Using Satellite, Aerial and Ground-Level Imagery to Check Algorithm-Generated GPS Trip Data

Resubmission Date: October 28, 2009

Word Count: 2,602

Mark Ojah
Texas Transportation Institute
Texas A&M University System
3135 TAMU
College Station, TX 77843-3135
Tel. (979) 845 9932   Fax (979) 845 7548
Email: m-ojah@tamu.edu

ABSTRACT

Household travel surveys are conducted in metropolitan areas around the world. These surveys provide data inputs for local and regional transportation planning and modeling efforts. Advances in GPS technology and mapping software have created opportunities to improve travel surveys. GPS devices provide more accurate trip information and offer insight into previously unavailable details such as route choice and precise travel speeds, distances and times. But processing vast quantities of raw GPS data into useful trip information with trip purpose involves potential pitfalls. The objective of this paper is to document simple procedures for analyzing GPS travel data using a standard GIS software package and satellite, aerial and ground-level imagery. Guidance on splitting and combining algorithm-generated trips is provided, with several examples of how to identify missing and erroneous trip ends. The intended audience includes transportation researchers, officials and practitioners interested in accurate, innovative and cost-effective GPS data processing techniques.

INTRODUCTION

Household travel surveys are conducted in metropolitan areas around the world. These surveys provide data inputs for local and regional transportation planning and modeling efforts. GPS technology has been utilized in travel surveys for more than a decade to obtain accurate trip information and insight into previously unavailable details such as route choice and precise travel speeds, distances and times. Unlike conventional travel survey information, data recorded by passive GPS loggers is not subject to the vagaries of a participant's memory or their willingness to properly document and report travel. But processing vast quantities of raw GPS data into accurate trip information with trip purpose involves potential pitfalls. The objective of this paper is to document simple procedures for analyzing GPS travel data using a standard GIS software package and satellite, aerial and ground-level imagery. Guidance on splitting and combining algorithm-generated trips is provided, with several examples of how to identify missing and erroneous trip ends.

APPLICATION OF A TRIP-END ALGORITHM

Trip-end algorithms are a key component in GPS data processing. The function of a trip-end algorithm is to automate the preliminary process of analyzing GPS data by identifying probable locations where trips begin and conclude. Automation of this task is essential given the millions of rows of positioning data that may be generated in a GPS travel survey. GPS devices configured to pause recording when a vehicle is not in motion produce data logs where the difference in consecutive time stamps varies from one second to several hours, depending on the existence and length of stops. The challenge in creating an effective trip-end algorithm is to identify as many legitimate trip ends as possible, while minimizing the number of false positives. Routine stops at locations such as intersections and differences in individuals' driving habits and patterns, complicate this task. Trip-end algorithms developed for use with wearable GPS loggers (as opposed to in-vehicle devices) require greater sophistication. They must be capable of efficiently handling larger quantities of data and distinguishing trips from a variety of modes (personal vehicle, walk, bike and transit, for example).

The tradeoff for developing a refined algorithm that identifies many probable trip ends is that it may also insert a number of invalid trip ends that then must be manually edited out. Conversely, a simple algorithm may identify too few legitimate trip ends, leaving it to the analyst (or respondent in the case of Prompted-Recall surveys) to spot and manually insert missing trips ends. The algorithm used for this paper tends toward the latter. It was developed for use with an in-vehicle GPS device and is based on a dwell-time threshold of 120 seconds. This means that whenever the vehicle velocity was zero for over 120 seconds, a preliminary trip end was automatically inserted by the algorithm. This dwell-time threshold value is frequently used in trip-end algorithms for in-vehicle GPS devices as it is longer than most traffic-signal cycle times yet short enough to capture many brief, legitimate trip ends (*1, 2, 3, 4*).

Although trip-end algorithms are invaluable in processing GPS data, they are not capable of catching all trip ends and invariably identify some false trip ends. This necessitates the use of manual procedures to check and revise algorithm-generated trips for all GPS travel data. The importance of this task is reinforced by the growing complexity of travel behavior and the rising propensity for trip chaining, which appears to have lowered trip-end dwell times for some travelers.

MANUAL DATA CHECKING PROCEDURES

Visual inspection of algorithm-generated trips can be accomplished in a variety of GIS software packages such as ESRI ArcGIS™ and Caliper® TransCAD. These GIS packages offer a framework within which street networks, GPS-collected longitude/latitude trip information, and geocoded addresses can be stored, manipulated, and analyzed. Prior to inspection of the preliminary trips in ArcGIS, a shapefile of the street network is added to provide a spatial reference for the travel data that will be overlaid on it. In ArcGIS version 9.3.1 and higher, aerial imagery layers from Bing Maps (formerly Microsoft Virtual Earth) can also be also added to facilitate inspection of suspected missing or erroneous trip ends. GPS data from individual GPS devices are then imported and displayed as a point layer. A Visual Basic macro is utilized to automate the process of importing and formatting the algorithm-generated trips. Alternatively, the preliminary trip data can be manually imported and formatted according to the Trip Number field, with contrasting colors used to differentiate the trips.

The second-by-second GPS data are displayed in the GIS as a series of points over the road network with each trip's points shown in a different color. Closely bunched points indicate travel at relatively slow speeds, while those spaced further apart indicate higher rates of speed. Algorithm-generated trip ends are denoted by a change in point color and are often characterized by small gaps in the GPS trace and diversions or loops off of roads. Trips are reviewed one at a time in chronological order by zooming in on the GPS trace to a resolution where individual data points can be visually distinguished. Each trip path is followed and stops or diversions are scrutinized as potential missed trips ends. Because the trip-end algorithm described in this paper is based on a 120-second vehicle dwell time threshold, brief trip-end stops, such as those that might occur when someone is picked up or dropped off, may be missed. Figure 1 shows several examples (circled) of valid trip ends that were identified in ArcGIS, but whose dwell times were too brief to trigger trip-end insertion by the algorithm.

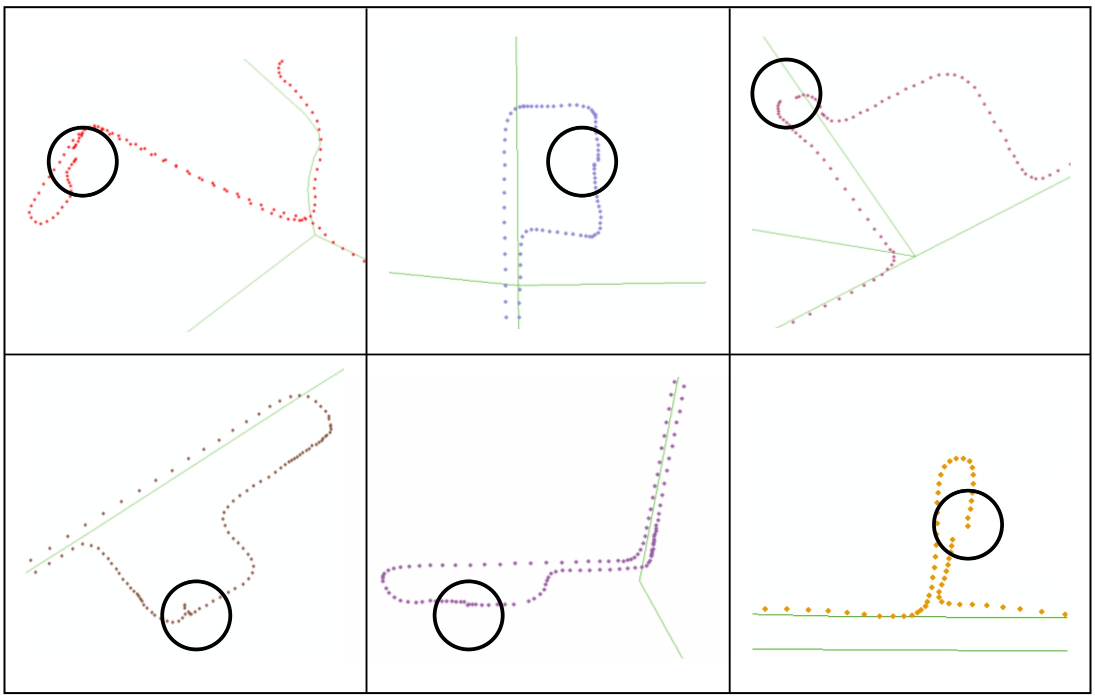

Figure 1 Missed Trip Ends Identified in a GIS.

Missed trip ends are sometimes less obvious than those depicted in Figure 1. The following process is used to identify missing trip ends that are difficult to detect. First, movement of the vehicle in the area of the suspected missed trip-end is assessed. If, based on the vehicle's path, location, speed and dwell time, a trip end is determined to have been missed, the algorithm's dwell-time threshold in the corresponding line of data in the underlying geodatabase is shortened. This manual correction procedure forces the algorithm to insert a trip end. ArcGIS versions 9.3.1 and higher support Bing Maps aerial, road and hybrid maps. This handy feature enables users to quickly view aerial imagery of the area in question when determining whether a trip end may have been missed or erroneously inserted by the algorithm. Figure 2 shows screenshots of a probable missed trip end in ArcMap - without and with the Bing Maps aerial imagery layer activated. The aerial imagery provides confirmation of the likelihood of a missed trip end and facilitates determination of the type of trip end (school).

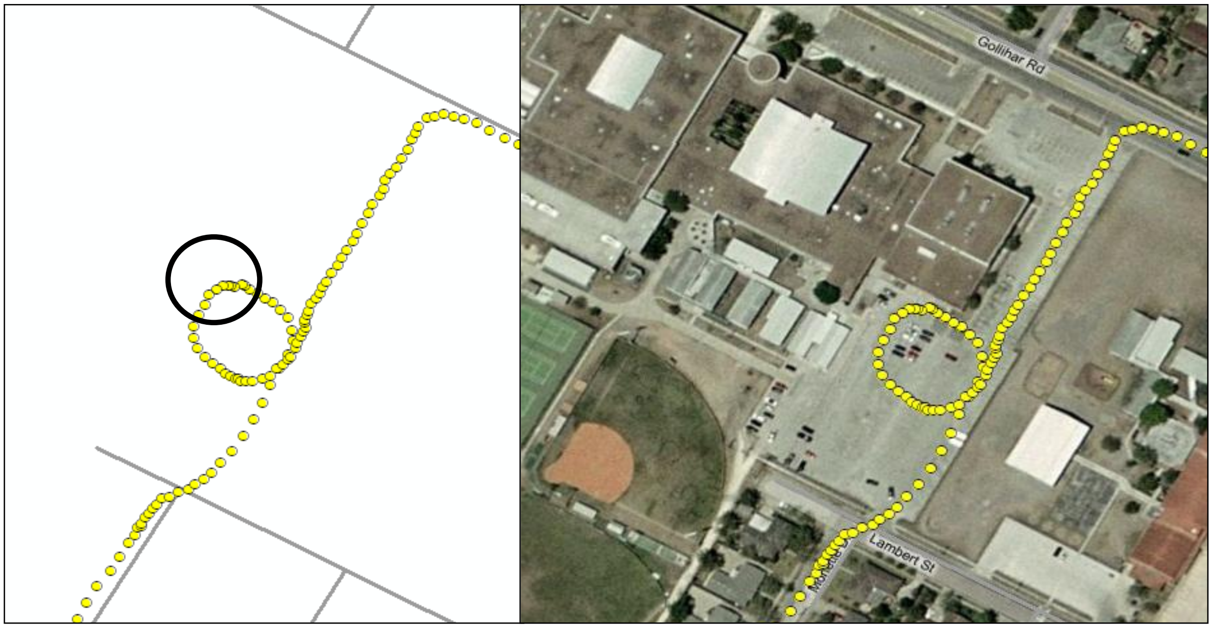

Figure 2 Use of Bing Maps Aerial Imagery in ArcGIS Version 9.3.1 for Missed Trip Confirmation.

The trip-end algorithm described in this paper results in a greater number of missed trip ends than false (erroneous) trip ends. However, in rare cases (such as left turns across busy streets or delays around at-grade rail crossings) the 120-second threshold may be too short and cause the algorithm to insert an erroneous trip end. Bing Maps imagery can be utilized to assess the feasibility/likelihood of these trip ends. The vehicle dwell time is checked using ArcMap's information tool. If the trip-end location is feasible and the dwell time is more than a few minutes, the trip end may be legitimate. Other free web-based mapping applications that offer satellite, aerial and ground-level imagery can also be leveraged to help determine trip-end validity. Coordinates for the points in question are simply pasted into these applications to view the location from the air or ground. Figure 3 shows an invalid trip end (long left-turn delay at signalized intersection) identified in a GIS and confirmed with Google Maps imagery.

Figure 3 Invalid Trip End Identified in a GIS and Verified Using Google Maps Satellite Imagery.

If the trip end is determined to be invalid, the algorithm's 120-second dwell time threshold for the relevant line of data in the underlying geodatabase is lengthened so that the trip end is removed by the algorithm and the trips are automatically combined and renumbered.

Many missed and erroneous trip ends are not readily apparent and require careful checking to be detected. A useful technique for checking GPS-derived trip data is to review algorithm-generated trips from the perspective of the path of least resistance. If, in the course of going from point A to point B on a single trip, a traveler diverges significantly from the shortest or most logical route, there is likely a missing trip end at or near the point of furthest divergence. This highlights the importance of examining GPS traces with sufficient resolution to detect subtle diversions and brief stops that may signal missed trip ends. Figure 4 shows three screenshots of a circuitous GPS trip trace between two points.

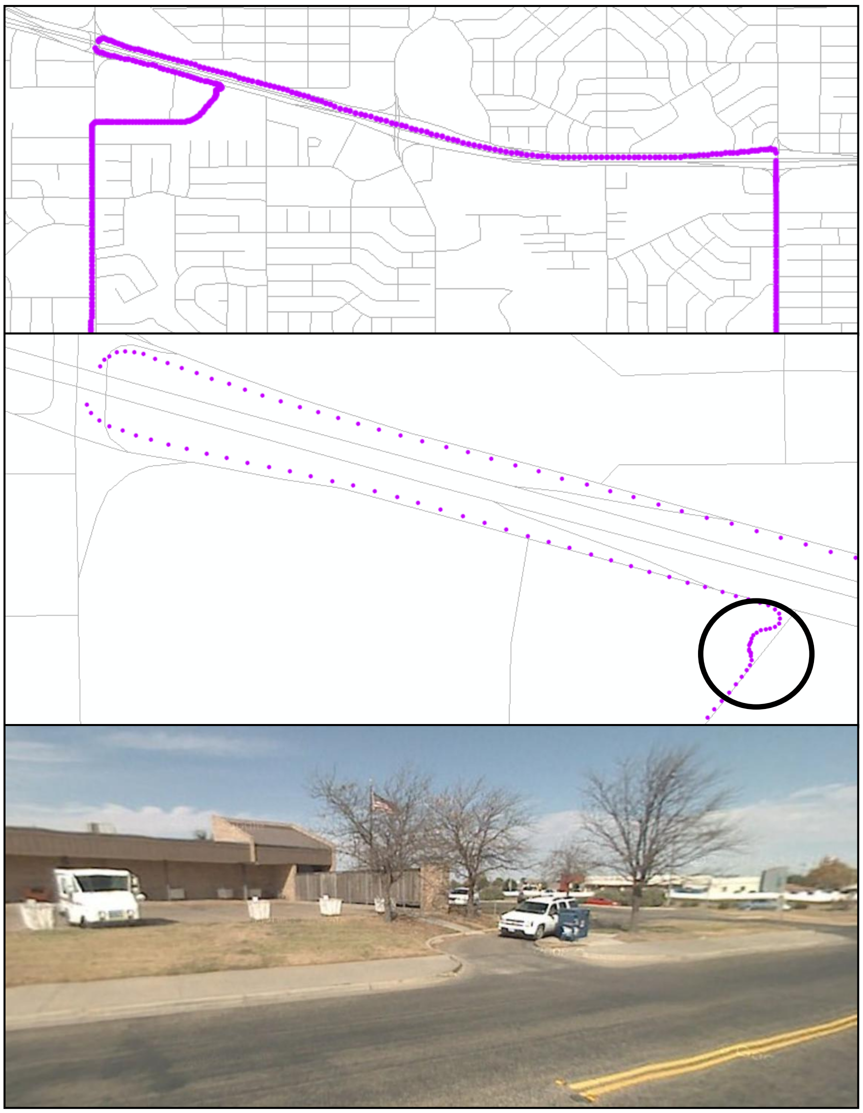

Figure 4 Missed Trip-End Identified in a GIS and Verified Using Ground-Level Imagery.

The first screenshot is a zoomed out view that does not permit individual data points to be distinguished with clarity. The second view shows the proper level of resolution for manually reviewing trip data. The satellite shadow of the freeway overpass is visible at the top left (this is not a trip end) and a slight jog in the GPS trace can be seen near the bottom right (this is the missed trip end). When the coordinates of this diversion are viewed using the Google Maps Streetview tool (ground-level imagery shown in third screenshot), the missed trip end is confirmed and trip purpose also becomes apparent (mailing letter at post office). Streetview ground-level imagery is currently available for cities in 14 countries and coverage is continually expanding (*5*).

Occasionally, trip paths that initially appear circuitous or illogical turn out to be valid upon closer inspection. Screenshots of these scenarios are shown in Figure 5. For example, a driver that wishes to merge onto a freeway may travel several blocks in the opposite direction to reach the closest on-ramp. Sometimes drivers appear to be lost, or reverse course en route without stopping (e.g. loop in second screenshot). In these instances, the analyst should resist the temptation to manually insert a trip end unless the data have been carefully examined using available imagery and a trip end is warranted.

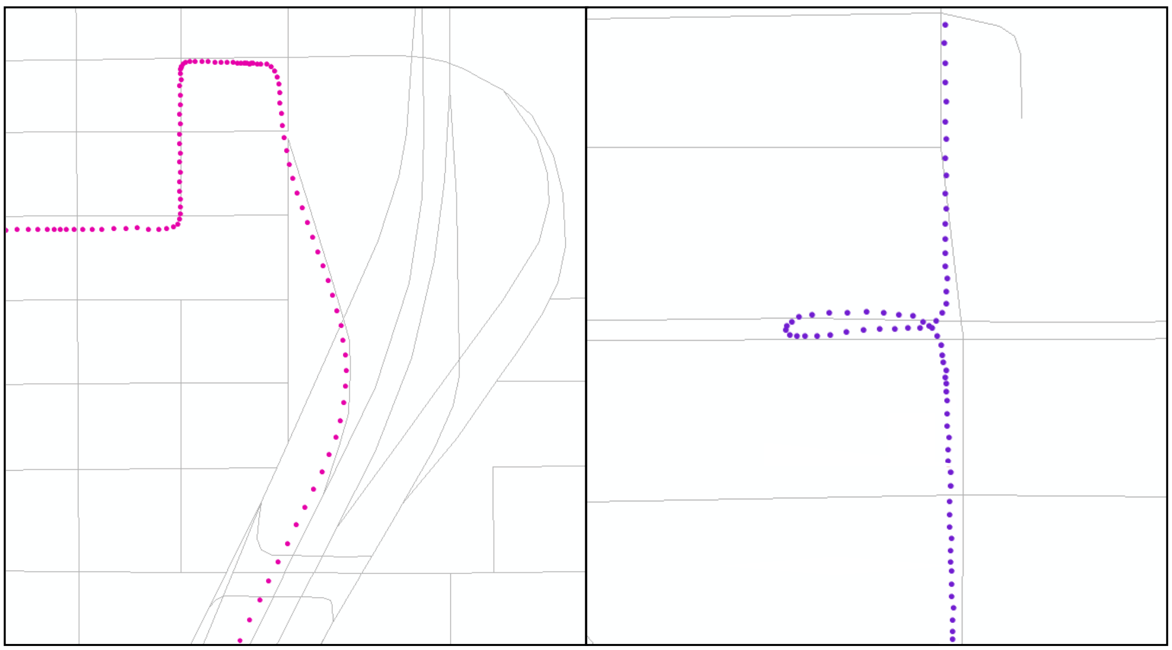

Figure 5 Misleading Trip Paths Shown in a GIS (Trips Do Not Require Splitting).

The development of new road infrastructure in an area or changes to the alignment of existing roads can cause GIS representations of the street network to become outdated. Vehicles that follow a new roadway may appear as though they diverted off of the street network displayed in a GIS. Obtaining recent street network maps and selectively using satellite, aerial and ground-level imagery can enhance the accuracy of trip information derived from GPS travel surveys. Figure 6 illustrates how incorrect manual trip-end insertion for a questionable portion of a GPS trace (actually new intersection) may be avoided by viewing recent aerial or satellite images of the location. Figure 7 shows an example of a legitimate missed trip end in a rapidly developing area. This missed trip end is confirmed using Bing Maps Aerial imagery and the extent of new development (bottom right portion of aerial image – development not shown in outdated vector map) is also apparent.

One of the hallmarks of quality GPS data is a trace characterized by trip ends that are in close proximity to subsequent trip starts. A common issue encountered when processing GPS data is signal loss resulting in data gaps or intermittent (spotty) data. Some researchers have reported development of advanced trip-end algorithms that are capable of determining whether signal loss occurred at the beginning of a trip and “repairing” the data if it did (*6*).

Figure 6 Possible Missed Trip End Revealed as New Intersection with Google Satellite Imagery.

Figure 7 Missed Trip End and Outdated Road Netwrok Shown in ArcGIS with Bing Aerial Imagery.

However, because no travel data are recorded during periods of signal loss, routes, speeds, and times cannot be known with certainty. This is particularly problematic in the case of larger data gaps, where one or more trip ends may have been missed.

The causes of GPS signal loss vary according to the device and procedures used but often include the urban canyon effect (signal blocked by adjacent buildings) or signal interference from overhead materials such as the roof of a vehicle or building, or a thick tree canopy. Selecting sensitive GPS devices with short signal acquisition times, properly installing in-vehicle equipment, and requesting that participants wait until a signal has been acquired (if device provides notification) before beginning travel, can help to mitigate this problem. Figure 8 shows an example of a sizable data gap and spotty data caused by signal loss. The dialog box in the first screenshot indicates the shortest path length of the data gap as determined with the measuring tool in ArcGIS. While small data gaps are common and can be assumed not to contain trip ends, frequent large gaps may render a GPS log unusable for travel analysis purposes.

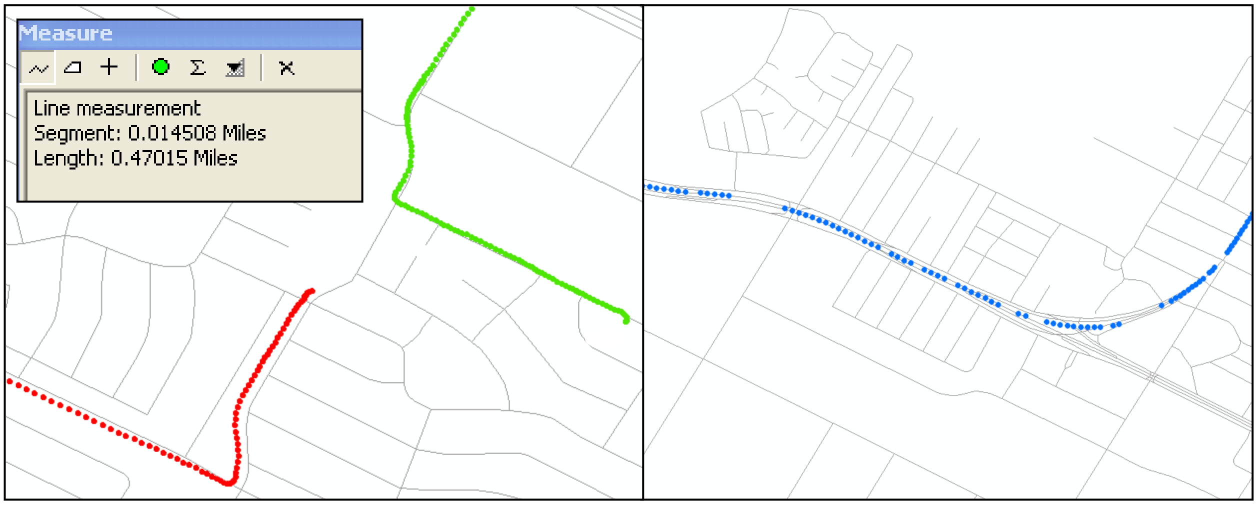

Figure 8 Data Gap and Spotty Data Shown in a GIS.

Assessing of the usability of a GPS data log with one or more significant gaps in the trace involves determining the size and time duration of each gap (shortest path distance), and the plausibility of direct travel across the gap. The functional classification of the roadway(s) in question and the existence of signalized intersections should be accounted for in this judgment. Notwithstanding the expense of conducting GPS travel surveys, GPS logs containing large and lengthy data gaps should not be utilized in trip rate calculations as there is considerable uncertainty with respect to number of trips that may have been made or missed.

CONCLUSION

Advances in GPS technology and the expanding array of tools available to users of GIS software have created new opportunities in the field of travel surveys. The integration of satellite, aerial and ground-level imagery into manual trip-end review procedures can enhance the quality of GPS-derived trip data by reducing the incidence of missed and erroneous trip ends. However, use of additional imagery and analysis tools is only recommended in cases where errors are suspected or the default street network is clearly outdated. Reliance on these tools to scrutinize all algorithm-generated trip data (whether it is questionable or not) can needlessly complicate and delay the manual trip review process.

It should also be noted that the techniques documented in this paper for identifying missed and erroneous trip ends are not intended for respondents reviewing their own trip data in a prompted recall survey. Additional trip inspection instructions may confuse these reviewers or prompt them to insert trip ends were they are not warranted. These procedures may, however, be useful for organizations that are conducting prompted-recall surveys and are interested in auditing a small percentage of participant-corrected GPS trip data with a view to determining the accuracy of that methodology and the potential need for trip-rate adjustment factors.

The trip-end algorithm documented in this paper generally results in approximately half of all vehicle logs requiring some form of manual revision (trip slitting, trip combing or both). Using the recommended data checking and correction procedures, the average processing time for a GPS vehicle log (single travel day) is 8-12 minutes. This is comparable to what has been reported by other researchers

(*6*), though actual processing times vary according to the amount of travel recorded by the device, the quality of the data, existence of missing or erroneous trip ends, and analyst experience.

ACKNOWLEDGEMENTS

This paper is based on research that was financially supported by TxDOT. The opinions expressed are solely those of the author and do not necessarily reflect the official views of TxDOT.

REFERENCES

1. Stopher, P. R., Q. Jiang and C. FitzGerald. Deducing Mode and Purpose from GPS Data. 11th TRB National Transportation Planning Applications Conference, Daytona Beach, May 2007.
2. Bricka, S. and C.R. Bhat. 2006. A Comparative Analysis of GPS-Based and Travel Survey-Based Data. Transportation Research Record, 1972, 9-20.
3. NuStats. Kansas City Regional Household Travel Survey: GPS Study Final Report. Austin, TX: Mid-America Regional Council, 2004.
4. Zmud, J. and J. Wolf. Identifying the Correlates of Trip Misreporting - Results From the California Statewide Household Travel Survey GPS Study, 10th International Conference on Travel Behavior Research, Lucerne, 2003.
5. Where is Street View available?
   http://maps.google.com/help/maps/streetview/where-is-street-view.html
   Accessed October 9, 2009.
6. Stopher, P. R., C. FitzGerald and J. Zhang. Search for a Global Positioning System Device to Measure Person Travel. Transportation Research Part C 16, 2008, 350–369.