Improved Method for Collecting Travel Time Information

Toby D. Rickman, Mark E. Hallenbeck, and Margaret Schroeder

A primary difficulty in evaluating most new traffic control and traffic operations systems that are developed, tested, and installed is to determine the impacts those systems have on traffic behavior. To collect the data necessary to evaluate the systems, researchers have traditionally used floating car surveys and other data collection techniques. However, it is costly to perform the number of floating car surveys required to accurately measure the reasonably small changes in travel times that individual vehicles accrue as a result of traffic control system improvements. A more cost-effective and potentially more accurate alternative to the floating car survey for collecting that travel time information is described. Observers with lap-top computers collect license plate information. A series of simple computer programs performs the required license plate matching and produces summaries describing the travel characteristics of the traffic stream. This method of data collection costs less than floating car surveys, provides a larger number of travel time runs for a given level of personnel involvement within a given period, and can provide additional information on the traffic stream being monitored—for example, origin-destination information—at no additional cost.

A primary difficulty in evaluating the impacts of many of the new traffic control and traffic operations systems that are developed, tested, and installed is to determine the impacts those systems have on traffic performance. Although simulations created with NETSIM, FRESIM, or some other model can play an important part in developing the system's control algorithms and operating parameters, simulation programs are necessarily limited to the use of "static" input data that have been previously collected. These data are usually adequate for system development, but they barely include the variability found in real-world traffic operations.

Thus, system improvements must be evaluated under real traffic conditions. This evaluation requires the collection of travel time information for vehicles passing through the study area. Before-and-after travel time studies measure the vehicle performance within the study area, and statistical techniques then determine the impacts of the new system.

The collection and evaluation of traffic performance data pose two major limitations on these important evaluations:

- Traffic performance is extremely variable, and
- Collection of traffic performance data is costly.

These two factors hamper engineers in collecting a sufficient amount of traffic performance data to measure the reasonably small gains in individual vehicle performance that result from most new systems.

For example, a 2-mi section of arterial may contain four or more traffic signals. Depending on how well the signals are timed, the road's level of congestion, and when in the cycle a vehicle enters the network, a vehicle may encounter between zero and four red lights while traversing the network. The delay the vehicle experiences directly relates to the number of red lights it meets, and these delays create a wide variation in travel times along the road section. As a result of this high level of variation, a large number of travel times is necessary to adequately determine the mean travel time of vehicles passing through the system. Only with a valid measure of the mean travel time and the distribution of the travel time can statements about improvements to service be made with confidence.

Depending on the number of stop lights and the level of congestion, one team of two persons in a vehicle can usually make about three peak-period, peak-direction travel time runs on a 3-mi stretch of arterial during one peak hour. If sample sizes larger than three are needed (and with large coefficients of variation, any level of acceptable statistical precision requires more than three runs), additional days of data collection are required, or additional vehicles and data collection teams must be used. Even when multiple days of data collection are performed to examine variation over time, with the small number of data points collected using the floating car method, it is difficult to know whether measured differences are caused by different conditions on subsequent days or by the inherent variation in travel times on that road.

Both additional days of data collection and additional crews collecting information substantially increase the cost of data collection. If congestion on the subject arterial varies throughout the peak hour (which is often common), the three runs collected by a single crew may have very different travel times, and many days of data collection may be necessary to collect enough data to adequately describe changes in the road's travel times during the peak.

The relatively high cost of collecting travel time information means that most new signal systems (and many other traffic operations systems) are not adequately evaluated when they are installed and tested in the field. However, because the authors needed large amounts of travel time information to evaluate a number of ongoing projects, they developed a variation on the traditional license plate matching survey to provide better travel time information at a lower cost than floating car surveys could provide. That method is described and compared with other available methods.
THE NEW DATA COLLECTION SYSTEM

Engineers have used license plate matching for many years to collect traffic data. Traditionally, observers wrote down license plates on paper and then keypunched them into a computer, or spoke them into a tape recorder and then transcribed them onto punch cards. Both systems had limitations. Not only was staff time required to collect data, but also a considerable amount of office time was required to transcribe the data into a computer-readable format.

The advent of inexpensive lap-top computers allows license plate information to be collected more cost-effectively and efficiently. At least two observers are equipped with inexpensive ($400 per computer, although more expensive and capable lap-tops can also be used) lap-tops that run a simple BASIC program. (Additional observers with computers proportionately improve the collected data.) The observers type the first four digits of the license of vehicles that pass their location and then press the computer’s RETURN key. The BASIC program accepts the input, reads the computer’s system clock, and adds the newest entry and the time of that vehicle’s passing to a file of four-digit license plates. This process continues until the data collection activity has finished.

Upon completion of the data collection effort, the files from all the lap-top computers are transferred to a microcomputer in the central office. A simple FORTRAN program is then run, which

- Matches license plate numbers from the various lap-top files,
- Compares the times of the two matched observations of each license plate,
- Computes the travel time between each set of observation points for each license plate, and
- Creates a file containing all of the matches and travel times.

Minitab (or any one of a number of computer programs) is then used to read and summarize the output file. (The output files can be read by almost any standard microcomputer statistics, analysis, or graphics package, including Minitab, SPSS PC, SAS, or Lotus 1-2-3.) Output from Minitab includes the mean, median, standard deviation, and a simple scattergram of the travel times between the two study locations. This information is used to examine the data for spurious matches, which are then removed from the data set with a simple text editor. Minitab is then run again to produce the final travel time statistics for the study area.

The entire computerized analysis process can take from 20 min to a few hours to complete, depending on the style of lap-top computer (whether it is MS-DOS compatible and uses standard disk drives, or whether it requires a modem or hardwire data transfer to move data to the central computer) and the number of additional graphing and statistical tests to be performed.

For simple freeway or arterial travel time studies, a two-person crew can collect the data. A good typist (roughly 60 wpm) can enter as many as 900 four-character license plates in 1 hr, provided volumes are heavy (about 1,500 vehicles per hour) and the observers have good visibility. For a freeway section with only one or two exits, two such output files usually result in more than 100 matched license plates, providing an accurate picture of travel times during the given period. When computerized license plate matching is compared with the three or four travel time runs possible with a two-person floating car survey, the advantages of the former method become apparent.

On arterials or freeway sections with many entry and exit points observers, and on complex traffic signal networks, the number of matches between any two observation points may be significantly lower than that achieved with shorter freeway sections and simple arterials. The decrease occurs because many vehicles observed at one end of a complex network stop in or depart the network and do not reach later observation locations. Nevertheless, even within complex networks, the study teams may obtain between 11 and 40 valid travel times per hour, considerably more than would be possible with the same personnel and the floating car methodology. With the computerized methods, the number of matches is simply a function of the percentage of license plates recorded at each location and the traffic volume flowing between the two study points.

MULTIPLE DATA COLLECTION POINTS

The license plate survey can also provide some additional information that is not readily collected with the floating cars. This information is related to origin-destination (O/D) data, which have traditionally been collected with the license plate matching technique. With multiple observers (more than two), the relative volume of vehicles traveling from one location to another can be measured. This measurement has been an important and traditional part of freeway license plate studies—for example, studies of which off-ramps people who enter at a particular ramp use). However, this type of information can be just as important in arterial networks.

On arterials, a variety of paths through the network can be measured simultaneously if observers with lap-tops are placed throughout the network (see Figure 1). Not only do matches indicate which paths vehicles are taking through the network and how well the path selected for the floating car survey is flowing, but also they indicate how well the network as a whole is functioning under the selected operating plan. This is particularly important for grid networks in which vehicles may follow many paths between their entry points and final destinations.

One advantage of the license plate survey is that it collects data from the real traffic stream, whereas the floating car survey measures travel time along an arbitrarily selected path at what the driver estimates to be the average speed. For simple arterials, this path is easily selected. For complex grids, neither the paths that motorists use nor the proportion of vehicles making specific sets of turning movements may be so readily identifiable.

Programs such as TRANSYT-7F attempt to optimize vehicular movement for an entire network. They try to balance the movements and delay experienced by all vehicles on all links in the network, not just movements and delays for vehicles on selected parts of the network. Therefore, to accurately measure the impacts of a TRANSYT-7F signal plan, a researcher should measure traffic performance throughout the
Advantages and Disadvantages of License Plate Matching

Until now only the good points of the computerized license plate matching technique have been examined. However, all methods have advantages and disadvantages. Three principle methods for collecting travel time and vehicle performance information are briefly compared in this section. The three methodologies compared are:

- Floating car survey,
- License plate matching with a voice recorder, and
- License plate matching with lap-top computers.

Floating Car Surveys

The advantages of the floating car survey are that:

- The driver can positively identify the travel time measured,
- Collectors can easily record the travel time to the intermediate points on the roadway, and
- The recorder can note the cause of any delay the vehicle experiences during the travel time run.

The disadvantages of a floating car survey are that:

- Extensive staffing is required to drive the vehicles and record the data if a large number of travel times are desired;
Either extensive staff time is needed to transcribe the data from the recording sheets into a computer usable format, or the data must be collected initially on a portable computer;

- Because of staffing requirements, this method is costly for the amount of data gathered;
- Because the funds are usually not available, an insufficient amount of data is commonly recorded to adequately describe the travel time as a distribution of times with a mean, median, and standard deviation; and
- Behavior of the vehicle's driver can influence the travel times collected.

Despite its drawbacks, this method is the most widely used for travel time studies because of the perceived reliability of the data, the ease with which a small number of data points can be compiled, and the simultaneous collection of both travel time data and a limited amount of delay information.

Tape Recorded License Plate Surveys

The advantages of using voice recorders for performing license plate matching are that the voice recorders are inexpensive and this method provides a cost-effective way to collect large amounts of data. Another advantage of a voice recorder is that the observers can record license plates without taking their eyes off the traffic stream.

The disadvantages of voice recorders are as follows:

- Recording problems: Background noise can hamper the data compilation process. Voice quality in general can be a major problem when voice recorders are used.
- Cost: Although the cost of the recording devices is small, the conversion of the data from the recordings into machine readable formats can be costly. The recorders must be played with a time counter while the license plate information is transcribed into a computer-readable form. Estimates of the time required to complete this process range from 3 to 7 hr per hour of data recorded (7).
- Synchronization: Coordinating the time stamp with the voice recorders can be a problem.
- Tape accuracy: The accuracy of the tape machine requires periodic time marks on the tape during the data collection.
- Spurious matches: Incorrect license plate matches can occur if only a portion of the license plate is recorded, which is common.

Computerized License Plate Surveys

The advantages of using lap-top computers in travel time studies are as follows:

- Data recording: Observers can easily record the data by entering the license plate on a computer keyboard. The computer accurately records the entered plate number along with the accurate time.
- Synchronization: The computers are easily synchronized in the office and do not rely on the use of different watches.
- Data formatting: The data transfer easily from the lap-top computer to a disk file on a desktop computer. More advanced computers record the data directly on a usable computer disk file. The computer can be programmed to record the data in a form the matching program can read. This process eliminates keypunching errors that can occur when license plates have to be transcribed onto a computer file for matching.
- Cost: As with the voice recorder method, the use of computers is a cost-effective way to collect a large amount of data with only a minimum number of people. Because the data do not need to be transcribed or recoded, the use of computers reduces the cost of personnel required both in the office and in the field.

The cost comparison of the different methods in Figure 2 indicates that computers have a tremendous advantage over a floating car survey and save 40 percent compared with the use of voice recorders (2).

The disadvantages of lap-top computers are that the machines are still relatively expensive ($200 to $400 for an inexpensive model without a disk drive and $700 to $1,200 for an MS-DOS–compatible, dual-disk-drive model); the license plate matching method does not obtain any delay information at specific locations along a route; and unless the entire license plate is recorded, spurious matches can occur. The problem of spurious matches is described below.

Spurious Matches

Data collectors often enter only a portion of the license plate in a license plate matching survey (3). Partial entry eases the data collector’s task by reducing the number of digits to identify and type. After attempting various numbers of digits, it was found that the collection of up to four digits was possible with lap-tops when the data collectors were not accomplished typists. Because more than one plate can have the same first

![FIGURE 2 Cost comparison of data collection methods. Costs include staff time to travel to study site, collect the data, and reformat the data into a computer-usable format and are based on experience during the I-405 CORFLO study for FHWA.](image)
four digits, different vehicles can be incorrectly matched at the end and beginning of the roadway section. (For example, license plates ABC 123 and ABC 199 would be incorrectly matched because only the first four digits of the plates, "ABC 1," would have been recorded.)

Several methods can be used to deal with the influence of spurious matches on the results of the data collection. These responses include

- Eliminating duplicate matches,
- Establishing a minimum and a maximum travel time expected during the study (4) and eliminating matches outside of this boundary,
- Using graphs to recognize unrealistic outliers,
- Comparing smaller time intervals so that the chance of spurious matches is reduced, and
- Recording more digits of the license plates.

Duplicates should not be eliminated until other data integrity checks have been performed. This is a simple step that a computer can easily do. It eliminates data that are strongly suspect (with at least a 50 percent possibility of being incorrect).

The process of establishing a minimum and a maximum travel time expected during the study (4) and eliminating matches outside this boundary is somewhat arbitrary and can influence the validity of the data. However, it can be a necessary process when large study times and large numbers of license plates are involved. It is also necessary for long arterial sections in which some vehicles may enter the network, stop at a store within the network, and then exit the network during the study period.

Minimums and maximums can be set by either some form of calculation (speeds should not exceed 100 mph or be below 2 mph) or can be the mean travel time plus some multiple of the standard deviation of the measured travel time.

A study of methods for dealing with spurious matches suggested the use of the median rather than the mean travel time to represent average traffic conditions (4). Use of the median value allows spurious matches to be ignored. This method does yield a value for the travel time that reasonably represents the expected value, as can be seen in Figure 3.

However, the problem with using only the median value without eliminating the spurious matches is that the standard deviation of the data cannot be determined. It is important to have an expected value and a measure of the standard error to be able to describe the data as a distribution of travel times.

Graphical methods can also help in recognizing spurious matches. The same data shown in Figure 3 are displayed in Figure 4 in a different format. The data shown in these figures were collected on a freeway section during 1 hr of license plate data collection. The majority of the data points are at 500 sec, which indicates a speed of 18 mph. The data points in the range of 2,000 sec represent vehicles that would have been traveling at 5 mph at the same time and in the same traffic stream as the other vehicles traveling at almost 20 mph. Obviously, these few data points represent spurious matches and can be eliminated from the data.

Another method for reducing the number of spurious matches compares smaller time intervals, which reduces the chance of spurious matches. Figure 5 shows 1 hr of data. The travel time is hard to distinguish because of the large number of spurious matches and the variation of traffic speeds over that fairly long time period. A 15-min segment during this same hour produced the data shown in Figure 6, in which the travel time is distinguishable at 320 sec. The three outliers in this 15-min segment are easily recognized and eliminated. This method is actually quite similar to placing constraints on the minimum and maximum travel times, but it allows those constraints to vary over time, just as traffic conditions may vary over time with congestion.
FIGURE 4  Travel time data (1 hr, 4 digits).

FIGURE 5  Travel time data (1 hr, 3 digits).
Recording more digits of the license plates is also a way to eliminate spurious matches. The number of spurious matches is statistically related to the uniqueness of the number recorded and thus the number of digits. In the case of license plates, if the whole license plate is recorded, spurious matches occur only because of errors in data entry. However, observers have difficulty recording the whole license plate, and so the question becomes, how many digits are needed? The data shown in Figure 4 are 1 hr of matches from the freeway section described earlier. Four digits of license plates were recorded. These data contrast with the data in Figure 5, also collected for 1 hr on the same section of highway, but with only three recorded digits. With three entered digits, spurious matches become difficult to distinguish from the real matches, but with four digits, the erroneous matches are easy to determine. This examination was repeated several times with different data sets. Through statistical analysis it was determined that four-digit entries provide the best combination of ease of data entry and a low level of spurious matching.

**LICENSE PLATE AND FLOATING CAR SURVEY COMPARISON**

As part of the development of the computerized license plate matching system, the floating car and license plate matching methodologies were compared. Parallel data sets were collected using both the computerized license plate method and floating car surveys. Four tests were conducted on three arterials in Bellevue, Washington, a suburb of Seattle. Streets involved in the comparison were NE Eighth Street, Bel-Red Road, and 148th Avenue.

Highlights of the resulting comparisons are shown in Table 1. In all cases, the number of staff needed for the license plate travel time study was equal to or less than the staff required for the floating car methodology.

As Table 1 shows, the license plate methodology collected more travel times than the floating car methodology in all cases (ranging from 1.83 to 15 times the floating car data collection). In the case of NE Eighth, the majority of vehicles entering the study area turned off the study arterial to access parking within the study area, rather than reaching the far end of the study area where the second observer was located. An observer located in the middle of the study area obtained a substantially greater number of matches than shown in this table, both from the entry point to the middle of the arterial and from the middle of the arterial to the exit point. These travel times confirmed the original NE Eighth estimate.

As expected because of the fairly high variation in travel times associated with arterial travel time runs, the travel times for the license plate survey usually differed slightly from the floating car survey (see Figure 7). This difference was not statistically significant in any of the test cases. One interesting finding of the comparison was that the standard deviation of the license plate travel times was higher than the standard deviation of the floating car runs.

This difference, it is believed, was caused by a number of factors, including the following:

- The same drivers made all of the floating car runs, and their driving habits may have caused the runs to be very similar.
- The floating car drivers were told to "drive the average speed." Therefore they tended to drive consistently from trip to trip and passed about as many cars as passed them. Thus no extreme travel times were likely to be included in the floating car runs.
- The travel time runs were made in a loop pattern. This means that the drivers entering onto the study arterial had to turn from a side street to enter traffic. Their entrance required a break in the oncoming traffic. As a result, the floating cars tended to enter the study area at the same point of the signal.
TABLE 1 COMPARISON OF FLOATING CAR AND COMPUTERIZED LICENSE PLATE TRAVEL TIME METHODS

<table>
<thead>
<tr>
<th></th>
<th>Mean Travel Times</th>
<th>Number of Travel Time Runs</th>
<th>t-statistic*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Floating Car</td>
<td>License Plate</td>
<td>Floating Car</td>
</tr>
<tr>
<td>Bel-Red Road Eastbound PM</td>
<td>590</td>
<td>590</td>
<td>5</td>
</tr>
<tr>
<td>148th Avenue Southbound PM</td>
<td>453</td>
<td>487</td>
<td>3</td>
</tr>
<tr>
<td>NE Eighth Eastbound PM</td>
<td>242</td>
<td>264</td>
<td>6</td>
</tr>
<tr>
<td>148th Avenue Southbound AM</td>
<td>247</td>
<td>257</td>
<td>5</td>
</tr>
</tbody>
</table>

* The Student's T statistic is used here to compare the mean travel times of the two travel time distributions. All t values are within the critical T value at the level of alpha = 0.005, and the associated degrees of freedom for each test. This indicates that there is no statistical difference between the two travel time methodologies.

FIGURE 7 March floating car versus license plate travel times.

cycle each trip. This situation probably resulted in more consistent trip times than if the floating car had reached the beginning of the study area at random times in the signal phasing.

T-tests were used to statistically compare the travel times collected with the two methods. In addition to the full arterial travel times, these tests were also applied to the travel times collected for partial trips through the arterial. Table 1 shows that there were no statistically significant differences between the floating car and license plate methods when the entire arterial was examined. In only 1 out of 11 cases were the travel times measured over the partial arterials statistically different, and this instance occurred most likely as a result of the floating car runs making or missing a single traffic light more often than the vehicles measured with the license plate method.

As a result of the above comparisons, it was concluded that the license plate survey methodology produces results that are statistically the same as the floating car results. However, the larger number of observations in the license plate survey data set means that the results produce a higher level of statistical confidence in the estimates of mean vehicular travel time than the results of the traditional floating car method.
DATA COLLECTION CONSIDERATIONS

Basic Considerations For Performing License Plate Studies

The data collection teams that used the lap-top computers for the license plate studies noted several considerations that would improve the ease and reliability of computerized license plate surveys:

- Read the license plate as the vehicle approaches. This is especially important at higher speeds on freeways.
- Use observation points above the traffic stream when possible. Often the angle of the approaching vehicles makes the license plate difficult to read if the person is not above the traffic while recording the data.
- Concentrate on one lane of traffic. If a person is not used to recording plates for each lane of traffic, then one lane should be selected to increase the probability of matches. This process does raise a question about the travel time in different lanes. Two studies conducted by the authors indicated that travel times can differ statistically from lane to lane.
- Read as many plates as possible. Some plates are dirty, mangled, or missing on the front bumper, but the point of the study is to match as many plates as possible, so the data collectors should try to record all they can.
- Errors can occur if the plate is read or entered incorrectly or the number is transposed. The data should be checked for errors if possible. For example, if one computer is recording capital letters and the other is not, then matches will not result unless the matching program can account for this.

Extra Considerations For Arterial Data Collection

Arterials pose a special problem for license plate matching data collection. The largest drawback of the license plate matching scheme is that data are not collected along the length of the arterial but are only collected where observers are stationed. In addition, the data collected are only cumulative data (reflecting the time needed to reach a specific point); the collected information does not reveal the number, location, or extent of specific delays that vehicles encounter as they travel between two points.

The loss of specific delay information makes placement of observers especially important. Observers must stand where the most useful data will be obtained. In addition to the basic criteria for placing observers, described above, another important consideration for arterials is whether the data should be collected before or immediately after an intersection. Data collected immediately after an intersection (Observer X in Figure 8) include the delay time at that intersection. Data collected before an intersection (Observer Y in Figure 8) do not include delay time at that intersection.

The difference between these two data collection points may or may not be important, depending on the objective of the data collection effort. If the intent is simply to compare before and after conditions, the only consideration is to ensure that the before and after studies collect data at the same point. Other than that, the inclusion of the delay at any particular intersection is important only if the delay at that intersection is important to the study. For example, if the signal is the last signal in a coordinated group of signals and the study is trying to measure the time required to pass through the coordinated lights, the delay time at the final intersection should be included in the travel time estimate.

Collecting data after vehicles cross the stop line (from Position X) is the generally accepted location. The difficulty with collecting from this location is that observers often have trouble reading the license plates of vehicles in the second and later rows of a platoon that starts from a green light at that intersection. The tightness of a platoon starting up from a...
green light also causes a large number of vehicles to pass the observer in a very short time. In many cases, the combination of these two factors significantly reduces the percentage of vehicles passing a point that are entered into the license plate matching files.

To increase the number of license plates entered into the file, licenses can be entered when the approaching vehicles reach the back of the queue at the intersection (or the stop line when there is no queue). This procedure allows data on the delay imposed by that signal to be lost, but it usually increases the number of vehicles entered into the matching file. This increase occurs because the vehicles reaching the back of the queue at an intersection tend to be more dispersed and slow as they approach. Both these factors improve the observer's ability to read and enter license plates. Choosing Position Y over Position X also allows the observer to walk in the direction of travel (from Point Y towards Point Z in Figure 6) as vehicles queue. This movement further improves the observer's view and increases the accuracy of the license plate data entry.

REFERENCES