Using GPS Data to Understand Variations in Path Choice

Oliver Jan, Civil Engineering and Mechanics, University of Wisconsin – Milwaukee, PO Box 784, Milwaukee, WI 53201
Alan J. Horowitz, Center for Urban Transportation Studies, University of Wisconsin – Milwaukee, PO Box 784, Milwaukee, WI 53201
Zhong-Ren Peng, Urban Planning, University of Wisconsin – Milwaukee, PO Box 413, Milwaukee, WI 53201

Abstract. A comprehensive set of GPS vehicle location data from Lexington households was analyzed to determine if such data can be helpful in improving path choice assumptions in traffic assignment models. Primarily, the portion of the data used consisted of a reconstruction of the street network and the lists of street segments in each path. Analysis was based on "matches" of trips, e.g., pairs of trips with similar origins and destinations. Matches were obtained for trips within households and for trips across households. Statistics used to compare trips in matches were a path deviation index and the percentage of identical links. It was found that the path chosen on a trip was quite sensitive to the location of the origin and destination and that the chosen path most often differed considerably from the shortest time path across the network. Paths for trips made by the same driver were very consistent over time; paths by different drivers showed more deviations even when the trip ends were the same or very similar. Recommendations are made as to how GPS data on path choice can be better collected in the future and for improvement of traffic assignment models.

INTRODUCTION

A path is a sequence of links (e.g., road segments) and nodes (e.g., intersections) that comprise a trip from an origin to a destination. Notions of path choice by travelers are fundamental to the traffic assignment step in travel forecasting models and to many traffic simulation models.

The current methods used by planners for modeling path choice in traffic assignment have been developed largely in the absence of objective empirical evidence of actual path choices. Theories of user-optimal equilibrium assignment and stochastic multipath traffic assignment have proven quite useful to planners, but those algorithms' underlying assumptions related to path choice have not received an adequate level of validation. Furthermore, these algorithms are most often applied to a network overlaid on a coarse zone system, and the implications of different levels of zonal aggregation on the validity of path choice assumptions are nearly unknown. There has been a recent interest in "microsimulation" for travel forecasting, which drastically reduces the level of spatial aggregation but greatly increases the amount of computation. The objective of this study is to explore the use of objective path choice data to begin to understand the differences between actual behavior and traffic assignment theory and practice.

A recent data collection effort in Lexington, Kentucky *(1)* employed the Global Positioning System (GPS) to track vehicles over an extended period of time. The data set is unique in its comprehensiveness, involving all trips for a single vehicle from 100 households and 216 drivers over a one-week period of time. More than 3000 trips are represented in this sample. These data allows analysis of actual path choices by drivers, analysis of the stability of path choice for the same driver taking the same trip at different times of day and different days of week, and comparisons of different drivers taking essentially the same trip. The data also allows comparisons of paths of trips with similar, but not identical, trip ends.

Of particular importance to this study is a set of derived data from the raw GPS data from Lexington that identifies the sequence of street segments for each trip. In network terms, each segment is a link. The Lexington network consists of about 13,000 separate street segments (or links) representing virtually every road in the metropolitan area. The raw GPS coordinates (longitude-latitude) had been matched to street segments so all the links in a path can be identified.

The Lexington data can be used to compare exact paths for sets of trips that have identical or similar ends. These sets of trips are referred to here as "matches". Given the large number of trips in the Lexington data,

there are many valid matches. Matches can then be classified as to the type of path deviations seen, compared to the theoretical shortest path, and analyzed quantitatively for the degree of deviation.

BACKGROUND

Factors in Path Choice

Most often several paths connect an origin and a destination, and the rationale for selecting a given path might be based on route attributes or personal characteristics of the traveler *(2)*. Route attributes refer to geographic network layout, road features and traffic features. Personal characteristics include the traveler's personality, trip purpose, budget and his/her background. These factors in path choice are summarized in Table 1, indicating that path choice is a very complex process.

Table 1. Path Choice Factors *(2)*

Traveler	Age, sex, life cycle, income level, education, household structure, race, profession, length of residence, number of drivers in family, number of cars in family, etc.	
Route	Road	Travel time, travel cost, speed limits, waiting time. Type of road, width, length, number of lanes, angularity, intersections, bridges, slopes, etc.
	Traffic	Traffic density, congestion, number of turns, stop signs and traffic lights, travel speed, parking, probability of accident, reliability and variability in travel time, etc.
	Environment	Aesthetics, land use along route, scenery, easy pick-up/drop-off, etc.
Trip	Trip purpose, time budget, time of the trip, mode use, number of travelers	
Circumstances	Weather conditions, day/night, accident en-route, route and traffic information, etc.	

Other studies stress the importance of network and geographical factors. Duffell and Kalombaris *(3)* conclude from their empirical study that minimum travel time is the most important criterion affecting travelers' path choice. They indicate that drivers are willing to take some extra travel distance to find the fastest route. Abdel-Aty and Jovanis's survey of path choice behavior *(4)* found the three top-ranking factors affecting travelers' path choice behavior are shortest travel time, travel time reliability and shortest travel distance. However, there are some other important factors such as number of traffic lights and stop signs and neighborhood security. Another similar motorist survey conducted in Houston suggests the top three factors affecting path choice are travel distance, travel time and congestion conditions *(5)*.

Three Paradigms of Path Choice in Traffic Assignment

The practice of travel forecasting is dominated by three paradigms of path choice: user-optimal equilibrium, stochastic multipath and microsimulation. Other path choice methods have occasionally been employed, for example it is possible to combine user-optimal equilibrium with stochastic multipath or to add better time-of-day resolution to user-optimal equilibrium. The success of the first two paradigms is due to their historical development and computational convenience, while microsimulation has not yet been proven.

The most popular of the paradigms is user-optimal equilibrium, and its path choice mechanisms are very difficult to understand, as they are not stated explicitly. The idea of user-optimal equilibrium is often attributed to Wardrop *(6)*, who postulated that an assignment is at equilibrium when no user can improve his/her travel time by unilaterally changing paths. Wardrop's principle implies that all users are assigned to a shortest path between their respective origins and destinations and that travel times and volumes are consistent with each other everywhere on the network. User-optimal equilibrium assignments can be multipath when two or more paths between an origin and destination have equal travel times -- a frequent occurrence. There have been a very large number of reports and papers published about user-optimal equilibrium. Major milestones on the development of algorithms were established by Beckmann, et al. *(7)*, who formulated the problem as a nonlinear optimization; by LeBlanc, et al. *(8)*, who provided an efficient algorithm for solving the nonlinear optimization problem; and by Powell and Sheffi *(9)*, who demonstrated

a simpler and more robust algorithm that did not depend on notions of nonlinear optimization. The consideration of traffic controls in user-optimal equilibrium is a recent innovation *(10)*.

The only restriction on path choice in a user-optimal equilibrium assignment is that a used path must also be a shortest path. Otherwise, routing can be quite indirect. The algorithm can achieve a set of shortest paths by any means from minor variations of one basic route to completely different routes. Without turn penalties, as used by some practitioners, paths can zig-zag wildly across the network. Strictly interpreted, user optimal equilibrium implies that travelers have and act upon perfect knowledge of the traffic system between their origins and destinations.

Stochastic multipath uses notions of probability theory to distribute trips across several likely paths between an origin and destination. A popular example of a stochastic multipath algorithm was developed by Dial *(11)*. Dial's algorithm finds a potentially large subset of paths between an origin and destination that do not backtrack. That is, any choice of link along a path must get the traveler farther from the origin (or closer to the destination, depending upon implementation). Paths can be of different lengths, and it is usually assumed that the shortest path would be assigned the greatest percentage of trips. Dial's path finding algorithm is attractive in that it requires only a little more computer resources than a traditional shortest path algorithm. Dial observed that the path finding algorithm could be run either from origin to destination or from destination to origin with distinctly different results. Tobin *(12)* argued that finding paths in the backward direction from destination to origin had a clearer behavioral interpretation. Running backwards the algorithm implies that travelers make choices at each intersection as to which link they select next. Travelers will consider all links that get them closer to the destination, and travelers will have a greater probability of choosing a link that results in a shorter trip to the destination.

Dial's algorithm suggests that travelers' knowledge of the traffic situation is either imperfect or that travelers are idiosyncratic. Minor path variations tend not to occur, as they are often eliminated as constituting backtracking. Dial's algorithm does not recognize delays caused by traffic control at nodes.

Microsimulation is a new technique, so generalizations about its assumptions and applications are not yet appropriate. Microsimulation has grown in interest since FHWA sponsored the development of TRANSIMS *(13)*. TRANSIMS is a Monte Carlo simulation where the decisions of every traveler are individually represented. TRANSIMS attempts to be entirely behavioral, where the model tries to make choices in the same manner as an actual traveler. Travelers within the model can respond to changes in the traffic system at each second of time. Although a microsimulation program could use a simple shortest path criterion for route choice, it could also implement much more sophisticated rules. In addition, a microsimulation like TRANSIMS should be able to eliminate most errors associated with spatial aggregation, because origins and destinations are at their nearly correct locations, not at an arbitrary center of a traffic analysis zone (TAZ).

Unlike microsimulation, traditional traffic assignment is performed between traffic analysis zones. All trips are assumed to begin and end at the centroid of a TAZ, that is at a node placed near the TAZ's center. Large TAZs can introduce substantial error in path finding because many trip ends can be seriously displaced from their actual locations. In an important study of an early travel forecasting model in Melbourne, Wildermuth *(14)* found a strong relationship between assignment error and TAZ size. However, reducing TAZ size beyond a certain point did not seem to improve the error. This TAZ size corresponded to a link-to-zone ratio of about 10. For very small TAZs assignment error was due to other factors. It is commonly argued that small displacements of trip ends do not greatly affect path choice, but this assumption needs further validation.

DATA PREPARATION

Table 2 is a summary of the portions of the Lexington data used in this research. The data comes from 2 files, street network data and personal trip data.

Table 2. Description of Data Used

Type of Data Sources	Attribute and Description		Units
Street Network Data	Record	Unique ID for each street segment (same as Maplink in another data set)	-
	Length	Distance of the segment	Mile
	Freesp_ab / Freesp_ba	Free speed on the segment for both directions	Mph
	Cap_ab / Cap_ba	Capacity of the segment on both directions	# vehicle
	Lon / Lat	Longitude and latitude information for the segment(4 numbers for the two ends)	Decimal degree
Personal Trip Data	HH_ID	Unique ID for each household	-
	DrvrNo	Unique ID for each driver in household	-
	TripNo	Unique ID for each trip recorded by GPS	-
	LinkNo	Unique ID for each link traveled during trip	-
	MapLink	Link ID from GIS map for current map link	-
	LinkFC	Functional class of link based on GIS map	-
	LinkTime	Start time of travel on current map link	00:00:00
	LinkDist	Distance traveled on link based on GIS calculations	Mile
	LinkDur	Duration traveled on link based on GIS calculation	minute

Network Creation

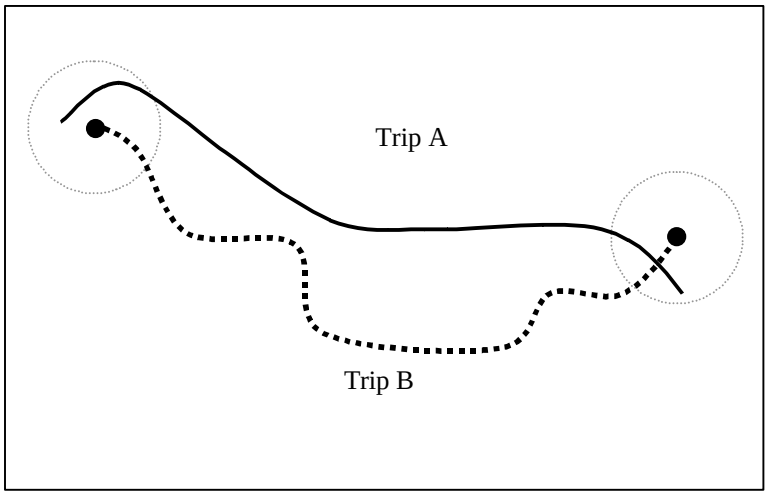
The Lexington database contained a GIS readable file that provided a visually accurate map of all the streets in the Lexington metropolitan area. For this research on path choice, it was also necessary to create a network version of this map, as this GIS data was not structured in a manner that could be easily analyzed with transportation planning models. A network was created from the list of street segments, which also contained coordinates of the end points of each segment. A node was created from each unique coordinate pair of segment ends. Then links (one link per segment) were attached to the newly created nodes. All relevant attributes from the segment file were then transferred to the links on the network. Longitude and latitude coordinates were converted to Cartesian coordinates with units of miles.

Numerous errors were encountered during the creation of the network. These errors appeared to be inherent in the coding of the street segments and are not artifacts of the method of converting the segment data to network form. The first type of error can be referred to as a looped link. Each of these links had the same beginning and ending node, and therefore had exactly the same beginning and ending coordinates in the original segment data. Looped links seemed to have been intentionally added to the GIS map to represent cul-de-sacs, cloverleaves, and other roughly circular street geometry. There were approximately 600 looped links. The second type of error can be referred to as a redundant link. These links connect to the same pair of nodes as some other link. About 1000 redundant links were found. Many of these links were probably intended to represent opposite directions on divided streets, but the reasons for the remaining redundant links remain unclear. In almost all cases, a redundant link lay almost on top of some other link in the file and those two links would be virtually indistinguishable from each other when working with raw GPS data.

Overall, about 1600 segments were dropped from the network, but none of the dropped segments had a serious impact on network connectivity. However, some of the dropped segments did cause problems when we tried to reconstruct the paths that drivers took for their trips. When the dropped segments became problematical for reconstructing a path, the path was eliminated from the analysis.

The Lexington data did not contain any information about traffic controls including cycle lengths and phasing, delays at nodes, delays to particular turning movement and street continuity. These omissions made it impossible to identify the effects of nodes on path choice.

Figure 1. Example of Two trips that Have Similar Origins and Destinations



Creating Matches

A match consists of two or more trips that have similar origins and destinations, as illustrated in Figure 1. Matches can occur entirely within a household or can occur across households. In order to avoid going back to raw coordinate data from the GPS, matches were created using the beginning and ending coordinates of the first and last segments in each trip. Unfortunately, the source data did not indicate which end of the segment best corresponded to the actual origin or destination of the trip. Consequently, both ends of a segment were used to identify matches.

Matches were created when segment ends were within an arbitrary threshold value of between zero and one-half mile. First and last segments of the trips being compared must match within this threshold, with either end of the first and last segments being allowed to match. Figure 2 illustrates four of the sixteen conditions that constitute a match.

In addition very short trips (less than 2 miles) were eliminated so as to remove those matches that might be heavily biased by large thresholds.

Sampling Matches

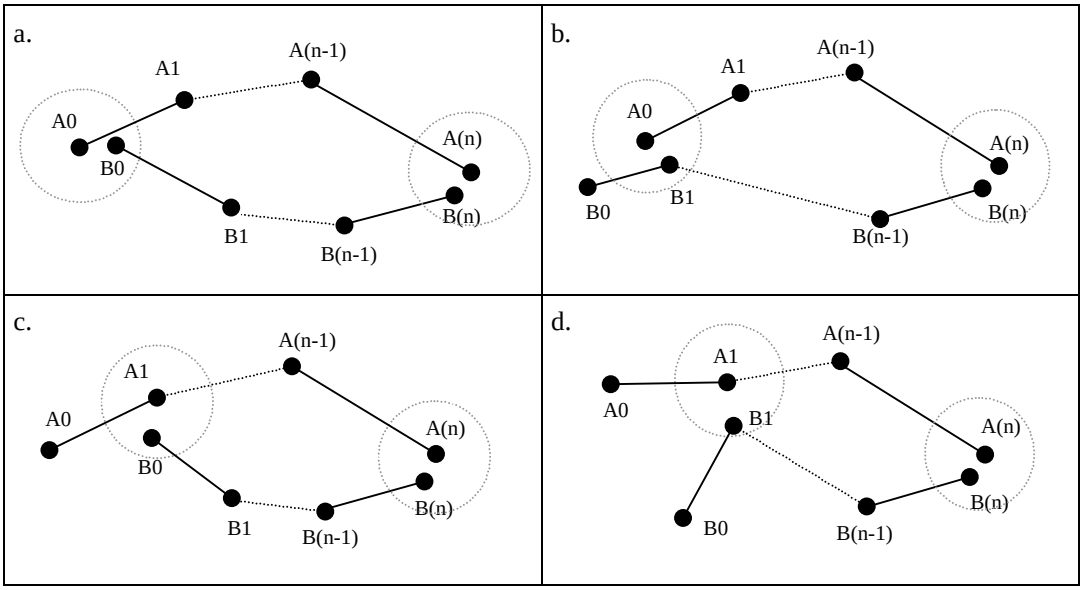
Not only were there errors due to the elimination of certain segments, but there were numerous errors in the source data due to incorrect identification of the segments in a trip. These errors were most apparent when comparing two absolutely identical trips, that is, two trips taken by the same person over the same path on successive days. In some cases, first and last segments were not identified properly; in other cases, segments were randomly skipped or added. For most trips, the errors where obvious when the trip was viewed as a whole, but it was judged extremely difficult to automate the process of cleaning the data. Consequently, much of the analysis of the matches was done by hand, thereby limiting the number of matches that could be investigated.

A large number of matches could be found, with the number of matches increasing with the size of the threshold. Beyond one extreme, there were over 400 matches of trips greater than 2 miles when the threshold was set to one mile. Of those matches that contained sufficiently clean trips, the following were randomly selected for further analysis.

31 matches within a one-quarter mile threshold and within the same household
11 matches within a one-quarter mile threshold and each trip from a separate household

10 matches within a one-half mile threshold and each trip from a separate household

Figure 2. Four of the Sixteen Conditions to Determine when Two Trips Form a Match



Travel Time Calculations

Three different travel times were calculated for each path in a match: network path time, GPS path time and shortest path time.

- Network path time is the sum of the calculated times on all links identified as being in the actual path from the trip's origin and destination. Link time is derived from the reported free speed (reduced by 15% to better approximate LOS C conditions) and the length of the link. Network path time will contain errors associated with random delays at traffic controls, errors associated with estimating the free speed, errors in adjusting free speed to actual traffic conditions and errors due to missing and superfluous segments in the path.

- GPS path time is the sum of the actual times on each link between the trip's origin and destination. Random errors in the time for each link could be substantial, but would tend to cancel over the course of a whole trip. GPS time also contains errors due to missing and superfluous segments in the path. GPS path time would properly account for delays due to traffic controls and congestion.

- Shortest path time is calculated with the same link times as network path time, but represents the shortest possible time between the origin and destination, regardless of the actual path. Shortest path time is highly artificial, but would be similar to the path time calculated as part of a traffic assignment step in a travel forecasting model. The shortest path time is not influenced by missing or superfluous links.

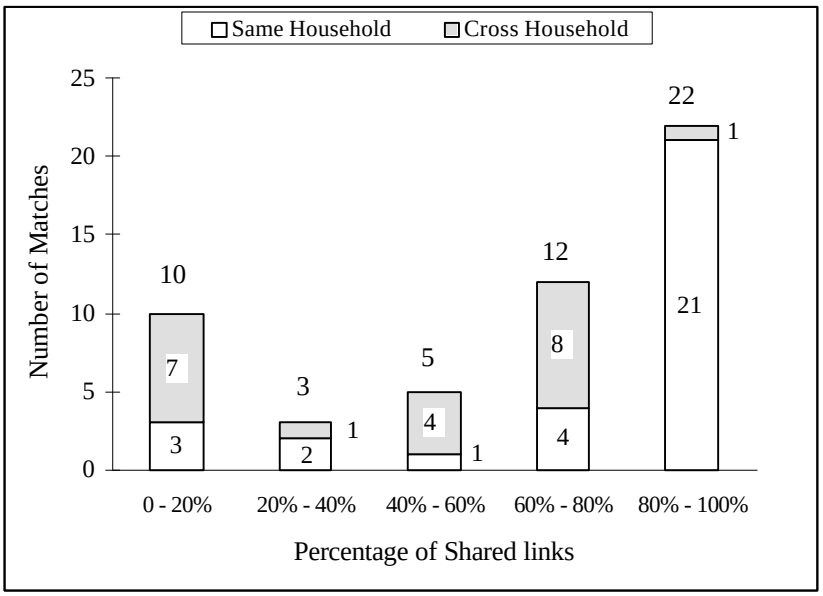
It should be noted that network path time and GPS path time are almost equally biased by missing and superfluous links, so any comparisons between these two values are still valid. Shortest path time could conceivably be longer than either network path time or GPS path time, if there are missing segments in the actual path.

MEASUREMENT OF PATH DEVIATION

Percentage of Shared Links

There are numerous paths from a trip origin to the destination. Some paths share links, others have no overlap. One obvious measure of path deviation pattern is the percentage of length of shared street links. The extent of sharing could be from zero percent to 100 percent. The histogram of Figure 3 illustrates the number of shared links by percentage range of the sampled matches. It can be seen that most pairs in matches share 60% or more links. Drivers from the same household share more links than drivers from different households for the same or similar origins and destinations.

Figure 3. Number of Matched Paths Categorized by Different Percentage of Shared Links and by Same/Cross Household



Path Deviation Index (PDI)

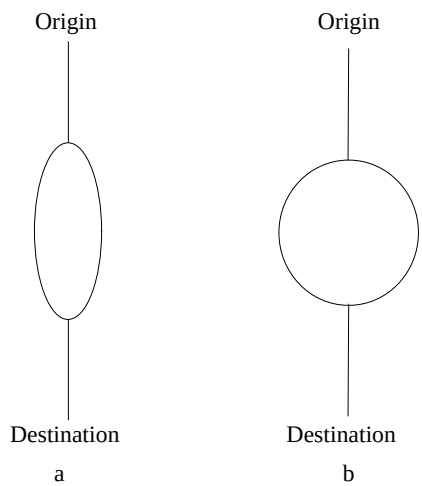
For those pairs of paths that do not share all links, some have a greater deviation than others. One shortcoming of the percentage of shared street links is that it cannot tell us the spatial extent of this deviation. For example, in Figure 4ab both pairs of paths share the same percentage of street links, but pair b shows more deviation than pair a. For pairs of trips within a match it is possible to calculate a path deviation index in order to measure the spatial separation between paths. The index of deviation is defined as the area between paths divided by the distance (in miles) along the shortest time path as shown in Equation 1. The area between paths was derived from one or more hand-drawn polygons, where the sides of the polygon were links in the network. The index of path deviation can be interpreted as the average distance in miles between the paths. So as to avoid any arbitrary selection of path length and to ensure common comparisons between paths, the shortest path is chosen for the denominator.

$$Path - Deviation\,Index = \frac{Encompassed\ Area\,(Square\,miles)}{Length\,of\ the\,shortest\ path\,(mile)} \qquad (1)$$

The minimum value of PDI is 0, but there is no theoretical maximum. A practical maximum is about $\pi r^2/2r$ (= $0.5\pi r$), where r is the radius of a circle that goes through the origin and destination or about half the length of the shortest path. It is possible to have a PDI of 0 for nonidentical paths in a match, if the links on one path are entirely a subset of links on another path. The value of the PDI depends on the length of the trips involved, sometimes making the comparison of PDI difficult for paths with different origins and

destinations. For example, the path deviation may be perceived as being greater for a three-mile PDI on a three mile path than the same PDI on a ten mile trip.

Figure 4. An Illustration of Path Deviation Patterns



Normalized PDI

To take into consideration of the relative deviation of paths a normalized PDI can be created. The PDI is normalized by the shortest path distance between an origin and destination, d.

$$Normalized\,PDI = \frac{PDI}{d} \times 100\% \qquad (2)$$

The normalized PDI is the percent of deviation between two paths over the shortest path. For example, if a normalized PDI is 30 percent, it means that the average distance between two paths is 30 percent of the length of the shortest path. With normalized PDI, all path deviation can be compared regardless of the length of the shortest path. For example, the normalized PDI could show that intercity trips may have less relative path deviation than intracity trips, because major highways or freeways are usually the only options for inter-city travel, whereas more alternative routes are available when traveling within a city.

STATISTICS OF THE PATH DEVIATION INDEX

The path deviation index can show the extent of consistency of path choices for the same driver over time, and the consistency of path choices of different drivers in the same neighborhood. Table 3 shows the calculated PDI for drivers within the same household and different drivers within ¼ and ½ miles of the same origin and destinations. This table and the subsequent taxonomy (next section) breaks the data into categories at a PDI of 0.3 miles, which is about the average dimension of a TAZ in a typical travel forecasting model in densely developed portions of a medium-sized city.

It can be seen from Table 3 that almost all of the trips made by the same driver between the same origins and destinations take the same path. This shows drivers have remarkable consistency in choosing the same path over time. For different drivers (and when the trip origins or destinations are within a close proximity), the path deviations are also small. With the small sample in this study, over half of trips take exactly the same path (with PDI equal to zero), but 40 percent take paths that are over 0.3 mile apart. As the proximity of origins or destinations become larger, the paths taken by different drivers deviate from each other more noticeably. In the small sample of this study for drivers within ½ mile threshold, 10 percent takes the same path, 80 percent takes different paths that are on average more than 0.3 miles apart.

The average PDI for the trips made by the same driver is only 0.08 mile, while the PDI for the trips made by different drivers within a quarter mile is 0.2 mile, and the PDI for the trips made by different drivers within half a mile is 0.52 mile.

Table 3. Selected Matches with Different Degrees of Path Deviation

PDI (mile)	Same Household	Different Households		Total
		Within ¼ threshold	Within ½ threshold	
Major (PDI>0.3)	2	4	8	14
Minor (PDI<0.3)	2	1	1	4
None (PDI=0)	27	6	1	34
Total	31	11	10	52

TAXONOMY OF PATH DEVIATION PATTERNS

It is helpful to refer to a taxonomy of path patterns to identify different types of path deviation patterns in terms of percentage of shared links and PDI. Based on the percentage of shared links, four types of path deviation patterns (type A, B, C, and D) can be categorized. Furthermore, based on the path deviation index, two kinds of path deviation can be subsequently categorized for types B, C and D: major deviated paths and minor deviated paths. If the PDI is less than 0.3 mile, the paths are assumed to be minor deviated paths; while if the PDI is larger than 0.3 miles, the paths are assumed to be major deviated paths.

Type A: trips that are completely overlapped with each other (with more than 95% of shared links from an origin to a destination, normalized PDI is equal to or close to 0);
Type B: trips with a large portion of overlap (with more than 50% but less than 95% of shared links),
Type B1: Paths with major deviation, i.e., PDI >= 0.3 mile;
Type B2: Paths with minor deviation, i.e., PDI < 0.3 mile;
Type C: trips with small overlap (with less than 50% of shared links but more than 5%);
Type C1: Paths with major deviation, i.e., PDI >= 0.3 mile;
Type C2: Paths with minor deviation, i.e., PDI < 0.3 mile; and
Type D: trips with essentially no overlap (with less than 5% of shared links),
Type D1: Paths with major deviation, i.e., PDI >= 0.3 mile;
Type D2: Paths with minor deviation, i.e., PDI < 0.3 mile;

The number of paths that belongs to each category is shown in Table 4. It can be seen that for drivers in the same household, most of their paths are in the B1 type. For trips made by drivers in different households, more paths are in the category of C and D, showing a larger degree of divergence.

Table 4. Number of Trips in each Type of Path Deviation

Path Deviation Type	Same Household	Different Household	Total
Type A	3	1	4
Type B1	0	2	2
Type B2	22	7	29
Type C1	1	6	7
Type C2	3	1	4
Type D1	2	4	6
Type D2	0	0	0
Total	31	21	52

A COMPARISON ANALYSIS OF ACTUAL PATHS AND THE SHORTEST PATH

Travel Time Comparison between Actual Paths and the Shortest Path

As discussed previously, there are three travel times for each path: shortest path time; network path time; and GPS path time. It can be seen from Table 5 that there is no significant difference between the calculated network path time and the actual GPS path time, indicating that link times in the network (on average) were accurate. As indicated previously, the shortest path time is not comparable.

The last column in Table 5 is the standard deviation of GPS path time calculated for the same household with the same origin and destination but traveled at different times. It shows that the variation of a GPS path time for the same trip at different times is very small. The average standard deviation of GPS path time for all households is only 1.8 minutes.

Table 5. Travel Time Comparison between Actual Paths and the Shortest Path

	Shortest Path (minutes)	**Network Path Time (Minutes)**	**GPS path time (Minutes)**	**GPS path time (standard deviation)**
Mean	10.61	11.43	11.12	1.81
Median	7.52	8.21	8.09	1.02
Standard deviation	6.18	7.42	7.67	1.83
Number of trips	31	83	83	31

Travel Path Comparison between Actual Paths and the Shortest Path

Some examples of the actual paths and the shortest path are shown in Figure 5. (Figures 5a and 5c were drawn with GIS software, while Figures 5b and 5d were drawn with a network editor.) It can be seen that shortest paths are almost always different from the actual paths. Few travelers took the same path as the shortest path; some have only minor deviation; and most travelers have major deviation from the shortest path. The available data does not allow an explanation as to why any traveler took the particular path. Data on node delay and a follow up survey would have been very helpful in understanding route choice behavior. Future data collection should make the effort to obtain such data.

CONCLUSIONS

This study shows that GPS is a viable tool to study travelers' path patterns. It can reveal important travel behavioral information that was impossible to discern with earlier conventional survey methods, such as interviews, questionnaires or driver simulators. The employment of GPS in analyzing path choices and path deviation has several advantages. First, the data is more accurate than that gathered with any other survey method. Second, the data are more convenient to collect. Finally, GPS records travel time and travel speed, which are useful in understanding travelers' route choices. However, one of the disadvantages of GPS is equipment cost. For a large-scale survey that involves thousands of GPS units, the cost could be very high. However, the cost of GPS equipment is declining rapidly.

GPS data themselves also do not provide information about the underlying reasons as to why a traveler chooses certain routes over others. A post follow-up interview is needed to gain insight of travelers' decision making process. Furthermore, a substantial effort is required to post-process the raw data to snap GPS point data to the street network. The accuracy issue in the GPS and the network data is also a concern. To achieve high accuracy, more powerful GPS units should be used, which could drive the cost even higher. In addition, better methods should be developed to integrate the GPS data with the street networks.

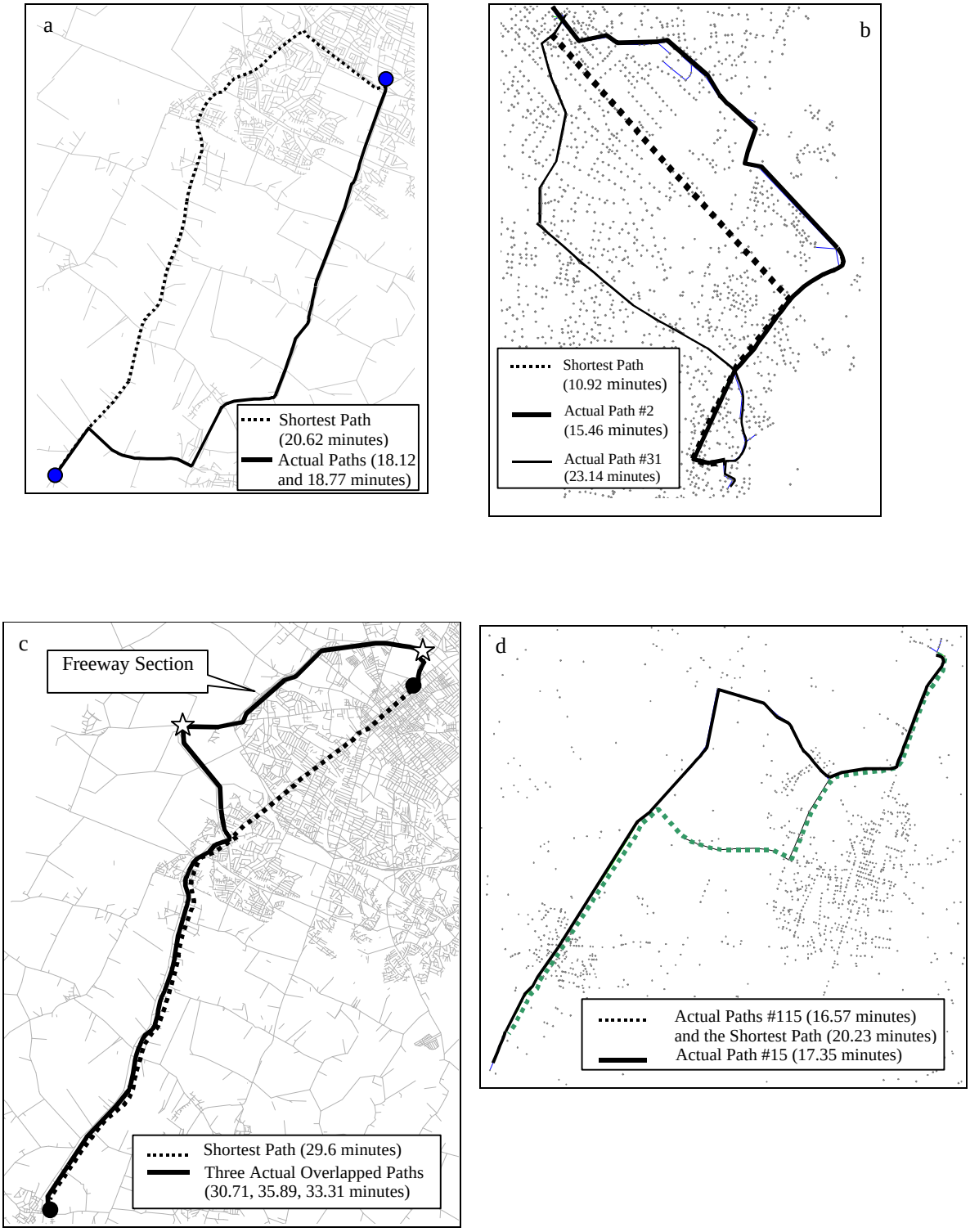


Figure 5. Examples of Actual Paths and Shortest Paths

This study found that travelers habitually follow the same path for the same trip. However, path deviation increases as origins or destinations become farther apart. For example, path deviation increases when the distance threshold for defining as the same origin or same destination increases from ¼ mile to ½ mile. This suggests that the size of a traffic analysis zone (TAZ) has an important impact on the quality of traffic assignment in transportation planning models. Further detailed study is needed to determine the optimal size of a TAZ to eliminate the error associated with displacement of origins and destinations in traffic assignments. Microsimulation or some other technique to achieve high spatial resolution may be required to get a good representation of path choice in travel forecasting models.

Another important finding of this research is that the actual travel time tracked by GPS is very close to the calculated network time based on the same path and to the shortest path time. But the actual travel path is often quite different from the shortest path. This result could have important implications to the design of path finding algorithms in travel forecasting models, especially on how delays at nodes are calculated.

REFERENCES

1. Battelle Transport Division, "Lexington Area Travel Data Collection Test", Final report prepared for the FHWA, 1997.
2. Bovy, P. and E. *Stern, Route Choice: Wayfinding in Transportation Networks*, Kluwer Academic Publishers, 1990
3. Duffell, J. and A. Kalombaris, "Empirical Studies of Car Driver Route Choice in Hertfordshire", Traffic Engineering and Control, Vol 29. Printerhall, London, 1988.
4. Abdel-Aty M. and P. Jovanis, "Using GIS Capabilities to Improve the Understanding of Route Choice Behavior", ITS, University of California, Davis, 1997.
5. Wohlschlaeger, S. "Factors Affecting Motorist Route Choice and Diversion Potential", 76th TRB Annual Meeting, 1997.
6. Wardrop, J. "Some Theoretical Aspects of Road Traffic Research", Proceedings Institution of Civil Engineers, Part II, 1, pp. 325-378, 1952.
7. Beckmann, M., C. McGuire and C.Winsten, *Studies in Economic of Transportation*, Yale University Press, 1955.
8. LeBlanc, L. J., E. K. Morlok, and W. P. Pierskalla, "An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem, *Transportation Research*, Vol. 9, No. 5, pp. 309-318, 1975.
9. W. B. Powell and Y. Sheffi, "The Convergence of Equilibrium Algorithms and Predetermined Step Sizes", *Transportation Science*, Vol. 16, 1982, pp. 45-55.
10. Claudio Meneguzzer, "Review of Models Combining Traffic Assignment and Signal Control", (ASCE) *Journal of Transportation Engineering*, March/April 1997, pp. 148-155.
11. Dial, R., "A Probabilistic Multipath Traffic Assignment Model Which Obviates Path Enumeration", *Transportation Research* 5, 1971.
12. Tobin, R., "An Extension of Dial's Algorithm Utilizing a Model of Tripmakers' Perceptions", *Transportation Research*, Vol. 11, pp. 337-342, 1977.
13. Shunk, G. "TRANSIMS Project Description", Travel Model Improvement Program, U.S Department of Transportation, 1994.
14. Wildermuth, Bruno R., D. J. Delaney, and K. E. Thompson, "Effect of Zone Size on Traffic Assignment and Trip Distribution," Highway Research Record N. 392, 1972.