Accuracy Issues with Route Choice Data Collection by Using Global Positioning System

JEAN WOLF, SHAUNA HALLMARK, MARCELO OLIVEIRA, RANDALL GUENSLER, AND WAYNE SARASUA

Advancements in global positioning system (GPS) technology now make GPS route choice data collection for travel diary studies and other transportation applications a reality. Opportunities abound for increased quantities of data, for improved quality of data, and for new data elements that were once considered too burdensome or expensive to capture. For example, automated travel diaries can electronically capture trip purpose, origin and destination location names, and driver and passenger names at the push of a button. An accompanying GPS receiver can accurately capture origin and destination locations, departure and arrival times, as well as trip lengths and travel routes. This wealth of data can be used to validate or calibrate travel demand models, for in-vehicle information systems analysis, and for modeling mobile source emissions across a given network. These data collection and processing advancements do have their costs, however. In fact, care and caution should be exercised when GPS technologies are selected and used to collect route choice data. The focus of this paper is on the accuracy issues related to route choice data collection and processing using GPS technology. Vendor specifications, observation techniques, data collection procedures, data postprocessing, and the importance of using a reliable and accurate geographic information system (GIS) database are examined in detail. Critical issues in the calculation of GPS accuracy are reviewed. Finally, recent experience in Atlanta is reported, and recommendations designed to reduce the introduction of error into automated route choice data collection are provided.

As technology spreads throughout the transportation field, researchers are automating numerous manual data collection processes. These advances generally reduce labor costs and manual recording errors. Automation of survey data collection allows researchers to collect new data streams without increasing respondent burden. Automation of data processing (including data analysis) reduces data transcription errors and further reduces overall research costs. The detailed data that are becoming available through automation open new modeling and analysis realms. Areas in transportation that have been swept up in the automation wave include traditional ones such as traffic signal timing and synchronization, travel speed studies, and traffic counts. In addition, emerging areas such as intelligent transportation systems (ITS) and complex computer modeling of transportation systems and vehicle emissions, which allow for policy decision analysis, are founded in automation. The use of such models places high demand on both the quantity and quality of transportation data needed for model calibration (*1*).

Currently, one specific data collection effort under evaluation for automation is the personal travel survey, which traditionally has been administered using mail-out, mail-back travel diaries. This paper-and-pencil interview (PAPI) method typically involves hundreds or thousands of residents within a given region recording each trip made on a given day. For each trip, the respondent is asked to record detailed information, including origin and destination names and addresses, trip purpose, type of location (i.e., land use), mode of transport, driver and passenger names (if the mode is personal vehicle), trip start and finish times, and trip length. With this high respondent burden, it is common for participants to misreport or omit trips, to significantly round off trip times and lengths, and to omit origin and destination addresses. Within the past decade, telephone interviews using computer software [e.g., computer-assisted telephone interview (CATI) method] have been combined with PAPI methods in an attempt to improve data accuracy. However, these combined methods are also subject to participant recording errors and omissions and to operator data entry errors. Furthermore, CATI methods alone are subject to respondent recall problems.

Even though current data collection methods are subject to numerous errors, the survey data collected are used to forecast regional travel patterns such as the number of daily trips, trip purpose frequencies, trip origin and destination selection, trip duration distributions, and temporal trip-making distributions. In addition, travel routes are not typically collected as part of the travel survey; rather, shortest-path algorithms are used to model the flow of trips across a given network. These forecasts are then used to make high-cost transportation planning decisions.

Several recent studies are evaluating the potential of automated travel diaries—that is, use of handheld computers to administer computer-assisted self-interview (CASI) travel surveys (*2,3*). The impetus for this method is to collect more data and more accurate data with lower survey costs. Data elements such as trip purpose, origin and destination names, and driver and passenger names can all be selected from predefined lists, and trip start times and finish times can be automatically collected with the push of a button. To further assist with the accurate capture of origin and destination locations, departure and arrival times, as well as trip lengths and travel routes, a global positioning system (GPS) component has been introduced in other recent automation assessments (*4,5*).

The advantages of using GPS for travel survey data collection are numerous: (*a*) trip origin, destination, and route data are 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 well as trip lengths; and (*d*) the GPS data can be used to verify self-reported data. However, there are also a few disadvantages in GPS data col-

School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0355.

lection: (*a*) some route segments may not be captured by GPS because they are obscured by tree canopies, urban canyons, or parking garages; (*b*) system or component failure (or theft) could result in the loss of some or all travel data; and (*c*) the accuracy (and thus usefulness) of the route choice data is still dependent on many system variables.

Although GPS technology has been successfully used in many applications, care and caution should be exercised when GPS systems are selected to collect route choice data. In addition, a reliable and accurate geographic information system (GIS) database is necessary for processing GPS data and interpreting results. The focus of this paper is on accuracy issues related to route choice data collection and processing using GPS and GIS technologies. The data streams that are available through automated travel survey techniques are described, and the accuracy issues in route choice that result from using the new equipment are discussed. Observation techniques, equipment accuracy, and data postprocessing issues are examined in detail. Recent experience in Atlanta is reported, and recommendations designed to reduce the introduction of error into automated route choice data collection are provided.

ROUTE CHOICE DATA COLLECTION

Route choice data can be used to validate or calibrate travel demand models, or for both uses, and to improve the imbedded traffic assignment algorithms. These data can also be used to model route choice in other transportation applications, such as those investigating the impact of advanced traveler information systems (ATIS) or in-vehicle information systems (IVIS) on driver route choice behavior. In mobile source emissions modeling, route choice plays a significant role in allocating emissions to the appropriate road segments.

Although route choice data are useful for a variety of purposes, traditional travel diary data collection methods, including both PAPI and CATI methods, have not typically collected route data because of the additional burden. Instead, survey respondents are usually asked to record or report the street address of each trip origin and destination. Within the transportation planning process, these addresses are then associated with their corresponding traffic analysis zones, from which travel routes are generated using shortest path (or lowest cost) algorithms.

Some travel diary studies have collected route choice data. Mahmassani et al. (*6*) report success in route choice collection by implementing a two-staged survey, in which the first stage screened for respondents willing to provide additional information in the second stage. Abdel-Aty et al. (*7*) summarize numerous studies collecting route choice data for the purpose of analyzing the impact of traffic information on route choice behavior. However, all of these studies used self-reported manual procedures for capturing actual route choice.

Another method used to capture route choice decisions is stated preference surveys, in which respondents are asked how they would behave (i.e., what route would they take) under a specific set of conditions (*8,9*). These surveys can be administered using manual or automated methods. One obvious advantage of this method is that it does not depend on recall; however, it may also not reflect true route choice decisions.

Recently, a few surveys have collected route choice data using GPS technology. A 1996 Michigan study used in-vehicle GPS receivers to compare route choice behavior and perceptions of three different in-vehicle navigation-assistance systems (*10*). Also in 1996, FHWA sponsored a project in Lexington, Kentucky, in which personal digital assistants (PDAs), or handheld computers, were equipped with limited-accuracy GPS receivers to collect travel survey data (vehicle mode only) and route choice data from 100 participants for a one-week period (*11*).

In addition, several ongoing travel diary projects are introducing a GPS component for route choice data collection. In Austin, Texas, the 1998 Transportation Household Travel Survey was conducted using PAPI methods for traditional travel data collection and an in-vehicle GPS device to record travel data for those trips made in the vehicle (*12*). A study in the Netherlands has developed a portable activity diary, also equipped with a GPS device, to capture travel activity for all modes of travel (*13*). At the Georgia Institute of Technology (Georgia Tech) in Atlanta, Georgia, a comprehensive vehicle instrumentation package for monitoring individual trip-making behavior has been developed (*14*). The Georgia Tech instrumentation package includes a handheld computer for logging general trip information, a GPS device to capture all spatial and temporal coordinates, and an onboard engine-monitoring system to capture a variety of vehicle and engine parameters on a second-by-second basis. The handheld diary and instrumentation package will be used by the Atlanta Regional Commission in the Year 2000 Atlanta Regional Household Travel Study to collect data for a subset of 4,000 households.

STATUS OF GPS TECHNOLOGY

There are many GPS products on the market, and the level of technical detail can easily be overwhelming when one is deciding which receiver and antenna best meet the needs of a particular application. Before a review of the range of GPS products available, a few GPS basics will be reviewed to explain or clarify terms to follow.

Background

The GPS is a satellite-based positional system conceived by the U.S. military in the 1970s. The system was implemented in phases in the 1980s and is now at full capability with a complement of 24 satellites. Applications of the system today include military use and civilian applications that require positional information (*15*). The GPS computes ground position by first measuring the signal travel times between a group of satellites and a ground-based receiver. Since radio signals travel at the speed of light, these travel times can be used to calculate the distances to the satellites. Finally, the position of the ground receiver and antenna can be calculated using triangulation.

Calculation of the antenna's position is accomplished using triangulation to solve four unknowns: the x-, y-, and z-coordinates and the difference between the satellites' clocks and the receiver's internal clock. A sphere for each satellite describes the calculated distance from the satellite to the receiver. That is, the receiver must lie somewhere on the surface of the sphere. Three satellites in view provide three spheres. The intersection of these three spheres yields two points, one on the surface of the earth and another one in outer space (which is automatically discarded). For the determination and elimination of any clock drift, four satellites must be in view to compute a three-dimensional position. A two-dimensional (x, y) position can be calculated by using the third satellite to resolve the clock differential rather than for determining the z-coordinate.

GPS data are subject to a number of sources of error, including satellite orbit errors, satellite clock errors, receiver errors, atmospheric and ionospheric errors, multipath errors, and selective availability (SA), which is the intentional introduction of satellite ephemeris errors by the U.S. Department of Defense. SA degrades the precision of GPS accuracy up to 100 m for real-time, nonmilitary users. SA is by far the biggest source of errors in GPS positioning (*16*). Some GPS receivers support differential correction, a feature that eliminates SA-introduced errors in either a real-time or postprocessing mode. Uncorrected GPS data are referred to as autonomous or raw data.

Differential correction is the technique of reducing GPS errors by collecting data with two units simultaneously. A base station receiver and antenna are placed at a known position. The satellite data collected at this known position allow the computation of corrections for the GPS signals. These corrections are then applied to the data collected by the other receiver, eliminating the effects of SA and other global errors. Differential corrections can either be sent by the base station to the receiver in real time through a radio link [real-time differentially corrected GPS (rtDGPS)] or be stored on the base station and applied later to the receiver's data using a process known as postprocessed differentially corrected GPS (ppDGPS).

Available GPS Technology

There are hundreds of commercially available GPS units on the market. Units range in price from a few hundred dollars for low-accuracy recreational units to tens of thousands of dollars for surveying-quality receivers. GPS receivers are designed for numerous uses, including marine, aircraft, and land use, and can come in a variety of forms, including Personal Computer Memory Card International Association (PCMCIA) cards, Oracle Enterprise Manager (OEM) boards, handheld computers, and "black-box" sensors. Antennas also come with a variety of features and configurations.

The many options can make it difficult to select a GPS receiver and antenna that best meet application needs while optimizing cost, accuracy, and other desired features. However, accuracy ranges tend to fall into three distinct groups: low-end receivers with 100-m "plus" accuracy levels (available for $200 to $500), mid-range receivers with 1- to 10-m-level accuracy (in the $2,000 to $5,000 price range), and high-end receivers with centimeter- or millimeter-level accuracy (often with prices exceeding $10,000). Table 1 gives a few of the GPS products initially considered by Georgia Tech from each group; manufacturer or vendor accuracy specifications are provided with and without differential correction.

GPS ISSUES WITH ROUTE CHOICE DATA ACCURACY AND LOSS

There are many areas for concern in the evaluation of GPS receiver accuracy. First, published accuracy specifications by vendors of GPS equipment are usually developed under optimal conditions. Field accuracy can vary significantly. Next, the survey methodology (i.e., observation technique) greatly affects achievable accuracy levels. Knowledge of these factors is essential when GPS equipment is used to collect positional data. Real-time data collection factors—such as the number of available satellites and their orientation (geometry), the impact of line-of-sight obstacles (such as tree canopies and tall buildings), and the loss of satellite lock on the receiver's internal algorithms for computing the location—have a major impact on data completeness and positional accuracy. Once the data are collected, postprocessing routines can improve the accuracy of the data but can also significantly reduce the amount of usable data. Finally, once the data have been processed and are ready for evaluation, most typically within a GIS, there are several methods used to calculate data accuracy, each with its own set of issues.

Vendor Specifications

As stated previously, vendors publish accuracy specifications of GPS receivers under optimal conditions rarely encountered in the real world. In addition, these conditions are seldom stated with the specifications. For example, one common practice of vendors is to report static GPS accuracy. Since this method involves capturing data at a fixed position over several minutes, the resulting average of point values can significantly improve the accuracy level. Although this specification is suitable for surveying applications, it is not very appropriate for route choice data collection.

Some vendors publish GPS accuracies that can only be achieved if selective availability is deactivated. In many cases, accuracy

TABLE 1 Representative GPS Products

Group	Vendor / Product	Price[a]	Antenna	Accuracy			
				Autonomous		Differential	
				w/o SA	w SA	rtDGPS	ppDGPS
Low	Garmin 35 LP TrackPack	$250	Passive	15 m	100 m	3-10 m	na
	Garmin GPS II Plus	$250	Passive	15 m	100 m	1-5 m	na
Mid	GeoResearch Workhorse	$1,795	Active	25 m	100 m	1-5 m	< 1 m
	Garmin Survey II	$3,060	Active	15 m	100 m	3-10 m	1-5 m
High	Ashtech GG24 (G/GPS)	$9,340	Active	10 m	16 m	* 1 m	* 0.5 cm
	3S GNSS-300 (G/GPS)	$8,900	Active	10 m	10 m	* 3-10 m	* 1-5 m
	Ashtech Z12 (GPS only)	$22,000	Active	15 m	100 m	< 1 m	0.5 cm

rtDGPS = real-time differentially corrected GPS
ppDGPS = postprocessed differentially corrected GPS
na = This feature is not available with this product.
[a]Prices include receiver and antenna, without accessories or discount.
* GLONASS data do not need correction; these values are for the GPS differential correction.

under selective availability conditions is provided as a footnote. This accuracy is somewhat deceiving because even surveying-quality GPS cannot achieve much better than 100-m-level accuracy if used autonomously. Furthermore, the orientation of satellites in the sky can significantly affect accuracy. Vendor specifications are usually based on a desirable number and geometric distribution of in-view satellites. In route choice data capture applications, this is not usually the case because of satellite line-of-sight obstructions such as tree canopies and urban canyons.

Observation Technique and Positional Correction

Achievable accuracy levels are different depending on the field survey and data processing techniques employed. The two most common survey techniques are static and kinematic. Static surveying involves the occupation of a point for a predetermined period of time. During this fixed time, data are collected by the receiver and processed later. This procedure results in better positional estimates and is quite suitable for fixed-position surveying applications.

Kinematic surveys, on the other hand, involve a moving receiver; this movement eliminates the possibility of averaging as used in static surveys. Only one observation is used to compute the position for each data point, resulting in lower accuracy levels. In transportation, this observation technique is most applicable to moving vehicles or persons carrying a GPS, which is exactly the case for route choice capture, automatic vehicle location, and road centerline determination.

Both static and kinematic GPS surveying can benefit from the use of differential correction. Postprocessed GPS data usually have higher accuracy results than rtDGPS data because with rtDGPS there is always some time delay in the process of computing the corrections, transmitting them to the other receiver, and receiving and applying them to obtain the coordinates. This delay does not exist in ppDGPS data because all the data files are placed on the processing computer.

The Russian satellite system, known as GLONASS, is designed to provide the same geographic coverage as the GPS satellites, but without selective availability. Therefore, if a complete GLONASS 24-satellite configuration were available for a GLONASS-capable receiver, differential correction would not be necessary for most applications. Several U.S. GPS vendors are offering a combined GPS-GLONASS receiver that uses both the GPS and GLONASS satellites in point calculations.

Real-Time Route Choice Data Collection

Several sources of accuracy error and data loss are present during the actual GPS data collection effort. As mentioned previously, the spatial configuration of GPS satellites, along with the number of satellites continuously in view and available to the receiver, has a major impact on data accuracy and completeness. These factors are more important in kinematic surveying since positional estimates are based on one-epoch (usually 1-s) intervals.

The quality of the satellite geometry used by the GPS receiver to compute a coordinate is expressed by the position dilution of precision (PDOP). The geometry of the satellites can be interpreted as the volume of the angular cone formed by the lines that go from the receiver's antenna to the satellites and the plane formed by the satellites. The more tightly clustered together the satellites, the higher the PDOP value. As volume increases within this solid, geometry is improved and the PDOP value decreases; good satellite geometry is characterized by a PDOP value lower than 4 (*15*).

The ability of the receiver to continuously track a minimum of four satellites is critical in kinematic surveys. Loss of lock occurs when the number of satellites drops below four and is usually the result of signal obstruction. GPS reacquisition of satellite lock typically takes place within 2 to 3 s after the line of sight is restored. Although loss of lock does not significantly affect autonomous GPS data, it can be very problematic for differential GPS. Postprocessing engines need a fair amount of continuous good-quality data, either immediately before or after the loss of lock, to resolve temporal ambiguities and to obtain a solution that can be propagated across the missing points at high levels of accuracy. High-end units such as the Ashtech Z12 require at least 2 min of good data between losses of lock; otherwise, the data collected between the signal losses cannot be processed and recovered.

Frequent losses of lock, such as those resulting from driving through intermittent tree canopies, may cause a total loss of positional information for the segment if the real-time processor or postprocessor cannot resolve the ambiguities. In downtown areas, the problem is even more serious because the sky view available to the GPS antenna is reduced. In some urban canyons, there may be locations where no satellites are visible.

Multipath errors are also problematic during route choice data collection. These errors are caused by the reflection of GPS satellite signals off buildings, walls, and other surfaces before signal arrival at the GPS antenna. The reflected signals confuse the receiver with conflicting data. Improper positioning of the receiver's antenna can further worsen multipath error. For example, if the antenna is placed near the edge of a vehicle's roof or trunk, the possibility is greatly increased that signals reflected by the pavement or nearby buildings reach the antenna along with signals that come straight from the satellites. To reduce the multipath error, Motorola Corporation advises antenna placement as far as possible from the edges of the surface on which it is mounted (*17*).

Calculation of Route Accuracy

Once GPS data have been collected, a baseline or benchmark route must be created against which the accuracy of collected points can be measured. Three common methods exist to create this benchmark route. The first option is to survey the entire test route for positional location using either conventional or static GPS survey. Although this method produces the most accurate benchmark, this type of reference is costly for a long test route and is logistically infeasible along real roadway segments. A second method is to drive the test route using a GPS receiver that offers high accuracy under kinematic conditions. Unfortunately, even high-cost equipment may not perform well under conditions that cause frequent losses of lock, such as when there are intermittent tree canopies. The third and most commonly used method is to create the benchmark route on a street network base map in a GIS. This method is by far the simplest and most cost-effective if a GIS system and base map are readily available.

Various studies have used routes within street networks as the reference for calculating spatial data accuracy. Common measures of accuracy include the number of points or the percentage of total points collected for the route or route segment that fall within specified distances from the benchmark route (*18,19*). Several issues exist that should be considered when a street network is used for accuracy calculations. First, the accuracy of the base map is crucial. Digital base maps are created using a variety of methods, including digitizing existing paper maps or aerial photographs. A significant number of street databases are based on original Census Bureau topographically integrated geographic encoding and referencing (TIGER) files, which were created for census-taking purposes. Although inexpensive and widely used, the main flaw of the TIGER database is spatial accuracy: 30- to 50-m errors in street location are common within these files. For a given metropolitan area, a number of street network databases are publicly or commercially available with varying levels of accuracy and completeness. Each database may have its own set of accuracy errors resulting from a number of sources, such as inaccurate digitizing of the network, map projection errors, and discrepancies in representation of links between end points.

A second issue encountered when street network base map routes are used is the location of the GPS data points relative to the baseline route on the digital map. Most digital street networks represent streets as a line, usually the centerline. However, streets have widths, an attribute that is not typically represented in a digital database. Consequently, even with accurate GPS positioning, data points will almost always appear offset some distance from the roadway centerline. For example, positional data collected for a vehicle traveling in the far right lane of a six-lane urban arterial roadway with 3.6-m lanes and a center two-way left-turn lane would automatically be offset from the centerline by 11 m, even with a spatially accurate street network and centimeter-level GPS accuracy. Consequently, none of the data points would fall within 10 m of the route. Errors can easily be compounded if the street network has spatial inaccuracies as well.

RECOMMENDATIONS

Component testing was completed for the Georgia Tech project in September 1998. The final project report was released in March 1999 (*14*). As a result of the GPS component tests, the research team offers the following recommendations to other researchers who plan to evaluate or implement GPS and GIS technology for the collection and processing of route choice data.

Step 1: Establish GPS Route Choice Data Accuracy Criteria

The first step is identification of the level of accuracy appropriate to the application. Different applications require different levels of accuracy and generally no "one size fits all" GPS package exists. For construction of a transportation road improvement, survey points must be accurately located, mandating centimeter or millimeter accuracy. In contrast, travel speed studies do not require the exact location of the vehicle along a roadway segment and location errors up to 100 m may be allowed (since speeds are calculated using internal Doppler calculations rather than second-by-second changes in individual x, y-coordinates).

Electronic travel diary studies require a midrange accuracy level, with acceptable errors up to 10 m. The primary need for GPS data accuracy in travel diary studies is to enable the accurate and efficient assignment of roadway segments within the GIS database for route choice recovery. Since most roadways are separated by more than 10 m, GPS points falling within 10 m of a segment can usually be assigned. With lower accuracy levels, however, difficulties may arise in matching the points to the correct roadway amid closely spaced segments. With equipment accuracy of 30 m, for example, it would be difficult to assign points falling between roadways spaced at 60 m to one or the other. Frontage roads can be especially difficult to match correctly.

The consequences of inaccuracies in an electronic travel diary study depend on the end use of the data. Electronic travel diary information can be used for a variety of applications, including measures of actual versus reported trip times and lengths, travel speeds, origin and destination studies, and route choice modeling and validation. A minute or so of inaccurate trip locations may be insignificant. However, with route choice modeling, trips misallocated from major arterials to local roads could significantly affect enhanced travel demand modeling. With the use of travel diary data for emissions modeling, as is being evaluated at Georgia Tech, misallocation of GPS points to roadway segments may cause the misallocation of emissions within the network.

Step 2: Obtain Understanding of All Sources of Error and Select Appropriate GPS Equipment

It is not an easy task to identify the correct GPS equipment for a given route choice data collection application. At each step of the process, there are numerous sources of errors, each of which can degrade the quality and completeness of the results. Every attempt should be made to minimize these errors through proper experimental design and data collection planning efforts. The vendor should be questioned regarding accuracy specifications and conditions. In the consideration of other GPS performance studies, the test route and conditions should be examined closely. The accuracy of the GIS database to be used for route choice data processing should also be assessed before the actual purchase of GPS equipment. From a practical standpoint, error avoidance is the key. Postprocessing route data by hand to eliminate detected errors is tremendously time-consuming.

To achieve the targeted accuracy level of 10 m, the Georgia Tech research team identified three possible GPS options: real-time kinematic GPS, postprocessing kinematic GPS, or the use of combined GPS-GLONASS receivers. The use of autonomous GPS receivers was ruled out because of poor accuracy levels when subject to SA. In addition, real-time differential correction was eliminated from further evaluation because of limited availability of radio frequencies in the Atlanta metropolitan area at the time.

The GPS-only choice (with postprocessing differential correction) has the advantage of lower capital costs but involves significant, if not prohibitive, data storage, postprocessing labor, and data processing requirements. On the other hand, Georgia Tech already has a base station set up and running, readily available for use in the project. The combined GPS-GLONASS option is also appealing since the collected data do not require differential correction to achieve the required accuracy. Preliminary research at Georgia Tech with these receivers, however, indicates that these advantages will only be realized when the full GLONASS satellite system is in place (research has revealed that fewer than 18 GLONASS satellites have been in orbit for the past 2 years).

Step 3: Create Representative Test Route and Obtain Best Benchmark Data Feasible

A test route was designed by the Georgia Tech research team to include representative route conditions of the Atlanta metropolitan region; consequently, this route includes the following roadway segments: urban arterials and freeways with clear sky views and with overpasses; arterials and collectors with heavy, intermittent tree canopies; and arterials through downtown with high-rise buildings (urban canyons). The Atlanta route is presented in Figure 1.

The research team chose to use a high-accuracy GPS receiver with ppDGPS to create the benchmark route against which accuracies of test GPS equipment could be measured. However, because of the numerous, intermittent signal obstacles located along the actual test route, the GPS postprocessing software used in the benchmark data collection effort could not calculate points along the route given its own constraints for generating highly accurate point data (at the millimeter level). The vendor is currently looking at possible modifications to this postprocessing software to allow for fuller data recovery.

Therefore, two existing databases were evaluated for feasibility in generating the benchmark route until more accurate kinematic survey methods become available. The most readily available street network was the TIGER database. The second database is a commercially available, geometrically corrected version of the TIGER database; it has improved spatial accuracy and more current spatial and attribute data than the TIGER files. A comparison of the two different databases is shown in Figure 2; significant spatial discrepancies are obvious. A GPS test run with differentially corrected data is also shown. It is apparent that the GPS data more closely follow the spatially corrected database, the use of which would result in more correct map matching (a primary goal of route choice data collection). However, one drawback of more accurate databases is that they can become prohibitively expensive if not already available. The trade-off between accuracy and cost, then, depends on the application. If the TIGER database is the only option, caution should be exercised given that the inherent error in the database itself could be larger than that existing in the GPS data.

Step 4: Optimize Data Collection Effort Within Known Conditions

If benchmark route data collection is the objective (rather than testing real-world driving patterns), there are several steps that can be taken to improve receiver performance and data accuracy. First, mission-planning software can identify the best time for maximum satellite availability in a specific location. Driving the test vehicle on each road segment in the lane that reduces signal obstructions is also desirable for improving satellite visibility. Weather conditions also affect receiver performance; rainy weather should be avoided if possible. The impact of tree canopy obstructions can be reduced significantly by collecting data during winter months. Finally, methods using differential correction (real-time or postprocessed) are highly recommended.

If the data collection effort is for equipment evaluation purposes, real-world conditions should be used (i.e., typical, nonoptimal). In addition, it is recommended that differential correction techniques be used, since this addresses errors introduced with SA. Test data collected on two runs reveal that differential correction does have a large impact on the quality of data available for route reconstruction via

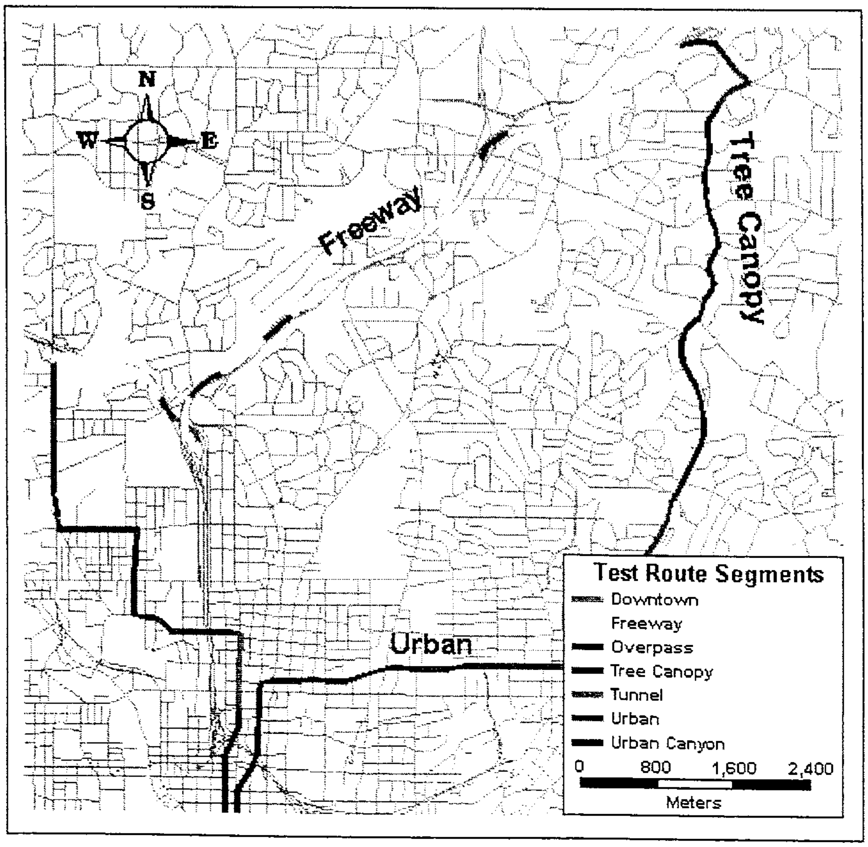

FIGURE 1 Atlanta route selected for testing GPS receiver performance and accuracy.

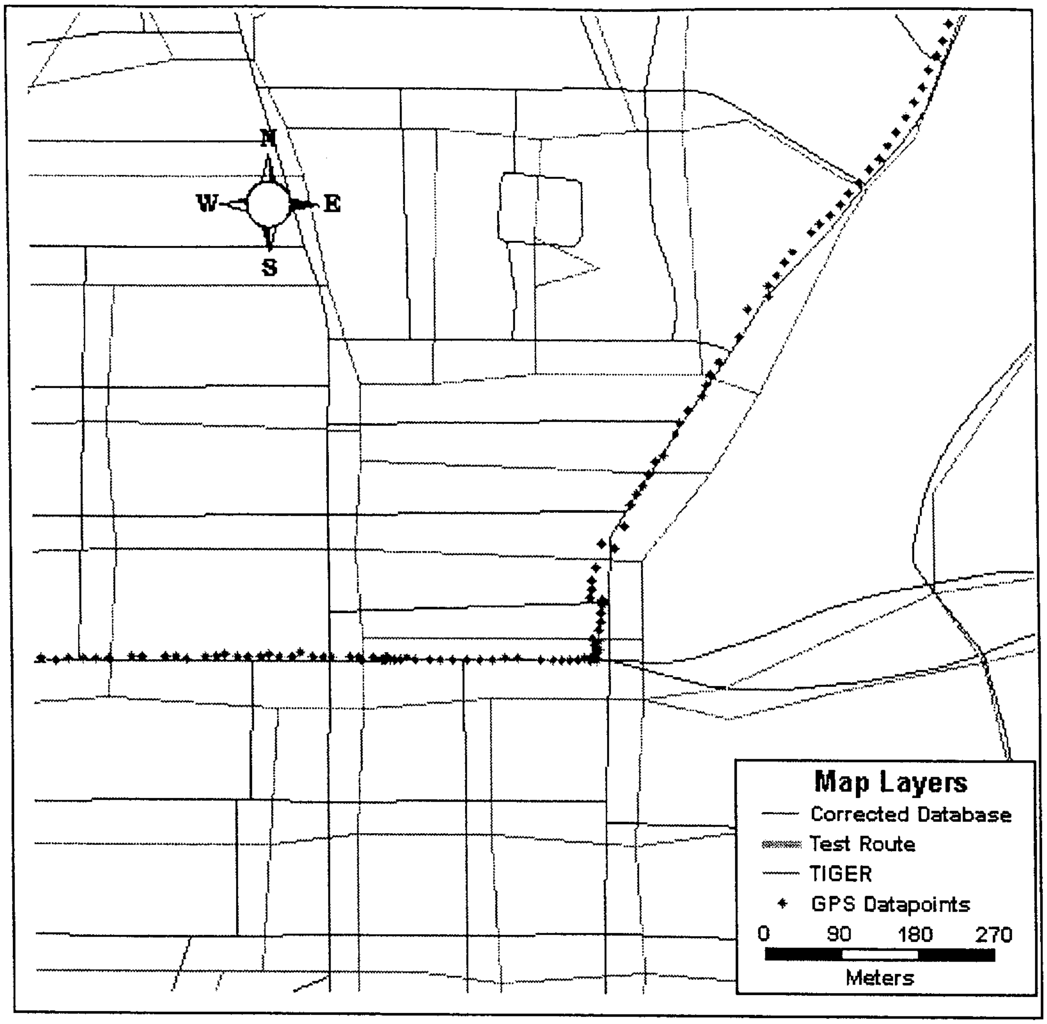

FIGURE 2 Overlay comparison of two street network databases with differently corrected GPS route.

map matching methods (see Figure 3). The two upper runs represent uncorrected data points, whereas the lower dark line represents the corrected points of both runs; it should be noted that these corrected runs lie exactly together and on the underlying street network.

Finally, if differential correction methods are desired, special precautions should be taken to ensure that the base station data are accurate. It is essential to choose a good site for the base station. The base station antenna should have an unobstructed view of the sky (starting at 15 degrees above the horizon), should be located on a rigid plane on a high-accuracy reference network (HARN) point, and should be located away from electromagnetic radiation sources like power lines and microwave towers. In rtDGPS applications, it is important that the radio link through which the corrections are broadcast cover the entire range where data are collected and be properly set up so that corrections arrive clearly and with minimal delay at the field receiver.

Step 5: Process Data Carefully

Quality assurance and control plans for data collection and processing should be developed before any field applications of GPS route choice collection are undertaken. Data must be coded and handled in a consistent manner so that meaningful comparative statistical analyses can be performed. If ppDGPS data have been selected, it is critical that base station data be saved safely until the final point corrections are performed.

Some manual postprocessing is necessary to verify the quality of the data and to determine which points should be discarded. This process is driven by the attributes for each calculated point: PDOP, the number of satellites used, and speed. Generally, data points with PDOP values above 4 are based on poor satellite geometry and should be discarded. In addition, points with a PDOP value of zero indicate that no data were available, and these points should also be discarded.

The number of satellites used to determine the point solution is also a good indication of the quality of the point being investigated and can be used as a threshold to filter bad points. At least three satellites must be tracked to obtain accurate x, y-position, although four are preferable. Although detection of these bad data points can be automated with software, it is also possible that some bad points have acceptable PDOP values and at least four satellites. Therefore, a visual inspection of the data using a GIS and underlying database is highly recommended for identification of other outliers.

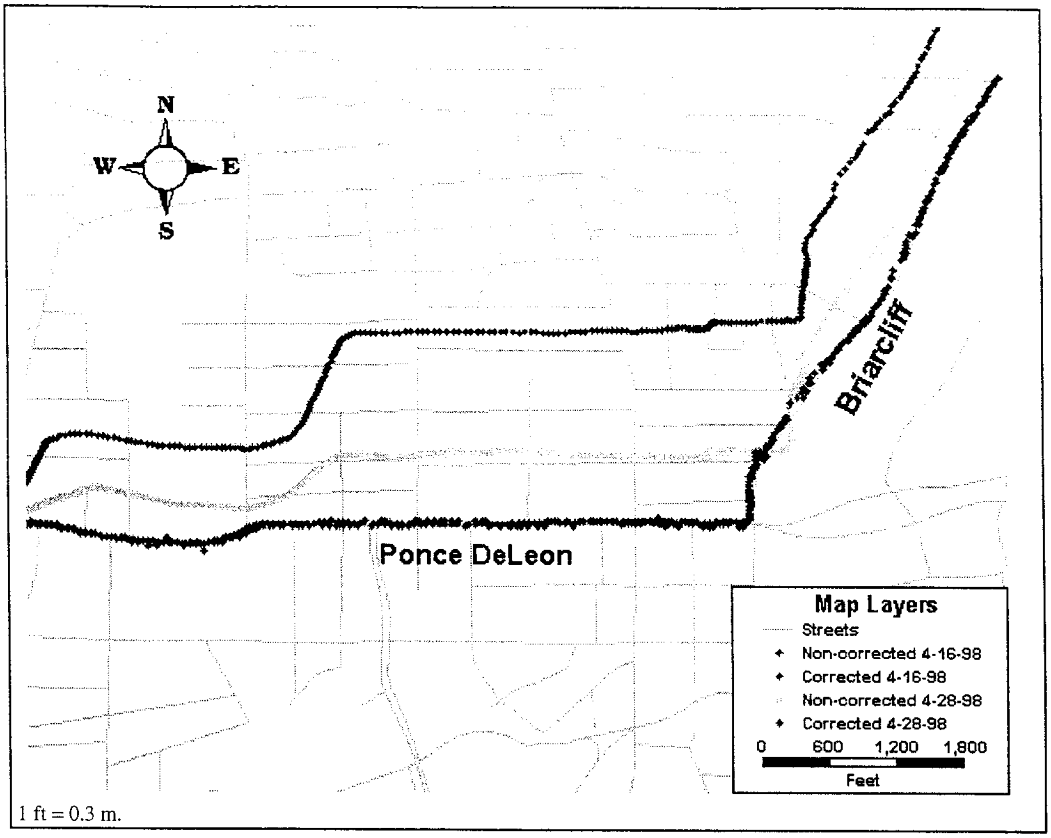

FIGURE 3 Comparison of two uncorrected and corrected test runs with geometrically corrected base map.

Speed calculations can also be used to detect bad data points. There are several ways in which speed can be calculated using GPS data. One method is directly derived from the calculated positions; since each point is time tagged, this calculation becomes a matter of dividing differences in distances by their corresponding time intervals. Speeds calculated in this method can vary widely because of the inaccuracy of the GPS computed points; such inaccuracy can produce unrealistic instantaneous estimates as high as 180 km/h (112 mph). These extreme speeds can be used to identify bad sample points since they reflect shifts in position that are unlikely to occur under the known survey conditions. Consequently, bad data points can also be identified by examining the second-by-second data and searching for erratic changes in speed.

Although detection of poor PDOP values, low numbers of satellites, or unrealistic speeds can be automated, a combined process using algorithms to detect potential bad points followed by visual inspection within a GIS of the points collected is highly recommended. Automated procedures could throw out points that are actually good. Conversely, it is also very possible that some automated procedures may not detect all bad data points, especially when the PDOP value and number of satellites are reasonable. Therefore, visual inspection of the data, along with other logic checks, is needed.

Automated map matching can continue to process bad data points, undetected. As a result, in areas where road segments are located fairly closely or where the data points collected have low accuracy levels, it is easy for map matching software to match the data to the wrong road segment. Therefore, a visual inspection of the routes determined by automated map matching is also highly recommended.

Step 6: Calculate and Interpret Results Cautiously

Given all the issues presented here, analysts should calculate, interpret, and present the results of route choice data collection efforts with caution. If the effort was performed with fixed lane assignments throughout the test run, offset adjustments can be made when accuracy levels are calculated. When results are presented, all parameters that may have had an impact on data accuracy and completeness should be provided, including an explanation of how the accuracy calculations were derived. Finally, because accuracy errors do exist, policy makers should exercise care in interpreting these results and should acknowledge uncertainty issues in any subsequent policy or planning studies that employ these data.

GEORGIA TECH GPS TEST RESULTS

Table 2 contains the Georgia Tech accuracy test results for the five receivers evaluated. These percentages represent the averages for all test runs, excluding the urban canyon segment of each. (The urban canyon segments yielded the poorest GPS performance across all receivers.) The calculations are based on the distances between each point collected and the benchmark route as defined within the geometrically corrected database. It should be noted that the receiver with differential correction, the GeoResearch Workhorse, outperformed all other receivers in the tightest accuracy buffers (<5, <10, and <30 m). It is also noteworthy that the low-end autonomous receivers performed well in the relaxed accuracy buffers (<30, <50, and <100 m) and that the combined GLONASS-GPS receivers had

TABLE 2 Accuracy Test Results: All Runs Combined Excluding Urban Canyon Segments

GPS	< 5 Meters	< 10 Meters	< 30 Meters	< 50 Meters	< 100 Meters	< 500 Meters
Garmin 35LP	15%	31%	83%	96%	99%	100%
Garmin II Plus	15%	30%	70%	89%	98%	100%
GR Workhorse	33%	63%	85%	90%	94%	96%
Ashtech GG24	15%	28%	60%	74%	88%	99%
3S GNSS-300	10%	19%	46%	63%	78%	88%

the lowest accuracy performance (most likely the result of the incomplete GLONASS satellite configuration).

From these results, the research team concluded that the Geo-Research Workhorse was the best receiver of those evaluated to meet the 10-m accuracy requirement for correct map matching in the dense Atlanta road network. One can see, however, that with relaxed accuracy constraints in the 30- to 100-m range, the low-end autonomous receivers are quite acceptable.

CONCLUSIONS

The use of GPS for automated route choice data collection shows great potential compared with manual methods, which have proved burdensome, time consuming, and error prone. GPS technology is now available at reasonable costs and with sufficient accuracy levels, making the possibility of large-scale route choice data collection a reality. In addition, the availability of GIS databases makes the automated processing of such data feasible. However, with the application of these technologies, care and caution should be applied with the use and interpretation of the route choice data obtained.

REFERENCES

1. Williams, B. G. Existing and Emerging Techniques for the Automatic Collection of Data for Transport Modeling. http://www.smithsys.co.uk/smith/public/tech/autocoll, Jan. 1997.
2. Kalfs, N., and W. E. Saris. New Data Collection Methods in Travel Surveys. In *Activity-Based Approaches to Travel Analysis* (D. F. Ettema and H. J. P. Timmermans, eds.), Pergamon Press, 1997.
3. Haubold, I., P. Jackson, K. W. Axhausen, and J. Polak. *Technology Assessment: Deliverable 1 of Technologies for European Surveys of Travel Behavior (TEST)*. June 1997.
4. Sarasua, W., and M. Meyer. New Technologies for Household Travel Surveys. In *Conference Proceedings 10: Household Travel Surveys: New Concepts and Research Needs*, TRB, National Research Council, Washington, D.C., 1996.
5. Czerniak, R., and J. Reilly. Applications of GPS for Surveying and Other Positioning Needs in Departments of Transportation. In *NCHRP Synthesis 258: Applications of GPS for Surveying and Other Positioning Needs in Departments of Transportation*, TRB, National Research Council, Washington, D.C., 1998.
6. Mahmassani, H. S., T. Joseph, and R.-C. Jou. Survey Approach for Study of Urban Commuter Choice Dynamics. In *Transportation Research Record 1412*, TRB, National Research Council, Washington, D.C., 1993, pp. 80–89.
7. Abdel-Aty, M. A., R. Kitamura, P. P. Jovanis, P. Reddy, and K. M. Vaughn. New Approach to Route Choice Data Collection: Multiphase, Computer-Aided Telephone Interview Panel Surveys Using Geographic Information Systems Data Base. In *Transportation Research Record 1493*, TRB, National Research Council, Washington, D.C., 1995, pp. 159–169.
8. Lee-Gosselin, M. Scope and Potential of Interactive Stated Response Data Collection Methods. In *Conference Proceedings 10: Household Travel Surveys: New Concepts and Research Needs*, TRB, National Research Council, Washington, D.C., 1996.
9. Polak, J., and P. Jones. Using Stated-Preference Methods to Examine Traveler Preferences and Responses. In *Understanding Travel Behavior in an Era of Change*, Pergamon-Elsevier, Oxford, 1995, pp. 177–207.
10. Eby, D., L. Kostyniuk, C. Christoff, M. Hopp, and F. Streff. *An On-the Road Comparison of In-Vehicle Navigation Assistance Systems: The FAST-TRAC Troika Study*. Four Project Reports. University of Michigan, Transportation Research Institute, 1997.
11. Wagner, D. *Lexington Area Travel Data Collection Test; GPS for Personal Travel Surveys*. Battelle Report. Office of Highway Information Management, FHWA, U.S. Department of Transportation, and Office of Technology Assessment, Sept. 1997.
12. Casas, J., and C. Arce; NuStats International. Trip Reporting in Household Travel Diaries: A Comparison to GPS-Collected Data. Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 1999.
13. Draijer, G., N. Kalfs, and J. Perdok. *Possible Application of Global Positioning System (GPS) for Collecting Travel Data*. Project Report. Transport Research Centre of the Ministry of Transport, Public Works, and Water Management, Netherlands, 1998.
14. 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-99005D. Georgia Transportation Institute and Georgia Institute of Technology, Atlanta, March 1999.
15. Hofmann-Wellenhof, B., H. Lichtenegger, and J. Collins. *GPS—Theory and Practice*, 3rd ed. Springer-Verlag, New York, 1997.
16. Zito, R., G. D'Este, and M. A. P. Taylor. Global Positioning Systems in the Time Domain: How Useful a Tool for IVHS? *Transportation Research C*, Vol. 3, No. 4, 1995, pp. 193–209.
17. Bullock, J.; Motorola Corporation. GPS Antenna Placement Analysis in an Automobile Simulation and Performance Test Results. Presented at the Intelligent Transportation Symposium, Washington, D.C., 1997.
18. Murakami, E., and D. Wagner. Using Global Positioning Systems for Measuring Household Travel in Private Vehicles. *Conference Proceedings: GIS-T Symposium*, March 1997, pp. 38–48.
19. Benjamin, M. Using GPS/GIS for the Collection of Motor Vehicle Activity Data. Presented at the Environmental Management and Technology Conference, Nov. 1997.

Publication of this paper sponsored by Committee on Urban Transportation Data and Information Systems.