

December 2003

# An Analytical Model to Evaluate Alternative Methods for Distance Calculation in GIS-Based Vehicle Routing Decision Support Systems

Julian Ray  
*Western New England College*

Follow this and additional works at: <http://aisel.aisnet.org/amcis2003>

---

## Recommended Citation

Ray, Julian, "An Analytical Model to Evaluate Alternative Methods for Distance Calculation in GIS-Based Vehicle Routing Decision Support Systems" (2003). *AMCIS 2003 Proceedings*. 321.  
<http://aisel.aisnet.org/amcis2003/321>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2003 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# AN ANALYTICAL MODEL TO EVALUATE ALTERNATIVE METHODS FOR DISTANCE CALCULATION IN GIS-BASED VEHICLE ROUTING DECISION SUPPORT SYSTEMS

Julian J. Ray  
School of Business  
Western New England College  
[jray@wnec.edu](mailto:jray@wnec.edu)

## Abstract

*Most commercial vehicle-routing decision support systems employ network optimization methods for calculating inter-stop distances over digital street data. Although potentially more accurate, these optimization methods are computationally intensive, increase overall problem complexity and require more sophisticated and expensive GIS systems to support the functions. Alternative approaches for calculating inter-stop street distances have been re-emerging in the literature as software developers seek to increase efficiency when solving these types of complex problems allowing larger problems to be solved in real-time. This research in progress describes an analytical model to determine if approximation methods are comparable, in terms of the quality of solutions generated, to computationally expensive optimization methods which have been generally adopted by software vendors.*

**Keywords:** GIS, vehicle routing, decision support systems

## Introduction

The proliferation of GPS and mobile communication technologies and services has generated a demand for systems which can perform real-time route-planning operations for both individual vehicles as well as vehicle fleets. Customer-focused location based services typically provide simple point-to-point route planning operations, returning maps and driving directions at the request of a user. More sophisticated vehicle routing and scheduling services are available for organizations managing fleets of service and delivery vehicles in real-time. Solutions available from companies such as UPS Logistics (MobileCast), SAIC (FleetOptimizer), and Descartes (Roadshow Dispatch) enable organizations dynamically assign vehicles to service customers based on current locations of vehicles and geographical distribution of customer demand. Business uses of such systems are myriad: telecommunications installation and service, courier services, home heating fuels, electronic commerce and para-transit systems are a few common examples. The demand for solutions which can address this class of vehicle routing has increased recently as organizations seek to integrate fleet management operations with Customer Relationship Management (CRM), online-retailing and Call Center solutions in an attempt to increase customer satisfaction by narrowing service delivery time-windows while simultaneously increasing percentage of on-time calls.

Central to the process of managing vehicles in real time is the need to quickly locate vehicles on a street network and estimate the distance and/or travel-time required for each vehicle to traverse from its present position to alternative service locations. Once distances between each available vehicle and each customer are known, customers can be assigned to vehicles such that an overall set of business objectives are best met. Business objectives are dependent on organization goals and can include single or combinations of objectives such as minimizing vehicle operation costs, maximizing measures of customer satisfaction, and maximizing revenue.

Distance estimations between locations are typically performed by software components, variously called in the literature time/distance calculators or Origin-Destination (O-D) generators, and provide primary input, in the form of one or more matrices, to higher-order vehicle route-planning algorithms. Two competing approaches to generating these input matrices are used by

software vendors: exact calculations performed using optimized shortest path calculations over a digital network of streets and highways, and approximations using geometric techniques. A recent survey of 24 technology companies providing fleet planning solutions report 19 vendors use “actual street distances” derived by optimizing shortest paths through digital street databases during the assignment and sequencing of stops on vehicle routes (Hall, 2002).

Once service stops have been assigned to a vehicle and a matrix of distances between stops generated, a vehicle route is generated by determining the most efficient ordering of stops for that vehicle. Calculating a vehicle route involves finding a ordering of  $n$  service stops  $(i_1, i_2, \dots, i_n)$  minimizing the quantity  $d_{i_1 i_2} + d_{i_2 i_3} + \dots + d_{i_n i_1}$  given a distance matrix  $\mathbf{D} = (d_{ij})$  where  $d_{ij}$  is the distance from stop  $i$  to stop  $j$  subject to a set of operational constraints such as meeting service delivery time windows. Clearly, the accuracy of calculations made by time/distance calculators should directly reflect on the quality of the resultant vehicle routes as matrix  $\mathbf{D}$  is the primary decision variable. More accurate distances provided by network optimizing software should, therefore, generate better vehicle routes as barriers to transportation, such as one-way streets, rivers, and parks, can increase inter-stop street level distances. These barriers are not taken into account by distance approximation methods leading to bias in the distance matrix and potentially sub-optimal vehicle routes.

This paper reports the preliminary results of a model designed to test whether the increased accuracy of optimizing distance calculators does indeed generate significantly higher-quality results for vehicle fleet management applications and, therefore, warrant their additional cost and computational overhead. The paper is organized as follows: Section 2 discusses various geometric and network-optimization models used to calculate distances over street networks; Section 3 describes the methods developed to test the different distance functions, and Section 4 discusses preliminary results.

## Distance Generating Models

Calculating actual distances between points over digital street databases requires generating a representation or “model” of the street system which can support network-optimization algorithms. In general, a digital street database is transformed into a system of connected nodes and edges; nodes typically correspond to street intersections, edges to the physical streets and highways. Weights assigned to edges are usually the lengths of the corresponding streets or average time taken to traverse the streets using a particular mode of transport. Techniques for transforming the digital street database to a network graph usually depend on the format of the digital street database and requirements of the time/distance calculation software. Laurini and Thompson (1992 Chapter 5) provide a general discussion regarding use of networks and graphs in GIS; ESRI (2002) and Intergraph (1998) discuss details of particular vendor transformations. For a general discussion of the relationships between GIS, vehicle routing, and their application to decision support refer to Keenan (1998).

Once a digital street database has been transformed into a system of connected nodes and edges, shortest-path algorithms can be used to find optimal routes through the system; an operation which can be both prohibitively time consuming and computationally expensive. The fastest known algorithm on GIS-based street databases has a worst-case computational complexity of  $O(m \log_{(n)})$  where  $m$  is the number of edges and  $n$  the number of nodes in the network (Zahn & Noon, 1998). For large street networks, such as those used to cover a single urban area such as Boston, Chicago, Washington D.C. or New York City, the size of the network in terms number of edges and nodes becomes a major performance factor. Networks as large as 1-2 million edges are common for modeling dense urban street networks while much larger networks are required for logistic operations covering geographic regions such as the North East. Even on state of the art hardware, calculating distance matrices in real-time using these methods for relatively small numbers of vehicles and customers is often intractable leading researchers and practitioners to search for alternatives. Theriault et. al. (1999), for example, estimate that generating the shortest-path matrix could take days or even weeks on the computer systems available to them. Similarly, Weigel and Cao (1999) note that generating complete shortest-path matrices is often a waste of computational resources as many of the stop-to-stop combinations would never be considered in the solution and, therefore, their computation is redundant.

A variety of estimation methods for generating distances between locations on street networks are discussed in the literature. Methods include direct Euclidean approximation (Rodriguez et. al. 1995), weighted geometric methods (Love and Morris 1972; Love, et. al. 1988 Chapter 10; Brimberg and Love 1992; Brimberg et. al. 1995; Campbell et. al. 2001) and neural network methods (Alpaydim et. al. 1996). These estimation functions have been variously developed to replace computationally expensive shortest-path optimizations or to negate the need for accurate digital databases of roads.

One commonly used function for estimating street distances, and the method selected for this analysis, is the weighted  $l_{(p)}$  norm. For two points,  $i$  and  $j$  on a Cartesian plane,  $(x_i, y_i)$  and  $(x_j, y_j)$ , the weighted  $l_{(p)}$  norm is calculated as:

$$d_{ij}^{kp} = k \left( |x_i - x_j|^p + |y_i - y_j|^p \right)^{1/p}$$

where  $k$  and  $p$  are parameters to be determined by fitting the model to a street network,  $p \geq 1$ . When  $p = 1$  the model reduces to determining  $k$ -times the recti-linear distance between the two points, often called the Manhattan-metric. When  $p = 2$ , the model determines  $k$  times the Euclidean distance between two points. The constant  $k$  is called the circuitry factor and models the differences between the actual street distance and the estimated street distance between points on the network and can be estimated using linear regression (Love et. al. 1988 Chapter 10).

## Method

To test the difference in quality of time/distance calculators on vehicle routes, a GIS base map was constructed from Geographic Data Technologies Dynamap Transportation data for Massachusetts (GDT 2002). The base map covers the eastern half of the State (7 counties). This area was selected because it provides a rich combination of both dense and sparse road systems and is typical of the size of base map necessary to support a vehicle fleet serving a large metropolitan area. Within the base map extent, three sample areas were selected to model different street-densities: an inner-city (urban) area, a lower density sub-urban area, and a rural area. The urban area selected is 5 mile-square area of inner-city Boston, the sub-urban sample area is a 10 mile square of Boston's western suburbs, while the 15 mile-square rural area is centered on the rural counties in the middle of the State. A network graph containing 754,448 directed edges and 301,823 nodes was constructed from the base map to support the shortest path algorithms. To provide starting locations for testing the distance functions, a thousand random points were generated within each sample area.

Before the tour analyses can be performed, the weighted  $l_{(p)}$  distance estimation functions have to be calibrated to the street network. As the density of streets in each sample area is purposefully different, each sample area has to be calibrated independently. The calibration is performed by generating 200 routes each with randomly selected starting and ending locations from the set of initial 1000 points in each sample area. For each of the 200 pairs of points the shortest path,  $l_{(1)}$ , and  $l_{(2)}$  values are generated using an initial value of  $k = 1$ . A linear regression analysis is used to predict the best value for the circuitry factor,  $k$ , for each of the two distance estimation functions in each of the sample areas.

Once the distance estimation functions have been calibrated they can be used to determine their utility in calculating vehicle tours. The test algorithm was run for 100 iterations for each sample area in turn. During each iteration, the algorithm randomly selects a tour length between 5 and 20 stops and then randomly selects a set of stops from the list of 1000 stops for that scenario. The first stop in the randomly selected list is used as the initial starting location for the vehicle. For each set of stops the inter-stop distance matrices are generated for each of the three distance models.

Once the matrices are generated, a Traveling Salesman Problem (TSP) solver is used to find the best ordering of stops on the vehicle route starting from the first randomly generated stop, visiting each of the stops in an optimal manner, and returning to the initial location. The TSP solver is run using each of the three distance matrices. The final length of each  $l_{(1)}$  and  $l_{(2)}$  generated tours are computed using the optimized distance matrix to allow the final tour length to be standardized against the optimized benchmark. Tour statistics are compiled for each scenario and a regression analysis performed to determine the relationship between the distance estimated and optimal vehicle tours.

**Table 1. Average Increase in Route Length Over Optimized Routes**

	Model	%age	Miles	R <sup>2</sup>
Urban	$I_{(2)}$	0.538	0.197	0.9971
	$I_{(1)}$	2.039	0.745	0.9892
Sub Urban	$I_{(2)}$	0.446	0.258	0.9983
	$I_{(1)}$	1.414	0.817	0.9934
Rural	$I_{(2)}$	0.570	0.520	0.9975
	$I_{(1)}$	1.763	1.611	0.9900

## Preliminary Results

Preliminary results, shown in Table 1, indicate that  $I_{(2)}$  provides the best and most consistent approximation model of the two tested so far. The  $I_{(2)}$  model generates an average 0.518% increase in tour length across all three sample areas compared to an average of 1.738% increase for  $I_{(1)}$ .  $I_{(2)}$  performs best in medium-dense sub-urban networks and least well in more-dense urban areas. In all three sample areas, difference in average vehicle tour lengths created from distance matrices generated by  $I_{(2)}$  and optimized methods are statistically significant within the error limits of the input data. Average time taken to create a distance matrix using optimized methods is 10.17 seconds per vehicle route (0.85 seconds per path tree) on a Pentium 1.6 GHz Intel processor and negligible for  $I_{(1)}$  and  $I_{(2)}$  approximation methods.

Results of the study to date indicate that estimation functions can be an effective and efficient alternative to calculating distance matrices using shortest path methods for vehicle routing decision support systems. Distance matrices constructed using these estimation models do not appear to significantly affect the quality of vehicle routes constructed using them. As construction of distance matrices forms a large part of the computational time and system resources necessary to solve fleet-planning problems, these results have significant impact on potential design of fleet decision support systems, particularly systems which are designed to operate in real-time.

## References

- Alpaydim, E., Altinel, I.K., and Aras, N. "Parametric Distance Functions vs. Non-Parametric Neural Networks for Estimating Road Travel Distances," *European Journal of Operational Research*, (92), 1996, pp. 230-243.
- Brimberg, J. and Love, R.F. "A New Distance Function for Modeling Travel Distances in a Transportation Network," *Transportation Science*, (26:2), 1992, pp. 129-137.
- Brimberg, J., Love, R.F., and Walker, J.H. "The Effect of Axis Rotation on Distance Estimation," *European Journal of Operational Research*, (80), 1995, pp. 357-364.
- Campbell, J.F., Labelle, A., and Langevin, A. "A Hybrid Travel Distance Approximation for a GIS-Based Decision Support System," *Journal of Business Logistics*, (22:2), 2001, pp 165-181.
- ESRI. "The ArcGIS Network Model." ArcObjects 8 Online Documentation, ESRI, Redlands, CA., 2000.
- GDT. *Dynamap Transportation V4.2*. Geographic Data Technologies, Lebanon, NH., 2002.
- Hall, R.H. "Vehicle Routing Software Survey: Change of Direction," *ORMS Today*, (29:1), 2002.
- Intergraph. 1998. *GeoMedia Network: User's Guide*. Intergraph Corporation, Huntsville AL., 1998.
- Keenan, P.B. "Spatial Decision Support Systems for Vehicle Routing," *Decision Support Systems*, (22), 1998, pp. 65-71.
- Laurini, R. and Thompson, T. *Fundamentals of Spatial Information Systems*, San Diego, California: Academic Press Ltd., 1992.
- Love, R. F., Morris, J.G., and Wesolowsky, G.O. *Facilities Location: Models and Methods*. New York: North-Holland, 1988.
- Rodriguez, M., Sirmans, C. F., and Marks, A. P. "Using Geographic Information Systems to Improve Real Estate Analysis." *The Journal of Real Estate Research* (10:2), 1995, pp. 163-173.

- Theriault, M., Venbdermissen, M-H, and Leroux, D. "Modeling Trip Length and Duration within GIS: Application to an O-D Survey." *Journal of Geographic Information and Decisions Analysis* (3:1), 1999, pp. 40-56.
- Weigel, D., and Cao, B. "Applying GIS and OR Techniques to Solve Sears Technician-Dispatching and Home-Delivery Problems," *Interfaces* (29:1), 1999, pp. 112-130.
- Zahn, F. B., and Noon, C. E. "Shortest Path Algorithms: An Evaluation using Real Road Networks." *Transportation Science* (32:1), 1998, pp. 65-73.