Establishing Monitoring Programs for Mobility and Travel Time Reliability (L02)
Strategic Highway Research Program 2
Transportation Research Board
National Academies

New York Case Study Validation: Travel Time Reliability Monitoring

Authors:

Tiffany Barkley, Berkeley Transportation Systems, Inc.
Rob Hranac, Berkeley Transportation Systems, Inc.
Alex Kurzhanskiy, Berkeley Transportation Systems, Inc.
Eric Mai, Berkeley Transportation Systems, Inc.
Berkeley, CA; February 2012

George List, North Carolina State University
Nagui Rouphail, North Carolina State University
Billy Williams, North Carolina State University
Raleigh, NC; February 2012

Acknowledgements:

The Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials sponsored this work. It was conducted in the second Strategic Highway Research Program (SHRP 2), which is administered by the Transportation Research Board of the National Academies. Bill Hyman and Mename Hedhli managed this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. MONITORING SYSTEM

STUDY DESCRIPTION

This is the fifth and final case study performed by the project team in order to validate the approaches to travel time reliability monitoring described in the Travel Time Reliability Monitoring Guidebook. The goal of each case study is to illustrate how agencies apply best practices in transportation performance monitoring and analysis. Typically, this includes monitoring system deployment, travel time reliability calculation methodology, and agency use and analysis of the data. However, in this use case, special focus is given to a new data source: vehicle probe data. In order to focus on practical issues associated with processing this data, no real-time travel time monitoring system or archived data user service was deployed as part of this case study. Rather, travel time reliability analysis was carried out offline, directly on the probe vehicle data set itself. This case study consists of the following sections:

- Monitoring System
- Methodological Advancement
- Use Case Analysis
- Lessons Learned

These sections map to the master system components, as shown below in Figure 1-1.

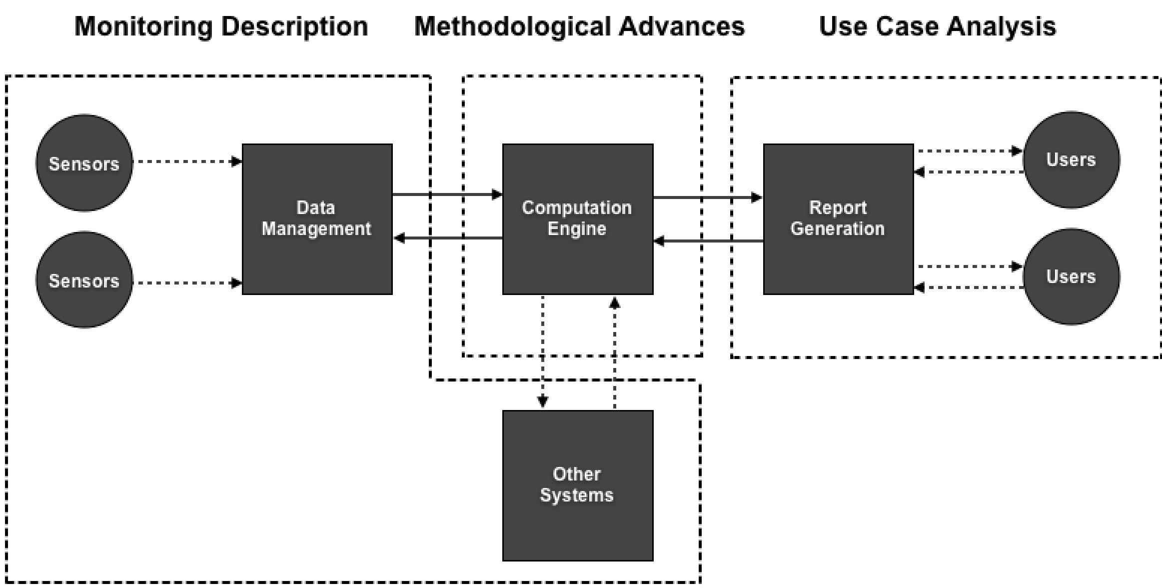

Figure 1-1: Travel Time Reliability System Description

This *monitoring system description* section details the reasons for selecting New York as a case study site and gives an overview of the setting. It briefly summarizes the archived probe vehicle data source and the underlying road network to which it corresponds, and gives an overview of approach that the team took to analyze that data.

The section on *methodology* describes the steps necessary to obtain Probability Density Functions (PDFs) of travel time distributions along a New York City route based entirely on probe data. Critically, this probe data is sparse and few probe vehicle runs traverse the entire route. Techniques are presented to preserve the correlation in speed measurements on consecutive links while synthesizing the aggregate route travel time PDF from segments of multiple probe vehicle runs.

The *Use Case Analysis* section is less theoretical, and more site specific. It is motivated by the user scenarios described in the Task 2/3 document, which are the results of a series of interviews with transportation agency staff regarding agency practice with travel time reliability. While the methodology section of this case study describes the steps necessary to process and interpret probe vehicle data, the use case section focus on a specific application of this methodology. This case study contains a single use case that focuses on three alternative methodologies for constructing travel time probability density functions (PDFs) from probe data.

Lessons Learned summarizes the lessons learned during this case study, with regards to all aspects of travel time reliability monitoring: sensor systems, software systems, calculation methodology, and use. These lessons will be integrated into the final guidebook for practitioners.

SITE OVERVIEW

The New York City site was chosen to provide insight into travel time monitoring in a high-density urban location. The 2000 US census revealed New York City's population to be in excess of 8 million residents, at a density near 26,500 people per square mile (1). Kings County (Brooklyn) is the second-most densely populated county in the United States after New York County (Manhattan) (1). New York City has a low rate of auto ownership; only 55% of households had access to an automobile in 2010 (2). For drivers of single occupancy vehicles, 53% of all commute trips take 30 minutes or more, with an average commute travel time of 31 minutes (2).

This site was also selected because it is covered by a probe vehicle data set, provided to the research team by ALK Technologies, Inc. This data is collected from mobile devices inside of vehicles, and consists of two types of data: (1) individual vehicle trajectories defined by timestamps and locations, and (2) link-based speeds calculated from each vehicle's trajectory. Probe vehicle detection technology provides high-density information about the entire path a vehicle travels, allowing travel times to be directly monitored at the individual vehicle level. In contrast, infrastructure-based sensors such as loop detectors measure traffic only at discrete points along the roadway, and do not keep track of individual vehicles as they travel. Probe data relies on the roadway's users to generate performance data, greatly reducing detection maintenance costs to the agency. These features make probe vehicles an attractive roadway data source to agencies.

The research team obtained this probe data for a region of New York City defined by a rectangular bounding box 25 miles long east-to-west and 40 miles long north-to-south. This bounding box covers Manhattan, The Bronx, and Brooklyn in their entirety, along with most of Queens (see Figure 1-2). Data from all roadway segments within this bounding box was obtained. Probe runs that crossed the boundary of the bounding box were truncated such that only the segments within the bounding box were included in the data set. Segments that had been truncated in this way were treated as unique trips.

In terms of format, the data obtained for this site was a static collection of raw traces and processed speed measurements, as collected by probe vehicles between May 19, 2000 and December 29, 2011. No real-time data was acquired or analyzed for this case study because it was not available. Unlike in other case study sites, an Archived Data User Service (ADUS) was not deployed. All data processing and visualization were carried out through custom routines run offline.

As with roadway data from other case studies, this probe data set was accompanied by a network configuration. The network configuration connects the traffic data to the physical roadway network through a referencing system. This configuration data is necessary for proper interpretation and analysis of the traffic data, such as computing route travel times from point speeds. For this probe data, the network configuration is made up of links defined by ALK. Links are unique to a roadway segment and direction and are less than 0.1 miles long on average. Due to limitations in GPS location accuracy, these links are not lane-specific; link data is interpreted as the mean speed across all lanes. The full data set obtained for this case study contains 180,061 links representing 14,402 roadway miles over the 1000 mi^2 area enclosed by the bounding box (see Figure 1-2).

DATA

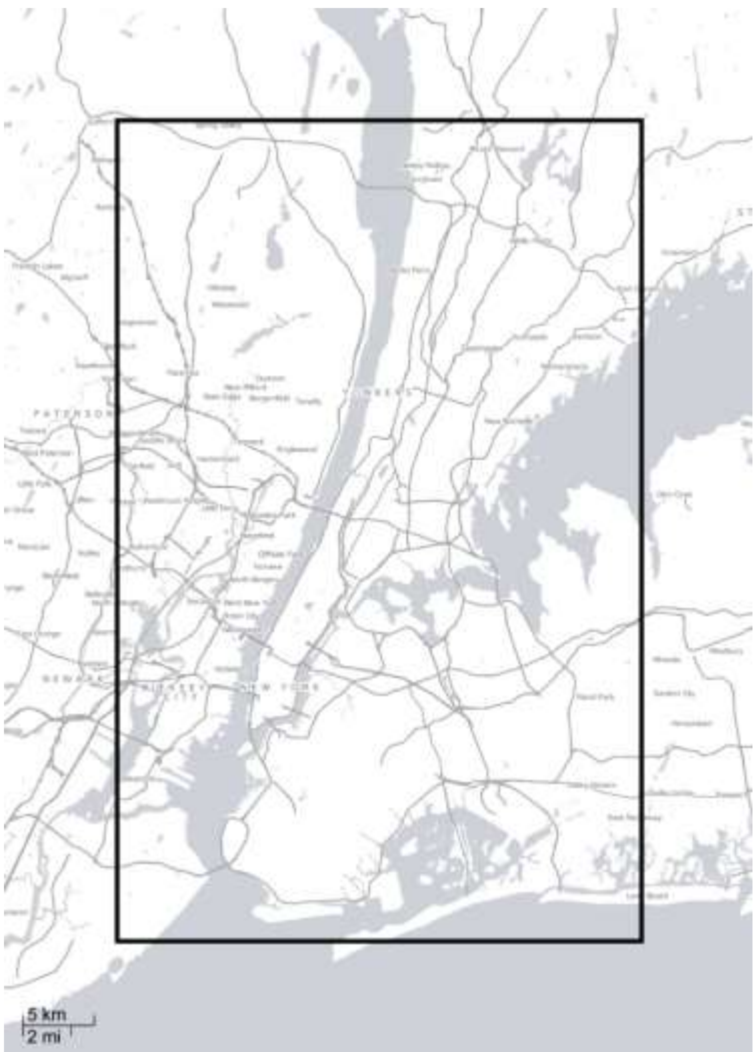

Figure 1-2: Site map with data bounding box

Three probe vehicle based data sets contribute to this case study. Each of these three data sets is based on the same original collection of probe vehicle runs, collected in the *Raw* dataset. The raw data contains unaltered GPS sentences, as originally recorded by the probe vehicles, and was not obtained by the research team. The second data set, called *Gridded GPS Track Data (GGD)*, contains most or all raw GPS points, matched to ALK's link-based network configuration. The third data set, called *One Monument,* is an aggregation of the GGD data set. The One Monument data contains a more manageable number of speed measurements that correspond with the vehicle's speed and timestamp at the midpoint of each ALK link.

Table 1-1: Probe vehicle data sets

Data set	Description	Number of data points	Uncompressed size
Raw	Untouched NMEA sentences	36,683,340 (or more)	4.19 GB (or more)
GGD	Data points reformatted and identified by ALK link	36,683,340	4.19 GB
One Monument	One vehicle measurement per link midpoint	4,282,136	0.48 GB

The Raw GPS data set is stored in the standard NMEA sentences originally recorded by the GPS device in the probe vehicle. A different file is typically created for each vehicle trip. The primary GPS data elements of interest for traffic analysis are location (latitude and longitude), speed, heading, and timestamp. GPS sampling frequency affects the temporal resolution of all three probe data sets. The data analyzed by the research team was based on GPS recorded every three seconds.

The GGD data set is produced through the cleaning and map-matching routines carried out on the raw GPS data, and contains speeds on links and travel times between links organized into trips. This data set is contained in a single file whose entries include timestamp, link ID, position along the link, speed, trip ID, and sequence within the trip. The organization into trips follows that of the GPS files. A gap greater than 4 minutes in a single GPS file is interpreted as the boundary between two trips made by the same vehicle. This preserves continuity in the data and ensures that only travel times (and not trip times) are represented. In this data set, each point is also map-matched to a single ALK link, and includes a value indicating how far along that link the point lies.

The One Monument data set aggregates each trip's data points into single time-stamped speed values for each ALK link that the trip traverses. This is a subset of the GGD data set. When there are multiple observations for the same link within a single trip, only the data point closest to the midpoint of the link is retained and its timestamp is interpolated to the time the vehicle likely passed the link's center point. The speed values in this data set are computed based on the total travel time along the link and the link's length, which effectively smooths out the instantaneous speeds over the link. This data set aids travel time analysis by greatly reducing the number of data points required to compute travel times over road segments for a single trip.

The ALK links themselves are defined in three configuration files, referred to as *Links*, *Nodes*, and *Shapes*. Each link lies within a cell of a rectangular grid, and is uniquely identified by the combination of its Grid ID and Link ID. Links are bounded on either end by nodes whose coordinates are defined in the *Nodes* file. The geometry of each link can be drawn from coordinates found in the *Shapes* file. Additionally, links are labeled with a class identifier, which corresponds to one of the following road types: interstate, interstate without ramps, divided road, primary road, ferry, secondary road, ramps, and local road. Local roads make up the vast majority of the links in the network configuration. Analysis in this case study was limited to interstate links. All links in the study area other than local roads are shown in Figure 1-3.

Figure 1-3: ALK links (without local roads)

DATA MANAGEMENT

Analysis of this probe data set was primarily carried out on the aggregated trip-link speeds present in the One Monument data. The aggregated speeds in this data set are similar in format to those TMC path-based data analyzed in the Atlanta case study. The Atlanta case study compared GPS trace data with video detector data, but only after it had been aggregated into link-based speed measurements. The complete GPS trace data in the GGD data set is the only data from any of the five case studies that traces the entire path of vehicle trips. Even though it is not analyzed directly in this case study, it deepens the analysis done on the *One Monument* data to enable the computation , as will be seen in the use case section.

The data was provided by ALK in flat files and managed by the project team manually through custom processing routines run offline. To focus on issues related to probe vehicle data processing, no additional data sources were considered in this case study.

REFERENCES

[1] U. S. Census Bureau. (2000). Population, Housing Units, Area, and Density: 2000 - State -- County / County Equivalent Census 2000 Summary File 1 (SF 1) 100-Percent Data. Retrieved February 22, 2012, from http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_00_SF1_GCTPH1.ST05&prodType=table.

[2] American Community Survey. (2010). ACS S0802: Means of Transportation to Work by Selected Characteristics, 1-Year Estimates. Retrieved February 22, 2012, from http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=5&ved=0CDwQFjAE&url=http%3A%2F%2Fwww.adcogov.org%2FDocumentView.aspx%3FDID%3D665&ei=3C9NT5nsOuOwiQKws4iZDw&usg=AFQjCNEOt2eTvtMIetEmYC9_631M00Cz1w

2. METHODOLOGY

OVERVIEW

The central goal of this use case is to advance the understanding of practical techniques for working with probe vehicle data in travel time reliability monitoring applications. To accomplish this, the research team analyzed a collection of probe vehicle data. This section first describes the study route, illustrating how the probe data set was assembled and processed for the route and explaining the implications of data density on the resulting analysis. The section then describes methods for identifying and visualizing congestion and travel time reliability from sparse probe data. Finally, it lays the groundwork for computing route-level travel time probability density functions, a methodological issue that is explored in depth in the Use Case chapter.

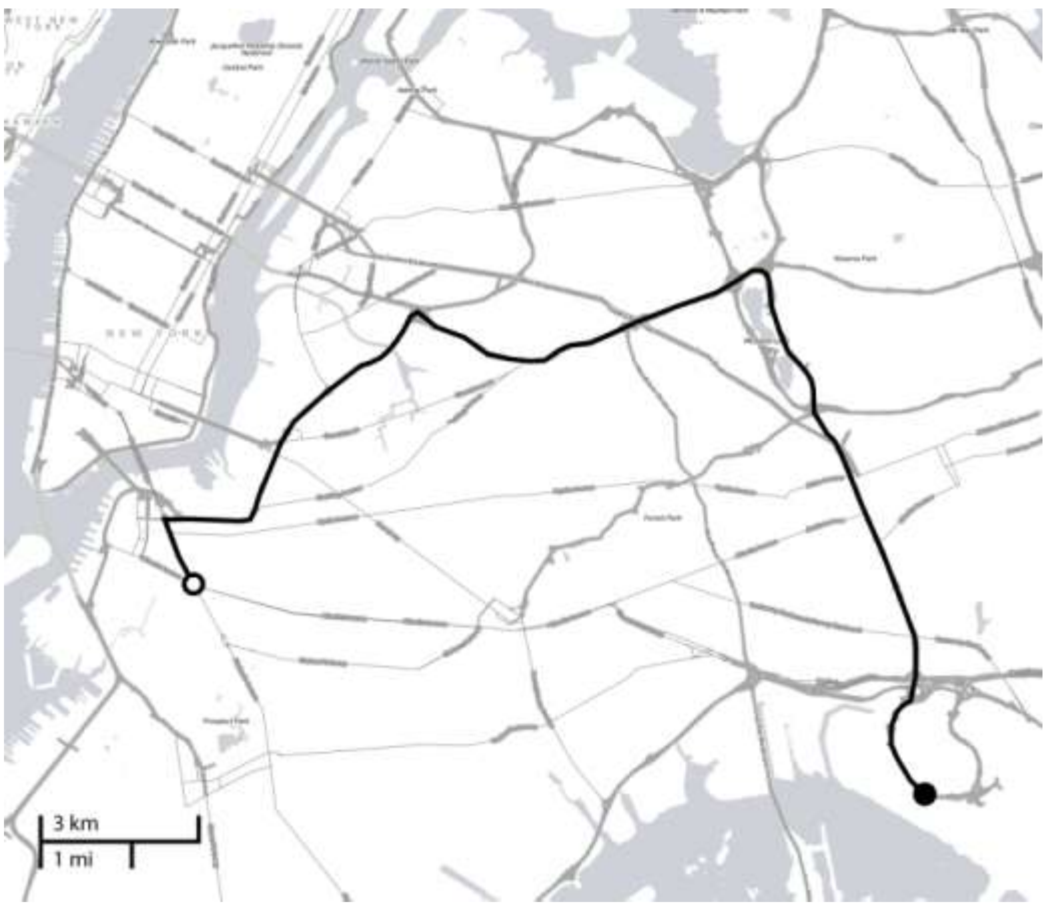

Figure 2-1: Study Route

SITE DESCRIPTION

The methodological steps in this section are conducted on a 17.4 mile route in New York City that travels from the densely residential Boerum Hill neighborhood of Brooklyn to JFK International Airport. This route was chosen because it lies within a well-connected roadway network, over which several alternate routes could be taken. This makes for a more interesting analysis, as drivers in the area likely base some of their travel decisions on the travel time and travel time reliability of this particular route. The route is also varied, traversing a series of arterials and three major freeways between Boerum Hill and JFK International Airport. The route begins at Atlantic Ave and Flatbush Ave., then travels over the Brooklyn-Queens Expressway (I-278 E), the Queens-Midtown Expressway (I-495 E), and the Van Wyck Expressway (I-678 S), ending near JFK International Airport's cell phone parking lot. This route is shown in Figure 2-1, with the origin identified in white and the destination in black.

To determine which ALK links make up this route, we begin by visually identifying the ALK grids that the route travels through. It is then possible to map all interstate-class links contained in the relevant grids and visually identify the links which make up the route. Upon the completion of this process, we find that the 17.4-mile long route is made up of 102 ALK links. The Grid IDs and Link IDs of these links are labeled with their order within the route and stored.

After the route links have been identified, it is possible to calculate the number of data points recorded for each link. Probe data is sparser during times when fewer vehicles are traveling (i.e., at night), making certain types of time-of-day analysis more difficult. Since each data point contains a timestamp, counts of data points by link and time of day can be obtained directly from the data. The timestamps must be converted from UTC time to local time (EST) (with adjustments made for daylight savings time) before the counts can be interpreted. Data availability on this route during the 11-year period of coverage is displayed in Figure 2-1.

As shown Figure 2-1, data coverage over the route is generally quite sparse, with the most densely covered link-hour containing 71 points. As such, analysis requiring data partitioning, such as comparing weekday and weekend speeds, will likely not yield rich results. The three freeway segments have the best data coverage, while coverage is sparser on the arterials near the origin, the freeway connectors, and the airport roads at the destination. Data coverage is highest in the evenings and around midday. Due to the sparseness of the data, no individual vehicle trips traversed the entire route from beginning to end.

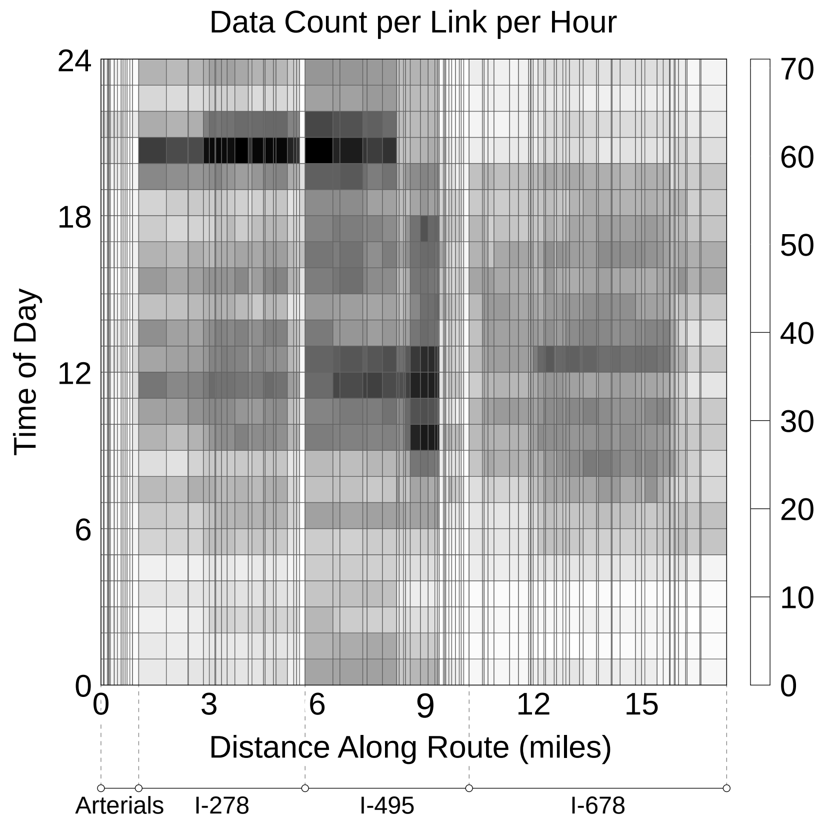

Figure 2-2: Quantity of data analyzed

METHODS

Since there were no travel time records for the entire route, methodologies had to be developed to construct the route travel time distribution piecemeal from the individual link data. The advantage of this approach is that it utilizes the entirety of the dataset, rather than a subset of long trips. Obtaining composite travel time distributions from vehicles that only traveled on a portion of the route is a complex process, primarily because, as this project has shown, travel times on consecutive links often have a strong linear dependence. This linear dependence must be accounted for when combining individual link travel times into an overall route travel time distribution. This is the core methodological challenge of this case study, fully explored in the Use Case chapter. The research team first approached this complex topic by examining probability density functions of speeds on an individual link, the results of which are presented in this section.

To understand the traffic conditions represented in the data set, we can plot time-of-day based speed distributions on a single link. Figure 2-3 depicts hourly probability density functions of speeds observed on the 38th link in the route (near the I-278 / I-495 interchange). From this visual, it is clear that most speeds fall between 45 and 65 mph, with the exception of the PM peak. From 2pm to 7pm, the speeds appear to be bimodally distributed, with a lower modal speed around 10 mph.

With the knowledge that mixed traffic conditions occur during the PM period on the 38th link in the route, we can analyze PM speeds along the entire route. Speed measurements on each link during the 3pm to 8pm commute period were obtained from the *One Monument* data set. To illustrate speed changes along the route in the PM period, the median PM speed for each link is plotted (see Figure 2-4). Each link has multiple speed measurements over the 11-year study period during these hours, so speeds between the 25th and 75th percentile for each link are shaded in gray to indicate the rough extent of each link's PM speed distribution. Speeds appear to dip in the middle of the freeway segments. Median speeds along the route outside of the PM period remain relatively high throughout the freeway segments, indicating PM period congestion.

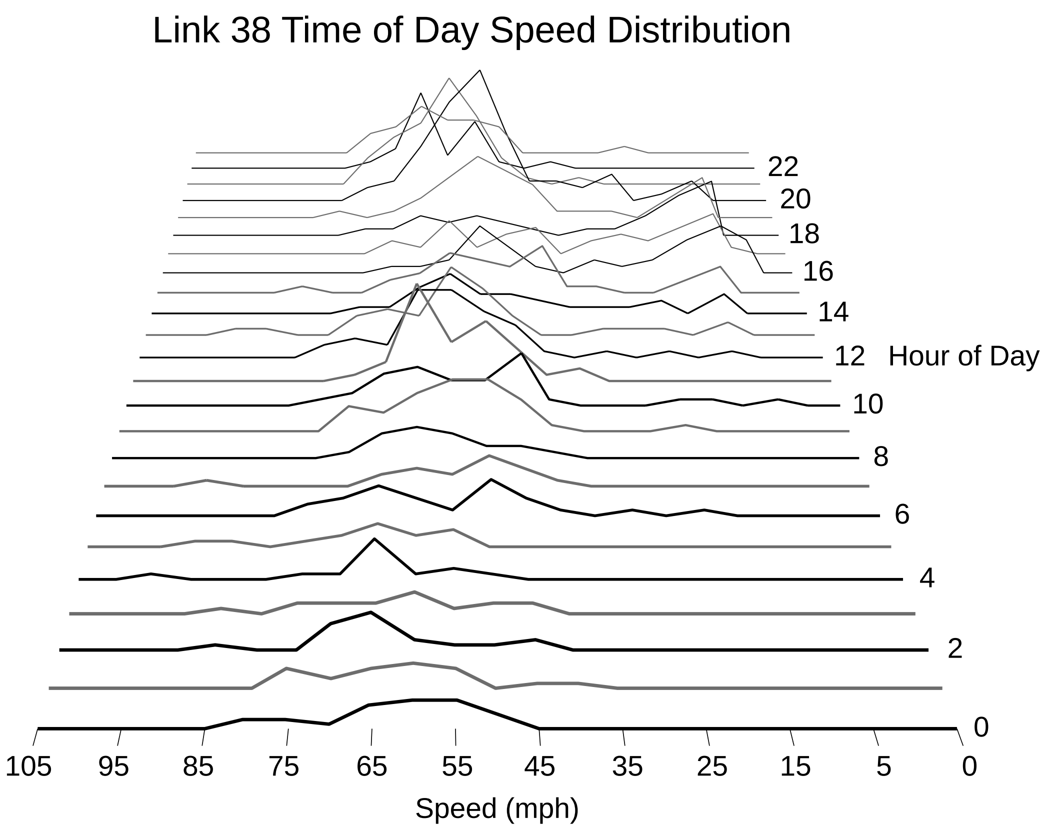

Figure 2-3: Time of day speed distribution on a link

Next we look at how speeds vary across the route throughout the whole day, again considering the entire speed distribution on each link-hour. Speed measurements on each link during each hour of the day are extracted from the *One Monument* data set and the 25th percentile, median, and 75th percentile speeds for each link-hour are computed. The variation of speeds along the route throughout the day is presented in Figure 2-5. Link-hours with no data (mostly at freeway interchanges and toward the end of the route at night) were marked with a speed of zero.

The speed data appears to show three triangular regions in the PM period of each freeway segment. These triangular regions indicate bottleneck regions of low speeds during the PM commute period.

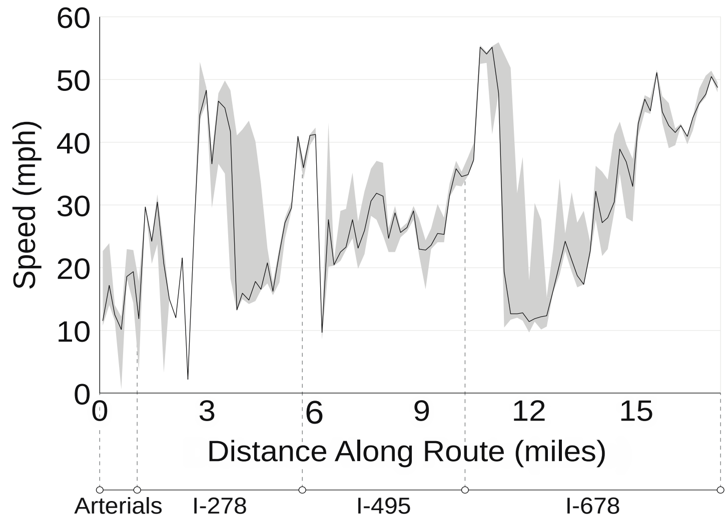

Figure 2-5: Speeds along route

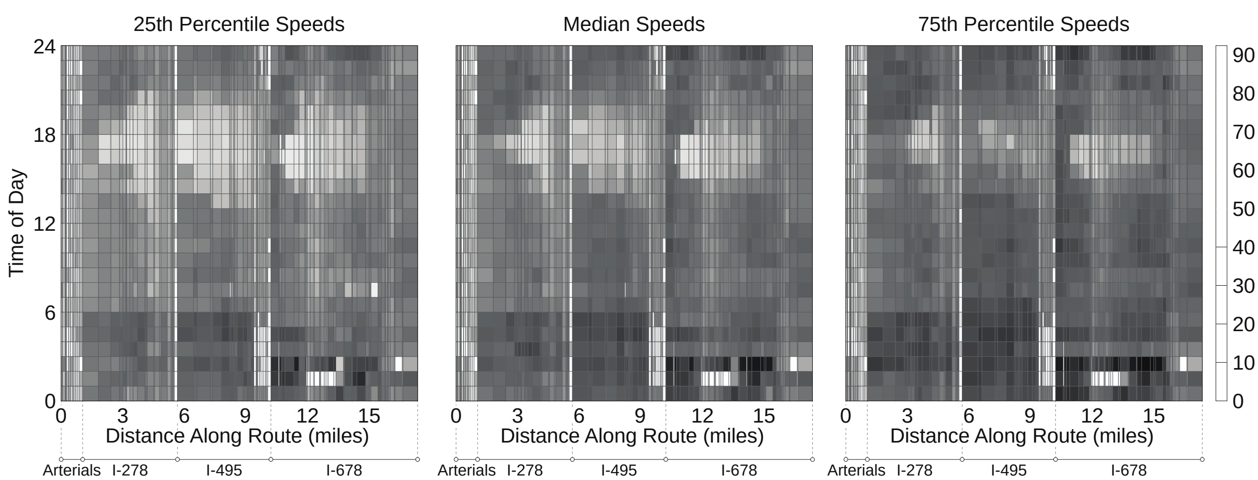

Figure 2-4: Quartile speeds along route by time of day

RESULTS

Using the quartile speeds for each link throughout the day, it is possible to simulate trip trajectories along the route for any slice of the speed distribution. We do this by first choosing a virtual trip start time, and then moving along the route link-by-link, simulating the arrival time at the next link based on the speed and length of the current link. The link speeds used to advance this simulation must correspond to the time of day in the virtual vehicle's trip.

Figure 2-6 shows the trajectory of trips simulated using PM period link mean speeds at 30-minute intervals. This type of time-space contour plot is practical in helping to identify locations or times that experience long travel times and seeing how unreliable conditions affect trips at different times of day. For example, the virtual trip departing at 5pm appears to experience more congested at the beginning of the I-678 segment more than later trips do. This gives it a longer travel time than it would have experienced had it departed 30 minutes later.

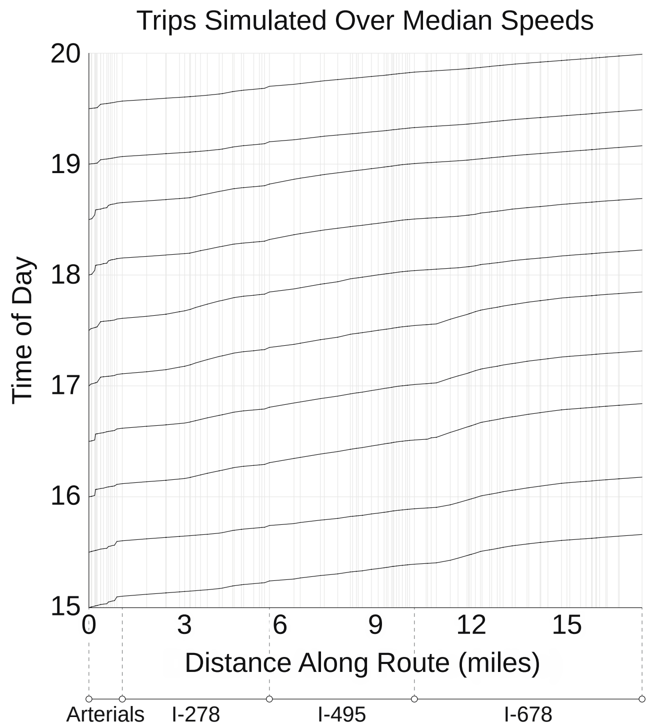

Figure 2-6: Virtual trips simulated over median link speeds

3. USE CASE

A single use case was evaluated in this case study. This use case is a site-specific application of the probe data processing and analysis techniques described in the Methodology section. The motivation for this use case is to generate and compare travel time distributions along a route at different times of day, using only probe data. The methodology chapter of this document describes a technique for simulating trips based on probe speed measurements; however, these simulated trips only apply to a particular slice of the speed distribution (such as the median speed). A more complex approach is needed to measure and illustrate the variation in speeds and travel times on a route at a given time.

This use case demonstrates three methods for obtaining route travel time distributions from probe-based speed data. For continuity, analysis is performed on the route describe in the Methodology chapter. The analysis in each of the three methods is performed on the One Monument data set. For this analysis, the most important variables in the data set are timestamp, speed, trip ID, and indexed position within a trip, if any (many trips are made up of a single point on the route). This yields two types of information useful for travel time analysis: (1) individual vehicle time-stamped link speeds, and (2) individual vehicle link travel times, as derived from the differences in the timestamps of consecutive trip points (for trips with more than one point). The methods differ in how they use these features of the data set to construct the travel time PDFs.

METHOD 1

The first method is the only method to use all available data elements in the One Monument probe data set to construct the route travel time PDF. It uses discrete link speeds as well as trip-based travel times to construct the travel time distribution in different time periods of the day.

Since the data coverage on the arterial links at the beginning of the route is so sparse, analysis is focused on the route beginning with link #17 (and continuing to JFK International Airport). The method is divided into two stages: a preparatory stage and a distribution construction stage.

Preparatory Stage

In the preparatory stage, we consider each link in the route, and identify trips that began on that link and traveled at least one link downstream on the route. The goal of this step is to calculate a link-startpoint to link-endpoint travel time for each multi-link trip in the dataset. Each One Monument data point contains a *LinkOffset* value that indicates the distance along the link that the speed value was taken (for, example 0.5 indicates that the data point was taken at the link's mid point). This trip travel time calculation method uses the data point timestamps to determine the travel time in between each trip's first and last link, and the link speed, length, and offset to extend that travel time to the start point of the first link and the end point of the last link. For a trip that travels from link 1 to link n, the trip travel time equation is:

$$TripTT = \frac{Length_1 * LinkOffset_1}{Speed_1} + \left(Timestamp_n - Timestamp_1\right) + \frac{Length_n * (1 - LinkOffset_n)}{Speed_n}$$

This step results in a set of travel times for each link that trips from that link to some downstream link. The travel times were divided up by time period (AM, midday, PM, and nighttime), and were then assembled into trip travel time distributions for each link and time period.

Distribution Construction Stage

The distribution construction stage builds up the full travel time distribution along the route link by link in four steps. Each iteration of the steps adds the subsequent downstream link into the route travel time distribution. The route travel time distribution is initialized as the travel time distribution on the first link on the route, as computed from all data points on the first link. The following four steps are then carried out sequentially down the route for all links:

1. Compute the travel time distribution for the current link using all data points measured on the link.
2. Add the travel time distribution for the current link to the route travel time distribution computed in Step 4 for the upstream link, assuming independence. To add two independent distributions of data, each point of the first data set must be summed with each point of the second data set. If the size of one dataset is *m* and the size of the other is *n*, the size of the dataset resulting from their sum is the product of the two sizes: *mn*. This is equivalent to convolving the probability density functions of the two independent distributions.
3. Obtain the set of travel times computed in the preparatory stage that end at the current link and merge their adjusted datasets to the dataset of the route travel time distribution computed in Step 2. The adjusted dataset will have been computed in Step 4 for a previous link.
4. For all trips that start at the downstream link, add the route travel time distribution computed in Step 3 to their travel time. This adjusts these travel times such that they represent the travel time distribution between the beginning of the route and the end of the trip.

The resulting travel time probability density functions computed using this method are shown for four time periods in Figure 3-1. The odd multimodal distribution of the 10pm to 12am travel times is due to a proportionally larger number of trip-based speeds than discrete link speeds at night. At other times of the day, the number of link speeds overwhelms the number of trip-based speeds, hiding the effects of individual trips .

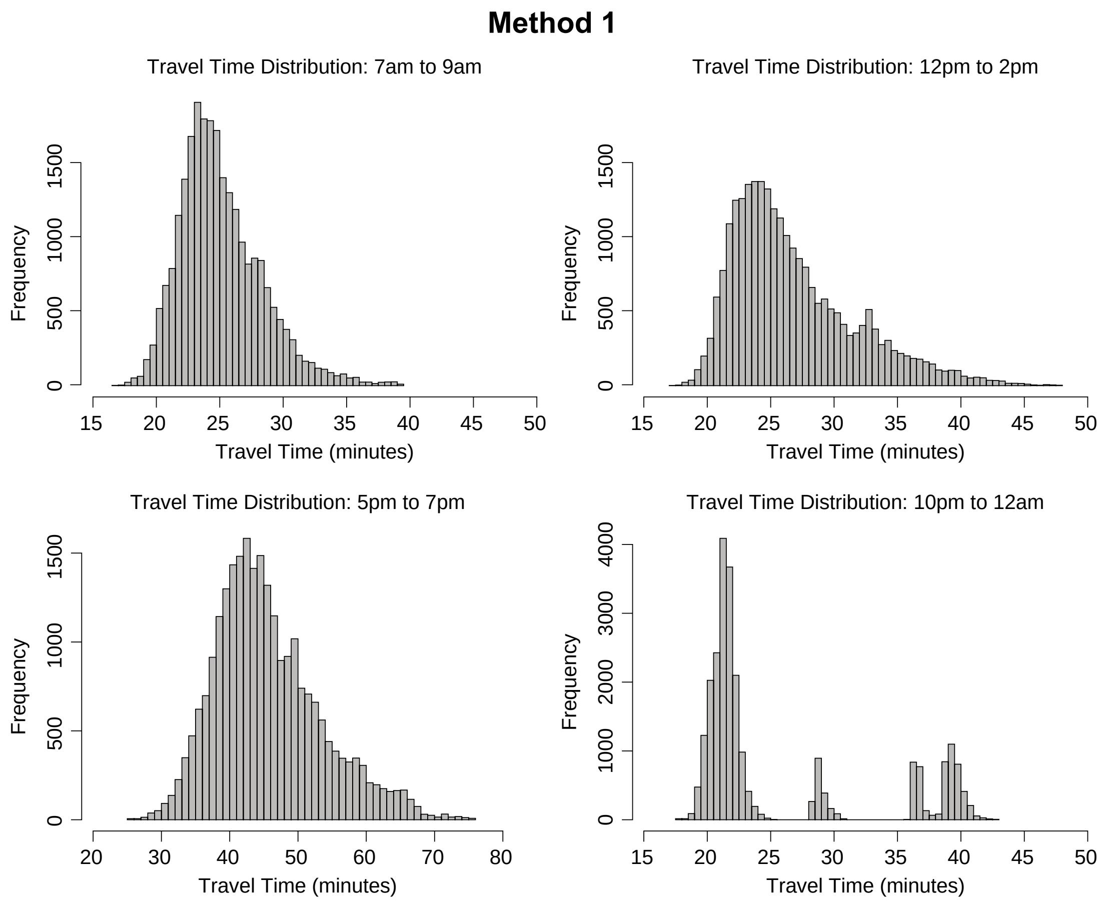

Figure 3-1: Route PDF generation method 1

METHOD 2

The second method for computing route travel time PDFs ignores the linear dependence between consecutive links and directly computes the route travel time distribution as if all link travel times were independent. This method is based entirely on directly-observed link speeds, discarding the timestamp differences between points in the same trip. It works by simply convolving the distributions of travel times on consecutive links down the route. For example, the frequency distribution of travel times on the first link is added to the frequency distribution of travel times on the second link, and so on until a full travel time distribution for the entire route is obtained.

This is the simplest route travel time PDF creation method considered in this case study. Here we treat every single measurement as independent of all others, ignoring all trip relationships between points. As in method 1, we compute travel time distributions for four time periods during the day. With the trip-based travel times discarded, the outlying spikes in the 10pm to 12am travel time distribution are no longer seen. The speeds between 5pm and 7pm appear to be shifted by roughly the same amount as seen in method 1. The 7am to 9am and 12pm to 2pm time periods appear to have very similar bimodality to that generated by method 1.

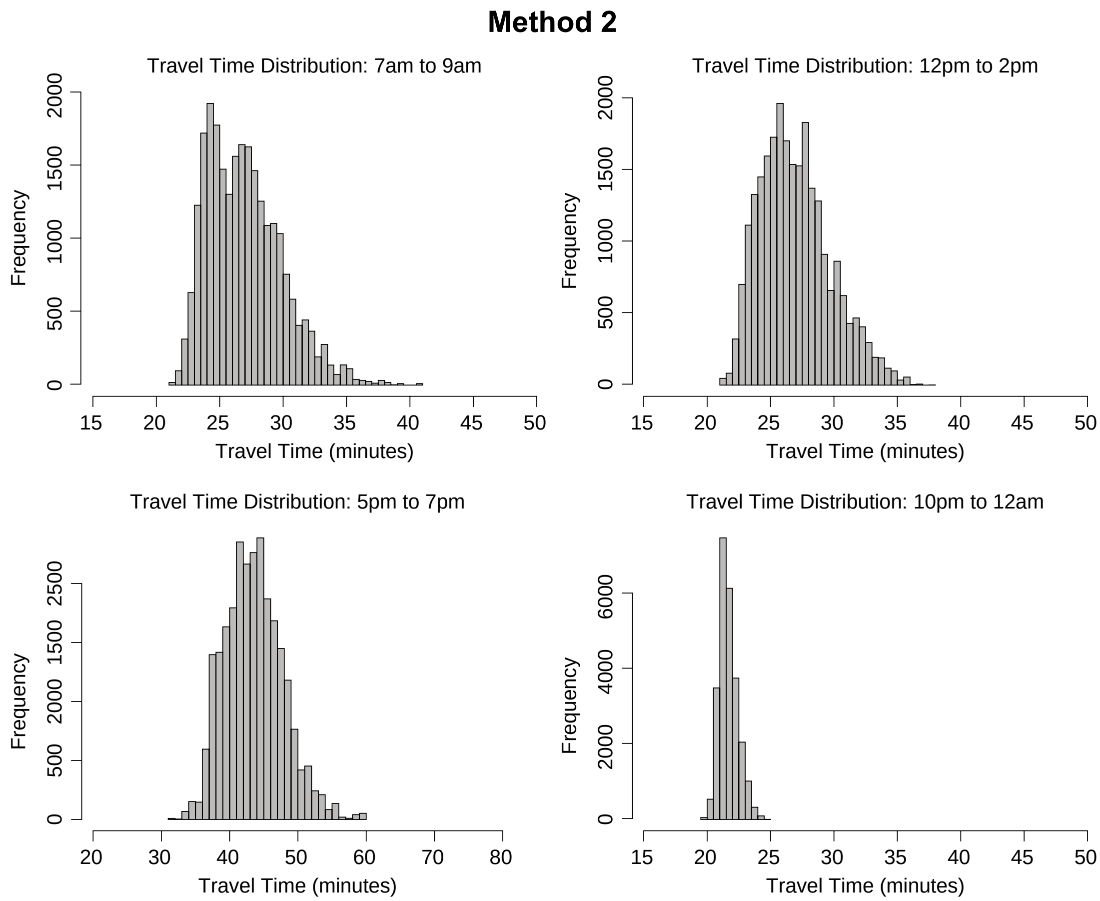

Figure 3-2: Route PDF generation method 2

METHOD 3

The third and final method developed for constructing route travel time PDFs computes and leverages the correlation between speeds on consecutive links within a trip. This method, which only requires speeds measure from trips that traveled on multiple links, uses the fewest One Monument data elements. It builds route travel time PDFs by simulating trips along a route, taking into account the measured data on each link as well as synthesized trips based on observed data and computed incident matrices. It builds up travel times link by link. As with the previous two methods, due to the lack of data on the arterials near the beginning of the route, we begin the route on link #17.

The method begins by computing incidence matrices for each pair of consecutive links. These incidence matrices describe the correlation in speeds between the two links. To construct the incidence matrices, evenly spaced bins are defined to group the speed data for each link. In this use case, 10 bins are used between 0 mph and 80 mph (each bin is 8 mph wide). A 2-D incidence matrix is created for each pair of consecutive links to capture the nature of the speed relationship between the two links within different bins. Speed bins on link #1 are represented in the incidence matrix's rows, and speed bins on link #2 are represented in its columns. Because 10 bins were used in this use case, all incidence matrices are 10 x 10.

Consider an incidence matrix for two consecutive links: link #1 and link #2. The incidence matrix describes the likelihood of a speed on link #2 occurring given a speed on link 1. The entry in the (*m*, *n*) cell of this incidence matrix contains the quantity of link #2 speed measurements that fell into the *n*th bin when the link 1 speed came from the *m*th bin. The counts in the cells of the incidence matrix become synthesized trip points for each observed data point on link #1.

For example, suppose a single link #17 speed observation falls within the 4^{th} speed bin, and the incidence matrix for links #17 and #18 lists two speeds in the 5^{th} bin and three speeds in the 4^{th} bin on link #18 following a 4^{th} bin speed on link #17. This single observed speed on link #17 has resulted in five pairs of speeds across links #17 and #18 (two between it and the 5^{th} speed bin, and three between it and the 4^{th} speed bin). These five speed pairs can be thought of as synthesized trips between the two links since they capture the correlation between speeds on the two consecutive links as well as the observed data. This process is repeated for each observed speed on link #17 and all synthesized trips over the first two links are recorded.

To continue the process on the next pair of links, #18 and #19, the speeds on link #18 resulting from the incidence matrix technique described above (there were 5 such speeds in the example) are combined with the directly observed speeds on link #18. This collection of speeds is then subjected to the same incidence matrix procedure to obtain synthesized link #19 speeds for each speed on link #18 that was either directly observed or synthesized from link #17's directly observed speeds. When the final link in the route is reached in this way, the speed on each link in each synthesized trip can be used to obtain its travel time. The distributions of these travel times calculated at different times of day are shown in Figure 3-3.

Since each preceding link speed generates multiple speeds for the following link, this method generates a large amount of data very quickly. To keep the travel time data set manageable, the growing data set of synthesized speeds was periodically reduced to a random sample whenever it grew too large to efficiently process.

The multimodal pattern seen in the 10pm to 12am data from method 1 is even more pronounced in travel times synthesized with this method. Both of these methods leverage individual trip travel times across multiple links. The low quantity of data at night exaggerates the influence of individual trips on the data, creating these spikes. This method produces very narrow travel time distributions that are offset slightly from those generated by the other two methods. Here, we see travel times shifted 5 minutes earlier during the am and mid day time periods, with dramatically fewer long travel times. The 5pm to 7pm travel time distribution is again the most widely distributed, but is shifted approximately 10 minutes later what is seen for this time period from methods 1 and 2.

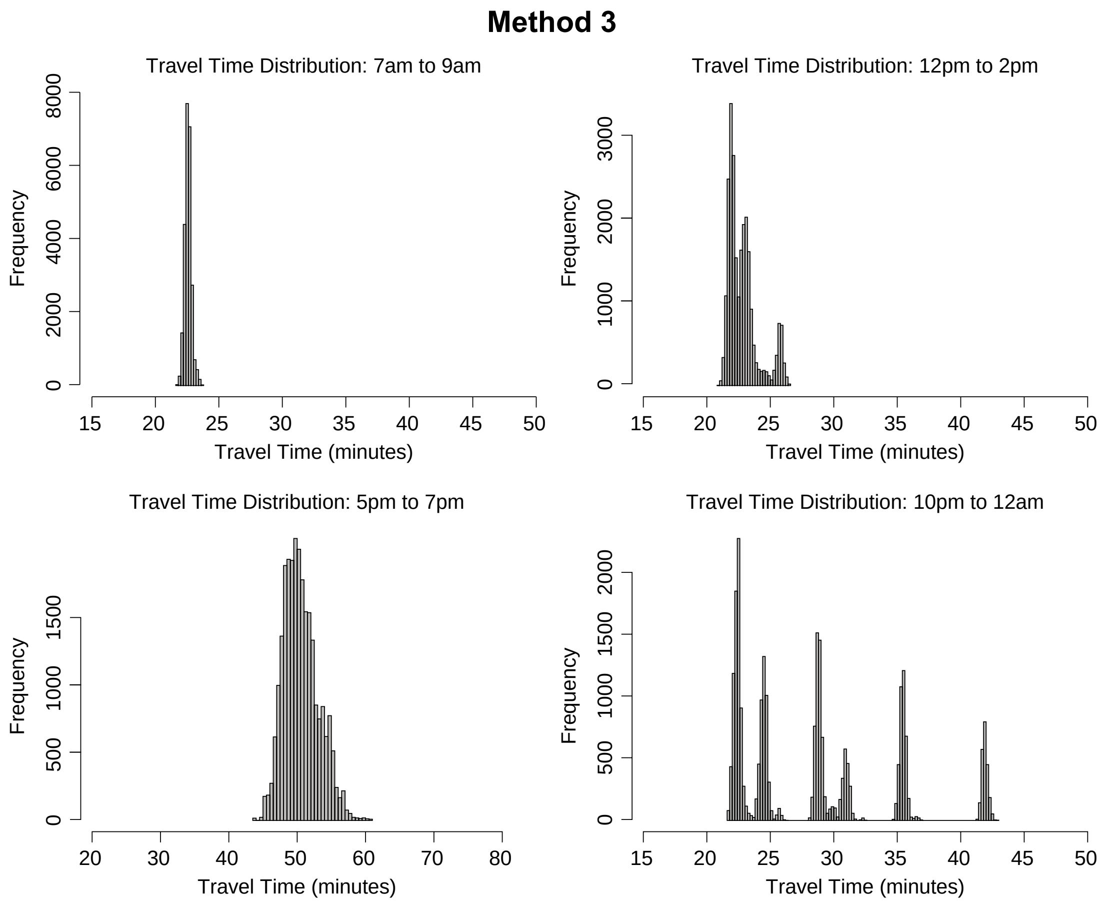

Figure 3-3: Route PDF generation method 3

Conclusions

Each of the three methods presented for assembling route travel time probability density functions from probe vehicle data is enabled by the techniques introduced in the methodology section. Constructing these PDFs requires identification of the data points corresponding to a particular route, separation of data by time of day when possible, and an understanding of the relationships between link speed distributions and route speed distributions. These tools, combined with the research team's findings related to speed correlations between consecutive links within a trip, led to the development of these three PDF-generation methods.

Methods 1 and 2 compared well with each other, while the results of method 3 differed in terms of travel time magnitude and variability. The differences in the shapes of the distributions across methods, particularly in the night-time period when data was sparse, demonstrates the strong influence of the correlations of speeds along consecutive links within a route. With most of the nighttime coverage made by full trips composed of two or more points, the timestamp-based travel times dominated the night-time data set. The modes of these unusually shaped distributions reveal individual trips in the data.

Although results were not validated with a different data source, the probability density functions generated using methods 1 and 2 appear to match expectations. An online trip planner estimates the travel time on this route to be 28 minutes, which generally agrees with the distributions seen here. They resemble typical route travel time distributions, even though no trips were observed traveling along the entire route.

It is possible to extract quantitative travel time reliability metrics from the time of day travel time distributions compiled and presented in this section. Knowing the distribution of travel times on a route enables the data user to compute any reliability metric, such as planning time or buffer time.

4. LESSONS LEARNED

OVERVIEW

This case study demonstrates that it is possible to obtain trip reliability measures based on probe data, even when that probe data is sparse. The travel time distribution for the route is constructed from vehicles that only travel on a portion of the route, and takes into account the linear dependence of speeds on consecutive links. This case study also contributes techniques for creating time-space contour plots based on probe speeds. These contour plots can be made to represent any measured speed percentile, so that contours for the worst observed conditions can be compared with typical conditions.

PROBE DATA CHARACTERISTICS

Much of this case study effort focused on understanding the aggregation steps used to convert data from GPS receivers into link-based speeds. Understanding the way raw GPS data is processed and aggregated is vital for proper interpretation of data elements. It also enables all components of the data set to be utilized to increase the richness of travel time PDFs.

As probe data finds wider adoption for travel time monitoring, it is important for users to understand that data from these sources is still sparse. This sparseness necessitates complex processes for determining travel time distributions on routes of interest. When GPS and other technologies reach a certain penetration rate in the population and more vehicles traverse entire route, the assemblage of route travel time distributions will be simplified. Currently, however, the construction of well-formed PDFs requires that every element in the data set (from speeds on single links to complex travel times across multiple links) should be used to generate the distribution.

Probe data sparseness also increases the minimum level of temporal aggregation that can be supported by the data set. For example, in this case study, the quantity of data was not sufficient to measure route travel time reliability at a granularity of five-minutes, which was the common reporting unit for case studies that relied on loop detector data. Instead, aggregation had to be done at the peak period, multi-hour level. Additionally, in this case study, weekend trips could not be removed from the data set, as there were not sufficient weekday data points to generate full PDFs. Finally, in order to generate the presented results, all data points collected over the 10 year span of the data set had to be used. In practice, this long time-frame does not allow for trend analysis. Transportation planners and operators often require an understanding of how route travel times vary on a day-by-day, week-by-week, and month-by-month basis.

One probe data characteristic that counteracts the sparseness problem is that data coverage is highest during the time periods, and at the locations, where the most vehicles are traveling on the roadway. These are the time periods and locations at which reliability monitoring is the most critical. As probe technologies become more common in vehicles, the availability of data points and route-level trip data will naturally increasing, resulting in richer data sets that can be analyzed at a finer-grained interval than were possible in this case study.