.8%	-0.2%	Coeff.	0.88	-5.37	65.62
				Prob.	0.000	0.000	0.000
C201	0.997	7.1%	0.2%	Coeff.	1.40	-6.53	45.57
				Prob.	0.000	0.000	0.000
R201	0.998	6.6%	-0.8%	Coeff.	2.16	-12.83	34.00
				Prob.	0.000	0.000	0.000
RC201	0.998	5.4%	-0.7%	Coeff.	1.94	-13.04	48.50
				Prob.	0.000	0.000	0.000
AVERAGE	0.998	4.8%	-0.3%	AVERAGE	1.44	-8.68	50.10

Table 7 Base Model – depot at the center- Variable demand and Time Windows

Instance	R2	MAPE	MAP		kl	kb	km
C101	0.999	3.0%	0.2%	Coeff.	1.20	-2.49	65.44
				Prob.	0.000	0.001	0.000
				St. Dev.	0.042	0.690	1.248
R101	0.999	2.8%	-0.2%	Coeff.	1.01	-4.72	63.24
				Prob.	0.000	0.000	0.000
				St. Dev.	0.057	0.402	1.299
RC101	1.000	2.1%	0.0%	Coeff.	0.84	-1.70	75.23
				Prob.	0.000	0.000	0.000
				St. Dev.	0.027	0.417	0.660
C201	0.998	5.7%	0.3%	Coeff.	1.35	-2.50	61.72
				Prob.	0.000	0.002	0.000
				St. Dev.	0.036	0.727	1.242
R201	0.999	4.9%	-0.5%	Coeff.	2.09	-9.89	54.14
				Prob.	0.000	0.000	0.000
				St. Dev.	0.030	0.561	0.847
RC201	0.999	3.8%	-0.3%	Coeff.	1.85	-6.63	71.01
				Prob.	0.000	0.000	0.000
				St. Dev.	0.025	0.602	0.726
AVERAGE	0.999	3.7%	-0.1%	AVERAGE	1.39	-4.66	65.13

Table 8 Adjusted Model – Depot at the center- Variable demand and Time Windows

Instance	C101	R101	CR101	C201	R201	CR201
Depot at the corner	136.5	104.9	130.1	135.1	104.9	130.1
Depot at the center	57.7	49.9	66.2	59.4	49.9	66.2

Table 9Average Distance between depot and customers

Instance	R2	MAPE	MAP		k1	k3
C101	1.000	3.7%	-2.2%	Coeff.	0.72	132.31
				Prob.	0.000	0.000
R101	0.999	1.9%	0.2%	Coeff.	0.88	106.84
				Prob.	0.000	0.000
RC101	1.000	1.9%	0.9%	Coeff.	1.04	121.20
				Prob.	0.000	0.000
C201	1.000	4.3%	-2.3%	Coeff.	0.75	129.56
				Prob.	0.000	0.000
R201	1.000	2.4%	-0.9%	Coeff.	0.90	103.60
				Prob.	0.000	0.000
RC201	1.000	4.4%	-2.0%	Coeff.	0.80	129.48
				Prob.	0.000	0.000
AVERAGE	1.000	3.1%	-1.0%	AVERAGE	0.85	120.50

Table 10 CVRP model – Depot at the Corner- Variable demand and NO time Windows

Instance	R2	MAPE	MAP		k1	K2	K3
C101	1.000	1.3%	-0.3%	Coeff.	0.46	-0.34	134.70
				Prob.	0.000	0.188	0.000
R101	0.999	2.3%	-0.4%	Coeff.	0.79	-3.94	106.19
				Prob.	0.000	0.000	0.000
RC101	1.000	1.9%	-0.1%	Coeff.	1.02	-5.77	119.21
				Prob.	0.000	0.000	0.000
C201	1.000	2.0%	-0.4%	Coeff.	0.53	-0.72	131.00
				Prob.	0.000	0.022	0.000
R201	1.000	2.5%	-0.7%	Coeff.	0.76	-2.66	103.77
				Prob.	0.000	0.000	0.000
RC201	1.000	2.6%	-0.5%	Coeff.	0.60	-1.15	130.31
				Prob.	0.000	0.021	0.000
AVERAGE	1.000	2.1%	-0.4%	AVERAGE	0.69	-2.43	120.86

Table 11 Base Model – Depot at the Corner- Variable demand and Time Windows

Instance	R2	MAPE	MAP		kl	kb	km
C101	1.000	1.2%	-0.1%	Coeff.	0.45	1.16	140.10
				Prob.	0.000	0.000	0.000
				St. Dev.	0.011	0.217	0.334
R101	0.999	1.8%	-0.3%	Coeff.	0.78	-1.86	112.88
				Probl.	0.000	0.003	0.000
				St. Dev.	0.045	0.594	1.050
RC101	1.000	1.4%	0.0%	Coeff.	0.98	-1.69	130.60
				Prob.	0.000	0.000	0.000
				St. Dev.	0.025	0.367	0.585
C201	1.000	2.1%	-0.1%	Coeff.	0.51	0.83	137.28
				Prob.	0.000	0.003	0.000
				St. Dev.	0.011	0.269	0.389
R201	1.000	1.8%	-0.3%	Coeff.	0.73	-0.97	110.89
				Prob.	0.000	0.000	0.000
				St. Dev.	0.012	0.227	0.324
RC201	1.000	2.2%	-0.1%	Coeff.	0.57	0.76	137.27
				Prob.	0.000	0.054	0.000
				St. Dev.	0.016	0.386	0.644
AVERAGE	1.000	1.7%	-0.1%	AVERAGE	0.67	-0.30	128.17

Table 12. Adjusted Model – Depot at the Corner- Variable demand and NO Time Windows

Instance	R2	MAPE	MAP		k1	k3
C101	0.999	3.8%	1.9%	Coeff.	1.25	133.46
				Prob.	0.000	0.000
R101	0.999	2.4%	1.0%	Coeff.	0.60	125.00
				Prob.	0.000	0.000
RC101	0.999	4.3%	2.3%	Coeff.	1.17	126.38
				Prob.	0.000	0.000
C201	0.998	5.9%	1.1%	Coeff.	1.49	116.68
				Prob.	0.000	0.000
R201	0.996	10.2%	6.2%	Coeff.	1.80	92.69
				Prob.	0.000	0.000
RC201	0.998	7.0%	3.9%	Coeff.	1.69	113.33
				Prob.	0.000	0.000
AVERAGE	0.998	5.6%	2.7%	AVERAGE	1.34	117.92

Table 13 CVRP model – Depot at the Corner- Variable demand and Time **Windows**

Instance	R2	MAPE	MAP		k1	K2	K3
C101	1.000	2.1%	-0.1%	Coeff.	1.34	-8.57	128.32
				Prob.	0.000	0.000	0.000
R101	0.999	3.2%	-0.3%	Coeff.	0.99	-8.07	112.16
				Prob.	0.000	0.000	0.000
RC101	1.000	2.8%	-0.3%	Coeff.	1.78	-13.65	109.45
				Prob.	0.000	0.000	0.000
C201	0.999	5.6%	-0.4%	Coeff.	1.46	-5.98	115.45
				Prob.	0.000	0.000	0.000
R201	0.999	5.3%	-0.9%	Coeff.	2.06	-12.38	88.21
				Prob.	0.000	0.000	0.000
RC201	0.999	4.2%	-0.7%	Coeff.	1.85	-12.18	110.27
				Prob.	0.000	0.000	0.000
AVERAGE	0.999	3.9%	-0.4%	AVERAGE	1.58	-10.14	110.64

Table 14 Base Model – depot at the Corner- Variable demand and Time Windows

Instance	R2	MAPE	MAP		kl	kb	km
C101	1.000	1.6%	0.0%	Coeff.	1.29	-3.56	143.07
				Prob.	0.000	0.000	0.000
				St. Dev.	0.041	0.571	1.154
R101	1.000	2.7%	-0.2%	Coeff.	1.20	-4.72	115.74
				Prob.	0.000	0.000	0.000
				St. Dev.	0.146	0.649	2.777
RC101	1.000	1.7%	-0.1%	Coeff.	1.78	-5.12	127.11
				Prob.	0.000	0.000	0.000
				St. Dev.	0.059	0.478	1.092
C201	0.999	4.5%	-0.2%	Coeff.	1.42	-1.81	132.33
				Prob.	0.000	0.047	0.000
				St. Dev.	0.039	0.888	1.330
R201	0.999	4.1%	-0.5%	Coeff.	2.00	-7.69	107.37
				Prob.	0.000	0.000	0.000
				St. Dev.	0.033	0.621	0.943
RC201	1.000	3.1%	-0.2%	Coeff.	1.77	-6.01	131.78
				Prob.	0.000	0.000	0.000
				St. Dev.	0.028	0.647	1.785
AVERAGE	0.999	2.9%	-0.2%	AVERAGE	1.58	-4.82	126.23

Table 15 Adjusted Model – Depot at the Corner- Variable demand and Time Windows

Instance	Cust. (n)	Routes (m)	$\Delta(n)$		$\Delta(m)$	
			No TW	TW	No TW	TW
C201	20	1	5.3	15.5	52.4	34.5
		5	6.5	18.3		
	60	1	3.2	8.2	57.1	46.0
		15	3.9	10.1		
R201	20	1	7.4	20.9	43.0	20.4
		5	8.8	24.3		
	60	1	4.0	10.5	48.6	34.7
		15	4.9	12.8		
RC201	20	1	6.3	22.9	61.5	33.7
		5	7.5	26.7		
	60	1	3.7	11.6	66.9	49.5
		15	4.5	14.1		

Table 16. Type 2 Problems, Central Depot, Distance Increase per Additional Customer or Route

Instance	Cust. (n)	Routes (m)	$\Delta(n)$		$\Delta(m)$	
			No TW	TW	No TW	TW
C201	20	1	5.0	15.9	126.9	103.8
		5	6.0	18.7		
	60	1	3.0	8.5	131.3	115.9
		15	3.6	10.5		
R201	20	1	6.6	20.0	99.1	75.2
		5	7.8	23.2		
	60	1	3.5	10.0	104.1	88.8
		15	4.3	12.2		
RC201	20	1	5.6	21.8	125.8	96.1
		5	6.8	25.3		
	60	1	3.3	11.1	130.7	111.2
		15	4.1	13.5		

Table 17. Type 2 Problems, Corner Depot, Distance Increase per Additional Customer or Route

Instance	R2	MAPE	MAP		kl	kb	km
C1	0.997	6.6%	0.3%	Coeff.	0.51	1.48	62.64
				Prob.	0.000	0.000	0.000
R1	0.996	6.1%	0.4%	Coeff.	0.80	-1.40	56.53
				Probl.	0.000	0.000	0.000
RC1	0.998	5.2%	-0.1%	Coeff.	0.63	0.31	75.06
				Prob.	0.000	0.111	0.000
C2	0.997	6.4%	0.2%	Coeff.	0.58	0.37	62.93
				Prob.	0.000	0.026	0.000
R2	0.996	6.3%	0.5%	Coeff.	0.84	-1.75	55.83
				Prob.	0.000	0.000	0.000
RC2	0.998	5.2%	0.0%	Coeff.	0.65	0.39	74.00
				Prob.	0.000	0.033	0.000
AVERAGE	0.997	6.0%	0.2%	AVERAGE	0.668	-0.099	64.499

Table 18. Adjusted Model – Depot at the center- Variable demand and NO Time Windows- Operational Estimate

Instance	R2	MAPE	MAP		kl	kc1	kc2
Distance	0.999	4.2%	-0.1%	Coeff.	0.78	0.25	49.52
				Prob.	0.001	0.709	0.000
Time	0.994	7.9%	-0.6%	Coeff.	0.40	-0.60	2.88
				Prob.	0.000	0.000	0.000

Table 19. Adjusted Model – Estimating distance and durations in Bankstown distribution area

LIST OF FIGURES

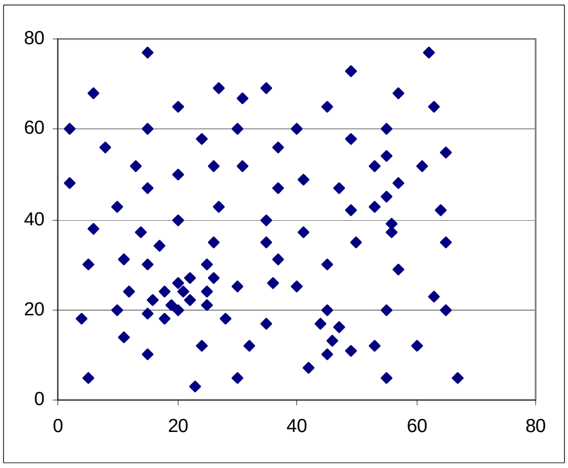

Figure 1 Random Problem (R)

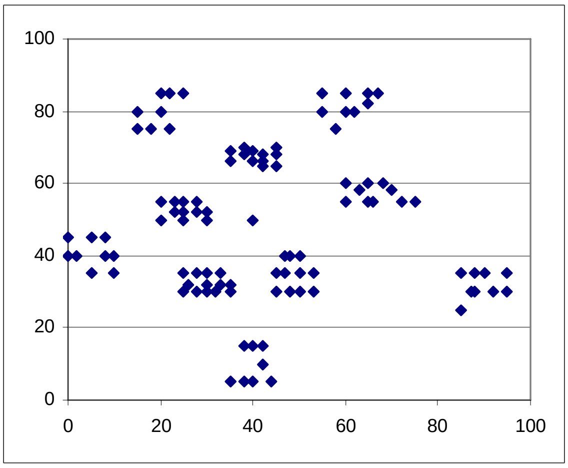

Figure 2 Clustered Problem (C)

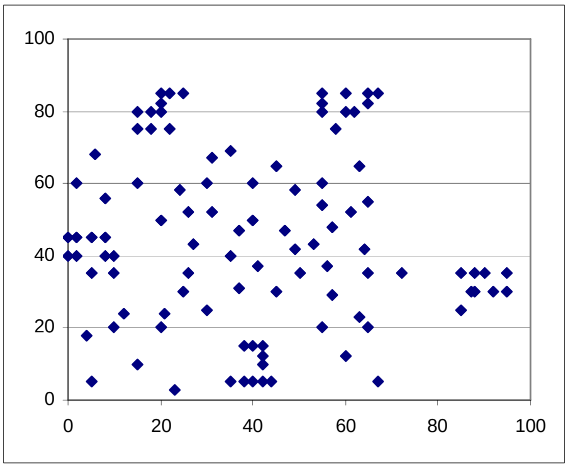

Figure 3 Random-Clustered Problem (RC)

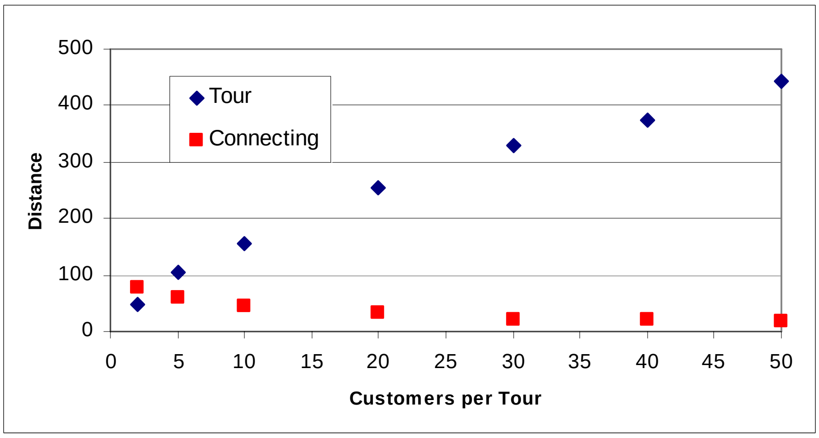

Figure 4. Tour and Connecting distances in TSP problems – R1 problem

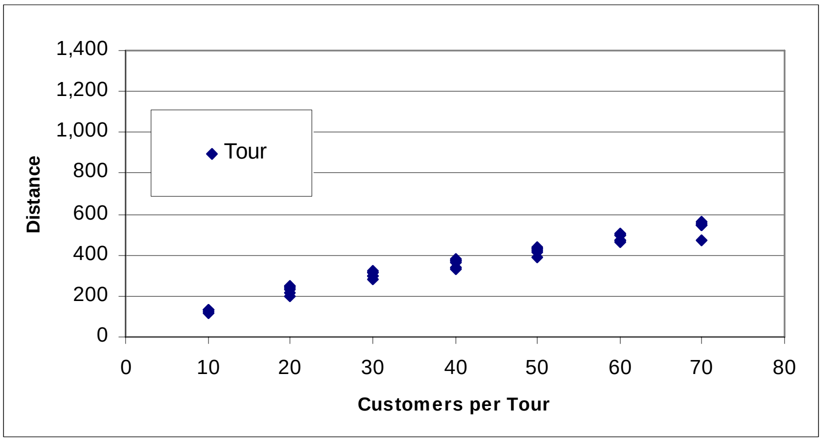

Figure 5 Tour distances and number of customers in VRP problems– R1 problem

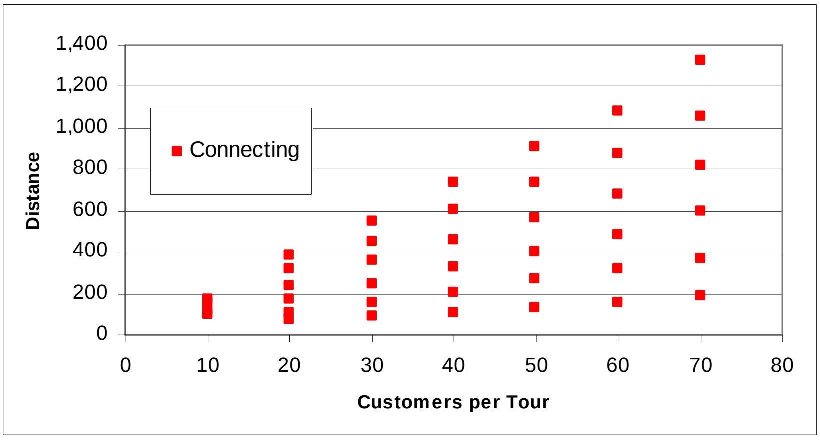

Figure 6 Connecting distances and number of customers in VRP problems– R1 problem

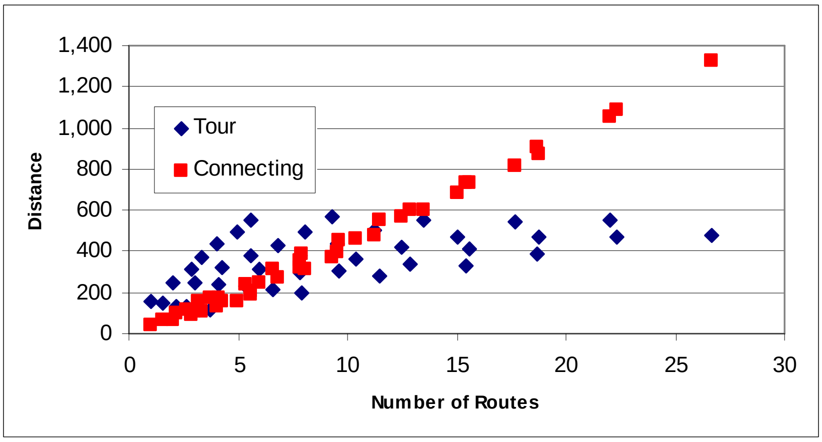

Figure 7. Tour and Connecting distances vs. routes in VRP problems– R1 problem

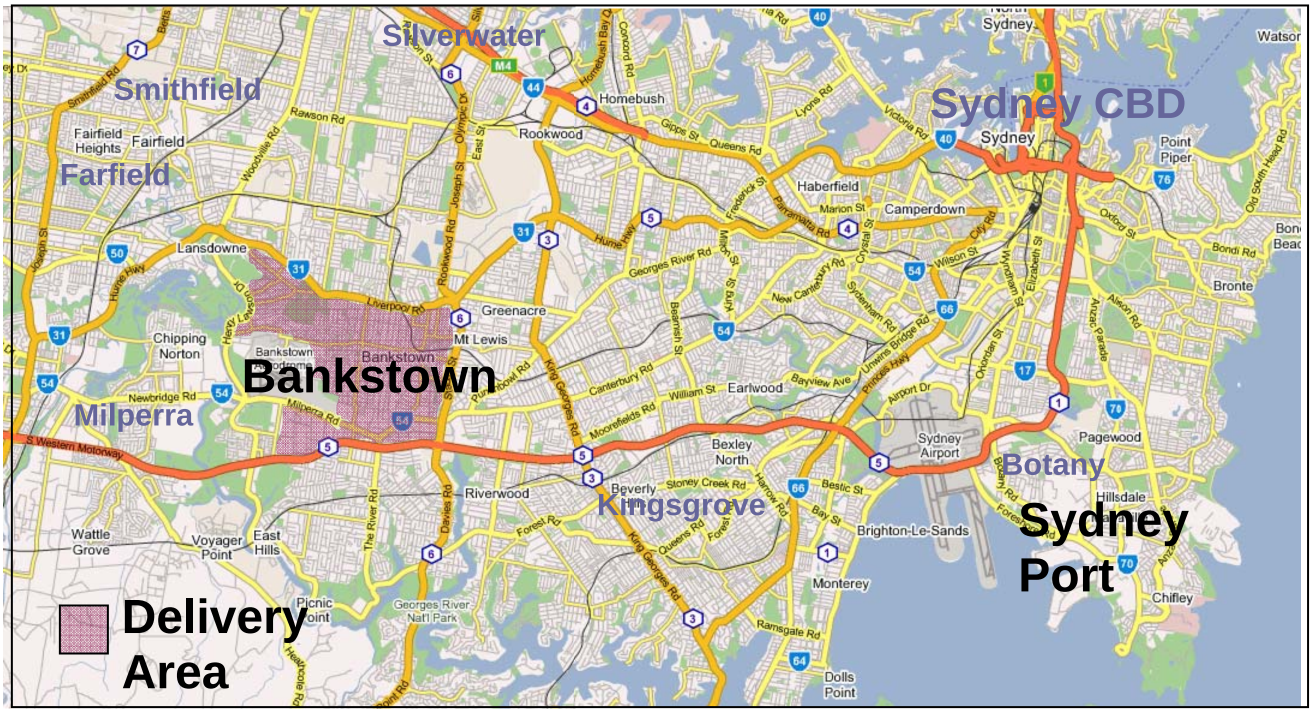

Figure 8. Relative Location of the Port of Sydney and Delivery Industrial Areas (Toll Freeways in thick orange, main arterial routes in thick yellow)[3]

[3] Map adapted from Google maps (http://maps.google.com/)

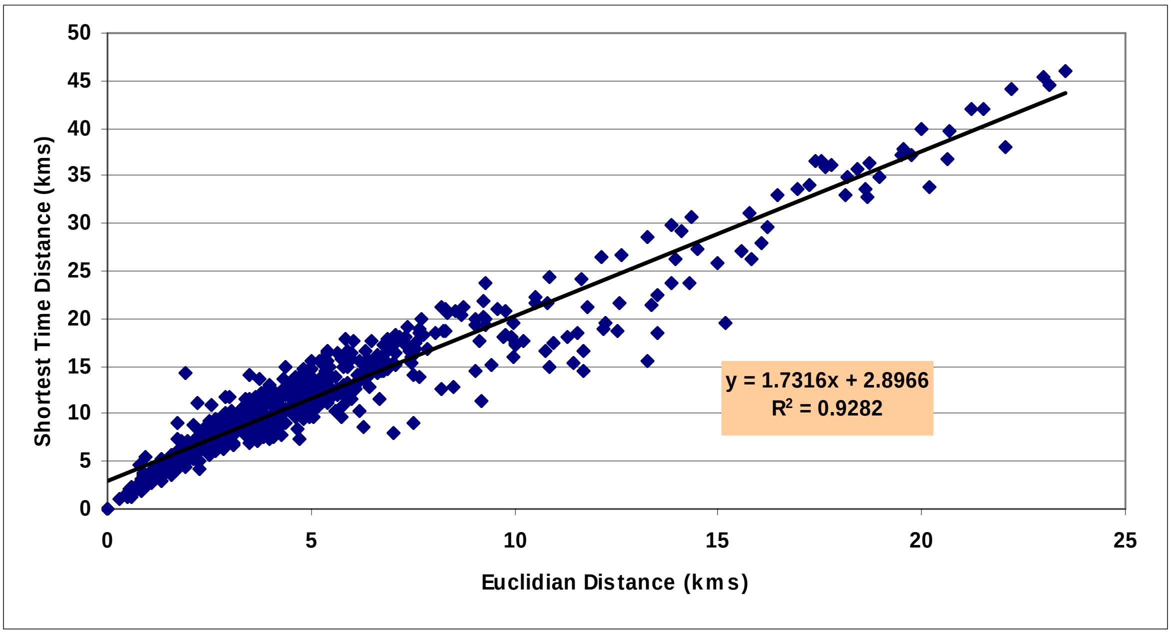

Figure 9. Euclidian Distance vs. Shortest Time distance among customers and depot-customers

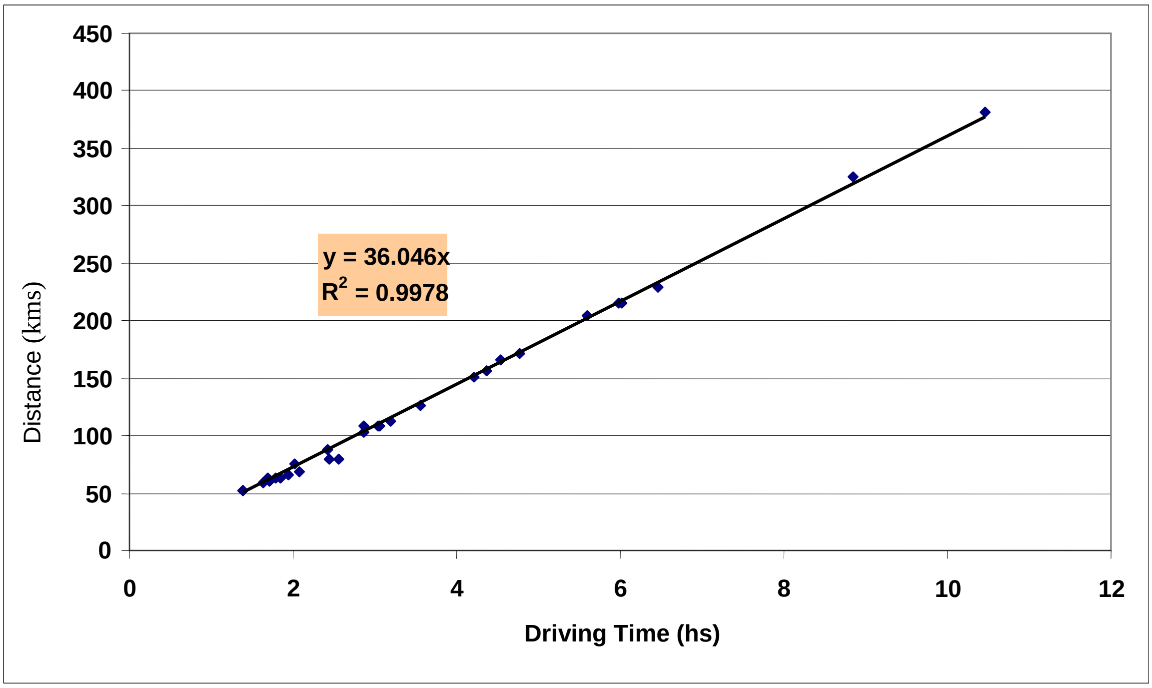

Figure 10 Distance Traveled and Time Driven