10-2009

Practical Approximations to Quantify the Impact of Time Windows and Delivery Sizes on Freight VMT in Urban Areas

Miguel A. Figliozzi

Portland State University, figliozzi@pdx.edu

Recommended Citation

Figliozzi, Miguel. Practical Approximations to Quantify the Impact of Time Windows and Delivery Sizes on Freight VMT in Urban Areas. OTREC-RR-09-07. Portland, OR: Transportation Research and Education Center (TREC), 2009. https://doi.org/10.15760/trec.13

OTREC

FINAL REPORT

Practical Approximations to Quantify the Impact of Time Windows and Delivery Sizes on VMT Multi-stop Tours

OTREC-RR-09-07
October 2009

PRACTICAL APPROXIMATIONS TO QUANTIFY THE IMPACT OF TIME WINDOWS AND DELIVERY SIZES ON VMT MULTI-STOP TOURS

Report Type

OTREC-RR-09-07

by

Miguel Figliozzi, Assistant Professor
Department of Civil & Environmental Engineering
Portland State University

for

P.O. Box 751
Portland, OR 97207

October 2009

Technical Report Documentation Page

1. Report No. OTREC-RR-09-07	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Practical Approximations to Quantify the Impact of Time Windows and Delivery Sizes on VMT Multi-stop Tours		5. Report Date April 2009
		6. Performing Organization Code
7. Author(s) Miguel Figliozzi		8. Performing Organization Report No.
9. Performing Organization Name and Address Portland State University Civil and Environmental Engineering P.O. Box 751 Portland, OR, 97207		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address Oregon Transportation Research and Education Consortium (OTREC) P.O. Box 751 Portland, Oregon 97207		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. 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.

17. Key Words Vehicle Routing Problem, Distance Estimation, Simulated Experiments, Case Study	18. Distribution Statement No restrictions. Copies available from OTREC: www.otrec.us	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 42	22. Price

ACKNOWLEDGEMENTS

DISCLAIMER

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

vi

EXECUTIVE SUMMARY

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 and the impacts of time-window constraints on the average vehicle routing problem (VRP) distance traveled have not yet been studied in the literature. Previous research has focused on the estimation of distances for either the traveling salesman problems (TSP) or the capacitated vehicle routing problems (CVRP). The purpose of this research was to determine how well a given formula approximates the *average* VRP distance for n actual customers given N potential customers and a variable customer demand (locations, demands, time-windows, etc.) in a service area.

The research provides a parsimonious approach with theoretical justification for the impacts of time windows to develop progressively more descriptive linear regression formulas for the average VRP distance. A total of 15 formulas are proposed and tested using the well-known 56 Solomon benchmark problems with varying numbers of customers, variable customer demands, and with and without time-window constraints. One of the regression models is then tested on actual customer distribution data from a freight forwarding company based in Sydney, Australia, and serving customers in the local suburb of Bankstown.

The statistical results obtained from the analysis of the Solomon problem instances and the Bankstown data show very high R^2 values and low mean absolute error (MPE and MAPE) for all models tested, with improvement in these values as models were made increasingly descriptive. Descriptive statistics for all regression predictors also were very statically significant beyond a 99% confidence level. The research results are widely applicable for strategic planning and analysis of transportation and logistics problems with varying customer demands and constraints, including cases where the number of routes is not known a priori.

1.0   INTRODUCTION

In many logistics problems it is necessary to estimate the distance that a fleet of vehicles travels 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. The transportation decisions associated with high-value, high time-sensitive products are the most demanding activities in terms of transport service requirements and usually require service within hard time windows (Figliozzi, 2006). Time windows also are a key constraint for JIT[1] production systems as well as emergency repair work and express (courier) delivery services. Time windows have a significant impact on decreasing the efficiency of routes, reducing service areas, and significantly increasing distance travelled (Figliozzi, 2007).

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. The existing body of literature has mostly focused on the estimation of distances for either the traveling salesman problems (TSP) or the capacitated vehicle routing problems (CVRP)[2]. All experimental studies have focused on the approximation of the length of specific TSP or VRP instances (i.e., given an a priori known set of customer demands, how well a given formula approximates the real distance of *one specific instance*).

This research has a different objective such that given *N* potential customers and a variable customer demand (locations, demands, time-windows, etc.) in a service area, this study explores how well a given formula approximates the *average* distance of VRP solutions for *n actual* customers and different levels of routing constraints. An intuitive and parsimonious mathematical framework to estimate average distances in VRP problems with hard time-window constraints is first provided. Subsequent to this, the ubiquitous case of a depot or distribution center (DC) serving up to *N* potential customers in the facility's delivery region is examined.

In many practical situations, not all potential customers request a visit on the same day. Thus, in practice, *n* may be significantly smaller than *N*. There also may be a significant variation in the number of customers visited per day of the week, such as on early weekdays vs. weekends. The amount to be delivered or picked up also may 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

---

[1] JIT stands for Just in Time, a make-to-order form of production that minimizes inventory on hand and relies on frequent and time-sensitive deliveries.
[2] Henceforth, the acronyms TSP, VRP and CVRP will be used in this paper

routes are designed daily and the number of routes and distance needed depends on the available freight.

The paper is organized as follows: Section 2.0 provides a literature review; Section 3.0 presents asymptotic results for the VRP and expressions to estimate the additional number of routes due to time-window constraints as well as justification for the approximation formulas to be tested; Section 4.0 describes the experimental design; the experimental results are analyzed and discussed in Section 5.0, along with an analysis of a real-world application; lastly, concluding comments are provided in Section 6.0.

2.0   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 (it seems redundant to include "problems" since TSP and VRP both end with that word). Comprehensive reviews of solution methods for TSP and VRP **problems** (see earlier comment) are found in Gutin and Punnen [1] and Toth and Vigo [2], 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 whole set asymptotically converges, with a probability of one, to a constant k multiplied by the square root of the number of points and the area, i.e. $k\sqrt{nA}$ when $n \to \infty$ such that

$$TSP(V^n) \approx k\sqrt{nA} .$$

(2.1)

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 in equation (2.1) has been estimated at $k = 0.765$ (Stein, 1978). For reasonably compact and convex areas, the limit provided by Beardwood et al. converges rapidly (Larson and Odoni, 1981). Jaillet (1988) estimated the constant $k = 0.97$ for a Manhattan metric.

Approximations to the length of CVRP were first published in the late 1960s and early 1970s (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 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:

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

(2.2)

$CVRP(V^n)$ is the total distance of the CVRP 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 known a priori and can be calculated as n/C. Daganzo's approximation 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 Daganzo's approximation were derived assuming $C > 6$ and $N > 4C^2$. Daganzo's

5

approximation works better in elongated areas as the routes were formed following the "strip" strategy. Robuste et al. (2004) tested Daganzo's approximation using simulations and elliptical areas; they propose adjustments based on area shape, vehicle capacity, and number of customers. A dissertation produced by Erera (2000) extended the usage of continuous approximations to estimate the distance of detours and routes in stochastic version of the CVRP.

Daganzo (1987a, 1987b) provides another analytical model of the VRP with time-window constraints. Daganzo divides a day into time periods or bins of equal length and then clusters customers in rectangles. Each customer is then placed in a balanced time period or bin, consistent with his or her time window; this allows a simplification of the problem as individual customer time-window characteristics are now associated with a time period. Using this time bin-cluster, first-route second approach, Daganzo analyzes main routing tradeoffs and determines that distance traveled is a function of the square root of the number of time periods and that lower distances are possible when routes are allowed to overlap. Different approximations are provided if the dominant constraint is either vehicle capacity or route duration. Although Daganzo's formulas are useful and intuitive they are not easily applied to estimate VRP distance since his approach does not guarantee feasibility. Unfortunately, no systematic method or general expression for clustering and determining the number of periods that guarantees balanced periods and feasible routes is provided.

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 eight different length/width ratios ranging from 1 to 8 and circular sectors with eight 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 five 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 linear regressions, found the best fitting parameters. The mean absolute percentage error (*MAPE*) was the benchmark to compare specifications. Chien finds that for the best model

$$TSP(V^n) \approx 2.1\bar{r} + 0.67\sqrt{nR} ,$$

(2.3)

the lowest *MAPE* is equal to 6.9% with an R^2 value of 0.99. In equation (2.3) $TSP(V^n)$ is the total distance of the TSP problem serving n customers. The area of the smallest rectangle that covers the customers is denoted R. The use of R instead of the total area A containing all customers may not be convenient for planning purposes when there may be many possible subsets of customers that are not known a priori. Chien also estimated the previous models for each of the 16 different regions; R^2 and *MAPE* statistics 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 they also used neural networks to identify better approximations. To test the accuracy of different models, they tested TSP **problems** in rectangular areas with eight 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) proposed approximations that make use of the geometric information proportioned by the ratio length/width of the rectangle (the shape factor S); length and width are 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:

$$TSP(V^n) \approx \left[0.83 - 0.0011(n+1) + 1.11S/(n+1)\right]\sqrt{nA} \qquad (2.4)$$

$$R^2 = 0.99$$

$$MAPE = 3.71$$

$$TSP(V^n) \approx 0.41\bar{r} + \left[0.77 - 0.0008(n+1) + 0.90S/(n+1)\right]\sqrt{nA} \qquad (2.5)$$

$$R^2 = 0.99$$

$$MAPE = 3.61$$

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, respectively. R^d 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:

$$TSP(V^n) \approx \left[0.87 - 0.0016(n+1) + 1.34S/(n+1)\right]\sqrt{nR^d} \qquad (2.6)$$

$$R^2 = 0.99$$

$$MAPE = 3.88$$

$$TSP(V^n) \approx 1.15\bar{r} + \left[0.79 - 0.0012(n+1) + 0.97S/(n+1)\right]\sqrt{nR^d} \qquad (2.7)$$

$$R^2 = 0.99$$

$$MAPE = 3.70$$

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 advantage against regression models. However, the models are less parsimonious and the terms harder to interpret in geometric terms.

The literature review indicates that previous research efforts have concentrated on the estimation of distances for TSP and CVRP *instances* that lack time-window constraints. To the best of this author's knowledge, there is no published research that estimates the impact of time windows and varying levels of customer demand on *average* VRP distances.

3.0   FORMULATION OF VRP APPROXIMATIONS

3.1   PROPOSED FORMULAS FOR APPROXIMATING VRP DISTANCES WHEN THE NUMBER OF ROUTES IS KNOWN

3.1.1 Approximations of VRP distances without time windows

The total distance travelled can be broken down into: (1) a distance between the depot and customers, herein denoted a "connecting distance," and (2) a distance between different customers, herein denoted as "local or tour distance." As the number of routes increases, the local distance tends to decrease whereas the connecting distance increases. The intuition behind these distance changes is evident after analyzing the variations in the corresponding number of links. 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 $n-m$.

This research tests several approximations or models to estimate the average VRP distances. The first two approximations use only information about the number of customers served and the number of routes needed. Model 1 is based on Daganzo's theoretical expression. The second approximation (Model 2) incorporates the tradeoff between connecting and local links. The approximation models tested are:

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

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

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

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

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

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

The parameters k_l, k_m, and k_b are estimated by linear regression. The following is a summary of the practical significance of these parameters: In Model 1 and Model 2, the number of routes, m, is given and only the parameter k_l has to be estimated (i.e., only the local distance factor is estimated). In Model 3 and Model 4, the parameter k_m is included, which estimates the connecting distance and captures increases in connecting distances as m increases or as the depot moves away from customers. In Model 5 and Model 6, the parameter k_b is included, but provides a less intuitive meaning. However, this factor provides information about the effects of a centrally located depot versus a depot with a corner location, which is explained in more detail in section 5.1.

3.1.2 Naïve approximations of VRP distances with time-window constraints

The addition of time-window constraints is introduced into the VRP distance approximations in this section. Model 1 and Model 2 from section 3.1.1 and an additional six models provided below are tested. Let m_d denote the number of routes needed to meet customer demand D using vehicles of capacity C, then:

$$m_d = D/C.$$

Let m_t denote the number of additional routes needed to meet the requirements of customers with time windows and let n_t be the number of customers with time-window constraints and let the fraction of customers with time windows be $p_t = n_t/n$. The additional proposed models are as follows:

Model 7: $VRP(V^n) \approx k_l\sqrt{An} + k_d m_d\, 2\bar{r} + k_t\, p_t^2\, 2\bar{r}$

Model 8: $VRP(V^n) \approx k_l\sqrt{An} + k_d m_d\, 2\bar{r} + k_t\, p_t\, 2\bar{r}$

Model 9: $VRP(V^n) \approx k_l\sqrt{An} + k_d m_d\, 2\bar{r} + k_t\, p_t\, \sqrt{n}\, 2\bar{r}$

Model 10: $VRP(V^n) \approx k_l\sqrt{An} + k_d m_d\, 2\bar{r} + k_t\, p_t n\, 2\bar{r}$

Model 11: $VRP(V^n) \approx k_l\sqrt{An} + k_d m_d\, 2\bar{r} + k_t\, p_t\, n\sqrt{n}\, 2\bar{r}$

Model 12: $VRP(V^n) \approx k_l\sqrt{An} + k_d m_d\, 2\bar{r} + k_t\, p_t\, n^2\, 2\bar{r}$

The parameter k_l is defined as in section 3.1.1. The additional regression parameters k_d and k_t represent the effects of customer-demand levels and time-window constraints, respectively. To the best of the author's knowledge, there are no practical or theoretical guidelines regarding the functional form that connects m_t with the number of time-window constraints, total number of customers, and the influence of demand levels. Model 7 through 12 assume that the contribution of time-window constraints and demand levels to the number of routes needed is linear and additive such that

10

$$m = k_d m_d + k_t m_t \,. \tag{3.1}$$

In particular, Model 7 assumes that m_t increases with the square of the percentage of customers with time windows, i.e. $m_t \approx p_t^2 = (n_t/n)^2$ whereas Model 8 has a linear relationship between m_t and the percentage of customers with time windows, i.e. $m_t \approx p_t = n_t/n$. Model 9 through 12 respectively assume that:

$$m_t \approx p_t \sqrt{n} = n_t / \sqrt{n} \,, \tag{3.2}$$

$$m_t \approx p_t n = n_t \,, \tag{3.3}$$

$$m_t \approx p_t n \sqrt{n} = n_t \sqrt{n} \,, \tag{3.4}$$

$$m_t \approx p_t n^2 = n_t n \,. \tag{3.5}$$

In all models the contribution of the demand levels and time-windows constraints is null when customer demand is zero ($m_d = 0$) or when the number of time-windows constraints is zero ($n_t = 0$). The proposed models are evaluated using numerical experiments in Section 5.0. In the next section, a theoretical approach is developed for the time window-constrained VRP in an effort to provide a more intuitive meaning to the regression parameters describing the effects on average route distance.

3.2 CHARACTERIZING THE IMPACTS OF TIME-WINDOW CONSTRAINTS

This section introduces a probabilistic approach to capture the impact of time windows on distance traveled for VRP instances that serve $N = \{1, 2, ..., n\}$ customers. Associated with each customer $i \in N$ there is a quintuplet $(x_i, q_i, s_i, e_i, l_i)$ that represents, respectively, the coordinates, demand, service time, earliest service starting time, and latest service ending time. The depot quintuplet is denoted $(x_0, q_0, s_0, e_0, l_0)$ with $q_0 = 0, s_0 = 0$ and $e_0 = 0$. Travel speed is constant and equal to one unit of distance per unit time. The distance between each customer $i \in N$ and the depot is denoted $d(x_i)$; feasibility conditions include $d(x_i) \le l_i$, $2d(x_i) + s_i \le l_0$, and $q_i \le Q$. Customers with time windows are drawn from a probability measure ν with bounded support. Without loss of generality, attributes of the quintuplet are scaled and shifted so they belong to the real interval $[0,1]$. The coordinates x_i are independently and identically distributed according to a distribution with compact support in $\Re^2$, $[0,1] \times [0,1]$; the customer parameters (q_i, s_i, e_i, l_i) are drawn from a joint probability distribution Φ with a continuous density function ϕ. The support

of ϕ is the feasible subset of $(x_1, x_2, x_3, x_4) \in [0,1]^4$. It is also assumed that customer locations and their parameters are independent of each other.

Customers without time windows are drawn from the same probability measure but their time windows are relaxed, i.e. (e_i, l_i) is replaced by (e_0, l_0). The "relaxed" probability measure is denoted μ, whose support is the feasible subset of $(x_1, x_2) \in [0,1]^2$ with $x_3 = 0$ and $x_4 = 1$. The expected number of routes needed to serve n customers with and without time windows is denoted $m_v(n)$ and $m_\mu(n)$, respectively.

Known results for the CVRP (Bramel et al., 1992) indicate that:

$$\lim_{n \to \infty} \frac{CVRP^*(n, \mu)}{n} = 2\gamma_\mu E(d) \tag{3.6}$$

where $\gamma_\mu > 0$ is a constant that depends only on μ, $E(d)$ is the expected distance between the depot and customers, and $CVRP^*(n)$ is the best VRP solution for travel distance. The ratio $1/\gamma_\mu$ is the average number of customers per route. Similar results can be derived for the VRP with time windows (Bramel and Simchi-Levi, 1996, Federgruen and Van Ryzin, 1997):

$$\lim_{n \to \infty} \frac{VRP^*(n, v)}{n} = 2\gamma_v E(d). \tag{3.7}$$

The first lemma provides a useful bound for the additional number of routes due to time-window constraints.

Lemma 1. The contribution of time windows to the distance traveled is bounded. Asymptotically, the number of additional routes due to time-window constraints can be expressed as kn, being k a constant such that $0 \le k \le 1$.

Proof. Asymptotically, the contribution of time windows to the distance travelled per customer can be expressed as $2n E(d)(\gamma_v - \gamma_\mu)$. The increase in the number of routes due to time windows, denoted $m_{v\mu}$, can be approximated by $m_{v\mu}(n) = n(\gamma_v - \gamma_\mu) = m_v(n) - m_\mu(n)$. There cannot be more routes than customers, hence $1 \ge \gamma_v$. Time windows, additional constraints, cannot reduce the VRP distance; hence $(\gamma_v - \gamma_\mu) \ge 0$.

The increase in the number of routes when time constraints are added is modeled probabilistically. Given any two customers $i, j \in N$ there is a probability P_{ij} that a vehicle can successfully visit customer j after visiting customer i without violating $j's$ time window. In general, P_{ij} is a random variable that will depend on the probability measure v. The goal is to find an expression that provides the average number of additional vehicles needed due to time-

window constraints, i.e. $m_{v\mu}(n)$. An exact solution using P_{ij} is likely to be intractable and, to the best of the author's knowledge, there is no general analytical expression that can be used to estimate the impact of time-window constraints on VRP distances.

To model $m_{v\mu}(n)$, the concept of an average probability of successfully sequencing any given customer with time-window constraints is introduced; this average success probability is denoted P_v; $b = 1/\gamma_\mu$ is denoted as the average number of customers per route or "bin" without time-window constraints. The probability associated to finding a feasible route with $c \le b$ customers, each with time-window constraints, can be expressed as:

$$P(c) = (P_v)^{c-1} \tag{3.8}$$

By definition $P(1) = 1$ because it is assumed that all customers can feasibly be served from the depot. When $b = 1$ the number of routes is simply $m = n$. When $b = 2$, the number of expected routes needed to serve n customers can be expressed as the weighted sum of routes with one and two customers:

$$\frac{n}{2} P(2) + n[1 - P(2)], \tag{3.9}$$

and generalizing for any b:

$$E[m_v(n)] = \sum_{c=1}^{c=b} \frac{nP(c)}{c} \prod_{j=c+1}^{b} [1 - P(j)]. \tag{3.10}$$

A similar expression can be found in the work of Diana et al. (2006), which estimated demand-responsive transit fleet sizes. The expected number of additional routes due to time-window constraints, $E[m_{v\mu}]$, can be expressed as:

$$E[m_{v\mu}(n)] = \left[\sum_{c=1}^{c=b} \frac{nP(c)}{c} \prod_{j=c+1}^{b} [1 - P(j)] \right] - \frac{n}{b} \tag{3.11}$$

Lemma 2. The expected number of additional routes due to time-window constraints, $E[m_{v\mu}(n)]$, is a continuously decreasing function of P_v.

Proof. The complete proof is presented in the appendix in Section 8.0; a sketch of the proof is presented in this section. The sum of weight factors $w(c)$

$$w(c) = P(c) \prod_{j=c+1}^{b} [1 - P(j)] \tag{3.12}$$

sum to one such that

$$\sum_{c=1}^{c=b} w(c) = \sum_{c=1}^{c=b} P(c) \prod_{j=c+1}^{b} [1 - P(j)] = 1. \tag{3.13}$$

As P_v increases from zero to one, the weight factors are shifted from $c=1$ to $c=b$, hence, the sum

$$\sum_{c=1}^{c=b} \frac{n\, w(c)}{c} \tag{3.14}$$

decreases as P_v increases.

Lemma 3. The expected number of additional routes due to time-window constraints $E[m_{v\mu}(n)]$ is bounded between $(0, n - n/b)$. The value of $E[m_{v\mu}(n)]$ is a fraction of the number of customers.

Proof. By substitution, it can be shown that $E[m_{v\mu}(n)] = n - n/b$ when $P_v = 0$ and $E[m_{v\mu}(n)] = 0$ when $P_v = 1$. Since $E[m_{v\mu}(n)]$ is a decreasing function it is bounded between $(0, n - n/b)$.

Theorem 1. A routing problem with customers drawn from a probability measure v has a unique P_v such that $E[m_{v\mu}(n)] = n(\gamma_v - \gamma_\mu)$ as $n \to \infty$.

Proof. Asymptotically, the additional number of routes is $m_{v\mu}(n) = n(\gamma_v - \gamma_\mu)$ with $0 \le (\gamma_v - \gamma_\mu) \le 1$. Due to Lemmas 2 and 3, $E[m_{v\mu}(n)]$ is a continuously decreasing function. Hence, there is a unique P_v such that $E[m_{v\mu}(n)] = n(\gamma_v - \gamma_\mu)$.

Corollary. The value of $1 - P_v$, the average probability of "failing" to sequence a customer with time-window constraints provides a measure, in a scale $(0,1)$, of the impact of time-window constrains on VRP distance. As $1 - P_v$ increases the relative impact of time-windows constraints on the number of routes and the distance traveled increases.

3.3 APPROXIMATING VRP DISTANCES WHEN THE NUMBER OF ROUTES IS UNKNOWN AND WITH THE ADDITION OF TIME WINDOWS

This section provides an approximation to VRP distance assuming a distribution center that serves a set of $N = \{1, 2, ..., n\}$ customers on any given day or time period. The number of daily requests may vary, but it never exceeds a maximum number $\bar{n}$, i.e. $n \leq \bar{n}$. The total number of customers with time windows is denoted n_t, $n_t \leq n$, and the total demand is denoted $q_N = \sum_{i \in N} q_i$.

The focus of this research is the derivation of general approximations to the average distance traveled to serve a total of n customers with n_t time windows (i.e., constraints $1 \leq n \leq \bar{n}$ and $0 \leq n_t \leq n$). This average distance is denoted $VRP(n, n_t, v)$. Instances of daily demands are formed by joining n_t customers, drawn according to a probability measure v, and $n - n_t$ customers drawn according to probability measure μ. A customer has a time window if either $e_i > e_0$ or $l_i < l_0$.

The value of P_v is approximated as the value that minimizes the absolute value of the difference:

$$\min \left| m_v(\bar{n}) - \sum_{c=1}^{b} \frac{\bar{n} P(c)}{c} \prod_{j=c+1}^{b} (1 - P(j)) \right| \tag{3.15}$$

s.t.: $P(c) = (P_v)^{c-1}$, $0 \leq P_v \leq 1$, and $b \approx \tilde{b} = \bar{n}/m_\mu(\bar{n})$. From Theorem 1, it is guaranteed that there is only one P_v that minimizes the absolute value of equation (3.15). The value of $m_\mu(\bar{n})$ and $m_v(\bar{n})$ can be estimated by sampling from the respective distributions and determining the number of routes needed.

To estimate the number of additional routes due to time windows when $0 < n_t < n$, it is necessary to model how time windows are distributed among routes. Assuming a binomial distribution, the probability of having a route with k time windows out of c customers is:

$$binomial(k; c, P_t) = \binom{c}{k} (P_t)^k (1 - P_t)^{c-k} \tag{3.16}$$

where:

$$\binom{c}{k} = \frac{c!}{k!(c-k)!},$$

$k =$ number of successes in b trials,

c = number of independent trials, and

$P_t = n_t/n$ = the probability of success on each trial.

Then, the number of additional routes to serve a total of n customers with n_t time windows can be approximated as follows:

$$m_{v\mu}(n, n_t) = \left[\sum_{c=1}^{b} \frac{nP(c, P_t, v)}{c} \prod_{j=c+1}^{b} (1 - P(j)) \right] - m_{\mu}(n)m \tag{3.17}$$

where $P(c, P_t, v) = \sum_{k=0}^{c} binomial(k; c, P_t)(P_v)^{k-1}$.

Approximating the number of routes related to "bin-packing" constraints, such as vehicle capacity or tour duration, is relatively straightforward:

$$m_{\mu}(n) \approx \max\left(\left\lceil \frac{q_N}{Q} \right\rceil, \left\lceil \frac{n\tau}{l_0 - e_0} \right\rceil \right) \tag{3.18}$$

where τ is the sum of estimated travel time plus service time per customer.

Although asymptotic results indicate that the number of routes is the only essential factor to estimate VRP distances, the literature review has shown that the best approximations to VRP distance account for (a) a term related to the distance traveled between the depot and customers and (b) a term related to the distance traveled between customers. The proposed approximation also accounts for both types of distances, but adds(?) terms to estimate the additional impact of time windows.

Model 13: $VRP(n, n_t, v) \approx \tilde{k}_{\mu}\sqrt{nA} + 2\tilde{k}_{\lambda\mu}\sqrt{n_t A} + \tilde{k}_{\mu}^m \tilde{\gamma}_{\mu} rm_{\mu}(n) + 2\tilde{k}_{v\mu}^m rm_{v\mu}(n, n_t)$

The vector of coefficients $(\tilde{k}_{\mu}, \tilde{k}_{\lambda\mu}, \tilde{k}_{\mu}^m, \tilde{k}_{v\mu}^m)$ is estimated by linear regression. The coefficients $\tilde{k}_{\mu}^m$ and $\tilde{k}_{v\mu}^m$ are related to the distance generated by the number of routes needed; the coefficients $\tilde{k}_{\mu}$ and $\tilde{k}_{\lambda\mu}$ are related to the intercustomer distance, as in Beardwood et al. (1959). If $\tilde{k}_v \sqrt{n_t A}$ and $\tilde{k}_{\mu}\sqrt{nA}$ approximate the inter-customer distance with and without time windows respectively, then, $\tilde{k}_{v\mu}$ represents the *change* in inter-customer distance when time-window constraints are added such that

$$\tilde{k}_v \sqrt{n_t A} = \tilde{k}_{\mu}\sqrt{nA} + \tilde{k}_{v\mu}\sqrt{n_t A}, \tag{3.19}$$

where $\tilde{k}_{v\mu} = \tilde{k}_v - \tilde{k}_{\mu}$ and $n_t = n$. The other two remaining coefficients relate to the number of routes as follows:

$$\tilde{\gamma}_{\mu} m_{\mu}(n) \to n\gamma_{\mu} \quad \text{as} \quad n \to \infty,$$

(3.20)

$$\tilde{\gamma}_{\nu\mu} m_{\nu\mu}(n, n_t) \to n(\gamma_{\nu} - \gamma_{\mu})$$

(3.21)

as $n \to \infty$.

4.0   EXPERIMENTAL SETTING

One of the main goals of this research is to estimate the impact of customer-demand levels and time-window constraints on VRP distance. 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[3].

The well-known 56 Solomon benchmark problems for the VRP with hard time windows are based on six groups of problem instances with 100 customers. The six problem classes are named C1, C2, R1, R2, RC1, and RC2. Customer locations were randomly generated (problem sets R1 and R2), clustered (problem sets C1 and C2), or mixed with randomly generated and clustered customer locations (problem sets RC1 and RC2). Problem sets R1, C1, and RC1 have a shorter scheduling horizon, tighter time windows, and fewer customers per route than problem sets R2, C2, and RC2, respectively.

Some studies have focused on the derivation or testing of asymptotic estimators of the TSP lengths (Bearwood et al., 1959, Ong and Huang, 1989). Hence, experimental tests have mostly included a large number of customers per route. However, real-world routes have a relatively small number of customers per route due to capacity, time windows, or tour-length constraints (Figliozzi et al., 2007). For example, in Denver over 50% of single and combination truck routes include less than six stops (Holguin-Veras and Patil, 2005) and 95% of the truck routes include less than 20 stops. As such, this research work tests the approximations using instances that range from one customer per route to over 35 customers.

This research focuses on approximations to average VRP distance as a function of the number of customers to be served in a given day or instance (n), the amount of freight to be delivered (D), and the number of customers with time-window constraints (n_t). The following real-world conditions must be considered:

- not all potential N customers may request a visit on the same day and the number of customers served per day (n) may be significantly smaller than N.

- The total amount to be delivered or picked up (D) also may vary on a daily basis, ranging from one to several pallets.

- The number of customers with time-window constraints (n_t) also may vary on a daily basis.

---

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

As previously stated, while variability in the amount and characteristics of the day-to-day demand exists, the VRP **problem** analyzed in the following sections is neither dynamic nor stochastic as 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/distances needed depends on the available freight. From an operational standpoint for the day-to-day design of routes, the daily customer demand is known one 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. From a planning perspective (e.g., for fleet design) the average number of routes needed is not known in advance and depends on the specific values of n, D, and n_t.

Random samples of the Solomon problems are used to examine the accuracy of models. 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 4-1(3.21) to each subset of customers. Applying the factors in the second row of demand factors in Table 4-1, the customers have similar demands as in the original Solomon problems (the row characterized by all ones [1]). The resulting problems using the highest demand multipliers (last row of Table 4-1) 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 in order to test the approximations when problems are highly constrained and have a large number of routes and a small number of customers per route. Similarly, for each sample, out of the n customers a random subset of time windows is turned off; P_t takes on values from the set {0.0, 0.2, 0.4, 0.6, 0.8, 1.0}. In all cases the routes' durations were limited by the depot time window. 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) different levels of time-window constraints.

Table 4-1(3.21): Demand Factors

Problem		C1	R1	CR1	C2	R2	RC2
Vehicle Capacity		**200**	**200**	**200**	**700**	**1000**	**1000**
Max. Demand		**50**	**41**	**40**	**41**	**41**	**40**
Demand Level	0	0	0	0	0	0	0
	1	1	1	1	1	1	1
	2	1.6	1.78	1.8	3.6	5.68	5.8
	3	2.2	2.56	2.6	6.2	10.36	10.6
	4	2.8	3.34	3.4	8.8	15.04	15.4
	5	3.4	4.12	4.2	11.4	19.72	20.2
	6	4	4.9	5	14	24.4	25

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 also were 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).

To evaluate the prediction accuracy, the *MAPE* and the *MPE* (Mean Percentage Error) are used and 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.

5.0   ANALYSIS AND DISCUSSION OF EXPERIMENTAL RESULTS

All the regression results presented in this section are obtained by forcing the intercept or constant term to be zero; this is consistent with previous studies by Chien (1992) and Kwon et al. (1995). The results analysis is organized as follows: Section 5.1 provides results for the CVRP (i.e., no time window constraints); Section 5.2 provides results for the time window-constrained VRP approximations in two subsections separating results for the naïve approximations and the approximations for when the number of routes is unknown a priori. Lastly, a distance component analysis is provided in Section 5.3.

5.1   COMPUTATIONAL RESULTS: MODELS WITHOUT TIME-WINDOW CONSTRAINTS (CVRP)

Results for CVRP instances (i.e., no time windows and the depot located at the center) are shown in Table 5-1. 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 three decimals are displayed for the R^2 values, and two for MPE and MAPE values.

As can be seen from Table 5-1, all tested models have very good R^2 values. However, models with more terms such as Model 5 and Model 6 have a superior *MAPE* performance. The models that adjust the tour distances using the term $(n-m)/n$ (Model 2, Model 4 and Model 6) have a superior *MAPE* performance over their counterparts with the same number of estimated coefficients (Model 1, Model 3 and Model 5).

In the next section, the models with time-window constraints are analyzed which impact model prediction ability and the coefficients describing customer spatial distribution.

23

Table 5-1: Comparison of models 1 to 6 with centrally located and corner depots (no time windows)

MODEL	Statistic	Central Depot			Statistic	Corner Depot		
		R2	MPE	MAPE		R2	MPE	MAPE
Model 1	Average	0.966	1.40%	6.00%	Average	0.968	6.20%	12.00%
Model 1	Min	0.933	-0.80%	4.20%	Min	0.954	3.20%	6.20%
Model 1	Max	0.986	3.50%	7.30%	Max	0.982	10.20%	17.90%
Model 2	Average	0.991	1.50%	4.70%	Average	0.984	4.90%	7.90%
Model 2	Min	0.986	-1.20%	3.10%	Min	0.977	2.50%	5.20%
Model 2	Max	0.994	4.20%	6.50%	Max	0.990	9.20%	12.70%
Model 3	Average	0.999	1.00%	4.00%	Average	0.994	6.80%	11.00%
Model 3	Min	0.998	-0.90%	2.20%	Min	0.987	3.50%	5.80%
Model 3	Max	1.000	3.50%	6.40%	Max	0.998	12.30%	17.80%
Model 4	Average	0.999	-0.70%	3.20%	Average	0.997	4.30%	6.90%
Model 4	Min	0.999	-2.60%	1.70%	Min	0.994	1.60%	2.80%
Model 4	Max	1.000	1.60%	4.50%	Max	0.999	8.80%	12.10%
Model 5	Average	0.999	-0.40%	3.10%	Average	0.998	-0.30%	4.80%
Model 5	Min	0.999	-0.70%	2.00%	Min	0.997	-0.80%	2.80%
Model 5	Max	1.000	-0.10%	4.30%	Max	0.999	0.20%	7.10%
Model 6	Average	1.000	-0.10%	2.40%	Average	0.999	-0.10%	3.70%
Model 6	Min	0.999	-0.30%	1.50%	Min	0.998	-0.50%	2.10%
Model 6	Max	1.000	0.10%	3.40%	Max	1.000	0.30%	5.70%

5.2 COMPUTATIONAL RESULTS FOR TIME WINDOW-CONSTRAINED MODELS

5.2.1 Results of naïve approximations of VRP distances

The following is an analysis of the naïve approximations of the VRP distance with time-window constraints from section 3.1.2. The models are tested with centrally located and corner-located depots. Results for the depot located at the center are shown in Table 5-2. As with the models analyzed in section 5.1, the average distance per sample size is the dependent variable. Model fit R^2, MAPE, and MPE are displayed for pooled data consisting of Model 1 and Model 2 and Model 7 through 12.

As with the models without time-window constraints, all eight approximations with time windows have high R^2 values. Model 2 adjusts the tour distances using the term $(n-m)/n$, and has a superior *MAPE* performance compared to its counterpart, Model 1, with the same number of estimated coefficients. The models with more parameters have a similar or superior performance in this measure. Model 11, with the parameter $m_t \approx p_t n\sqrt{n} = n_t \sqrt{n}$, has the best approximating power, followed by Model 10 ($m_t \approx p_t n = n_t$), and Model 12 ($m_t \approx p_t n^2 = n_t n$).

Table 5-2: Approximation Quality by Problem Class (Pooled data)

Problem	R2	MAPE	MAP
C101	0.996	5.60%	-0.60%
R101	0.999	3.40%	-0.30%
RC101	0.999	3.30%	-0.50%
C201	0.995	5.80%	-1.80%
R201	0.997	5.20%	-0.20%
RC201	0.999	4.00%	-0.70%

The data for all 12 models with time-window constraints are displayed in Table 5-3. Using a corner depot, the same trends are observed as with the models without time-window constraints. However, there is a significant improvement in the predicting power of Model 1 and Model 2 while the remaining models have similar *MAPE* values. As previously observed, Model 11 has the best approximating power, followed by Model 10, Model 12, and Model 2, in this order. According to Model 11, the impact of time-window constraints on the number of routes is a function of the number of time-window constraints and the square root of the total number of customers. For a given n_t, if the number of customers increase by a factor x, the increase in the number of routes due to time-window constraints increases by a factor $\sqrt{x}$.

Although Model 7 and 12 provide good approximations, they do not account for the interrelation between the number of routes generated by customer demands and time windows. In instances that are highly constrained, *both* in terms of customer demands and time windows, the following constraint may be binding or even violated:

$$m \geq k_d m_d + k_t m_t$$

(5.1)

Let m_{max} denote the maximum number of routes that can be achieved by the combination of customer demands and time-windows constraints (a proxy is needed because, by definition, m is an unknown). The term $(1 - m_d)/m_{max}$ is introduced to avoid the violation of the constraint $m \geq k_d m_d + k_t m_t$ when customers' demands are high and all customers have tight time windows[4]. Model 14 is the result of adjusting Model 12:

$$\text{Model 14: } VRP\left(V^n\right) \approx k_l \sqrt{An} + 2k_d m_d \bar{r} + 2\frac{\left(1 - m_d\right)}{m_{max}} k_t p_t n \sqrt{n\bar{r}}$$

---

[4] Alternatively, the term $(1 - k_t m_t)/m_{max}$ can be introduced if time-window constraints generate more routes than demand constraints. However, the term $k_t m_t$ is unknown and an iterative approach is required. On the other hand, the value of $k_d \approx 1$; hence $(1 - m_d)/m_{max}$ provides a reasonable approximation, without iterations, when demand constraints are at least as relevant as time-window constraints.

25

Table 5-3: Comparison of models 1 to 12 with centrally located and corner depots and time-window constraints

MODEL	Statistic	Corner Depot			Central Depot		
		R^2	MPE	MAPE	R^2	MPE	MAPE
Model 1: $VRP(V^n) \approx k_l \sqrt{An} + 2m\bar{r}$	Average	0.955	3.40%	8.70%	0.985	0.80%	3.30%
	Min	0.941	0.00%	5.60%	0.970	-1.00%	1.70%
	Max	0.977	6.20%	13.20%	0.994	2.80%	4.50%
Model 2: $VRP(V^n) \approx k_l \dfrac{n-m}{n} \sqrt{An} + 2m\bar{r}$	Average	0.985	3.20%	5.40%	0.981	1.40%	2.90%
	Min	0.976	0.90%	3.20%	0.965	-0.20%	1.50%
	Max	0.989	6.60%	9.30%	0.995	3.00%	4.70%
Model 3: $VRP(V^n) \approx k_l \sqrt{An} + k_m m$	Average	0.998	2.70%	5.60%	1.000	-1.00%	3.10%
	Min	0.996	1.00%	2.40%	0.999	-2.30%	1.90%
	Max	0.999	6.20%	10.20%	1.000	0.90%	4.40%
Model 4: $VRP(V^n) \approx k_l \dfrac{n-m}{n} \sqrt{An} + k_m m$	Average	0.999	3.00%	5.00%	1.000	0.10%	2.10%
	Min	0.997	1.00%	3.50%	0.999	-1.00%	1.70%
	Max	0.999	6.10%	9.00%	1.000	1.20%	2.60%
Model 5: $VRP(V^n) \approx k_l \sqrt{An} + k_b \sqrt{A/n} + k_m m$	Average	0.999	-0.40%	3.90%	1.000	-0.40%	2.10%
	Min	0.999	-0.90%	2.10%	0.999	-0.70%	1.30%
	Max	1.000	-0.10%	5.60%	1.000	-0.10%	2.60%
Model 6: $VRP(V^n) \approx k_l \dfrac{n-m}{n} \sqrt{An} + k_b \sqrt{A/n} + k_m m$	Average	0.999	-0.20%	2.90%	1.000	-0.10%	1.70%
	Min	0.999	-0.50%	1.60%	0.999	-0.30%	1.20%
	Max	1.000	0.00%	4.50%	1.000	0.00%	2.20%
Model 7: $VRP(V^n) \approx k_l \sqrt{An} + k_d m_d\, 2\bar{r} + k_t p_t^2\, 2\bar{r}$	Average	0.991	13.00%	7.40%	0.989	13.70%	8.00%
	Min	0.984	10.50%	5.30%	0.982	7.20%	3.20%
	Max	0.995	16.20%	10.10%	0.996	20.00%	13.00%
Model 8: $VRP(V^n) \approx k_l \sqrt{An} + k_d m_d\, 2\bar{r} + k_t p_t\, 2\bar{r}$	Average	0.991	12.50%	6.90%	0.989	13.30%	7.40%
	Min	0.985	9.90%	4.70%	0.982	7.10%	3.00%
	Max	0.995	15.40%	9.50%	0.997	19.40%	12.30%
Model 9: $VRP(V^n) \approx k_l \sqrt{An} + k_d m_d\, 2\bar{r} + k_t p_t \sqrt{n}\, 2\bar{r}$	Average	0.994	9.50%	5.00%	0.994	9.60%	5.20%
	Min	0.990	6.80%	2.60%	0.990	5.90%	2.10%
	Max	0.997	12.20%	7.60%	0.997	14.10%	9.10%
Model 10: $VRP(V^n) \approx k_l \sqrt{An} + k_d m_d\, 2\bar{r} + k_t p_t n\, 2\bar{r}$	Average	0.996	7.30%	2.80%	0.996	7.00%	2.70%
	Min	0.994	5.90%	0.80%	0.995	5.10%	1.10%
	Max	0.998	9.60%	5.40%	0.998	9.40%	5.30%
Model 11: $VRP(V^n) \approx k_l \sqrt{An} + k_d m_d\, 2\bar{r} + k_t\, p_t n\sqrt{n}\, 2\bar{r}$	Average	0.996	7.00%	1.90%	0.997	6.70%	1.80%
	Min	0.994	5.90%	0.20%	0.995	5.00%	0.40%
	Max	0.998	8.50%	4.30%	0.998	8.00%	3.80%
Model 12: $VRP(V^n) \approx k_l \sqrt{An} + k_d m_d\, 2\bar{r} + k_t\, p_t n^2\, 2\bar{r}$	Average	0.996	7.40%	1.80%	0.996	7.30%	1.70%
	Min	0.994	6.40%	0.20%	0.994	5.10%	0.40%
	Max	0.998	8.40%	4.00%	0.998	8.80%	3.50%

Table 5-4 compares these models for centrally located and corner depots. The results indicate that the additional terms do provide a better approximation both in terms of *MAPE* and *MAP* values.

Table 5-4: Comparison of Model 11 and adjusted Model 14 approximation quality for corner and centrally located depots

Model	Statistic	Central Depot			Corner Depot		
		R2	MAPE	MAP	R2	MAPE	MAP
Model 11	Average	0.997	6.70%	1.80%	0.996	7.00%	1.90%
	Min	0.995	5.00%	0.40%	0.994	5.90%	0.20%
	Max	0.998	8.00%	3.80%	0.998	8.50%	4.30%
Model 14	Average	0.996	6.50%	0.80%	0.997	6.50%	0.70%
	Min	0.994	4.80%	-0.70%	0.996	5.60%	-1.20%
	Max	0.998	8.00%	2.60%	0.998	7.50%	3.20%

The term $(n-m)/n$ seems to provide an explanatory advantage but cannot be used in Model 7 to Model 12 because m is unknown. However, the adjustment provided by $(n-m)/n$ can be simulated if $k_d m_d + k_t m_t$ is used as a proxy for m. This hypothesis is tested using the best approximating Model 14. Model 14 is estimated and with the calibrated parameters, k_d and k_t, and a new Model 15 is estimated with the previously estimated parameters k_l', k_d', and k_t'.

$$\text{Model 15: } VRP(V^n) \approx \frac{k_l'}{n}\left(n - k_d m_d - \frac{(1-m_d)}{m_{max}}k_t p_t n\sqrt{n}\right)\sqrt{An} + 2k_d'\bar{r}m_d + 2\frac{(1-m_d)}{m_{max}}k_t' p_t n\sqrt{n\bar{r}}$$

Table 5-5 compares Model 14 and Model 15 for centrally located and corner depots. The results indicate that the proxy approximation of the term $(n-m)/n$ does improve the performance in terms of *MAPE* and *MAP* values. This is not surprising given that it has some desirable theoretical properties (Figliozzi, 2008): (a) when $n = m$ the estimated *local* distance is zero, whereas (b) when $n \gg m$ or $m = 1$ the local tour distance in the limit becomes similar to the expression suggested by Beardwood et al. (1959). The convergence to the final parameters k_l', k_d', and k_t' is very fast and another iteration does not change the value of the estimated regression parameters.

Table 5-5: Comparison of Model 14 and adjusted Model 15 approximation for corner and centrally located depots

Model	Statistic	Central Depot			Corner Depot		
		R2	MAPE	MAP	R2	MAPE	MAP
Model 14	Average	0.996	6.50%	0.80%	0.997	6.50%	0.70%
	Min	0.994	4.80%	-0.70%	0.996	5.60%	-1.20%
	Max	0.998	8.00%	2.60%	0.998	7.50%	3.20%
Model 15	Average	0.997	5.90%	0.30%	0.997	5.60%	0.00%
	Min	0.995	4.20%	-1.00%	0.996	4.50%	-1.40%
	Max	0.998	7.20%	2.20%	0.998	6.20%	0.90%

The estimated regression parameters for Model 14 disaggregated by problem type are shown in Table 5-6. It is reassuring that the values of the parameters k_l, k_d, and k_t broadly reflect the characteristics of the underlying problem types; Table 5-7 contains some summary statistics of the problem types demand and time-window distributions. The value of k_t displayed in Table 5-6 is "normalized" so it has the same order of magnitude as the other parameters k_l and k_d:

$norm(k_t) = \bar{n}\sqrt{\bar{n}}k_t$ where $\bar{n} = 40$ is the average number of customers across all samples. As expected, the values of k_t are higher for problems where time-window constraints are predominant (i.e., type 1 problems and, in particular, R1 and RC1 problems (see time-window lengths in Table 5-7)). The values of k_d are distributed around one and tend to be smaller when there is a higher time-window influence.

Table 5-6: Regression Parameters by Problem Type (Model 14)

Class	k_l	t-stat	$Norm(k_t)$	t-stat	k_d	t-stat
C1	0.52	11.0	4.85	20.7	1.07	73.5
R1	1.38	29.3	7.28	30.7	0.83	52.7
RC1	1.66	34.2	6.34	25.6	0.78	57.8
C2	0.79	26.4	3.01	19.9	0.99	104.1
R2	0.82	32.0	4.08	30.7	1.05	122.9
RC2	0.78	29.8	3.65	26.4	1.03	138.7

Table 5-7 Statistics of customer distributions by problem type

Class	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
RC1	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

5.2.2 Results of VRP approximations when the number of routes is unknown

The average probabilities $1 - P_\nu$ of failing to connect any two customers due to time-window constraints are shown in Table 5-9. The values of $1 - P_\nu$ reflect the characteristics of the underlying problem types. Type 1 problems where time windows are tight result in higher $1 - P_\nu$ values. Table 5-9 also provides an understanding of the relative impact of time-window constraints on distance traveled. As the level of demand increases, the relative size of the "bin" or vehicle capacity is reduced and there is a consequent reduction in the feasible number of customers per route. Hence, the impact of time-window constraints is reduced as capacity constraints become more "binding."

For Model 13 of Section 3, the estimated regression parameters disaggregated by problem type are shown in

Table 5-10. These parameters are obtained by pooling the data of all different demand levels per problem type (i.e., using one set of parameters $\left(\tilde{k}_\mu, \tilde{k}_{\lambda\mu}, \tilde{k}_\mu^m, \tilde{k}_{\nu\mu}^m \right)$ for all instances). It is reassuring that the regression parameters are not only statistically significant but also reflect the characteristics of the underlying problem types. The values of $\tilde{k}_\mu$ are lowest and highest for clustered and random problems, respectively. In all cases the coefficients $\tilde{k}_{\lambda\mu}$ are significant and positive, which suggests that time-window constraints increase the distance traveled between customers. The coefficients $\tilde{k}_{\lambda\mu}$ follow a similar trend as the $\tilde{k}_\mu$ coefficients; lowest and highest values for clustered and random problems, respectively. As expected, the values of $\tilde{k}_\mu^m$ are slightly less than one but significantly different than zero. The type C2 coefficients demonstrate that although $\tilde{k}_{\nu\mu}^m$ is zero, $\tilde{k}_{\lambda\mu}$ can be positive and significant (i.e., time-window constraints increase the inter-customer distance but do not affect the number of routes that is determined by capacity constraints).

The approximation quality is high, particularly for random and random-clustered problems. The values of MAPE range from 3.4% to 5.6% with an average of 4.5% for the pooled data. As expected, a better fit can be obtained if a regression is run for each demand level. Approximation quality, as evaluated by MAPE, improves significantly as shown in Table 5-12.

Table 5-8: Demand Factors

Problem		C1	R1	CR1	C2	R2	RC2
Vehicle Capacity		200	200	200	700	1000	1000
Max. Demand		50	41	40	41	41	40
Demand Level	0	0	0	0	0	0	0
	1	1	1	1	1	1	1
	2	1.6	1.78	1.8	3.6	5.68	5.8
	3	2.2	2.56	2.6	6.2	10.36	10.6

4	2.8	3.34	3.4	8.8	15.04	15.4
5	3.4	4.12	4.2	11.4	19.72	20.2
6	4	4.9	5	14	24.4	25

Table 5-9. Average Probability $1 - P_s(v)$

		C1	R1	RC1	C2	R2	RC2
Demand Level	0	13.60%	25.40%	20.50%	0.00%	9.10%	10.60%
	1	13.60%	25.40%	20.50%	0.00%	9.10%	10.60%
	2	1.20%	23.10%	7.00%	0.00%	1.00%	1.00%
	3	0.00%	9.80%	1.40%	0.00%	0.00%	0.00%
	4	0.00%	2.60%	0.60%	0.00%	0.00%	0.00%
	5	0.00%	1.80%	0.00%	0.00%	0.00%	0.00%
	6	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 5-10. Estimated Regression Coefficients by Problem Class

Problem	$\tilde{k}_{\mu}$	t-stat	$\tilde{k}_{\lambda\mu}$	t-stat	$\tilde{k}_{\mu}^{m}$	t-stat	$\tilde{k}_{v\mu}^{m}$	t-stat
C101	0.67	24.3	0.25	5.6	0.81	48.5	1.31	6.49
R101	0.89	54.6	0.32	8.0	0.80	78.5	1.02	24.42
RC101	0.70	40.8	0.21	4.1	0.94	108.2	0.98	16.51
C201	0.70	37.5	0.29	10.1	0.78	55.7	-----	-----
R201	0.99	48.1	0.41	17.1	0.73	57.9	1.44	16.84
RC201	0.77	48.8	0.34	12.5	0.88	106.9	1.54	25.87

Table 5-11. Approximation Quality by Problem Class (Pooled data)

Problem	R2	MAPE	MAP
C101	0.996	5.6%	-0.6%
R101	0.999	3.4%	-0.3%
RC101	0.999	3.3%	-0.5%
C201	0.995	5.8%	-1.8%
R201	0.997	5.2%	-0.2%
RC201	0.999	4.0%	-0.7%

Table 5-12. Average Approximation Quality by Problem Class (By Distribution)

Problem	R2	MAPE	MAP
C101	0.998	3.9%	-0.7%
R101	0.999	2.5%	0.4%
RC101	1.000	1.9%	-0.4%
C201	0.994	5.5%	0.3%
R201	0.999	3.1%	-0.4%
RC201	0.999	2.6%	-0.3%

5.3   DISTANCE COMPONENT ANALYSIS

The fraction of connecting distance, f_c, can be obtained as the ratio between the connecting distance and total VRP distance, using Model 2:

$$f_c = \frac{2\bar{r}m}{2\bar{r}m + k_l \dfrac{n-m}{n}\sqrt{An}}$$

(5.2)

The fraction of connecting distance is a function of a coefficient λ and the number of customers and routes:

$$f_c(\lambda,n,m) = \frac{1}{1 + \lambda \dfrac{n-m}{m\sqrt{n}}},$$

(5.3)

where

$$\lambda = \frac{k_l\sqrt{A}}{2\bar{r}}.$$

(5.4)

As intuitively expected, expression (5.3) renders $f_c = 1$ when $n = m$ or $\lambda = 0$. The coefficient λ is the dimensionless "service area-depot" factor that reflects the relative size of the distribution area in relation to the average distance between the depot and its customers. Hence, λ is highest for centrally located depots and tends to zero as the average distance between the depot and its customers increases.

For large n and $n \gg m = 1$, the decrease in connecting distance is proportional to the term:

$$f_c(\lambda,n,m) \approx \frac{1}{1 + \dfrac{\lambda}{m}\sqrt{n}},$$

(5.5)

with $n \gg m$.

Hence, for the TSP with large n :

$$f_c(\lambda,n,1) \approx \frac{1}{\lambda\sqrt{n}},$$

(5.6)

31

with $n >> 1$.

For a TSP, the parameter $f_c \to 0$ when $n \to \infty$; this is congruent with Beardwood et al. (1959) which indicates with probability one that the average inter-customer distance tends to $k_l \sqrt{A}/\sqrt{n}$ when $n \to \infty$. Figure 5-1 shows the relationship between f_c and $(n-m)/(m\sqrt{n})$ for the R2 subsets. There is a clear division between VRP (to the right) and TSP points (to the left) in Figure 5-1. This is expected from the analysis of expression (5.3). For a given λ, when more constraints are added $m \to n$ and $f_c \to 1$; when the number of routes remains constant and more customers are added, $n \to \infty$, then $f_c \to 0$. As expected, the corner depot λ is smaller than the center depot λ, hence, for the same $(n/m-1)$ the fraction of connecting distance f_c tends to be higher for the corner depot points.
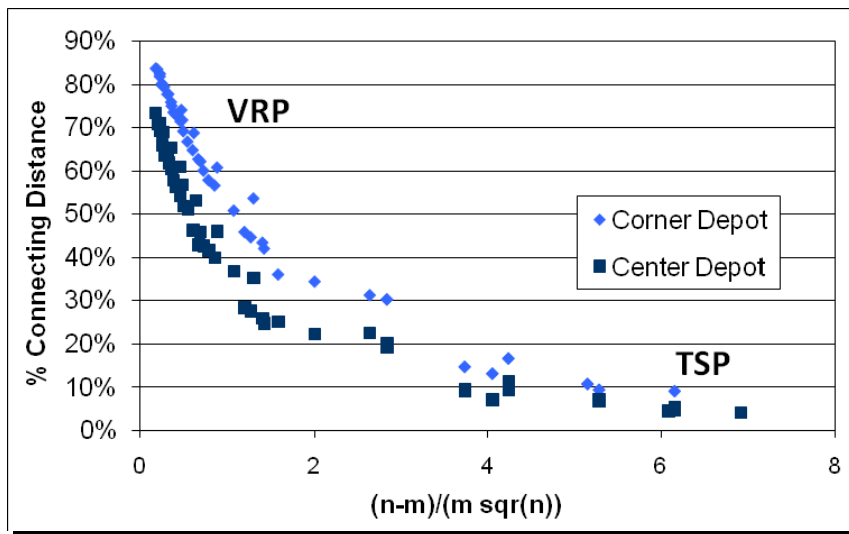

Figure 5-1: TSP and VRP Fractions of Connecting Distance

The analysis of f_c may be used to simplify the estimation of distances in theoretical logistics models. Even in TSP instances, the connecting component prevails over the local tour component for λ and n where $f_c(\lambda, n, 1) \approx 1$. For situations where $f_c(\lambda, n, m) \approx 1$, the local tour term can be safely ignored:

$$VRP(V^n) \approx 2k_m m \bar{r},$$
(5.7)

whereas for situations where $f_c(\lambda, n, m) \approx 0$ it is the connecting distance term(s) which can be safely ignored:

$$VRP(V^n) \approx k_l \sqrt{An}.$$
(5.8)

5.4   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 also can be used to estimate distance, costs, and times in real-life 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. [22].

Model 4 was tested with customers located in the industrial suburb of Bankstown, with 30 customers distributed in an irregular area of 39.5 squared kilometers (Figure 5-2). The delivery area is bordered by the Bankstown local airport to the west, a freeway to the south, and secondary highways to 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 one or two or, exceptionally, close to 30. 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.
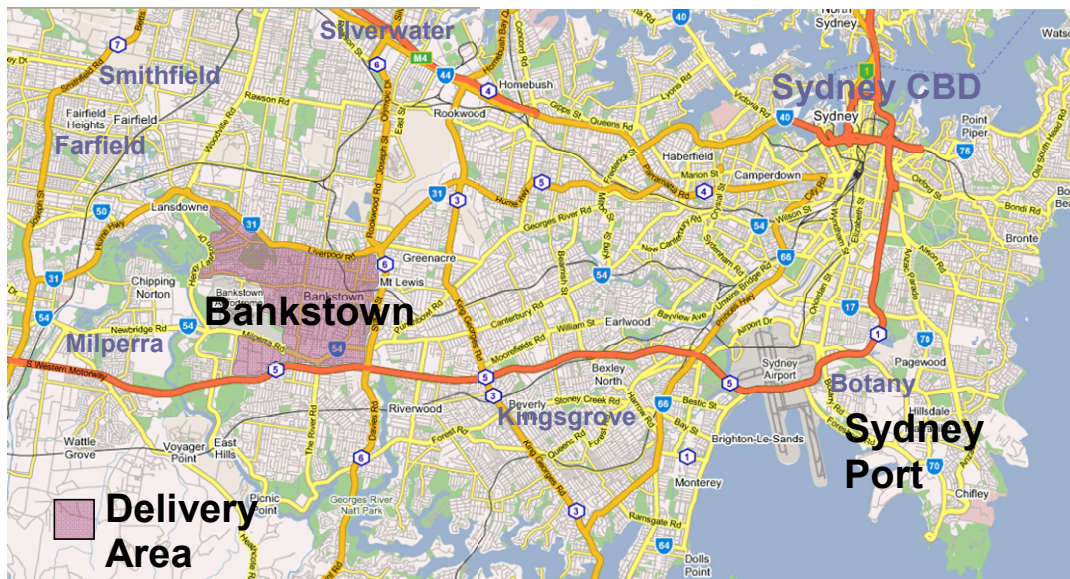

Figure 5-2: Relative Location of the Port of Sydney and Delivery Industrial Areas

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 and avoid 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 for freeways, 35 km/hour for main connecting streets (four lanes or more with traffic lights), and 25 km/hour for local streets. With this speed information, a matrix of shortest travel times between customers and the depot was constructed using the urban highway network and geographic information system (GIS) software.

Figure 5-3 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 suburban customers while the longer distances are mostly depot customers. The R2 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 depot-customer pairs. From existing customer data, an average service time of 45 minutes per customer is used.
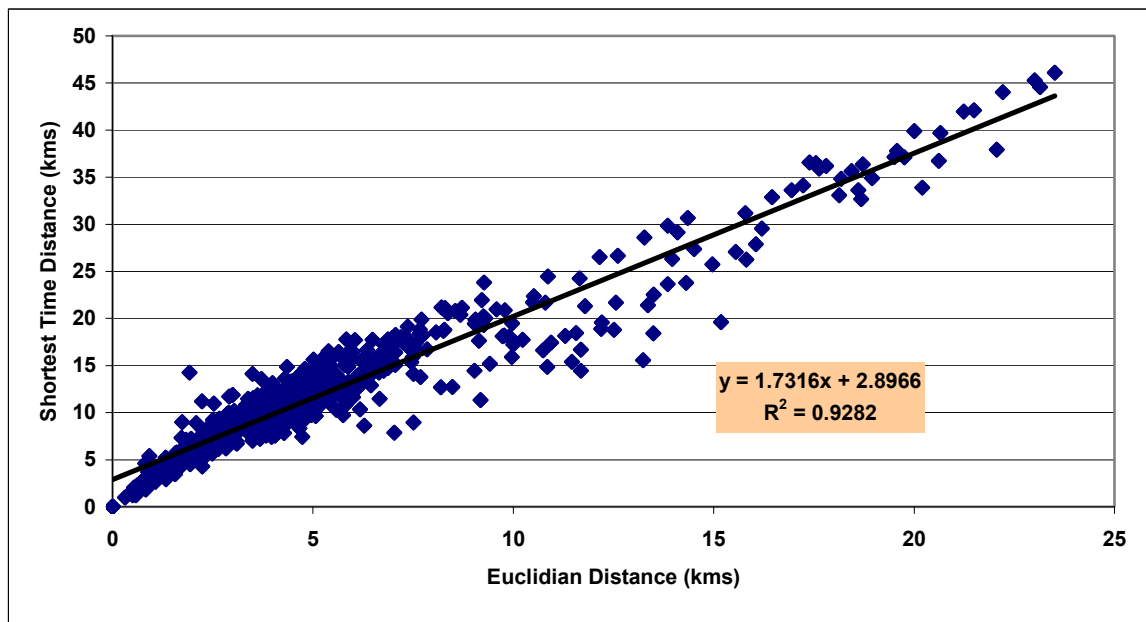

The regression equation shown on the chart: $y = 1.7316x + 2.8966$, $R^2 = 0.9282$

Figure 5-3 Euclidian Distance vs. Shortest Time Distance among Suburban Customers and Depot Customers

Three different routing scenarios were constructed: (a) no constraints or TSP case, (b) with a tour duration constraint of eight hours, and (c) adding four-hour time windows per customer. The number of routes varied from one route in the TSP instances to five 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 5-13. 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 Table 5-4). Model 4 was used to approximate times and distances due to the different travel speeds; the connecting distance between the depot and customers does not always

follow the same type of highway. These results are encouraging and show that the proposed models may have useful applications in urban networks and modeling applications [23, 24] . 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.
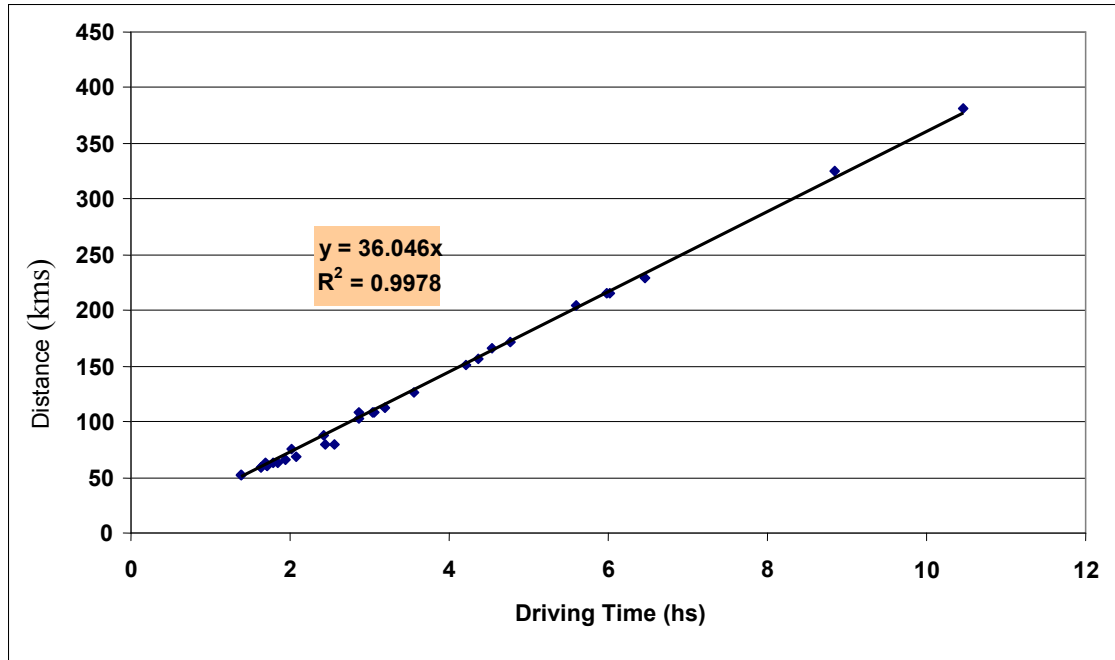

Figure 5-4: Distance Traveled and Time Driven

Table 5-13: Real-life network distance and time estimation (Model 4)

Instance	R2	MPE	MAPE	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

6.0   CONCLUSION

This is the first research effort to study approximations to the average length of VRP when there is variability in the number of customers, time-window constraints, and demand levels. These approximations are intended for the strategic and planning analysis of transportation and logistics problems, when the number and location of customers vary daily and are not known a priori. Several approximations are proposed and successfully tested using instances with different patterns of customer spatial distribution, time windows, customer demands, and depot locations. A relatively simple approximation formula can be used to approximate reasonably well the impact of the number of time-window constraints and the demand level on the number of additional routes.

Adjustment factors were successfully introduced to enforce the total number of routes constraint and to account for the decrease in local distance when more routes are needed. The adjustments not only improve the approximation quality in terms of *MAPE* and *MAP*, but also render parameters that reflect the characteristics of the underlying routing problems. The VRP distance can be broken down in three distinct components: local, demand, and time-window distances. The level of demand and time-window constraints are successfully used to predict the number of routes needed. The analysis of tradeoffs between distances travelled, number of customers, number of routes, and depot/delivery area factors allows for the determination of conditions that correspond to either VRP or TSP routes. The distance analysis for VRP and TSP limit cases provide a theoretical justification to simplify the distance expressions in limit cases where the connecting distance or the local tour component prevails.

7.0   REFERENCES

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

BRAMEL, J., COFFMAN, E. G., SHOR, P. W. & SIMCHI-LEVI, D. (1992) Probabilistic analysis of the capacitated vehicle routing problem with unsplit demands. *Operations Research,* 40(6)**,** 1095-1106.

BRAMEL, J. & SIMCHI-LEVI, D. (1996) Probabilistic analyses and practical algorithms for the vehicle routing problem with time windows. *Operations Research,* 44(3)**,** 501-509.

CHIEN, T. W. (1992) Operational Estimators For The Length Of A Traveling Salesman Tour. *Computers & Operations Research,* 19(6)**,** 469-478.

CHRISTOFIDES, N. & EILON, S. (1969) Expected Distances In Distribution Problems. *Operational Research Quarterly,* 20(4)**,** 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(4)**,** 331-350.

DAGANZO, C. F. (1987a) Modeling Distribution Problems With Time Windows.1. *Transportation Science,* 21(3)**,** 171-179.

DAGANZO, C. F. (1987b) Modeling Distribution Problems With Time Windows.2. - 2 Customer Types. *Transportation Science,* 21(3)**,** 180-187.

DIANA, M., DESSOUKY, M. M. & XIA, N. (2006) A model for the fleet sizing of demand responsive transportation services with time windows. *Transportation Research Part B,* 40(8)**,** 651-666.

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.*

FEDERGRUEN, A. & VAN RYZIN, G. (1997) Probabilistic analysis of a generalized bin packing problem and applications. *Operations Research,* 45(4)**,** 596-609.

FIGLIOZZI, M. (2008a) An Iterative Route Construction and Improvement Algorithm for the Vehicle Routing Problem with Soft and Hard Time Windows. *Applications of Advanced Technologies in Transportation (AATT) 2008 Conference Proceedings.* Athens, Greece, May 2008.

FIGLIOZZI, M. A. (2006) Modeling the Impact of Technological Changes on Urban Commercial Trips by Commercial Activity Routing Type. *Transportation Research Record 1964***,** 118-126.

FIGLIOZZI, M. A. (2007) Analysis of the efficiency of urban commercial vehicle tours: Data collection, methodology, and policy implications. *Transportation Research Part B,* 41(9)**,** 1014-1032.

FIGLIOZZI, M. A. (2008b) Planning Approximations to the Average Length of Vehicle Routing Problems with Varying Customer Demands and Routing Constraints. *Transportation Research Record 2089***,** 1-8.

FIGLIOZZI, M. A., KINGDON, L. & WILKITZKI, A. (2007) Analysis of Freight Tours in a Congested Urban Area Using Disaggregated Data: Characteristics and Data Collection Challenges. *Proceedings 2nd Annual National Urban Freight Conference, Long Beach, CA. December*.

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(6)**,** 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(10)**,** 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(2)**,** 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(2)**,** 254-265.

STEIN, D. (1978) An asymptotic probabilistic analysis of a routing problem. *Mathematics Of Operations Research,* 3(2)**,** 89-101.

WEBB, M. (1968) Cost functions in the location of depots for multiple-delivery journeys. *Operational Research Quarterly,* 19(3)**,** 311-320.

8.0   APPENDIX

Lemma 2. The expected number of additional routes due to time-window constraints, $E(m_{v\mu})$, is a continuously decreasing function of P_v.

$$E[m_{v\mu}(n)] = \left[\sum_{c=1}^{c=b} \frac{nP(c)}{c} \prod_{j=c+1}^{b} [1 - P(j)] \right] - \frac{n}{b}, \tag{8.1}$$

Proof. This is a continuous function because it is a linear combination of continuous functions of the variable P_v.

The weight factors

$$w(c) = P(c) \prod_{j=c+1}^{b} [1 - P(j)], \tag{8.2}$$

are applied to each feasible route with c customers per route. Developing the sum of weight factors and denoting $P(c) = P_c$ for the sake of brevity

$$\sum_{c=1}^{c=b} P_c \prod_{j=c+1}^{b} (1 - P_j) = P_b + (1 - P_b) \left[\sum_{c=1}^{c=b-1} P_c \prod_{j=c+1}^{b-1} (1 - P_j) \right], \tag{8.3}$$

Since $P_1 = 1$ as all customers can be served from the depot without violating time-window constraints, for $b = 2$ expression (8.3) is equal to one:

$$\sum_{c=1}^{c=2} P_c \prod_{j=c+1}^{2} (1 - P_j) = P_1(1 - P_2) + P_2 = 1, \tag{8.4}$$

For $b = 3$, expression is also equal to one:

$$\sum_{c=1}^{c=3} P_c \prod_{j=c+1}^{3} (1 - P_j) = P_1(1 - P_2)(1 - P_3) + P_2(1 - P_3) + P_3 = P_3 + (1 - P_3)[P_1(1 - P_2) + P_2]$$

$$= P_3 + (1 - P_3) \left[\sum_{c=1}^{c=2} P_c \prod_{j=c+1}^{2} (1 - P_j) \right] = 1. \tag{8.5}$$

By induction, for any positive integer b, expression (8.3) is always equal to one such that

$$\sum_{c=1}^{c=b} P_c \prod_{j=c+1}^{b}(1-P_j)= P_b +(1-P_b)\left[\sum_{c=1}^{c=b-1} P_c \prod_{j=c+1}^{b-1}(1-P_j)\right]=1 \tag{8.6}$$

Developing the sum and replacing $P_c =(P_v)^{c-1} = P^{c-1}$ for the sake of brevity, for any sum up to $b=i$:

$$\sum_{c=1}^{c=i} \frac{nP(c)}{c}\prod_{j=c+1}^{i}[1-P(j)]=$$

$$n\left[\frac{P^{i-1}}{i}+\frac{P^{i-2}}{i-1}\left(1-P^{i-1}\right)+\frac{P^{i-3}}{i-2}\left(1-P^{i-1}\right)\!\left(1-P^{i-2}\right)+...\right.$$

$$\left....+\frac{P^1}{2}\left(1-P^{i-1}\right)\!\left(1-P^{i-2}\right)..\left(1-P^2\right)+\frac{P^0}{1}\left(1-P^{i-1}\right)\!\left(1-P^{i-2}\right)..\left(1-P^2\right)\!\left(1-P^1\right)\right] \tag{8.7}$$

Making the terms $1-P^j, j=1,...,i-1$ common factors in equation (8.7) gives

$$n\left[\frac{P^{i-1}}{i}+\left[(1-P^{i-1})\left[\frac{P^{i-2}}{i-1}+\left[(1-P^{i-2})\left[...\ ...+\left[\frac{P^2}{3}+\left[(1-P^2)\left[\frac{P^1}{2}+[(1-P^1)\frac{P_1}{1}]\right]...\right]\right]\right]\right]\right]\right] \tag{8.8}$$

For any increase in P_v, any $1-P^j, j=1,...,i-1$ will have an increase. However, any increase in P^j will reduce the term $1-P^j$ that multiplies the sum of the $j-1, j-2,....,2,1$ terms. Since the sum of the weight factors remains constant and equal to one, as P_v increases the weight applied to the terms:

$$n\frac{P^{j-2}}{j-1},n\frac{P^{j-3}}{j-2},...,n\frac{P^1}{2},n\frac{P^0}{1}$$

decreases whereas the term $n\dfrac{P^{j-1}}{j}$ increases for any $j=1,...,i-1$. Hence, as P_v increases $E(m_{v\mu})$ decreases. In particular, as P_v increases the term with the largest index always increases whereas the term with index one always decreases.

OTREC

**OREGON TRANSPORTATION RESEARCH
AND EDUCATION CONSORTIUM**

P.O. Box 751
Portland, OR 97207

OTREC is dedicated to
stimulating and conducting
collaborative multi-disciplinary
research on multi-modal surface
transportation issues, educating
a diverse array of current
practitioners and future leaders
in the transportation field, and
encouraging implementation of
relevant research results.