Paper No. 01-3255

Elimination of the Travel Diary

Experiment to Derive Trip Purpose from Global Positioning System Travel Data

Jean Wolf, Randall Guensler, and William Bachman

Several recent pilot studies combined Global Positioning System (GPS) technology with travel survey data collection to evaluate opportunities for improving the quantity and accuracy of travel data. These studies used GPS to supplement traditional data elements collected in paper or electronic travel diaries. Although many traditional trip elements can be obtained from the GPS data, trip purpose has remained an important element, requiring the use of a diary to continue. Presented are the results of a proof-of-concept study conducted at the Georgia Institute of Technology that examined the feasibility of using GPS data loggers to completely replace, rather than supplement, traditional travel diaries. In this approach, all GPS data collected must be processed so that all essential trip data elements, including trip purpose, are derived. If this processing is done correctly and quickly, then the computer-assisted telephone interview retrieval call could be shortened significantly, reducing both respondent burden and telephone interview times. The study used GPS data loggers to collect travel data in personal vehicles. The GPS data were then processed within a geographic information system (GIS) to derive most of the traditional travel diary elements. These derived data were compared with data recorded on paper diaries by the survey participants and were found to match or exceed the reporting quality of the participants. Most important, this study demonstrated that it is feasible to derive trip purpose from the GPS data by using a spatially accurate and comprehensive GIS.

Transportation planners use travel demand models to estimate changes in transportation activity over time. Travel surveys, or travel diary studies, are used to collect the input and calibration data used to derive and validate travel demand models. Consequently, data collected from thousands of households across the region are analyzed to estimate current travel demand and to predict future travel demand. The accuracy and completeness of the household travel data obviously have a critical impact on model results.

Household travel data collection methods have evolved over time. Early surveys were conducted by using paper-and-pencil interview methods in the form of mail-out–mail-back surveys with in-home interviews. During the 1980s and 1990s, most travel surveys replaced the mail-back portion of data retrieval with computer-assisted telephone interviews (CATI). Most recently, computer-assisted-self-interview methods, in which respondents record their responses directly into a computer (desktop, laptop, or handheld), are being implemented. Full automation of travel survey data-collection processes should produce more data and more accurate data.

J. Wolf, GeoStats, Suite 310, 530 Means Street NW, Atlanta, GA 30318. R. Guensler, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332. W. Bachman, Center for Geographic Information Systems, Georgia Institute of Technology, Atlanta, GA 30332.

Global Positioning System (GPS) technologies, which provide second-by-second position data with accuracies of 3 to 5 m, as well as highly accurate velocity and time data, introduce a new level of comprehensiveness and accuracy to travel surveys. The potential advantages of using GPS to supplement travel survey data collection are numerous: (*a*) trip origin, destination, and route data are automatically collected without burdening the respondent for the data; (*b*) routes are recorded for all trips, allowing for the postprocessing recovery of unreported or misreported trips (including linked trips); (*c*) accurate trip start and end times are automatically determined, as are trip lengths; and (*d*) the GPS data can be used to verify self-reported data.

Although several studies have been conducted to evaluate the use of GPS with either paper travel diaries or electronic travel diaries, these studies considered GPS data only as a supplement to the traditional data elements collected in the diary itself. However, it might be possible that the travel diary could be completely eliminated from the travel survey process by using GPS data loggers and then processing the collected data in a manner that generates most, if not all, traditional trip data elements. A proof-of-concept study was conducted at the Georgia Institute of Technology (Georgia Tech) to evaluate the feasibility of this approach. The complete description of this study can be found in the dissertation by Wolf (*1*). The focus of this paper is the derivation of trip purpose, which is the one trip element that has forced some level of respondent reporting in GPS travel studies to date.

BACKGROUND

Several recent studies investigated the use of automated diaries with GPS, including the 1996 FHWA-sponsored Lexington, Kentucky study with 100 respondents (*2*) and a 1999 travel survey conducted in the Netherlands with 150 respondents (*3*). These two projects were the first to combine electronic travel diaries (ETDs) with GPS receivers to gain exact temporal and spatial details of each trip. The Austin, Texas, 1998 household travel survey included a passive GPS component to supplement its traditional survey data (*4*). Studies under way in Quebec City use passive GPS receivers combined with computerized activity scheduling surveys to record detailed spatial-temporal activity patterns and the underlying decision-making processes of individuals within a household (*5*). Researchers at Louisiana State University, Baton Rouge, have been testing a process in which GPS receivers are sent to recruited survey participants and installed in each vehicle of the household (*6*). Research conducted at Georgia Tech focused on GPS data accuracy and equipment functionality necessary to support a variety of data-collection needs, including travel surveys and vehicle emissions studies (*7–9*).

These studies fall into two general categories—electronic travel diaries with GPS and passive in-vehicle GPS systems. The primary purpose of GPS in the electronic travel diary is to augment the electronic trip data that are entered by the study participants. The intent of passive in-vehicle GPS systems is to conduct a passive audit of in-vehicle travel that can be compared in a postprocessing step to the recorded travel diary of the respondent to validate the reported data and to determine trip underreporting rates. Household surveys scheduled for 2000 and 2001 for California (statewide and regional), Ohio statewide, and the Atlanta region included GPS data collection for one or both purposes.

The analyses performed on the GPS data collected in these pilot studies centered on the opportunities offered by the data to match or exceed traditional travel survey data quality. Table 1 presents five of the studies mentioned earlier and the scope of the analysis work performed to date. For example, the Battelle report on the Lexington study includes detailed comparisons of GPS-recorded trip details with travel reported by the same participants during a recall interview conducted a short time after the study period (*2*). This same GPS data set was used by researchers at the University of South Florida to examine GPS-based data-collection methods for capturing multistop trip-chaining behavior (*10*) and by researchers at the University of Wisconsin for using GPS data to evaluate path choice assumptions in trip assignment models (*11*).

THIS STUDY

Although enhancing the accuracy and completeness of travel data, recent ETD projects with GPS have not necessarily reduced respondent burden due to additional equipment-operating responsibilities. Furthermore, passive GPS data loggers create additional data-processing work for survey researchers as they attempt to reconcile reported and recorded data streams. A research topic in the area of GPS travel survey data collection that has not been explored is the possibility of using GPS data to completely replace the travel-diary recording and retrieval process. If in-vehicle GPS data were collected and processed so that most traditional trip details could be derived, then travel diaries could be eliminated from the travel survey process. The transformed GPS data could be easily integrated into CATI operations so that household data retrieval phone calls could validate derived travel data and could collect any missing elements.

To assess the feasibility of this approach, a proof-of-concept study was designed and conducted at Georgia Tech in early 2000. Thirty survey participants were recruited to use GPS data loggers in their personal vehicles for 3-day periods. Twenty-four of the participants were also asked to keep a paper trip diary (in the form of a memory jogger) for the same survey days. The GPS data were processed and translated into trip logs; these derived trip logs were then compared with manual trip logs recorded by the same participants.

This study assessed the ability of this new process to replicate reported travel behavior. Determination of success or failure was based on comparisons made between the reported and the derived travel data. Because previous GPS travel data-collection studies investigated the use of GPS data to identify trips, along with trip start times, trip durations, and trip routes and distances, the research focused on the one trip characteristic that has not been evaluated in previous GPS travel data-collection studies—trip purpose.

Scope

Because of the types and quantities of equipment available for this data-collection effort, only vehicle-based trips were targeted for analysis; this simplification allowed for the development of three comparable in-vehicle GPS data loggers. In addition, developing a robust process for in-vehicle trips was a logical first step to developing the appropriate logic for processing GPS collected during all modes of travel. To capture all modes of travel, a personal GPS data logger would be required. To obtain a household-level solution by using GPS equipment, a household-level approach would be needed; that is, all persons or vehicles would need to be equipped with personal GPS data loggers. Finally, for vehicle-dependent households and regions, such as the case in Atlanta, a vehicle approach to household travel data collection is reasonable.

Data Element Analysis

The primary data elements collected in the travel diary component of the 1990 Atlanta Household Travel Study (*12*) were reviewed, as

TABLE 1 GPS-Based Travel Data-Collection Analyses

Study / Analyst	Data Analyses
Lexington / Battelle	1) Compared ETD & GPS-captured trips versus trips reported by CATI recall interviews for same trips; elements analyzed include individual trips, travel times, and trip lengths 2) Evaluated GPS data accuracy and completeness; map matching software was considered to identify trip links
Lexington / UWi	1) Compared derived travel paths with shortest paths
Lexington / USF	1) Evaluated use of GPS-based data collection methods in capturing multi-stop trip chains
Austin / NuStats TTI	1) Compared passive GPS-collected data with travel diary / telephone interview reported travel data
Netherlands / Ministry of Transport	1) Compared GPS performance across different travel modes 2) Assessed GPS system usage by survey respondents
Quebec City / Laval University	1) Tested feasibility of recording multi-week vehicle travel using GPS 2) Developed GIS algorithms to identify trip ends, travel times, road usage, and speed
Atlanta / Georgia Tech	1) Tested accuracy levels and performance characteristics of different GPS equipment with respect to ability to successfully map match GPS data to underlying GIS network database

NOTE: UWi = University of Wisconsin; USF = University of South Florida; TTI = Texas Transportation Institute.

were the data elements available in the standard GPS messages (*13*). GIS databases available at the Center for GIS at Georgia Tech were inventoried for their applicability to the GPS data-processing tasks required for this study. Finally, the essential trip elements were matched to their potential GPS or GIS data sources. The following list contains the preliminary matching of the key travel diary elements to the GPS elements that are most likely to provide the desired travel diary field and to the GIS components that are expected to greatly assist the travel data derivation process:

- Trip origin address: GPS latitude and longitude;
- Trip start time: GPS first second of movement;
- Trip destination address: GPS latitude and longitude;
- Trip finish time: GPS last second of movement;
- Travel distance: GPS points and GIS links;
- Trip purpose: GIS origin and destination land uses;
- Travel mode(s): vehicle only for this study; and
- Mode details: not available.

Because this study focuses on deriving traditional travel diary elements from second-by-second GPS data, the availability and use of accurate GIS databases were identified as a critical component for success. Consequently, the requirements for the GIS databases included spatial accuracy in all roadway centerlines for the study area, land use characteristics at the parcel level for the study area, and integrated aerial photography for visual examination of suspect data.

Finally, this study acknowledged beforehand that personal vehicle mode details, such as driver and passenger identification and number of occupants, cannot be derived from GPS or GIS data. Although some insight into these variables (e.g., typical vehicle-driver pairs and carpooling behaviors) could easily be gathered during the CATI recruitment call, the only methods with which to obtain the identity of the driver and the number of vehicle occupants per trip are respondent interviewing or use of video within the vehicle.

DATA COLLECTION

Thirty survey participants initially were recruited into this study; all recruits were affiliated with the Georgia Tech School of Civil and Environmental Engineering. The first six participants were used to test the GPS data logger equipment installation and use procedures. The remaining 24 participants were given both the GPS data-logging equipment and paper diaries and were asked to use both to record and capture all in-vehicle travel for the duration of their assigned survey period. The paper diaries were simple memory joggers; space was provided for the participant to record each trip and the corresponding trip details—date, start time, finish time, destination, address or cross roads, purpose, and distance traveled. The survey periods were approximately 3 days, implemented as two waves per week, in either a Friday-through-Sunday or a Tuesday-through-Friday cycle.

Three GPS data-logging packages were deployed. The packages contained different GPS and differential GPS (DGPS) receivers so that analyses could be made regarding the ability to accurately derive travel data based on varying accuracy levels available in the GPS data. Because the data collection occurred before May 1, 2000, the date on which President Clinton announced the immediate termination of selective availability (SA), two greatly different levels of position accuracy were achievable at the time of this study. (SA is the intentional degradation of GPS signal accuracy by the U.S. Department of Defense for national security purposes.) Two packages used GPS receivers only, which could capture GPS position data to within 30 to 100 m of the true position (because of the impact of SA), whereas one package contained both a GPS and a DGPS receiver to achieve a 3- to 10-m accuracy level by correcting for SA-induced and atmospheric delay errors.

These equipment packages were developed by using available, off-the-shelf hardware. The Palm IIIx personal organizer was used in each package as the GPS data logger, and consistent power supplies were used for each package—the vehicle's power supply powered the rooftop GPS equipment via a cigarette lighter adapter, and two AAA batteries powered each Palm device. A small camera bag contained the Palm IIIx and DGPS receiver (if applicable).

Data-Collection Procedures

Respondents were given simple instructions for installing and operating the GPS equipment package. A member of the research team performed the installation while the participant observed and reviewed the instructions. Once installed, the GPS receiver would attempt to acquire satellite signals as soon as the vehicle was started and would automatically stop receiving data as soon as the vehicle was turned off. Data Logger, the GPS data-logging application, was installed on each Palm device. The respondents were instructed to power on the Palm each time they started their vehicles, tap on a start icon to begin logging, and tap on a stop icon when the trip was complete.

Data Logger was programmed to log second-by-second GPS data while movement was detected. To reduce storage requirements, Data Logger was written to stop logging GPS data if 60 s of nonmovement were detected and to restart as soon as movement was detected again. To reduce processing burden on the Palm, zero or near-zero speeds were used as nonmovement indicators. The second-by-second data elements recorded by Data Logger are date, time, latitude, longitude, speed, heading, DGPS flag, number of satellites used, horizontal dilution of precision (HDOP), and DGPS age. There are also a few user-defined parameters, including nonmovement speed thresholds for GPS and DGPS modes, a countdown period before the data logging stops once a nonmovement condition is detected, and the number of hours offset between universal time and local time.

On completion of the survey period, a research team member accompanied the participants to their vehicles to remove the equipment and to collect the paper trip diary and any feedback from each participant. The equipment was returned to the research office, the data were downloaded to a PC via the Palm HotSynch cradle, the transfer was confirmed via visual inspection of the file contents, and then the data were cleared from the Palm device. Each equipment package was then reassembled and redeployed for the next survey participant.

Data-Collection Results

From March 23 through April 28, 2000, 30 survey participants collected GPS data with three types of GPS data loggers installed in their vehicles for periods of 3 days. The first six participants were used to pretest the installation and usage procedures for the equipment packages and therefore were not given paper diaries. These pretest participants are not included in the trip derivation analyses and statistics. However, their qualitative observations about equipment installation, functionality, and concerns are included in the corresponding summaries. In addition, the trips logged by these six packages were used to develop and calibrate the trip-detection macros using manual (i.e., visual inspection) processes.

Of the 24 respondents who were given both a GPS equipment package and a paper diary, 4 experienced equipment problems and therefore did not complete the data-collection effort. Three of these four failures appeared to be related to a faulty Palm IIIx that did not power down correctly. Consequently, these four participants were removed from the analysis. Finally, one respondent did not return his paper diary despite repeated reminders. As a result, 19 respondents both successfully collected travel data with the GPS data logger and returned a completed paper trip diary. These are the samples that are the basis of the GPS-data-to-trip-details analysis.

DATA PROCESSING

After the GPS data were collected and transferred successfully to a PC, the GPS comma-delimited text files were ready for processing. These GPS trip files contained the second-by-second date, time, position, speed, heading, HDOP, satellite, and DGPS information for every trip made by each survey participant—assuming that the participant started the GPS data logger for each trip and that the power and data cables were connected securely.

The four primary steps defined to transform the raw GPS second-by-second data into trip-level details were trip detection, land use and address assignment, trip purpose derivation, and travel route determination and distance calculation

The trip-detection step identified potential trip ends within the GPS data stream by searching for time periods of nonmovement. This step produced a trip-ends file that contained a trip number and the starting and ending coordinate information, including date, time, and position, for each trip detected. This file was then used to identify the nearest approximate land use codes and addresses for the trip ends (i.e., the GIS-based land use and address assignment step) and to identify the individual travel routes and distances (i.e., the travel route determination and distance calculation step). The results of the land use and address assignment step were then used to derive trip purpose. Data obtained from each process were combined with the information in the trip-ends file to completely describe the derived trip. Results of these steps were summarized at the trip level by respondent and compared to the recorded paper diary trip data.

Trip Detection

The trip-detection step identified potential trip ends within the GPS data stream by using a preset interval for "no detected vehicle movement." A TransCAD macro detected these gaps in vehicle movement, by using a minimum time threshold of 120 s. If this threshold is set too high, the macro will miss short-duration stops. If the threshold is set too low, the macro will misidentify periods of extended traffic congestion or signal-related delays as stops. Thresholds of 90 s and 60 s were also tested, but the 120-s threshold yielded the best predictions of true trip ends (*1*). The output of the macro was a trip-ends file for each participant, containing the trip number and the starting and ending coordinate information, including date, time, and position, for each trip detected.

After the trip-ends files were generated, each participant's detected trips were compared to reported trips. Results fell into four general classifications:

- Class 1: complete match. The 120-s threshold resulted in a complete match of detected versus reported trips (seven participants).
- Class 2: detection of trips not reported. The trip-detection software identified additional trips beyond those reported on the paper diary (three participants).
- Class 3: nondetection of short-duration stops. The trip-detection software did not identify passenger or item pickups or drop-offs (three participants).
- Class 4: potential problems. The trip-detection results varied greatly from the reported trips, in both overdetecting and underdetecting trips (six participants).

Table 2 summarizes the results of the trip comparisons made between the reported trips and the detected trips using the 120-s stop threshold. Notes in the far right column reflect the comparison results

TABLE 2 Differences Between Reported and Detected Stops

Study ID	Pkg	# Trips Paper	# Trips Macro	Class	Notes
11	P3	10	10	1	Complete match
13	P1	11	11	1	Complete match
15	P4	3	3	1	Complete match
16	P1	15	15	1	Complete match
19	P1	3	3	1	Complete match
21	P4	6	6	1	Complete match
25	P1	13	13	1	Complete match
9	P4	14	15	2	One extra trip detected – unclear if travel delay or actual stop
12	P4	7	8	2	One extra trip detected – missed trip
20	P3	12	13	2	One extra trip detected – missed trip
7	P1	21	19	3	Two drop off / pick ups not detected
17	P3	20	14	3	Six drop off / pick ups not detected
28	P1	21	21	3	Two drop off / pick ups not detected
					Two extra trips detected – one missed trip and one travel delay
18	P4	5	11	4	Equipment & cabling problems
23	P3	4	3	4	Procedural problems
24	P4	6	8	4	Equipment & cabling problems
27	P4	8	10	4	Equipment & cabling problems
29	P3	12	9	4	Procedural & cabling problems
30	P4	12	15	4	Equipment & cabling problems

for each participant. Because this paper is focused on the derivation of trip purpose and not trip detection, the details of the entire trip-detection process and analysis are not included here (*1*).

Land Use and Address Assignment

After the trip ends were identified within the GPS data stream, the study evaluated the use of a digital land use inventory to derive most, if not all, the individual trip purposes from the GPS trip-end data. It was hypothesized that a geographically referenced land use database was the only element needed for success; that is, a simple point-in-polygon analysis could be used to transfer land use descriptions from the polygon-based land use inventory to the point-based GPS trip ends. Trip purposes could then be estimated by evaluating the origin and destination land use pairs.

Although many urban regions maintain a good land use inventory, Atlanta is not one of them. The land use inventory for Atlanta is under development and contained several hurdles to overcome for this experiment. Primarily, the land use database is actually a series of tax assessor property databases that were designed for estimating property taxes, not for land use or activity determination. Second, the database consisted mostly of center points, not polygons. Because polygons represent property boundaries, they are much more accurate than a database of property center points alone, which requires a manual GPS point-to-property-point matching process. To overcome this hurdle, supplemental information was made available to clarify the land use at the trip ends. The supplemental data included 1993 digital orthophoto quarter quadrangles (DOQQs), Fulton County property boundaries, road centerline data, and a Haines name and address database. All these databases were combined within ArcView (GIS software by ESRI, Inc.), allowing them to be displayed simultaneously with any other geographic data (i.e., GPS coordinates). This setup allows an analyst to display, query, zoom, pan, and spatially manipulate any or all of the data at one time.

To prepare for the GIS needs of the SMARTRAQ project, researchers at Georgia Tech's Center for Geographic Information Systems have been developing a land use inventory for the Atlanta metropolitan region. The heart of this inventory is the 1999 county tax assessors' databases that were acquired from Property Data Systems, Inc. (PDS). These databases (by county) are flat files of properties with more than 100 property variables (physical dimensions, structure descriptions, land values, zoning, owner information, land use characteristics, etc.). Geographic property coordinates are not part of the database, but it does contain street addresses and tax map page coordinates (relative to unique tax map pages). Three techniques were used to generate geographic coordinates for each property. First, records with the best accuracy levels (within 10 m, for Fulton and Gwinnett Counties only) were linked to polygon-based geographically referenced databases acquired from the counties themselves. Second, tax map pages were georeferenced by using land lot maps for DeKalb and Cobb Counties only (within 50 m). Third, and least accurate (within 100 m), tax map pages were georeferenced by using the addresses of the properties. The final product is a database of property polygons when available and property center point coordinates when no polygon data were available.

The PDS property database does not contain ideal land use classifications for use in this project. Ideally, the land use inventory should be structured like the land-based classification standards (LBCS) developed by the American Planning Association. Although the Atlanta land use inventory eventually will be structured in this manner, analysis is restricted to the myriad land use (zoning-category) codes employed by each county. However, there was a business code that provided more insight into the actual land use and, when combined with the land use codes, allowed for the accurate assignment of land use in many cases. Mixed land uses had generic business codes such as strip mall, shopping center, or office building and were processed separately.

As a result of the variation in land use data sources and accuracy, as well as the current development stage of the comprehensive parcel database, the investigation of land use and street address was predominantly a manual GIS process:

1. All the trip-end records were combined into a single file that enabled the simultaneous display of all end points in the GIS. A status field was added to the composite file, and each trip end was given a value of 1, indicating that the record was not yet processed.
2. All the trip ends that fell within the boundaries of Fulton or Gwinnett County were batch processed by using a point-in-polygon analysis technique in ArcInfo. All trip ends that fell within a property's boundaries were assigned the land use and street address for that property. For these records, the status field was changed from 1 to 2, indicating that the assignment process was completed.
3. All the trip ends were then displayed in the GIS along with county boundaries and roads. The color for each trip end displayed on the screen represented the value of the status field.
4. A trip end or trip-end cluster (i.e., three or more trip ends in the same general location) with a status of 1 would be selected and examined. Note that clusters are most logically associated with common destinations, such as work or home.
5. All the layers in the GIS were then turned on to display their relative position. The investigator would first use the aerial photo (DOQQ) to determine the trip-end location and probable land use. The road database could be used to identify the closest intersection.
6. On the basis of the relative pattern of parcel points, the appropriate parcel was selected, with its land use and address displayed. If the investigator thought that the right parcel had been identified, the land use and address were transferred to the corresponding trip end(s) record within the database. The status would then be changed to 2.
7. If a trip end was processed but the investigation was inconclusive, a code of 3 was assigned to the corresponding record in the database.

Steps 4 through 7 were repeated until all trip ends in the database were evaluated.

Figure 1*a* shows a cluster of trip-destination and origin points for Participant 7 that was selected for analysis. The underlying street network is also displayed, along with the corresponding street name. Figure 1*b* shows the property center points for that area added to the GIS display. Note the offset between the page of property points and the underlying street network. The points should be shifted upward and slightly to the left. In Figure 1*c*, a visual assignment of the property-points offset identifies the most likely candidate for the parcel associated with the GPS trip-end points. Finally, in Figure 1*d*, the aerial photograph provides the complete picture of the streets, houses, and wooded areas. Because six trips originated at this location, and the land use was identified as residential, single family, this location is most likely the participant's home.

In Figure 2, the trip-end points and the GPS travel-path points are displayed. Figure 2*a* shows the travel route leading to the end of one trip for Participant 28 and the travel route leading away from that stop. In Figure 2*b*, the underlying street network is displayed. The individual parcel boundaries are included in Figure 2*c*. Figure 2*d* includes

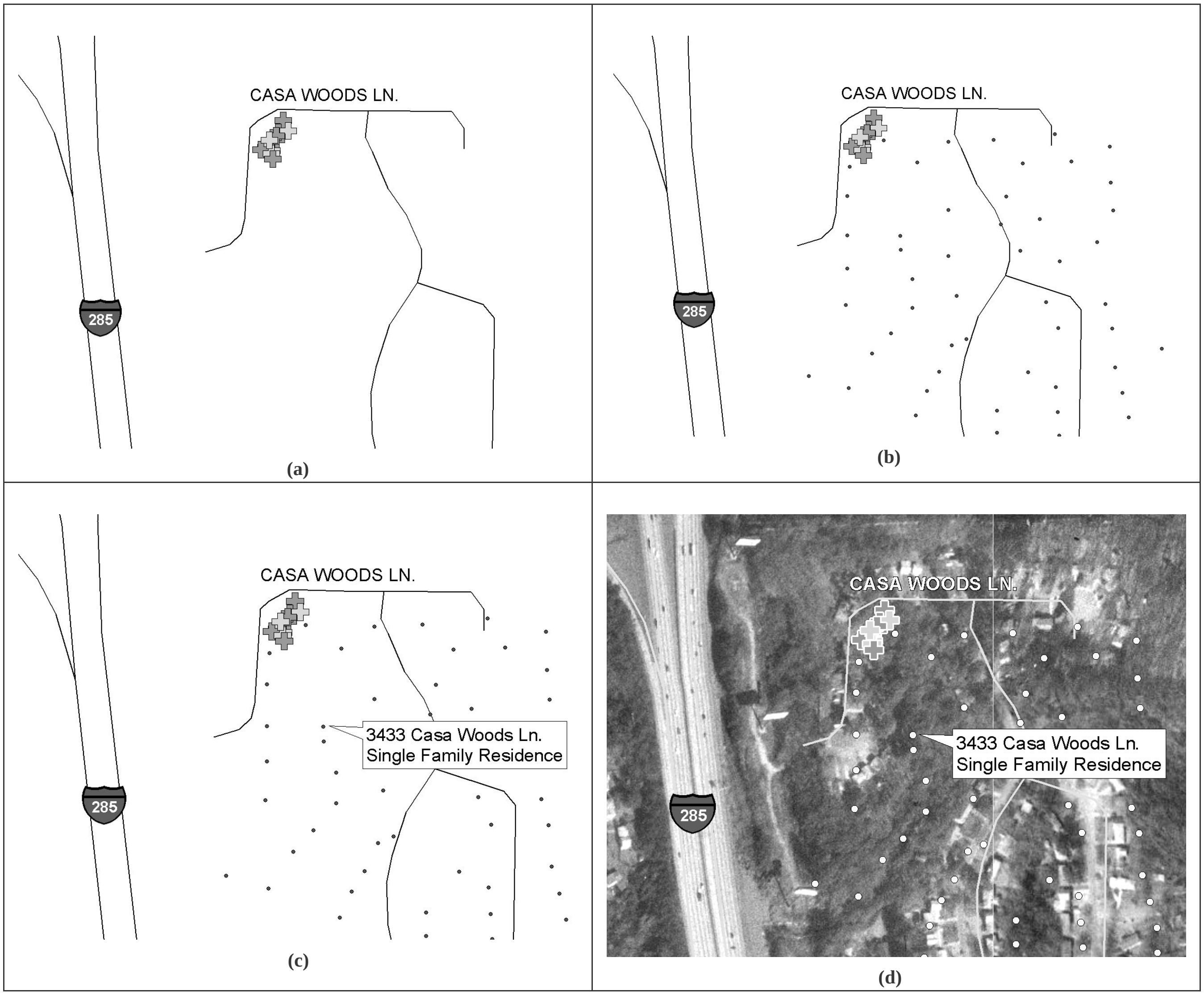

FIGURE 1 Land use determination and address assignment, home.

the property center points and the identification of the trip destination. Because the land use description for this destination is retail, multioccupancy, it is likely that this was a shopping-related trip.

This process was applied to the entire 414 trip ends obtained for the 207 trips detected by the total sample of 19 participants. However, once it was determined that six of the participants had experienced significant equipment problems or frequently had forgotten to turn on the GPS data logger, which resulted in the creation of many false, abbreviated, and omitted trips, the 56 trips detected by those participants were removed from further land use, address, and trip-purpose analysis (*1*). Consequently, only the 302 trip ends for the remaining 151 trips were used for all further analysis.

Trip-Purpose Derivation

The next step in the process was to use the derived land use and address information to determine the trip purpose. To support this task, a primary (or default) trip purpose was identified for each land use description, along with secondary and tertiary trip purposes, if applicable, by using the trip purposes from the 1990 Atlanta regional household survey. Table 3 lists the standardized land use codes and corresponding trip purposes as defined in this research effort. Two additional trip purposes, Eat and Convenience store/gas (typically coded as Other in travel diary studies), were added to provide more specific land use detail.

Specific trip purposes were necessary to accommodate both time of day and duration of stay for the same land use code. For example, if a participant went to the airport for 30 min and then traveled somewhere else in the same vehicle, the trip purpose would be Drop off/pick up. However, if the person arrived at 8:00 a.m. and departed at 5:00 p.m., this would be considered Go to work. Of course, this could have been a day trip to another location via air travel, which should be coded as Change mode. However, the exact physical location of the parked vehicle during the day can provide the necessary supporting detail to differentiate between a remote, employee parking lot and a daily traveler parking lot. The Other trip-purpose code was not used in this experiment because of the need to identify follow-up actions, but such a code could be used in an automated assignment process.

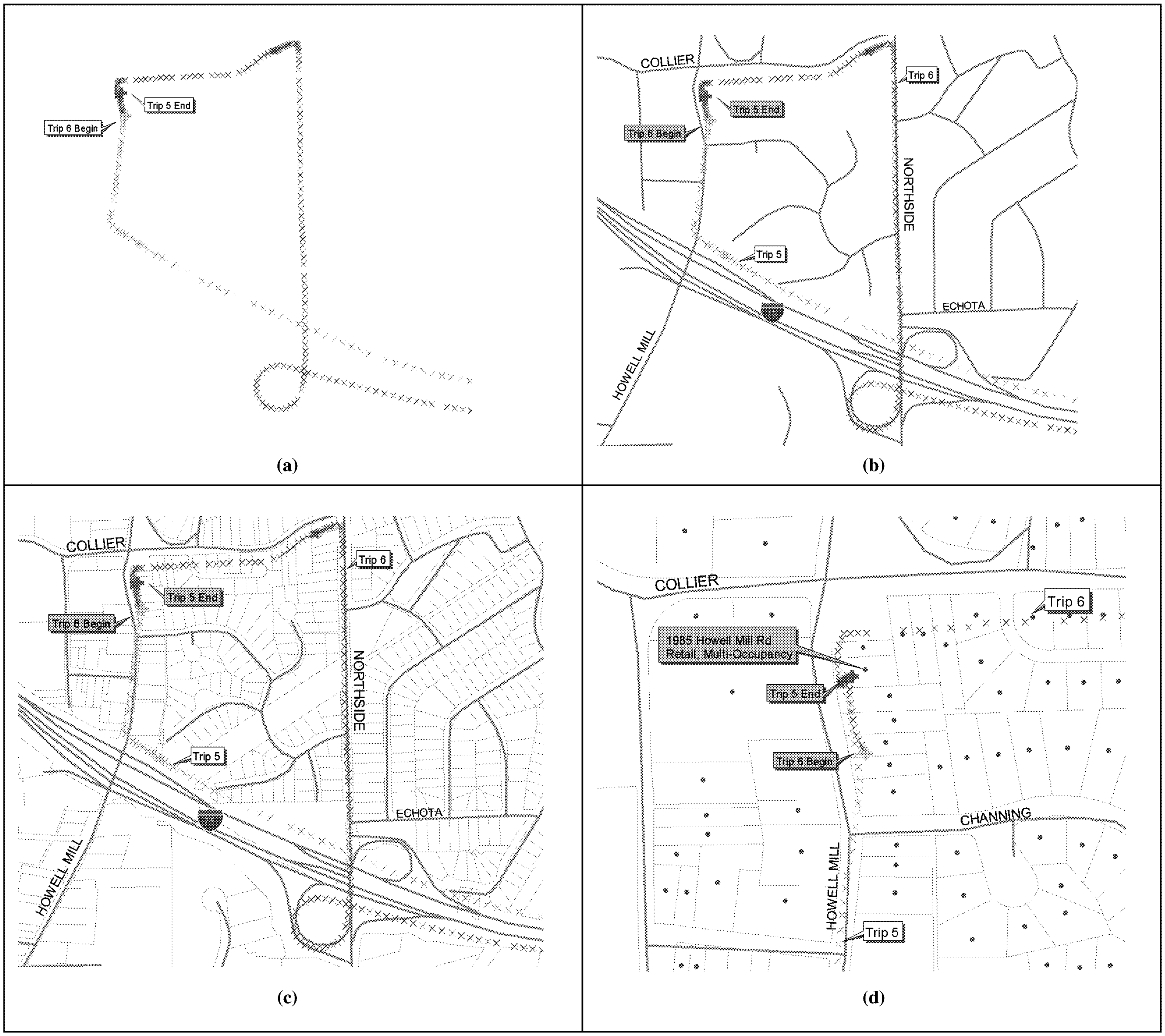

FIGURE 2 Land use determination and address assignment, retail.

A purpose of this experiment was to determine the types of trip purposes that can be assigned automatically. Examination of the land use codes revealed that some codes could not be given an automatic trip purpose. This is especially true for all mixed-use land parcels, such as shopping centers and strip malls, in which there is a variety of businesses serving a range of purposes, including shopping, eating, personal business, and refueling. In addition, there were some ambiguous land use codes, such as vacant lot, improved government exempt, and warehouse, which could indicate a property that has been miscoded or on which new development has recently occurred. Finally, some trips had a destination land use code Unknown–on road (for unknown), which indicates that the final GPS data point was located on the road network and not within property boundaries. For these trips, it is likely that either the GPS data stream terminated prematurely because of cabling problems or the vehicle parked on the street. The code 99 was assigned to all land uses that were mixed use, ambiguous, or unknown, and it was determined that these would require clarification during the CATI household data retrieval call.

After the relationships between land use and trip purpose were defined, the trips for each participant were processed to assign the appropriate trip purpose based on each trip's destination land use. The previous trip's purpose also was examined as a logic check to this assignment process. In addition, the arrival time of day at the destination and the activity duration at the destination were factored into this assignment (as discussed previously in the airport parking example).

FINDINGS

The equipment packages deployed for the pilot study proved to have many more problems than anticipated. The off-the-shelf units and cabling used for this study were not optimized for durability. Of the

TABLE 3 Standardized Land Uses and Corresponding Trip Purposes

Land Use Description	Purpose 1	Purpose 2	Purpose 3
Airport	Drop Off / Pick Up	Change Mode	Go to Work
Bank	Personal Business		
College or University	Go to Work	Go to School	Drop Off / Pick Up
Commercial	99		
Convenience Store / Gas	Conv Store / Gas		
High School	Drop Off / Pick Up	Go to School	
Improved Gov't Exempt	99		
Residence, Multifamily	Return Home	Social / Recreation	Drop Off / Pick Up
Office Building	Personal Business	Go to Work	
Shopping Mall (Deck)	Shop		
Post Office	Personal Business		
Religious, Church	Social / Recreation		
Residence, One Family	Return Home	Social / Recreation	Drop Off / Pick Up
Restaurant	Eat		
Restaurant - Fast Food	Eat		
Retail, Multiple Occupancy	99		
Retail, Single Occupancy	Shop		
Shopping Center	99		
Strip Mall	99		
Supermarket	Shop		
Unknown - on Road	99		
Vacant Exempt Land	99		
Vacant Land / Railroad	99		
Vacant Lot, Commercial	99		
Warehouse	99		

24 participants who were given both the GPS data logger and the paper diary, 5 were eliminated during the data-collection period because of Palm-related power problems or application errors. Another six participants were dropped during the trip-detection step when numerous data errors indicated problems with equipment performance, cabling connections, and user operation. Improved equipment packages are necessary for commercial deployment.

Beyond the problems experienced with the equipment, however, the process itself worked quite well. The 13 remaining participants reported 156 trips during their 3-day survey periods. The trip-detection macro identified 151 trips, including 146 of the 156 trips reported by participants, plus 3 missed trips and 2 unreported stops that are most likely traffic signal delays. Ten short-duration passenger pickup or drop-off trips went undetected, but the data characteristics of these trips were very similar and will enable the implementation of procedures in the next phase of research to automate the detection of these types of trips (*1*).

The land use and address assignment step, performed with a spatially accurate GIS that included a land use inventory, also was successful. Land uses and addresses were found for 145 trip destinations; the other 6 trips terminated on the road network and could not be assigned a land use. These six unknown destinations were the result either of premature termination of the GPS data stream because of loose cable connections or of on-street parking at the destination. Improved GPS data logger design will greatly reduce the cabling-related errors, leaving only on-street parking to need follow-up CATI clarification.

The trip-purpose derivation step also performed well with respect to determining trip purpose by GPS-derived data, specifically from the trip-destination coordinate information (including land use and address), the arrival time at the destination, and the activity duration time (i.e., the time gap between trips). Table 4 contains a summary of the trip-purpose derivation analysis for each participant, along with notes on code 99 and misidentified trip purposes. Of the 151 trips detected, Go to Work and Go Home trips accounted for approximately 54 percent of the total trips made. Of the 39 trips (26 percent) coded as 99, 26 were trips made to mixed-use shopping centers, strip malls, or commercial zoning; 6 were trips made to ambiguous land uses, such as vacant lot, government exempt, or warehouse; and 6 were trips that terminated prematurely because of equipment failure, which were coded with an Unknown–on road land use. Assuming that better cabling will be provided in a commercial deployment, only 22 percent of the total trips would require follow-up questions during the CATI retrieval call.

One significant issue found in these results is that 10 trips (7 percent) resulted in incorrect trip-purpose assignment. All these trips were misidentified as a result of inaccurate land use assignment. These land use assignment errors resulted from GPS position errors (e.g., uncorrected GPS data or premature termination of data stream), inaccurate parcel boundaries in GIS database, inaccurate assignment of parcel to the GPS trip end, or inaccurate coding of land use in the parcel database. If these error sources are not eliminated, inaccurate land use assignments will result in undetectable trip-purpose errors in the final trip-data derivation and thus threaten the accuracy of the

TABLE 4 Trip-Purpose Derivation Analysis

Study ID	Total Trips Detected	Purpose: Go to Work	Purpose: Go Home	Purpose: 99	Wrong Purpose	Notes on Trip Purpose
11	10	3	4	1	2	both errors are for the same location; WalMart coded as an office building
13	11	3	4	2	0	
15	3	1	1	1	0	
16	15	6	5	2	1	gas station coded as shopping center
19	3	0	2	1	0	missed trip -- round trip to store
21	6	2	2	1	0	
25	13	2	4	3	1	post office coded as retail, single occupancy
9	15	1	5	5	3	post office coded as shopping center; publix coded as convenience store last trip terminated on roadway
12	8	3	3	2	0	both 99's shopping centers
20	13	2	4	6	0	4 shopping centers; day care center coded as vacant land
7	19	2	6	5	1	residential code instead of store; GPS equipment / position error
17	14	3	4	4	0	two 99's coded as UNK - on road indicating cabling problem
28	21	3	6	6	2	parking garage coded as restaurant publix coded as restaurant; three 99's coded as UNK - on road; 2 are MARTA drop offs, the other most likely a traffic delay
Totals	**151**	**31** **20.53%**	**50** **33.11%**	**39** **25.83%**	**10** **6.62%**	

NOTE: UNK = unknown; MARTA = Metropolitan Atlanta Rapid Transit Authority.

results. However, the removal of SA and continued improvements to the land use database should greatly reduce these errors. The next phase of research will concentrate in part on this issue.

CONCLUSIONS

Unlike other research that examined the feasibility of using GPS data to supplement traditional or electronic travel-diary methods, this research examined the complete replacement of the paper or electronic diary with a GPS data stream. By comparing the GPS-derived travel data with reported diary data, the analyses presented in this paper confirm that the proposed process to replace trip diaries in household travel surveys with GPS data loggers is feasible. This pilot study demonstrated the capability to accurately identify trip ends within a GPS data stream, to perform land use and address look-ups with a spatially accurate and comprehensive GIS, and to assign individual trip purposes from the derived land use and address data. Whereas other GPS-related travel survey studies examined the ability to identify trip ends, travel routes, and travel distances from the GPS data stream, the research presented here is the first to demonstrate the feasibility of deriving trip purpose by combining the GPS point data with a spatially accurate GIS land use database. In addition, this study identified the symptoms of data problems related to equipment malfunction, receiver acquisition delays, and manual trip-reporting errors.

It should be noted that this study was conducted with several worst-case elements, including the degradation of GPS position accuracy caused by SA, the lack of data cleaning after data collection, and the oversimplification of trip-detection logic (based on vehicle nonmovement time gaps only). Even with these inaccuracies, the analysis performed in this study demonstrated that it is possible to replace travel diaries with GPS data loggers. Simple improvements to the process, including more-rugged GPS data logging equipment, the application of data-cleaning techniques, and enhanced trip-detection logic, should provide even better results in the next phase of this study. Also, the removal of SA on May 1, 2000 (after the study data were collected), will have a significant, positive impact on the performance of this process. GPS point accuracy improved from a range of 30–100 m to 3–20 m from May 1 to May 2. Finally, the collection of the most common trip-destination addresses (such as home, work, school, shopping, and dining) during the recruitment call could greatly enhance the automated trip-purpose assignment process.

A few concessions were made in this study, including a scope limited to in-vehicle trips, travel-mode details (such as driver identification and the number of passengers) left underived, and mixed-use land use codes requiring CATI follow-up. Although this research focused on in-vehicle trips only, the results could be extended to other modes of travel as well with the use of personal GPS data loggers. Because there are no GPS or GIS-related methods with which to determine the driver of the vehicle and the number of passengers, these details either could be collected during the CATI retrieval call or could be estimated from generalized household travel-pattern information collected during the recruitment call. Obtaining up-to-date land use codes for recently developed properties could help reduce the number of mixed-used codes within the land use database.

This study recognized that there may always be a need for certain follow-up questions regarding the derived travel data during the CATI household data retrieval call. In addition to exact driver identification and the number of passengers in the vehicle for each trip, mixed or ambiguous land uses in a region's land use inventory will require follow-up clarifications. Finally, destinations for which the driver parked in common areas, such as public parking lots or on-street parking, will require further destination information to be obtained from the survey participant. As long as the percentage of trips requiring supplemental follow-up questions is minimal, and as long as the follow-up is conducted while the participant's travel memory is fresh, the replacement of the manual diary with GPS data and a CATI follow-up survey should result in significantly improved travel behavior data quality.

Once implemented, the use of passive GPS data collection offers important benefits over traditional methods, including (*a*) a reduction in respondent burden, by eliminating the travel-diary instrument and by shortening the telephone interview length; (*b*) a related reduction in telephone interview costs; (*c*) the extension of survey periods to multiday, multiweek, or multiyear periods; (*d*) improved accuracy and completeness of existing inputs to travel-demand models (specifically for trip generation); and (*e*) the collection of new travel-data elements (routes and speeds) that enable trip assignment model validation, calibration, or update.

Finally, as research continues in the emerging field of activity-based travel analysis (which requires even more information from survey respondents), it appears that the approach presented in this paper would collect fewer, rather than more, of the activity details needed. However, by reducing respondent burden in the collection of basic trip details that can be derived from GPS data, it now may be possible to ask survey participants more questions targeting specific activity modeling input needs, such as activity constraints, opportunities, and priorities.

REFERENCES

1. Wolf, J. *Using GPS Data Loggers to Replace Travel Diaries in the Collection of Travel Data.* Ph.D. dissertation. Georgia Institute of Technology, Atlanta, 2000.
2. Wagner, D. P. *Report: Lexington Area Travel Data Collection Test; GPS for Personal Travel Surveys.* FHWA, U.S. Department of Transportation, 1997.
3. Draijer, G., N. Kalfs, and J. Perdok. Global Positioning System as Data Collection Method for Travel Research. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1719,* TRB, National Research Council, Washington, D.C., 2000, pp. 147–153.
4. Casas, J., and C. Arce. Trip Reporting in Household Travel Diaries: A Comparison to GPS-Collected Data. Presented at 78th Annual Meeting of the Transportation Research Board, Washington, D.C., 1999.
5. Doherty, S. T., N. Noël, M. L. Gosselin, C. Sirois, M. Ueno, and F. Theberge. Moving Beyond Observed Outcomes: Integrating Global Positioning Systems and Interactive Computer-Based Travel Behaviour Surveys. In *Transportation Research Circular E-C026: Personal Travel: The Long and Short of It.* TRB, National Research Council, Washington, D.C., 2001, pp. 449–466.
6. Stopher, P. R., and C. G. Wilmot. Some New Approaches to Designing Household Travel Surveys: Time Use Diaries and GPS. Presented at 79th Annual Meeting of the Transportation Research Board, Washington, D.C., 2000.
7. Wolf, J., S. Hallmark, M. Oliveira, R. Guensler, and W. Sarasua. Accuracy Issues with Route Choice Data Collection by Using GPS. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1660,* TRB, National Research Council, Washington, D.C., 1999, pp. 66–74.
8. Wolf, J., R. Guensler, S. Washington, W. Sarasua, C. Grant, S. Hallmark, M. Oliveira, M. Koutsak, R. Thittai, R. Funk, and J. Hsu. *Development of a Comprehensive Vehicle Instrumentation Package for Monitoring Individual Tripmaking Behavior—Final Report.* GTI-R-99005. Georgia Transportation Institute, Georgia Institute of Technology, Atlanta, 1999.
9. Wolf, J., R. Guensler, L. Frank, and J. Ogle. The Use of Electronic Travel Diaries and Vehicle Instrumentation Packages in the Year 2000 Atlanta Regional Household Travel Survey: Test Results, Package Configurations, and Deployment Plans. Presented at 9th International Association of Travel Behaviour Research Conference, Queensland, Australia, 2000.
10. Yalamanchili, L., R. M. Pendyala, N. Prabaharan, and P. Chakravarthy. Analysis of Global Positioning System–Based Data Collection Methods for Capturing Multistop Trip-Chaining Behavior. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1660,* TRB, National Research Council, Washington, D.C., 1999, pp. 58–65.
11. Jan, O., A. J. Horowitz, and Z.-R. Peng. Using Global Positioning System Data to Understand Variations in Path Choice. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1725,* TRB, National Research Council, Washington, D.C., 2000, pp. 37–44.
12. *1990 Household Travel Study: Final Report.* Atlanta Regional Commission, Atlanta, Ga., 1993.
13. *NMEA 0183 Standard for Interfacing Marine Electronic Devices. Version 2.30.* National Marine Electronics Association, New Bern, N.C., 1998.

Publication of this paper sponsored by Committee on Travel Survey Methods.