Planning Approximations to the Average Length of Vehicle Routing Problems with Varying Customer Demands and Routing Constraints

Miguel Andres Figliozzi

Portland State University

Maseeh College of Engineering and Computer Science

figliozzi@pdx.edu

Revised November 15, 2007

Number of words: 4591 + 8 Tables + 3 Figures = 7341

Planning Approximations to the Average Length of Vehicle Routing Problems with Varying Customer Demands and Routing Constraints

Miguel Andres Figliozzi

Portland State University

figliozzi@pdx.edu

ABSTRACT

This paper studies approximations to the average length of Vehicle Routing Problems (VRP). The approximations are valuable for strategic and planning analysis of transportation and logistics problems. The research focus is on VRP with varying number of customers, demands, and locations. This modeling environment can be used in transport and logistics models that deal with a distribution center serving an area with daily variations in the demand. The routes are calculated daily based on what freight is available. New approximations and experimental settings are introduced. Average distance travelled is estimated as a function of the number of customers served and the number of routes needed. Approximations are tested in instances with different customer spatial distributions, demand levels, number of customers, and time windows, Regression results indicate that the proposed approximations can reasonably predict the average length of VRP problems in randomly generated problems and real urban networks.

KEYWORDS: Vehicle Routing Problem, Distance Estimation, Simulated Experiments, Case Study

1. Introduction

In many logistics problems it is necessary to estimate the distance that a fleet of vehicles travel to meet a set of customer 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. s (VRP) changes with the number of customers served and the number of routes required. In particular, this research focuses on the ubiquitous case of a depot or distribution center (DC) serving up to N potential customers in the DC's delivery region. In many practical situations, not all potential customers request a visit on the same day. The number of customers served per day, n, may be significantly smaller than N. There may also be a significant variation in the number of customers visited per day of the week, e.g. early weekdays vs. weekends. The amount to be delivered or picked up may also vary on a daily basis, e.g. from one to several pallets, as might additional requirements such as time window constraints. The daily customer demand is known a night in advance, hence, each daily route and sequence of customers depends on what freight is available on a particular day for delivery or pick-up. Although, there is variability in the amount and characteristics of the day to day demand, the VRP problem analyzed in this paper is neither dynamic nor stochastic since all the information related to the customers' demands is known before the vehicles leave the depot or distribution center. The routes are designed daily and the number of routes/distance needed depends on the available freight.

Despite the growing implementation of customer-responsive and made-to-order supply chains, the impact of variations on the number of customer requests and demands on average VRP distance traveled has not yet been studied in the literature. All experimental studies have focused on the approximation of the length of specific traveling salesman problem (TSP) or VRP instances; i.e. given an a-priory known set of customer demands, how well a given formula approximates the real distance of *one specific instance*. This research has a different objective: given N potential customers and a variable customer demand (locations, demands, time-windows, etc.) in a service area, how well a given formula approximates the *average* distance of VRP solutions for different levels of n and routing constraints.

The paper is organized as follows: Section 2 provides a literature review. Section 3 presents and justifies the approximation formulas to be tested. Section 4 describes the experimental design. 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 V^n with n points distributed in an area A the length of the TSP tour through the set V^n asymptotically converges to:

$$TSP(V^n) = k\sqrt{nA} \tag{1}$$

The value of k is a constant. The asymptotic validity of this formula for TSP problems was experimentally tested by Ong and Huang (1989) using a nearest neighbor and exchange improvement heuristics. With an Euclidian metric and a uniform distribution of customers the constant term 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 compact and convex areas, the following approximation formula can be used:

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

Formula (2) requires a Euclidean travel metric or L_2 metric. Jaillet (1988) estimated the constant $k \approx 0.97$ for Manhattan travel metric or L_1 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 studied the correlation between route distance and customer-depot distances. Eilon et al. (1971) proposed several approximations to the length of the capacitated VRP or capacitated vehicle routing problem (CVRP) based on the shape and area of delivery, 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 rectangular 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.

$$CVRP(\boldsymbol{V}^{n}) \approx 2\bar{r}n/C + 0.57\sqrt{nA}) = 2\bar{r}m + 0.57\sqrt{nA}) \tag{3}$$

$CVRP(\boldsymbol{V}^{n})$ 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. Hence, the number of routes m is a priori known and can be calculated as n/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, vehicle capacity, and number of customers. Erera's dissertation (2000) proposes continuous approximations to estimate expected detour and distances in stochastic version of the capacitated vehicle routing problem.

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. Chien finds that the lowest *MAPE* for the best model is equal to 6.9%.

$$TSP(\boldsymbol{V}^{n}) \approx 2.1\bar{r} + 0.67\sqrt{nR}) \qquad\qquad R^{2} = 0.99 \qquad MAPE = 6.9 \qquad (4)$$

Chien used the area of the smallest rectangle that covers the customers; this area is denoted R. Expression (4) is not convenient for planning purposes when there may be many possible subsets of customers that are not known a priori. 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) compared equation (4) with two additional approximations 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:

$$TSP(\boldsymbol{V}^{n}) \approx [0.83 - 0.0011(n+1) + 1.11S/(n+1)]\sqrt{nA} \qquad R^{2} = 0.99 \qquad MAPE = 3.71 \qquad (5)$$

$$TSP(\boldsymbol{V}^{n}) \approx 0.41\bar{r} + [0.77 - 0.0008(n+1) + 0.90S/(n+1)]\sqrt{nA} \quad R^{2} = 0.99 \qquad MAPE = 3.61 \qquad (6)$$

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. R^{1} is defined as the area of the smallest rectangle that covers the customer and the depot. With the depot located at the center of the rectangle the results obtained are as follows:

$$TSP(\boldsymbol{V}^{n}) \approx [0.87 - 0.0016(n+1) + 1.34S/(n+1)]\sqrt{nR^{1}} \qquad R^{2} = 0.99 \qquad MAPE = 3.88 \qquad (7)$$

$$TSP(\boldsymbol{V}^{n}) \approx 1.15\bar{r} + [0.79 - 0.0012(n+1) + 0.97S/(n+1)]\sqrt{nR^{1}} \quad R^{2} = 0.99 \qquad MAPE = 3.70 \qquad (8)$$

It can be observed that *MAPE* slightly increases when the depot is located at the center of the rectangle. Kwon et al. (1995) also used neural networks to find a model that better predicts TSP length. They concluded that the capability of neural networks to find "hidden" relationships provides a slight edge against regression models. However, the models are less parsimonious and the terms harder to interpret in geometric terms.

3. Approximations Proposed and Tested

The previous section has 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. Increased accuracy can be obtained if additional terms related to the shape of the region and customers are added, as in Kwon et al. The third term within the brackets in the approximations (5) and (6) proposed by Kwon et al. (1995) has the form:

$$\sqrt{nA}/(n+1) \approx \sqrt{A/n}$$

Although Daganzo (1984) and Robuste et al. (2004) propose distance formulas for the CVRP, the number of necessary routes or vehicles, *m*, is known a priori. Daganzo (1984) and Robuste et al. (2004) assume demands that are factors of the vehicles' capacities, with this assumption the number of vehicles can be accurately determined by using $m = n/C$. In this research it is also assumed that the number of routes needed is known a-priori.

Given *n* customers and *m* routes, there is a relationship between the number of links that *connect* the depot and the first/last customer of each route and the number of *local* inter-customer links. Any solution to a TSP with *n* customers uses $n+1$ links, where $n-1$ links are local and 2 links are connecting. If capacity and/or window constraints are added, the resulting VRP has $m \geq 1$ routes. In general, for *m* routes and *n* customers any solution to a VRP uses

$n + m$ links. In general, for any given n and m the number of *connecting* links is $2m$ and the number of *local* links is $n - m$.

Six approximations or models to estimate the length of VRP instances are proposed:

Model 1: $VRP(\boldsymbol{V}) \approx k_l \sqrt{An} + 2\bar{r}m$ (9)

Model 2: $VRP(\boldsymbol{V}) \approx k_l \dfrac{n-m}{n}\sqrt{An} + 2\bar{r}m$ (10)

Model 3: $VRP(\boldsymbol{V}) \approx k_l \sqrt{An} + k_m m$ (11)

Model 4: $VRP(\boldsymbol{V}) \approx k_l \dfrac{n-m}{n}\sqrt{An} + k_m m$ (12)

Model 5: $VRP(\boldsymbol{V}) \approx k_l \sqrt{An} + k_b \sqrt{A/n} + k_m m$ (13)

Model 6: $VRP(\boldsymbol{V}) \approx k_l \dfrac{n-m}{n}\sqrt{An} + k_b \sqrt{A/n} + k_m m$ (14)

The parameters k_l, k_b, and k_m are estimated by linear regression. The term $(n-m)/n$ is proposed in this research to modify the local tour distance in models 2, 4, and 6. This term has some desirable properties: (a) when $n = m$ the estimated *local* distance is zero, whereas (b) when $n \gg m$ or $m = 1$ the local tour distance tends to the expression suggested by Beardwood et al. (1959). This research proposes the use of these ideas to reflect the trade-offs between n and m and improve the accuracy of the average VRP distance estimation as a function of n and m.

The term $k_b \sqrt{A/n}$ may be significant to estimate distances for the TSP when n is small and $m = 1$ as suggested by the results of Kwon et al. (1995). The term $k_m m$ estimates the connecting distance and captures increases in connecting distance as m increases or as the depot moves away from the customers.

The proposed models, expressions (9) to (14), are evaluated using numerical experiments in Section 5. The next section describes the experimental setting.

4. Experimental Setting

This research utilizes the classical instances of the VRP with time windows proposed by Solomon (1987) to test the approximations. The Solomon instances include distinct spatial customer distributions, vehicles' capacities, customer demands, and customer time windows. These problems have not only been widely studied in the operations research literature but the datasets are readily available[1].

In the Solomon problems there are 100 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 of 200 units, allowing fewer customers per route than the remaining sets. In contrast, problem sets R2, C2, and RC2 have vehicle capacity 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 time windows and route sizes are in the order of 30 customers per route. The first instance of each problem class is used in this research.

Random samples of the Solomon problems are used to examine the accuracy of approximations (9) to (14). Out of $N = 100$ possible customers in a service area A, a problem or instance is formed by a subset of n randomly selected customers. 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 1 to each subset of customers. Applying the factors in the second row of demand factors in Table 1, the customers have similar demands as in the original Solomon problems. The resulting

---

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

problems using the highest demand multipliers (last row of Table 2) are such that some customers are truckload (TL) or almost TL customers. Increasing some customer demands to or close to the TL level was done test the approximations when problems are highly constrained and have a large number of routes and a small number of customers per route. On the other hand, the situation of having a large number of customers per route is obtained when the demand factor is zero (first row Table 1). In all cases the routes' durations were limited by the depot time window. Lastly, in all Solomon problems, customers' time windows are different in width and start time. This adds an additional layer of variability. Hence, for each problem class or set, variability is introduced in three distinct ways: a) different subsets of customer locations, b) different levels of customer demands, and c) non-uniform time windows.

Most studies have focused on the derivation or testing of asymptotic estimators of the length TSP problems (Bearwood et al., 1959, Ong and Huang, 1989). Hence, experimental tests have mostly included a large number of customers per route. However, real-live routes have a relatively small number of customers per route due to capacity or tour length constraints. For example, in Denver over 50% of single and combination truck routes include less than 6 stops (Holguin-Veras and Patil, 2005); 95% of the truck routes include less than 20 stops. This research work tests the approximations using instances that range from 1 customer per route to over 35 customers per route.

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 were 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. To the best of the author's knowledge there is no published research that reports *MAPE* and simulation results for CVRP or VRP with time windows.

All problem instances in this research were solved with a VRP improvements heuristic that has obtained the best published solution in terms of number of vehicles (Figliozzi, 2007). The solution quality of this heuristic is clearly superior than the performance of savings or construction heuristics used in previous research efforts such as Ong and Huang (1989) or

Robuste et al (2004). The overall solution quality of the heuristic used, i.e. total number of vehicles needed to solve the 56 Solomon problems, is approximately 4% over the best known solutions (Braysy and Gendreau, 2005a, Braysy and Gendreau, 2005b).

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 $MPE \leq MAPE$. 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.

5. Analysis and Discussion of Experimental Results

Results for CRVP instances, i.e. no time windows, and the depot located at the center are shown in Table 2. Herein, all the regression results were obtained forcing the intercept or constant term to be zero; this 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. Model fit R^2, MAPE, and MPE are displayed for models 1 to 6. The average, maximum, and minimum correspond to the first Solomon problem in each of the six problem types (R1, C1, RC1, R2, C2 and RC2). For the sake of clarity, only 3 decimals are displayed.

In Table 2, all six approximations have very good R^2 values. However, models with more terms such as 5 and 6 have a superior *MAPE* performance. The approximations that adjust the tour distances using the term $(n-m)/n$ (models 2, 4, and 6) have a superior *MAPE* performance

than their counterparts with the same number of estimated coefficients (models 1, 3, and 5 respectively)

Table 3 reveals the impact of time windows on the accuracy of average distance estimation. These results are obtained using the same instances used previously to obtain Table 2 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 all six models. The increases in *MAPE* range from 50% to 175% for Model 3. As observed in Table 2, the approximations that adjust the tour distances using the term $(n-m)/n$ still have a superior *MAPE* performance than their counterparts. Further, the performance of approximations 2, 4, and 6 is relatively better with time windows when compared to approximations 1, 3, and 5. This can be explained by the higher number of routes needed when time windows are introduced, when m is higher the term $(n-m)/n$ plays a more significant role.

Time windows also affect the value of the estimated local tour parameters k_b. Table 4 shows the value of the parameter k_b for approximation 2 and customers *without* time window constraints. The value of k_b changes with the spatial distribution of the customers; the value of k_b is highest for randomly distributed instances and lowest for clustered instances. This is intuitively correct as the value of k_b is a proxy for the average distance between customers in a local tour between the first and last customer of a route. Table 5 shows the value of the parameter k_b for approximation 2 and customers *with* time window constraints. Comparing with the values in Table 4, all parameters k_l show an increase that is highly statistically significant. This is intuitively correct since time window constraints do not allow the formation of compact routes, hence, the average distance between customers in the local tours almost doubles on average.

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. Tables 6 and 7 show the results with and without time windows and the depot at a corner. Despite the change in the depot location, the same trends are

still observed: a) the models adjusted by $(n-m)/n$ do perform better in terms of *MAPE* than their counterparts, b) time windows decrease the predictive accuracy of the models, and c) with time windows the parameter k_l increases. With the corner depot, all three models perform better in terms of *MPE* and *MAPE* than with a centrally located depot. The same phenomenon can be observed in the experimental results of Kwon et al. (1995) for TSP distances.

Regarding the value of k_m, the value of this coefficient is closer to the corresponding value of $2\bar{r}$ when the depot is not centrally located. With a central depot, the values of the parameter k_m in approximations 3 to 6 is approximately within $\pm 20\%$ from the corresponding value of $2\bar{r}$; with a corner depot, the values of the parameter k_m in approximations 3 to 6 is approximately within $\pm 10\%$ from the corresponding value of $2\bar{r}$. Hence, the average distance from the depot to the customers, $2\bar{r}$, still appears in approximations 3 to 6 but under the form of the estimated coefficient k_m.

The comparison of the results in Tables 2, 3, 6, and 7 indicate that Model 6 is clearly superior in terms of *MPE* and *MAPE* across all experimental settings. However, the higher accuracy requires the estimation of a higher number of parameters (three). In addition, the interpretation of the term $k_b\sqrt{A/n}$ is not straightforward. This is compounded by the change of sign in the estimated k_b parameter, from positive in all instances with a centrally located depot to negative in some instances with a corner located depot. The sign change takes place in both models 5 and 6.

Model 2 is clearly superior if parsimony and interpretability in addition to accuracy are taken into account. Model 2 is simple and easily interpreted as well as robust when time windows are introduced. The coefficient k_b is in all cases easily interpreted, highly significant, and positive. Further, Model 2 outperforms Model 4 which uses two regression coefficients with time windows and cornered depot. This can be attributed to the influence of the term $(n-m)/n$ which plays a larger role when a higher number of routes is required (more routes are required when time windows are introduced and when the depot is moved away from the customers).

6. Real-life 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 models, they can also be used to estimate distance, costs, and times in real-live 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 random-clustered Solomon problems. The company's customers are in the hundreds but they are not visited every day. The freight forwarding company consolidates less than container (LTC) shipments and customers are visited only if a consignment has arrived before the distribution cutoff time. Further details about the tour characteristics can be found in Figliozzi et al. (2007).

Model 4 was tested with customers located in the industrial suburb of Bankstown with thirty customers distributed in an irregular area of 39.5 squared kilometers (see map in Figure 1). 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 on the connecting freeway. To test Model 4, 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 congestion, road characteristics, and speed limits. In this application the travel speeds used are: 65 km/hour in freeways, 35 km/hour in main connecting streets – four lanes or more with 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 urban highway network and geographic information system (GIS) software. Figure 2 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 of 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 per customer is used.

Three different routing scenarios were constructed: (a) no constraints or TSP case, (b) with a 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 regression was estimated using the consolidated data from all three scenarios. The results are shown in Table 8. The network distance traveled is well approximated with a *MAPE* of 4.2%. The prediction of travel time or driving time in hours has a *MAPE* of 11.7%. The good *MAPE* percentage is not surprising given the good correlation between distance traveled and time driven (see Figure 3). Model 4 was used to approximate times and distances due to the different travel speeds; the connecting distance between depot and customers does not always follow the same type of highway. These results are encouraging and show that the proposed approximations may have useful applications in urban networks. While these results are promising, from this example it is impossible to generalize the results. Further research efforts are necessary to study the accuracy of VRP distance approximation in cities with different layouts and highway networks.

7. CONCLUSIONS

This research studies approximations to the average length of vehicle routing problems when there is variability in the number, level and locations of customer demands. These approximations are intended for strategic and planning analysis of transportation and logistics problems, when the number and location of customers vary daily and are not known a priori.

A new parsimonious, intuitive, and effective approximation is proposed and successfully tested using instances with different patterns of customer spatial distribution, time windows, customer demands, and depot locations. It was found that time windows negatively affect the accuracy of the approximations. Time windows not only increase travel 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-life urban network with encouraging results.

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.

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.

ERERA, A. (2000) Design of Large-Scale Logistics Systems for Uncertain Environments. *Ph D dissertation, University of California-Berkeley*.

FIGLIOZZI, M. A. (2007) An Iterative Construction and Improvement Algorithm for the Vehicle Routing Problems with Hard and Soft Time Windows. *Working Paper, Portland State University*.

FIGLIOZZI, M. A., 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.

HOLGUIN-VERAS, J. & PATIL, G. (2005) Observed Trip Chain Behavior of Commercial Vehicles. *Transportation Research Record 1906***,** 74-80.

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.

ONG, H. L. & HUANG, H. C. (1989) Asymptotic Expected Performance Of Some Tsp Heuristics - An Empirical-Evaluation. *European Journal Of Operational Research,* 43**,** 231-238.

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	C1	R1	CR1	C2	R2	CR2
Vehicle Capacity	200	200	200	700	1000	1000
Max. Demand	50	41	40	41	41	40
Demand Factors	0	0	0	0	0	0
	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 1 – Truck capacity and customer demand data by problem type

MODEL	Statistic	R^2	MAP	MAPE
Model 1	Average	0.966	1.4%	**6.0%**
	Min	0.933	-0.8%	4.2%
	Max	0.986	3.5%	7.3%
Model 2	Average	0.991	1.5%	**4.7%**
	Min	0.986	-1.2%	3.1%
	Max	0.994	4.2%	6.5%
Model 3	Average	0.999	1.0%	**4.0%**
	Min	0.998	-0.9%	2.2%
	Max	1.000	3.5%	6.4%
Model 4	Average	0.999	-0.7%	**3.2%**
	Min	0.999	-2.6%	1.7%
	Max	1.000	1.6%	4.5%
Model 5	Average	0.999	-0.4%	**3.1%**
	Min	0.999	-0.7%	2.0%
	Max	1.000	-0.1%	4.3%
Model 6	Average	1.000	-0.1%	**2.4%**
	Min	0.999	-0.3%	1.5%
	Max	1.000	0.1%	3.4%

Table 2 – Model fit comparison with a central depot and no time windows

MODEL	Statistic	R^2	MAP	MAPE
Model 1	Average	0.968	6.2%	**12.0%**
	Min	0.954	3.2%	6.2%
	Max	0.982	10.2%	17.9%
Model 2	Average	0.984	4.9%	**7.9%**
	Min	0.977	2.5%	5.2%
	Max	0.990	9.2%	12.7%
Model 3	Average	0.994	6.8%	**11.0%**
	Min	0.987	3.5%	5.8%
	Max	0.998	12.3%	17.8%
Model 4	Average	0.997	4.3%	**6.9%**
	Min	0.994	1.6%	2.8%
	Max	0.999	8.8%	12.1%
Model 5	Average	0.998	-0.3%	**4.8%**
	Min	0.997	-0.8%	2.8%
	Max	0.999	0.2%	7.1%
Model 6	Average	0.999	-0.1%	**3.7%**
	Min	0.998	-0.5%	2.1%
	Max	1.000	0.3%	5.7%

Table 3 – Model fit comparison with a central depot and time windows

Instance	Coeff.	t-stat.	St. error
C101	0.62	41.58	0.01
R101	0.87	37.40	0.02
RC101	0.79	25.82	0.03
C201	0.64	57.71	0.01
R201	0.90	48.85	0.02
RC201	0.80	33.79	0.02
Average	0.77	40.86	0.02

Table 4 – Local tour regression coefficient without time windows (Model 2)

Instance	Coeff.	t-stat.	St. error
C101	1.30	56.24	0.02
R101	1.39	45.80	0.03
RC101	1.06	45.17	0.02
C201	1.32	69.79	0.02
R201	1.89	53.47	0.04
RC201	1.74	63.38	0.03
Average	1.45	55.64	0.03

Table 5 – Local tour regression coefficient with time windows (Model 2)

MODEL	Statistic	R^2	MAP	MAPE
Model 1	Average	0.985	0.8%	**3.3%**
	Min	0.970	-1.0%	1.7%
	Max	0.994	2.8%	4.5%
Model 2	Average	0.981	1.4%	**2.9%**
	Min	0.965	-0.2%	1.5%
	Max	0.995	3.0%	4.7%
Model 3	Average	1.000	-1.0%	**3.1%**
	Min	0.999	-2.3%	1.9%
	Max	1.000	0.9%	4.4%
Model 4	Average	1.000	0.1%	**2.1%**
	Min	0.999	-1.0%	1.7%
	Max	1.000	1.2%	2.6%
Model 5	Average	1.000	-0.4%	**2.1%**
	Min	0.999	-0.7%	1.3%
	Max	1.000	-0.1%	2.6%
Model 6	Average	1.000	-0.1%	**1.7%**
	Min	0.999	-0.3%	1.2%
	Max	1.000	0.0%	2.2%

Table 6 – Model fit comparison with a corner depot and no time windows

MODEL	Statistic	R^2	MAP	MAPE
	Average	0.955	3.4%	**8.7%**
Model 1	Min	0.941	0.0%	5.6%
	Max	0.977	6.2%	13.2%
	Average	0.985	3.2%	**5.4%**
Model 2	Min	0.976	0.9%	3.2%
	Max	0.989	6.6%	9.3%
	Average	0.998	2.7%	**5.6%**
Model 3	Min	0.996	1.0%	2.4%
	Max	0.999	6.2%	10.2%
	Average	0.999	3.0%	**5.0%**
Model 4	Min	0.997	1.0%	3.5%
	Max	0.999	6.1%	9.0%
	Average	0.999	-0.4%	**3.9%**
Model 5	Min	0.999	-0.9%	2.1%
	Max	1.000	-0.1%	5.6%
	Average	0.999	-0.2%	**2.9%**
Model 6	Min	0.999	-0.5%	1.6%
	Max	1.000	0.0%	4.5%

Table 7 – Model fit comparison with a corner depot and time windows

Instance	R2	MAPE	MAP	Coefficient	k_l	k_m
Distance (kms)	0.999	-0.5%	4.2%	Estimated	0.80	49.51
				t-stat	4.158	48.317
Time Driven (hrs)	0.988	5.9%	11.7%	Estimated	0.028	1.25
				t-stat	2.838	13.088

Table 8 – Real-life network distance and time estimation (Model 4)

List of Figures

Figure 1 –Relative Location of the Port of Sydney and Delivery Industrial Areas

 (freeways in red and main highways in yellow)[2]
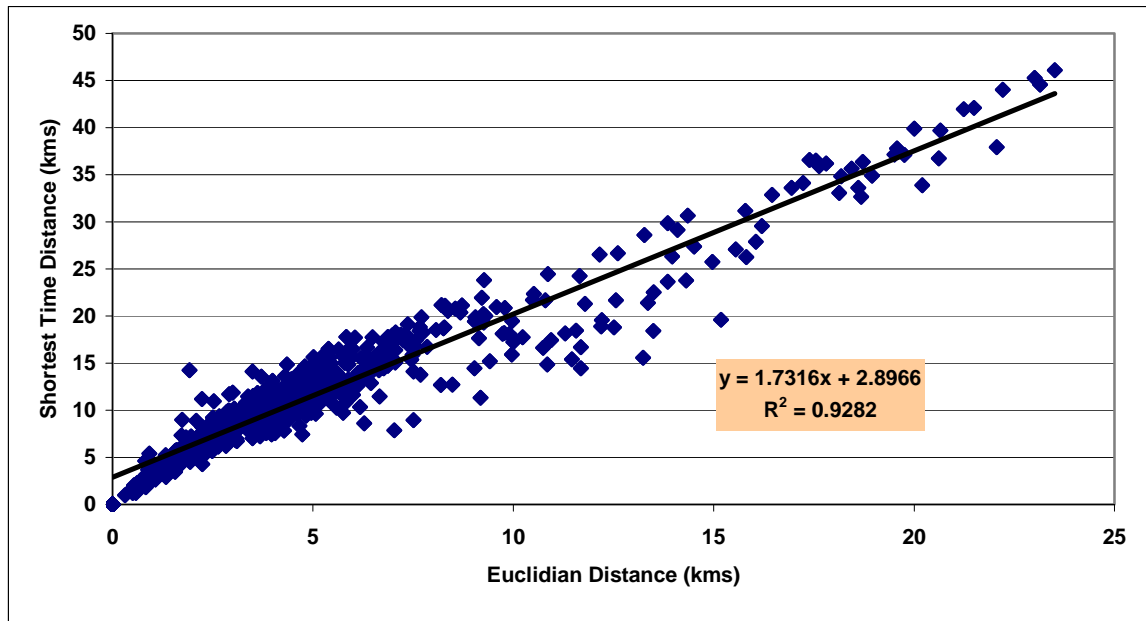
---

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

Figure 2 – Euclidian Distance vs. Shortest Time distance among customers and depot-customers
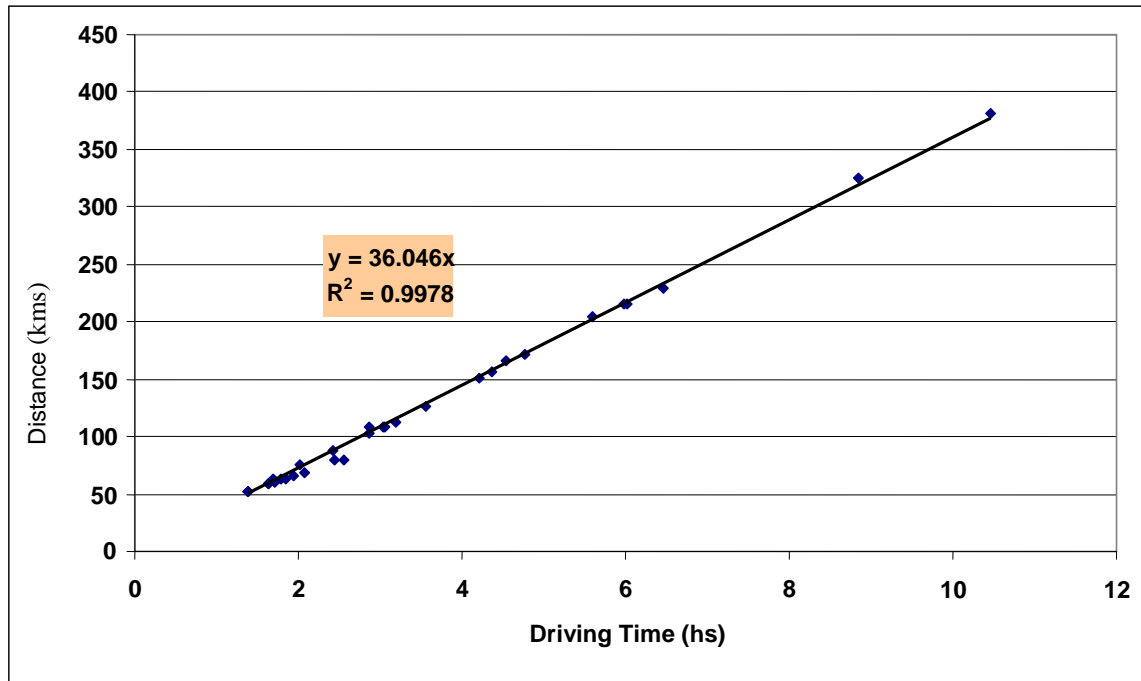
y = 36.046x
R² = 0.9978

Figure 3 – Distance Traveled and Time Driven