Center for Applied Demography & Survey Research
University of Delaware

Study and Calculation of Travel Time Reliability Measures

by

David P. Racca
Daniel T. Brown

Prepared for the Delaware Department of Transportation

August 2012

Transportation Studies

TABLE OF CONTENTS

Page

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF TABLES (Continued)

Table **Page**

LIST OF FIGURES

LIST OF FIGURES (Continued)

Figure **Page**

Executive Summary

Travel speeds and their corresponding travel times on Delaware's road network provide valuable performance measures of interest to transportation planners and operators. Of related interest is the variability of expected travel times, since lower variability implies greater predictability of travel conditions, and hence greater reliability. Delaware's Department of Transportation (DelDot) provided funding to use the GPS coordinates of state fleet vehicles to explore roadway reliability measures. Though DelDot funded this research, the authors are solely responsible for its design and execution.

Vehicles owned and operated by the State of Delaware are currently monitored using global positioning system (GPS) equipment to assist with location of vehicles, vehicle status, and a range of other applications. There are over 2,000 vehicles on the system, each broadcasting locations, speeds, and vehicle status at two minute intervals whenever the ignition is on. These measures are captured and yield more than 2 million measures per month. The program has been in place since 2007.

This report successfully generates travel time and reliability measures for two corridors in New Castle County. We find that State Fleet Vehicle GPS data is a valuable statewide resource to estimate and monitor travel times, average speeds, and travel time reliability for various times of day, days of week, and seasons. The data could support many applications, including the study of traffic flows at intersections, vehicle routing support, research into the relationship of volume, capacity and speed, and an examination of the relationship between land use and transportation system improvement. .

Producing programs to compile and process the State Fleet vehicle GPS was a large effort and involved a couple weeks of computer processing time for each month of data. However these steps were a starting point which can be improved and automated in the future.

A method of synthetically generating trips based on small segment chains was also developed and provides a method for calculating travel time and variability where less occurrences of a specific trip of interest are available within the source data. The value of the State vehicle GPS data can be vastly extended using this approach allowing for the calculation of mean travel times and travel time reliability measures for any trip/route of interest. Early results suggest that reliability measures could be created for particular areas of Delaware's road network, instead of just a single route with arbitrary beginning and end points.

Though this report relied on State Fleet data, the approaches it developed could be used with any GPS data source. With the proliferation of GPS enabled mobile devices, these methods could be much more important to understand in the future. This research demonstrates how traffic data can be referenced and integrated into existing roadway data.

Advantages of using the State Fleet data are:

- The number of GPS records being provided is very large and growing. The State Fleet data covers most roads in the state, and data is being continually generated each month.
- Initial data collection costs are covered by an existing State project to monitor fleet vehicles. Thus, there are no additional data collection costs.
- Speed, travel time, and reliability data would be built around a publicly available, state owned, road centerline file with no licensing cost.
- For the scale needed for performance measures the data are expected to produce sufficient and reliable estimates through time.
- The raw data and processed versions would be available to support a range of other planning and operations applications.

Disadvantages of using the State Fleet data are:

- Observations come from state-owned vehicles instead of random sample of vehicles on the road. Actual travel times would be expected to be less than observed travel times where drivers regularly exceed the speed limit, such as on limited access highways like I-95 and Route 1. For these types of roads, field counters may be a less biased source of information. However, a brief inspection of the State Fleet data did not seem to yield noticeably slower speeds on other arterials. Of course, measures from State Fleet vehicle data should be compared to other measures to understand any systematic differences.
- Two minute intervals between GPS measures can introduce errors that have to be screened in post processing steps. Interpolation is often necessary, and future research should test whether this step adds a significant source of bias. Of course, the benefits of higher frequency data come with the tradeoff of an increased data management burden and longer processing times.

Focused speed studies on specific road links can provide controlled measurement for particular routes. But when the State Fleet data are compared to current methods of obtaining travel times, the fleet data proves to be an extensive and powerful resource. Other sources of speed data are available from Bluetooth technology or commercial groups (e.g. TomTom and Navteq). These alternative data sources have the advantage of observing more vehicles per route and/or more frequent measurements per vehicle. Though some commercial products have already implemented web based interfaces and tout a very high number of observations, such data likely comes with proprietary restrictions on use and dissemination. Moreover, purchasing network wide summaries or processed source data directly from commercial groups may prove costly. State Fleet vehicle is free and already owned by the State of Delaware.

Introduction

Actual speeds and travel times on Delaware's road network provide valuable performance measures that can be used for transportation operations and planning. The public receives benefits not only from reducing congestion, but also by increasing predictability of travel times. For instance, a trip to work may normally take 30 minutes on average, but may take 45 minutes or longer due to such variable factors as congestion or accidents. Reducing the frequency of these random delays makes each trip more predictable, thereby benefiting the public. This research is an exploratory analysis of using GPS data from the State's fleet of vehicles to calculate performance measures of reliability.

In order to evaluate the performance of Delaware's roadways, technicians need to know traffic volumes, roadway capacity, travel speed, and incidents. Existing technology in place on Delaware's roads provide good measures of traffic volumes throughout the day in many areas. There are considerably less measures of speed. A relatively small number of devices measure traffic speed at fixed points in the roadway network. An ongoing study by the Delaware Center for Transportation, "Application of Global Positioning System (GPS) to Travel Time and Delay Measurements"[1] measures speed and travel time delay for each peak period (AM/PM) for major roadways in Delaware. Unfortunately, the study only creates two measurements per daily peak per year. More observations are needed to create performance measures of reliability.

[1] http://www.ce.udel.edu/dct/Publications_TrafficITS.html

Vehicles owned and operated by the State of Delaware are currently monitored using GPS equipment to assist with vehicle location, status, and a range of other applications. There are over 2,000 vehicles on the system, each broadcasting locations, speeds, and vehicle status at two minute intervals whenever the ignition is on. More than two million measurements are taken per month for roadways across the State, and the program has been in place since 2007. Access to this data has been given to the authors by the Delaware Office of Budget and Management, who direct the fleet monitoring program.

The main goals of this research are:

1. Examine the viability of using State Fleet vehicle data as an ongoing source of speed measurements on Delaware's road network.
2. Use State Fleet data to develop travel time reliability measures for two test corridors.
3. Previous research conducted by our Center developed a linear referencing system to integrate traffic data from GPS devices into a functional representation of Delaware's road network. This report demonstrates the benefit of that referencing system by combining multiple, independent datasets to one, highly detailed dataset.
4. Develop experience managing large traffic data sets to accommodate the increasing amount of information coming primarily from GPS-enabled devices.

Data

In this section, we review the State Fleet vehicle data underpinning our research. Important issues such as coverage, processing steps, and referencing techniques are detailed here.

Overview of the State Fleet Vehicle GPS Data

Over 2,000 State of Delaware public vehicles are equipped with GPS devices that broadcast diagnostic information every two minutes while in operation. These broadcasted signals include a time and date stamps, longitude, latitude, instantaneous speed, and the average speed since last measure. Each month, millions of measurements can be compiled to make up tens of thousands of trips. The State of Delaware works with fleet management company, Networkfleet (www.networkfleet.com) for the fleet tracking system. Using Networkfleet software, historical and real time information can be accessed, maps of the real time location of vehicles can be viewed, and summary reports can be generated.

Figure 1 GPS Measurements near or on Old Baltimore Pike

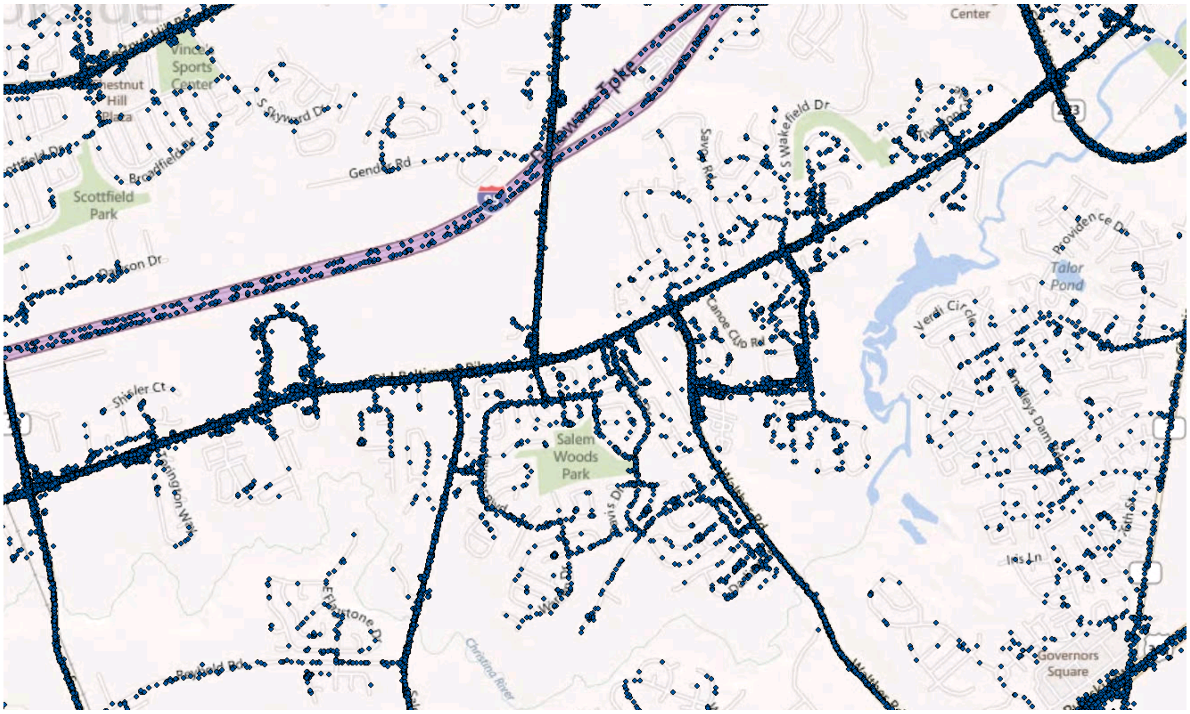
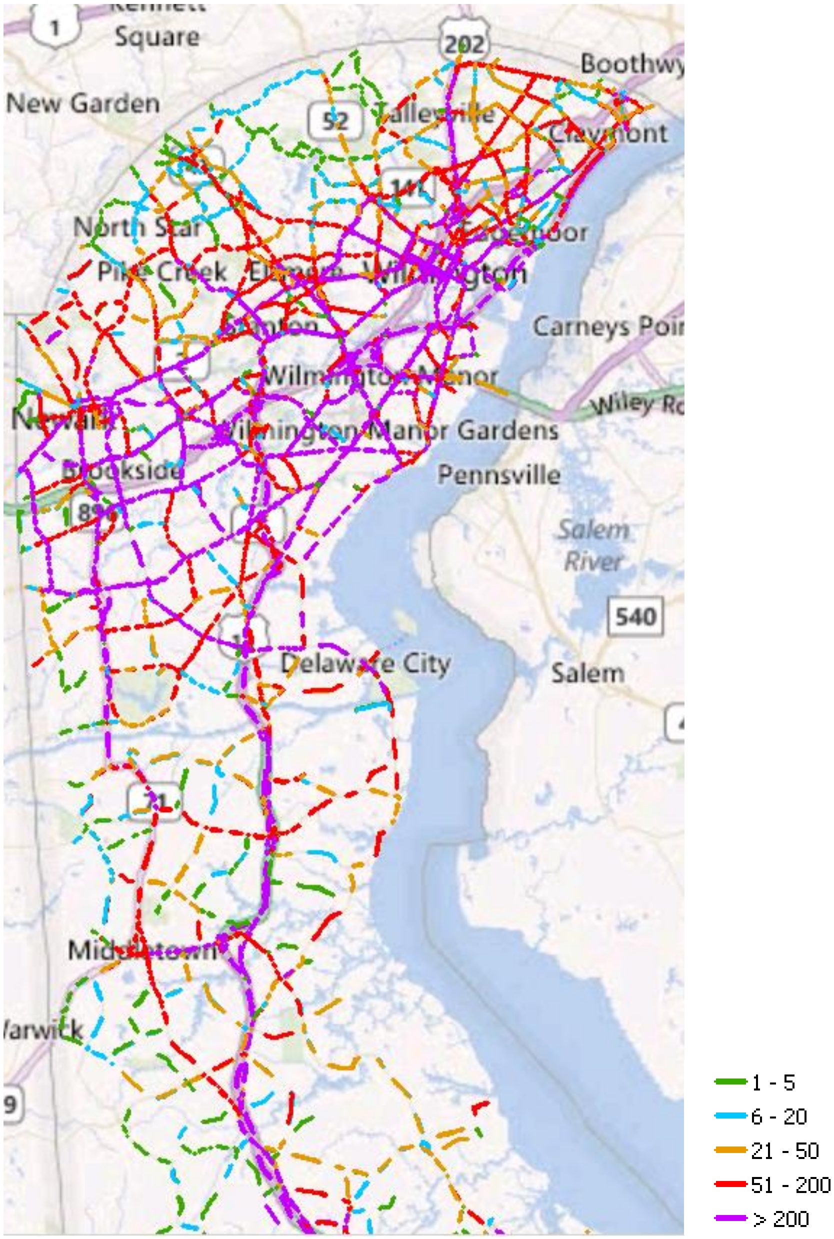


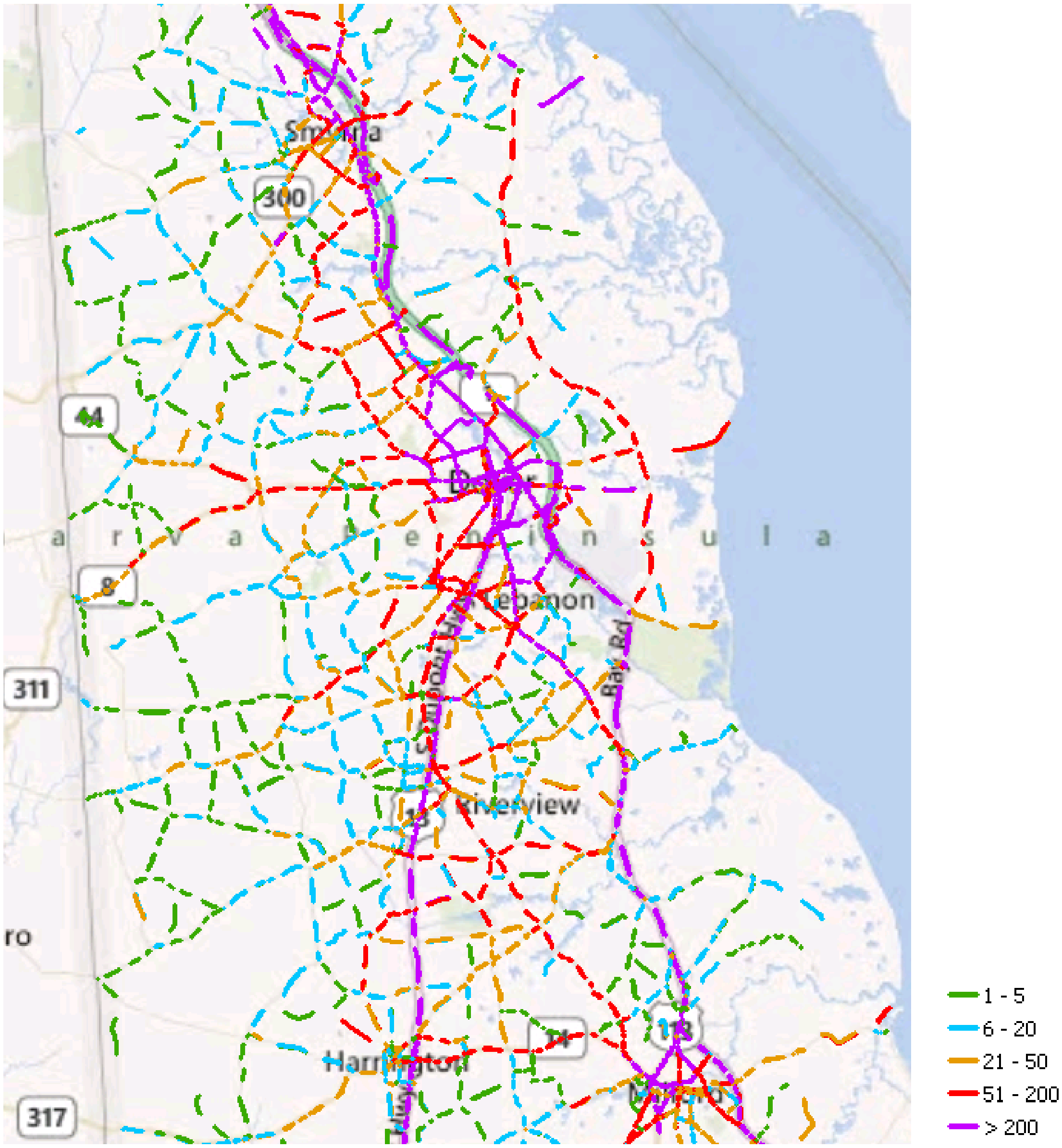
Figure 1 is a map of the main dataset used in this research. Each point on the map represents a single GPS reading from a state fleet vehicle on or near Old Baltimore Pike during one month. These GPS points are associated with portions of the roadway between intersections, which we call "road links", and are identified with specific trips of each vehicle. The entire Networkfleet dataset has similar coverage across the State. Figure 2 through Figure 4 shows one month of data (October 2011) for the links on the major roads in each county. For many roads, one month's worth of data provides hundreds of measures.

Figure 2 Number of GPS Measurements for each Major Road in New Castle County



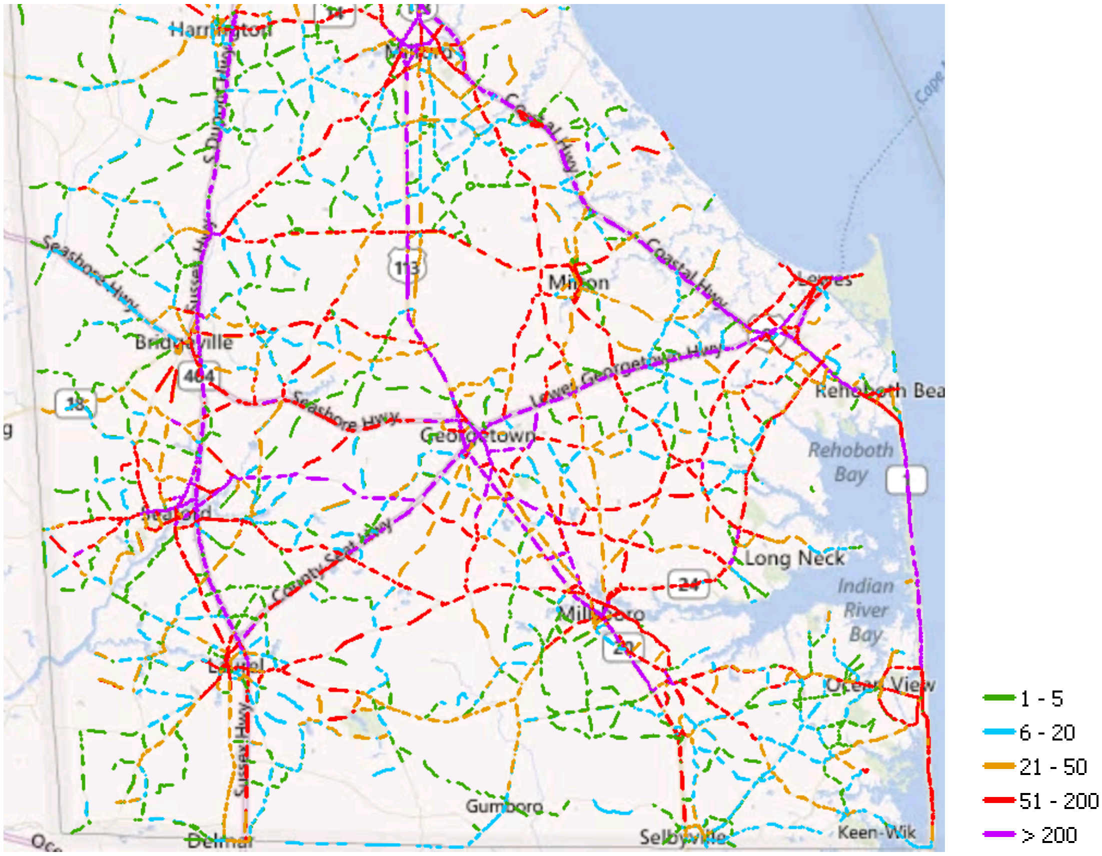
- For the month of October 11, New Castle County, State Fleet GPS Vehicle Data

Figure 3 Number of GPS Measurements for each Major Road in Kent County



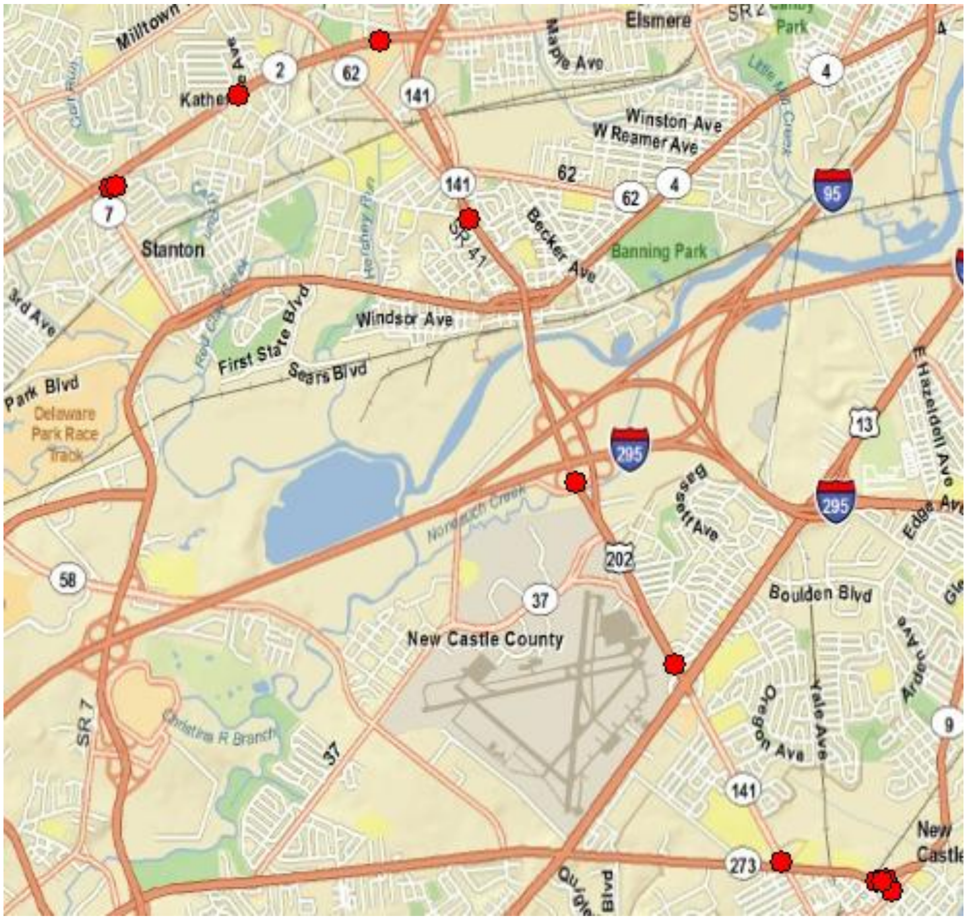
- For the month of October 11, Kent County, State Fleet GPS Vehicle Data

Figure 4 Number of GPS Measurements for each Major Road in Sussex County



- For the month of October 11, Sussex County, State Fleet GPS Vehicle Data

Figure 5 GPS Measurements of a Single Trip



Each trip in the data set is defined by the time a driver turns on the ignition to the time the driver turns off the ignition. Measurements are taken every two minutes during each trip. A sample of what a trip would look like is shown in Figure 5, where the red dots are the location of actual GPS measurements. With the current GPS equipment installed, Netfleet estimates that 90% of the reported positions are within 8 meters of their actual position.[2]

[2] The positional accuracy of the GPS readings could introduce errors if the reading is associated with the wrong road. However, this was not a major problem. Post processing steps that examine the GPS readings, expected routes, and speeds generally identify such errors. These errors can usually be corrected during the trip mapping process.

Processing State Fleet Vehicle GPS Data

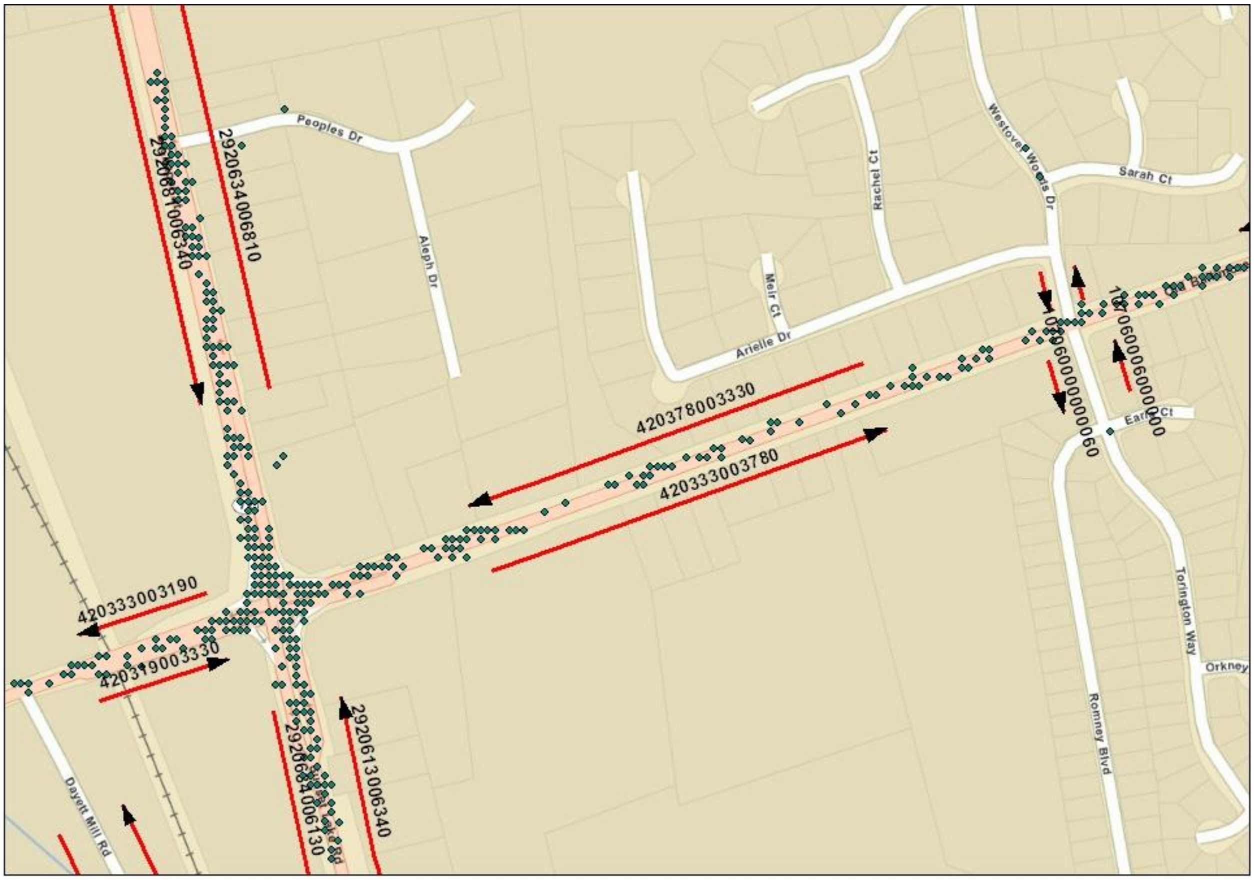
Historical data is obtained from Networkfleet by writing java programs to request data from their web services, which is then downloaded into xml formats. Compiling, examining, and processing this xml data into a useable dataset was very time intensive. Batch programs were written which convert the raw data into a tabular format for each observation.

Though the spatial component of this dataset is interesting, the primary goal of this research requires viewing the data as trips over defined links, covering particular origins and destinations. In order to make this point-to-link relationship, the data had to first be processed. The main processing steps were:

- Capture historical GPS data by querying Networkfleet for Delaware vehicles.
- Process GPS XML response data
- Create GPS point databases and GIS files
- Extract and associate GPS points with particular trips taken through time
- Build a trip and link based version of the GPS data
    - Estimate the path taken between GPS readings
    - Associating particular point measures with a particular road link and direction
- Where portions of a particular trip include road links with no actual measurement, interpolate speeds estimates and travel times between GPS measures
- Screen the data for errors and anomalies.

The end result of this process maps each observation to a particular road link and direction, as shown in Figure 6. The specific links and directions which a particular vehicle travels constitute its trip. The origin, destination, and intermediate links of each trip are collectively referred to as a route. Many trips may be made over one particular route.

Figure 6 March 2010 State Fleet GPS Points, Old Baltimore Pike



The processing was done using with tools that included java programming, SPPS statistical software, Excel spreadsheets, ARCGIS Desktop (ESRI), and ARC Workstation (ESRI). Processing the most likely path between GPS readings and building trip and road link databases was done using ARC Workstation (ESRI) and ARC AML batch coding. ARC AML coding is a higher level programming language that allows for complete control using link and node features. These initial steps took considerable amounts of computer processing time. Two computers dedicated to processing one month's worth of data took approximately two weeks. We anticipate that this processing time could be greatly reduced in the future if programming was streamlined and additional computing resources were employed.

Imperfect GPS positional accuracy, roughly +/- 25 feet, was not a substantial source of error. Steps in the above programming ensured that each GPS reading was associated with the correct road link and direction by examining subsequent links and nodes mapped during that trip. As illustrated by the sample trip in Figure 6, it is generally possible to construct the path taken between readings with a high degree of certainty.

Because signals were broadcast from GPS devices every two minutes, speed fluctuations between those signals are unknown. Therefore, average speeds were estimated by dividing the distance travelled between two subsequent GPS readings by the change in time between those measures (i.e. two minutes). That average speed was apportioned to any links that were completely traversed between adjacent GPS readings. For links that were partially traversed between adjacent GPS readings, average speeds were calculated as the sum of any interpolated distances on each link divided by the sum of the corresponding interpolated times on that link.

For each link, we expect that the average of interpolated speeds will approach the average of actual speeds taken over many observations. Of course, it would be desirable to compare interpolated values with actual values from other more detailed GPS data in the future. More frequent GPS observations will improve the accuracy of speed measurements over smaller road segments.

Referencing GPS and Transportation Data

Perhaps the most important issue in processing and managing traffic data is a standard means for referencing data to the transportation network. An immense amount of traffic data is being collected by field devices, traffic impact studies, permanent counts, vehicle probes, and other means. Moreover, that information comes in different formats, so integrating all of this data requires a consistent manner of referencing.

Typically, data points reference a GIS-based model of the road network that uniquely identifies each road segment or collection of segments.[3] How portions of the road are identified in these files is a crucial step in relating multiple data sources together. DelDot has historically used a linear referencing system (LRS) which groups continuous road segments into a single route and then assigns each of those segments a beginning and end mile marker.[4] In previous research, CADSR created a standard referencing scheme based on the Delaware LRS which identifies each link, milepoint information, and direction of traffic flow.[5]

The standard referencing scheme was particularly useful in this project. Processing outputs can be referenced with any file using the DelDot Centerline linear referencing system. GPS readings can therefore be mapped directly to a route and milepoint, independent of how links were constructed (segmentation scheme) or the cartographic representation. The referencing system also provided the ability to more accurately determine distances within links. The direction of each vehicle is identified in the standard referencing scheme, and direction is important when estimating trip time.

[3] The model used most often is the DelDot Centerline File.
[4] A discussion of linear referencing systems is beyond the scope of this report. In practice these systems offer another level of processing capability that is less dependent on any segment's cartographic representation in a particular road network model.
[5] Further discussion of the standard and its application can be found at www.cadsr.udel.edu.

Processing GPS data required that we identify which links the vehicle had travelled between readings, particularly in the case where subsequent GPS readings span two or more links. This implied that the centerline file we used during processing had to be routable, meaning that its connection of neighboring links and allowable turning movements had to reflect actual traffic patterns. The statewide centerline file that CADSR has developed based on the DelDot Centerline file and the standard identification system it employs is routable.

In principle, any road centerline file for Delaware that supported the LRS could be used to reference the traffic data, and DelDot has been working to develop its own standard routable centerline file over the last few years. Integrating the many historical and emerging data sources will require some standard method of referencing these sources to the road network. Commercial products, such as those available thru TomTom or Navteq, process their data based on proprietary digital maps. These products use their own identification system and must be licensed. At this time, integrating traffic data from a variety of sources is very limited due to these types of restrictions.

Once the GPS data has been prepared and processed, it is ready for direct analysis via queries. The next section of this report demonstrates how GPS data can be used to analyze the reliability of particular routes.

Travel Time Reliability

In this section we apply the GPS records to travel time reliability in Delaware. First, we discuss various measures of Travel Time reliability. Then we perform two case studies to evaluate performance measures for two particular routes in New Castle County. The first case study examines those trips that travelled on Old Baltimore Pike between Route 896 and Route 273. The second case study examines those trips that travelled on Route 40 between Route 896 and Route 13.

Definition of Travel Time Reliability

The US Department of Transportation Federal Highway Administration (FHWA) was the primary resource used to identifying useful measures of travel time reliability.[6] They define reliability to be "the consistency or dependability in travel times, as measured from day-to-day and/or across different times of the day." In some corridors travel times can vary greatly from expected travel time averages depending on weather, traffic incidents or other reason. If commuters planned their trips strictly on the median travel time they would be late half the time and early the other half of the time. Since drivers effectively pay a cost when arriving too early or too late, having reliability roadways is valuable. In fact, if public perceptions are most often formed during the most extreme experiences, reliability measures may better reflect roadway performance than simple averages of travel time.

[6] http://ops.fhwa.dot.gov/publications/tt_reliability

The FHWA has created four measures of travel time reliability. Rather than statistical measures of variability such as the standard deviation and the coefficient of variation, the FHWA measures are more readily understood by non-technical audiences and related to everyday experiences.

90th or 95th percentile travel times

For any particular route, 90% or 95% of the measured trips take less than this time.

Buffer Index

For a particular route, this describes the extra time that travelers must add to the average travel time to ensure that they will not be late. The buffer index is computed as the difference between the 95th percentile and the average travel time divided by the average travel time. The 95th percentile represents a worst case scenario, and the difference between it and the mean is the extra time a traveler should allow so as to not be late 95% of the time. This difference is called the “Buffer Time”.

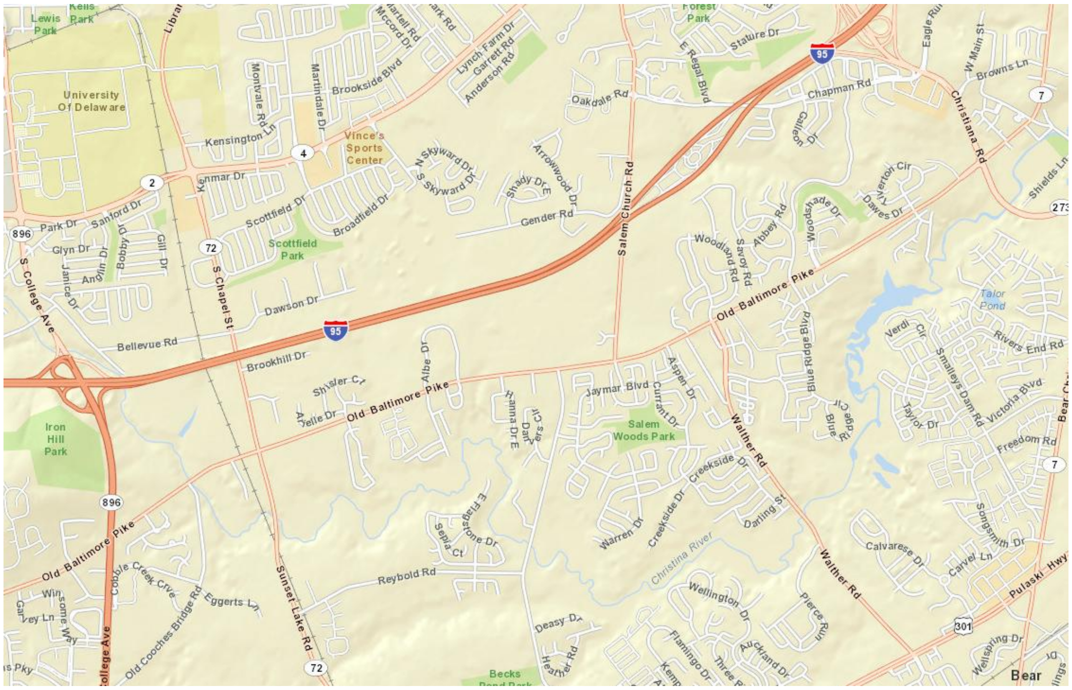
Planning Time Index

The total travel time that should be planned when and adequate buffer time is included. It compares near worst case travel time to travel time in light or free-flow traffic. It is computed as the 95th percentile travel time divided by the free-flow travel time.

Frequency of Congestion

Typically expressed as the percent of days or time that travel times for a particular route exceed X minutes or travel speeds fall below Y miles per hour.

Figure 7 Section of Old Baltimore Pike Under Analysis, without Links



Case Study: Old Baltimore Pike

The first case study analyzed in this report is the Old Baltimore Pike corridor between Route 896 and Route 273 (see Figure 7 and Figure 8). While any link on the corridor may contain several hundred measurements, the initial focus was to identify trips in the GPS database that traversed the entire portion of the study corridor. For the Eastbound path (shown with green directional links in Figure 8), 247 trips were identified in a year and a half of state fleet data. A portion of the relevant origin-destination table is exhibited in Table 1.

Figure 8 Section of Old Baltimore Pike Under Analysis, with Links

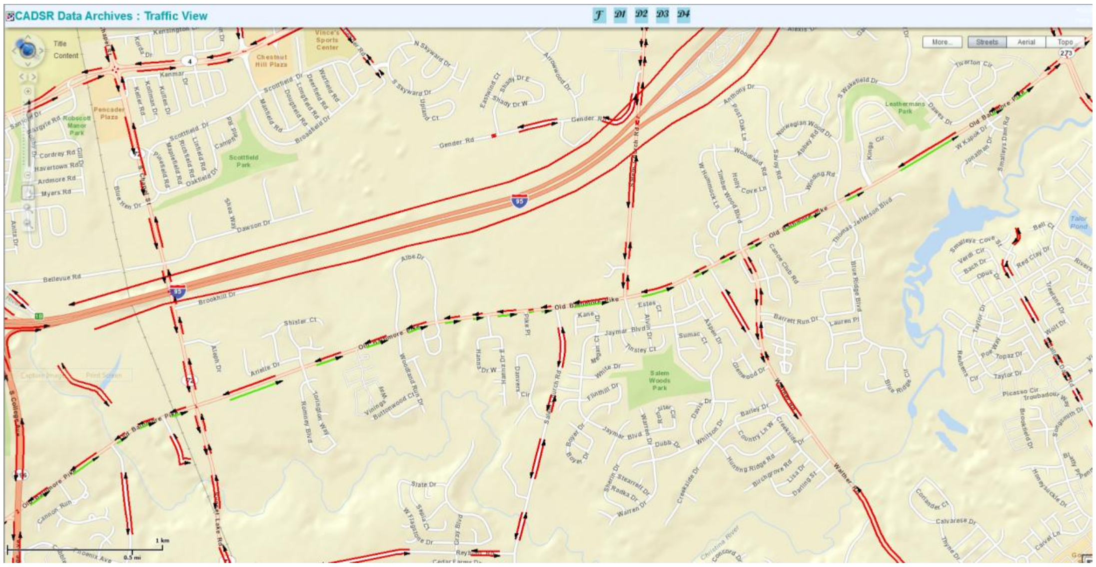


Figure 9 Beginning and End Links for Eastbound Trips on Old Baltimore Pike from Rt 896 to Rt 273

Beginning Link of Trip

End Link of Trip

Table 1 Example of an Origin-Destination Table for Eastbound Trips on Old Baltimore Pike: Rt 896 to Rt 273

tripid	travtime	blrsid	year	bmonth	bday	bhour	bmin	bsec	ptflagb	btrav	bspd	begadj	elrsid	emonth	eday	ehour	emin	esec	ptflage	endadj
3328	8.14	00004202490 02700S	11	1	6	10	33	45	interp	29	25	.28	0000420681006990S	1	6	10	41	30	actual	.11
10228	6.42	0000420249002700S	11	1	25	13	9	46	interp	20	37	.28	0000420681006990S	1	25	13	15	54	interp	.00
11073	6.79	0000420249002700S	11	1	19	12	30	0	actual	17	43	.27	0000420681006990S	1	19	12	36	31	interp	.00
42987	6.72	0000420249002700S	11	1	4	11	19	4	actual	18	40	.22	0000420681006990S	1	4	11	25	34	interp	.00
51142	8.23	0000420249002700S	11	1	10	18	5	54	interp	30	24	.28	0000420681006990S	1	10	18	13	51	interp	.00
59014	8.70	0000420249002700S	11	1	11	18	40	57	interp	30	24	.33	0000420681006990S	1	11	18	49	19	interp	.00
70866	9.94	0000420249002700S	11	1	7	9	17	3	interp	24	30	.33	0000420681006990S	1	7	9	26	18	actual	.36
70958	8.91	0000420249002700S	11	1	18	8	54	25	actual	24	30	.31	0000420681006990S	1	18	9	3	1	interp	.00
71469	10.30	0000420249002700S	11	1	16	18	16	22	interp	35	21	.38	0000420681006990S	1	16	18	26	17	interp	.00
87009	7.12	0000420249002700S	11	1	13	13	37	8	interp	25	29	.42	0000420681006990S	1	13	13	43	50	interp	.00
100452	7.32	0000420249002700S	11	1	25	13	18	31	interp	16	46	.33	0000420681006990S	1	25	13	25	30	interp	.00
131362	8.05	0000420249002700S	11	1	6	14	43	47	interp	46	16	.27	0000420681006990S	1	6	14	51	34	interp	.00
162186	7.40	0000420249002700S	11	1	17	8	22	35	interp	33	22	.85	0000420681006990S	1	17	8	29	8	interp	.00
176329	8.40	0000420249002700S	11	1	19	14	33	43	interp	20	36	.23	0000420681006990S	1	19	14	41	53	interp	.00
1028106	7.57	0000420249002700S	11	2	9	10	42	59	interp	20	36	.23	0000420681006990S	2	9	10	50	19	interp	.00
1039916	10.05	0000420249002700S	11	2	28	16	26	44	actual	20	37	.38	0000420681006990S	2	28	16	36	24	interp	.00
1066069	7.02	0000420249002700S	11	2	11	12	57	2	actual	20	37	.40	0000420681006990S	2	11	13	3	39	interp	.00
1067161	8.75	0000420249002700S	11	2	19	12	8	8	interp	21	35	.35	0000420681006990S	2	19	12	16	32	interp	.00
1071826	8.52	0000420249002700S	11	2	25	13	18	14	interp	25	29	.93	0000420681006990S	2	25	13	25	49	interp	.00
1117795	8.68	0000420249002700S	11	2	28	18	24	8	actual	20	36	.16	0000420681006990S	2	28	18	32	39	interp	.00
1185042	8.47	0000420249002700S	11	2	15	15	16	15	interp	29	25	.30	0000420681006990S	2	15	15	24	25	interp	.00
1230389	8.87	0000420249002700S	11	2	12	9	53	52	interp	24	30	.30	0000420681006990S	2	12	10	2	26	interp	.00
1237204	9.27	0000420249002700S	11	2	11	12	37	24	interp	33	22	.38	0000420681006990S	2	11	12	46	17	interp	.00
1244540	8.82	0000420249002700S	11	2	17	11	35	20	interp	21	35	.53	0000420681006990S	2	17	11	43	37	interp	.00

The beginning and end links of this particular route are shown in Figure 9 and in Table 1 under the "blrsid" and "elrsid" columns. Beginning and end times for the trip are also shown in the table. For each link in the chosen path only some measures are actual readings from the GPS devices. Speed values between two GPS measures are assigned the average vehicle speed between those measures.

Because the first and last GPS readings of a trip do not align perfectly with link endpoints, we adjusted each trip's travel time ("begadj" and "endadj") to reflect the hypothetical time it would take if those links were completely traversed. To make that adjustment, we partitioned each of the first and last links into two sections. One section was actually traversed during the trip, and the second section was not.[7] We assigned the section that was traversed with the interpolated speed between actual GPS readings, as usual. In some cases the adjustment was further studied by examining the distributions of actual speeds on the end segments. We assigned the section that was not traversed with the instantaneous speed of the first or the last GPS reading. Hypothetical travel times were now directly computable.

[7] Actual trips do not traverse the distance between the beginning of the first link to the position of the first GPS reading. They also do not traverse between the last GPS reading and the end of the last link.

Table 2 Example of an Eastbound Trip on Old Baltimore Pike: Rt 896 to Rt 273, by Link

year	month	day	lrsid	rtid	ptflag	speed	hour	minute	second
11	1	6	0003410343004180S	3328	actual	47	10	31	12
11	1	6	0003410418004840S	3328	interp	47	10	32	2
11	1	6	0003410484005230S	3328	interp	47	10	32	33
11	1	6	0003410523005570R	3328	actual	45	10	33	16
11	1	6	0000420249002700S	3328	interp	25	10	33	45
11	1	6	0000420270002950S	3328	interp	25	10	34	22
11	1	6	0000420295003050S	3328	interp	25	10	34	36
11	1	6	0000420305003190S	3328	interp	25	10	34	56
11	1	6	0000420319003320S	3328	actual	0	10	35	19
11	1	6	0000420333003780S	3328	interp	29	10	36	16
11	1	6	0000420378004000S	3328	interp	29	10	36	43
11	1	6	0000420400004180S	3328	interp	29	10	37	6
11	1	6	0000420418004340S	3328	actual	35	10	37	22
11	1	6	0000420434004470S	3328	interp	37	10	37	34
11	1	6	0000420447004550S	3328	interp	37	10	37	42
11	1	6	0000420455004700S	3328	interp	37	10	37	57
11	1	6	0000420470004740S	3328	interp	37	10	38	1
11	1	6	0000420474004820S	3328	interp	37	10	38	9
11	1	6	0000420482005080S	3328	interp	37	10	38	36
11	1	6	0000420508005340S	3328	interp	37	10	39	1
11	1	6	0000420534005460S	3328	actual	46	10	39	27
11	1	6	0000420546005590S	3328	interp	42	10	39	38
11	1	6	0000420559005710S	3328	interp	42	10	39	47
11	1	6	0000420571006020S	3328	interp	42	10	40	14
11	1	6	0000420602006150S	3328	interp	42	10	40	27
11	1	6	0000420615006690S	3328	interp	42	10	41	13
11	1	6	0000420669006810S	3328	interp	42	10	41	21
11	1	6	0000420681006990S	3328	actual	46	10	41	30
11	1	6	0000420699007090L	3328	interp	7	10	42	40
11	1	6	1000060740007740S	3328	actual	40	10	43	31
11	1	6	1000060774007760S	3328	interp	9	10	43	39

The processing of the path taken by vehicles captures the link sequence of the trip along with estimated times and speeds. Table 2 is an example of the exact path for a particular trip (route id 3328). Atypical trips, where for instance a car deviates from the desired path or where a service vehicle might stop for several minutes, are detected and removed during post processing to ensure proper link order and reasonable speed patterns. During the post processing step, the speed and route data above was merged with GIS-enabled volume counts from the TMC. Performance measures of reliability were now calculable.

Table 3 Travel Times and Reliability Measures for Eastbound Trips on Old Baltimore Pike: Rt 896 to Rt 273

Trip Category	Mean Time	95%	Buffer Index	Planning Time Index	Avg Hourly Volume	#Trips
All Trips	8.3	10.1	.22	1.3	323	247
Peak (AM/PM)	8.6	10.9	.27	1.4	639	52
AM Peak (7-9)	**8.6**	>10.8	.27	>1.4	601	18
PM Peak (16-18)	8.6	10.5	.22	1.4	676	34
Midday (9-16)	8.2	9.9	.21	1.3	491	120
Weekend (all day)	8.3	10.0	.27	1.3	270	50
Off Time (<7 & >17)	7.7	9.8	.24	1.3	147	25
All Trips 2010+Jan11	8.3	10.3	.24	1.4	317	115
All Trips 2011-Jan11	8.3	9.9	.19	1.3	325	132
Peak 2010+	8.9	12.9	.45	1.7	621	25
Peak 2011-	8.3	9.9	.19	1.3	648	27
Midday 2010+	8.3	10.3	.24	1.4	491	41
Midday 2011-	8.3	9.9	.19	1.3	493	71

- Links = (0000420249002700S to 0000420681006990S)
- "+" or "-" indicates addition or subtraction of January 2011 data
- The free flow time used in the Planning Time Index was the mean value of Off Time trips.

Table 4 Travel Times and Reliability Measures for Eastbound Trips on Old Baltimore Pike: Rt 72 to Rt 273

Trip Category	Mean Time	95%	Buffer Index	Planning Time Index	Avg Hourly Volume	#Trips
All Trips	6.4	8.1	.27	1.3	334	712
Peak (AM/PM)	6.7	8.5	.27	1.4	664	149
AM Peak (7-9)	**6.4**	7.8	.22	1.3	632	76
PM Peak (16-18)	7.0	9.8	.40	1.6	697	73
Midday (9-16)	6.4	7.9	.23	1.3	506	366
Weekend (all day)	6.5	8.3	.28	1.4	278	143
Off Time (<7 & >17)	6.0	7.8	.30	1.3	152	54
All Trips 2010+Jan11	6.6	8.4	.27	1.4	328	316
All Trips 2011 -Jan11	6.3	7.8	.24	1.3	336	396
Peak 2010+	7.1	9.8	.38	1.6	647	71
Peak 2011-	6.4	8.0	.25	1.3	674	78
Midday 2010+	6.5	8.3	.28	1.4	507	138
Midday 2011-	6.3	7.8	.24	1.3	508	228

- Links = (0000420333003780S to 0000420681006990S)
- "+" or "-" indicates addition or subtraction of January 2011 data.
- The free flow time used in the Planning Time Index was the mean value of Off Time trips.

Table 3 and Table 4 show reliability measures for eastbound trips on Old Baltimore Pike.

Table 5 Travel Times and Reliability Measures for Westbound Trips on Old Baltimore Pike: Rt 273 to Rt 896

Trip Category	Mean Time	95%	Buffer Index	Planning Time Index	Avg Hourly Volume	#Trips
All Trips	9.4	12.1	.29	1.3	357	248
Peak (AM/PM)	10.1	15.2	.50	1.7	652	43
AM Peak (7-9)	9.1	12.2	.34	1.4	485	26
PM Peak (16-18)	11.7	15+	.32	1.7	819	17
Midday (9-16)	9.2	12.1	.32	1.3	529	140
Weekend (all day)	9.0	11.7	.30	1.3	307	34
Off Time (<7 & >17)	9.2	10.9	.18	1.2	183	31
All Trips 2010+Jan11	9.1	12.0	.32	1.4	370	115
All Trips 2011 -Jan11	9.5	12.6	.33	1.4	351	132
Peak 2010+	10.3	14.8+	.47	1.7	670	17
Peak 2011-	10.1	15.1	.49	1.7	645	26
Midday 2010+	8.9	11.2	.26	1.2	553	72
Midday 2011-	9.5	12.8	.35	1.4	519	68

- Links = (0000420681006990S to 0000420270002490S)
- "+" or "-" indicates addition or subtraction of January 2011 data.
- The free flow time used in the Planning Time Index was the minimum average of weekend trips.

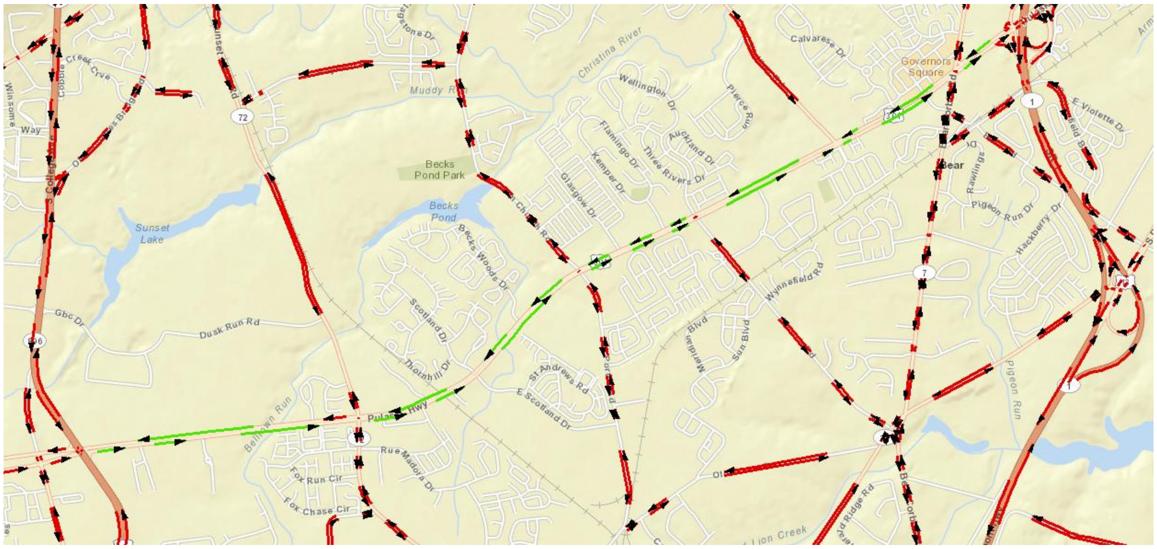
Table 6 Travel Times and Reliability Measures for Westbound Trips on Old Baltimore Pike: Rt 273 to Rt 72

Trip Category	Mean Time	95%	Buffer Index	Planning Time Index	Avg Hourly Volume	#Trips
All Trips	7.8	10.5	.34	1.4	382	403
Peak (AM/PM)	8.4	12.8	.52	1.7	652	77
AM Peak (7-9)	7.4	10.6	.43	1.4	485	46
PM Peak (16-18)	9.8	15.0	.53	2.0	819	31
Midday (9-16)	7.7	10.4	.35	1.4	529	234
Weekend (all day)	7.5	10.1	.35	1.3	307	45
Off Time (<7 & >17)	7.6	9.0	.18	1.2	183	47
All Trips 2010+Jan11	7.6	10.3	.35	1.4	370	174
All Trips 2011 -Jan11	7.9	10.7	.35	1.4	351	229
Peak 2010+	8.1	13.1	.62	1.7	670	17
Peak 2011-	8.5	13.5	.59	1.8	645	46
Midday 2010+	7.5	10.3	.37	1.4	553	72
Midday 2011-	7.9	10.5	.33	1.4	519	124

- Links = (0000420709006990S to 0000420332003190S)
- "+" or "-" indicates addition or subtraction of January 2011 data.
- The free flow time used in the Planning Time Index was the minimum average of weekend trips.

Table 5 and Table 6 show reliability measures for westbound trips on Old Baltimore Pike.

Figure 10 Section of Route 40 under Analysis: Rt 896 to Rt 13



Case Study: Route 40

The Route 40 corridor (see Figure 10) was the second area where we examined travel time reliability, and was travelled much less frequently by state vehicles than the Old Baltimore Pike corridor. We processed data for two months (October 2011 and March 2012) and found only 56 eastbound trips and 90 westbound trips for this corridor. Thus, we did not have sufficient observations to directly estimate travel time reliability for peak periods. Table 7 and Table 8 report reliability measures that were based off of limited data. Reliability indexes are high on this corridor even during midday periods.

Table 7 Travel Times and Reliability Measures on Eastbound Route 40: Rt 896 to Rt 13

Trip Category	Mean Time	95%	Buffer Index	PT Index	#Trips
All Trips	13.2	19.6	.48	1.8	56
Peak (AM/PM)	13.1	-			15
Midday (9-16)	13.6	24.7	.82	2.2	28
Off hours	12.5				13

- Links = (0000520233002630 to 0000520901009480)
- Speed data as derived from State of Delaware Vehicle GPS, from total of 56 observed trips
- “”PT Index” is the Planning Time Index. Freeflow value used in calculation of the Planning Time Index is 10.7 minutes which is the 5th percentile for all trips.

Table 8 Travel Times and Reliability Measures on Eastbound Route 40: Rt 13 to Rt 896

Trip Category	Mean Time	95%	Buffer Index	PT Index	#Trips
All Trips	12.9	19.7	.53	2.1	90
Peak (AM/PM)	14.6	-			15
Midday (9-16)	12.3	17.0	.38	1.8	51
Off hours	13.1	19.7	.05	2.1	24

- Links = (1000521021010680 to 1000521620017360)
- Speed data as derived from State of Delaware Vehicle GPS, from total of 56 observed trips
- “”PT Index” is the Planning Time Index. Freeflow value used in calculation of the Planning Time Index is 9.6 minutes which is the 5th percentile for all trips.

Route 40 is a good example of the problem with evaluating reliability based on actual GPS trips. Even though one month of data might yield between 500 and 1,000 measures for each link on a particular route, relatively few trips covered all links on that route. One option available to the researcher is to limit the focus just to those routes that have reasonable sample sizes in the data. Another option is to create synthetic trips by combining the travel times of multiple smaller road segments into one larger route. This latter option is discussed in the next section of the report.

Link Based Analysis

The preceding section identified trips that traversed long sections of Old Baltimore Pike and Route 40. As the length of those routes increased, the number of observations that completed this trip decreased. For example, only 247 trips were found to have travelled between Route896 and Route 273 on Old Baltimore Pike over the course of one year. However, 712 trips travelled the shorter distance on Old Baltimore Pike between Route 72 and Route 273. In the case of Route 40, two months of GPS data yielded only 56 trips between Route 896 and Route 13. Further disaggregating these trips by characteristic (e.g. Peak/Off Peak, Week/Weekend, AM/PM, etc.) exacerbates this problem.

This section discusses how one might create a trip synthetically by combining the distributions of actual observations for each link into one longer route.

Suppose we wanted the expected travel time for a route that consists of 20 neighboring links. The approach in the previous section first identified those trips that traversed all 20 links. The travel time of all trips completing this route was then averaged to find the expected time of that route. The first two columns in Table 9 report the mean travel time and sample size for trips travelling west on Old Baltimore Pike from Route 273 to Route 72. Alternatively, we could have taken these trips and averaged the travel time for each link in this route. The average trip time would simply be the sum of these link averages.[8]

[8] It is well known that the expected value of the sum of two random variables equals the sum of the expected values of those random variables.

Table 9 Travel Times and Reliability Measures, Westbound Old Baltimore Pike: Rt 273 to Rt 72

Trip Category	Mean Time for actual trips	#Trips	Mean Est From summing estimates at the Link level	Range of Sample Sizes per Link
All Trips	7.8	403	7.8	2338-4367
Peak (AM/PM)	8.4	77	8.5	344-1282
AM Peak (7-9)	7.4	46	8.1	344-1026
PM Peak (16-18)	9.8	31	10.3	166-290
Midday (9-16)	7.7	234	8.1	1581-2498
Weekend (all day)	7.5	45	7.8	228-513
Off Time (<7 & >17)	7.6	47	7.9	202-674
All Trips 2010+Jan11	7.6	174	7.6	948-1796
All Trips 2011 -Jan11	7.9	229	8.4	1722-3098
Peak 2010+	8.1	17	8.0	180-535
Peak 2011-	8.5	46	8.7	474-737

Hypothetically, suppose that *all travel times* for each of the links in this route could be generated by a constant, link-specific random distribution. Under this condition, we could have instead averaged the travel times for each link across *all observations in the data* rather than just across those trips identified as having completed the route. The expected value of the sum of these link-specific averages would asymptotically equal the expected total travel time of that route. However, if *some travel times* on any one of these links were generated by different, link-specific random distributions, then performing the latter calculation would asymptotically bias the procedure.

Column 3 in Table 9 shows the average trip time implied by aggregating across *all observed travel times* for each link in the specified route. Column 4 in that table shows the range of corresponding sample sizes for each link in that route. There are two main results. First, there are decidedly smaller sample sizes for complete routes, as expected. Second, the travel time implied by summing the average travel times for each link tends to be larger than the travel time for each route. Table 10 shows a similar result for eastbound trips on Old Baltimore Pike.

Table 10 Travel Times and Reliability Measures, Eastbound Old Baltimore Pike: Rt 896 to Rt 273

Trip Category	Mean Time	#trips	Mean Est From summing estimates at the Link level	Range of Sample Sizes per Link
All Trips	8.3	247	8.3	1152-3960
Peak (AM/PM)	8.6	52	8.8	1359-1121
AM Peak (7-9)	8.6	18	8.6	192-796
PM Peak (16-18)	8.6	34	9.3	157-369
Midday (9-16)	8.2	120	8.6	11050-2733
Weekend (all day)	8.3	50	8.6	185-302
Off Time (<7 & >17)	7.7	25	8.0	110-278
All Trips 2010+Jan11	8.3	115	8.4	1037-1750
All Trips 2011-Jan11	8.3	132	8.5	1483-2710
Peak 2010+	8.9	25	8.7	116-502
Peak 2011-	8.3	27	8.7	247-616

- Links = (0000420249002700S to 0000420681006990S)

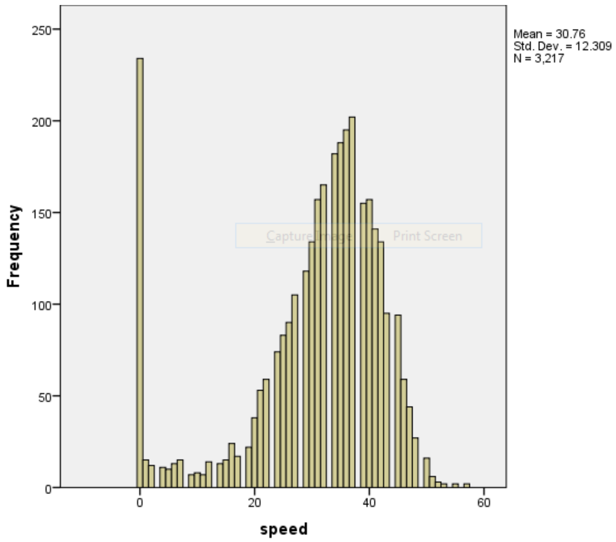
Table 9 and Table 10 indicate that directly summing across link-specific average travel times will likely bias the expected travel time for that route. For routes on Old Baltimore Pike, the sum across link-specific travel times makes the expected travel time seem longer than it actually is. This implies that the travel times for trips which did not traverse the particular route of interest were slower, on average, than the travel times for trips that did completely traverse the route. For this particular route, we can examine turning movements to explain why adding observations lengthens average travel time.

Figure 11 Intersection of Old Baltimore Pike and Salem Church Road

Consider for example the intersection of Old Baltimore Pike and Salem Church Road shown in Figure 11. The route in our case study assumes that vehicles do not turn at this intersection, but continue travelling straight on Old Baltimore Pike (segments of the eastbound trip are highlighted in green). However eastbound trips on Old Baltimore Pike that turn onto Salem Church Road must make a left turn, which means that the time these vehicles wait at the intersection may be longer than vehicles travelling straight.

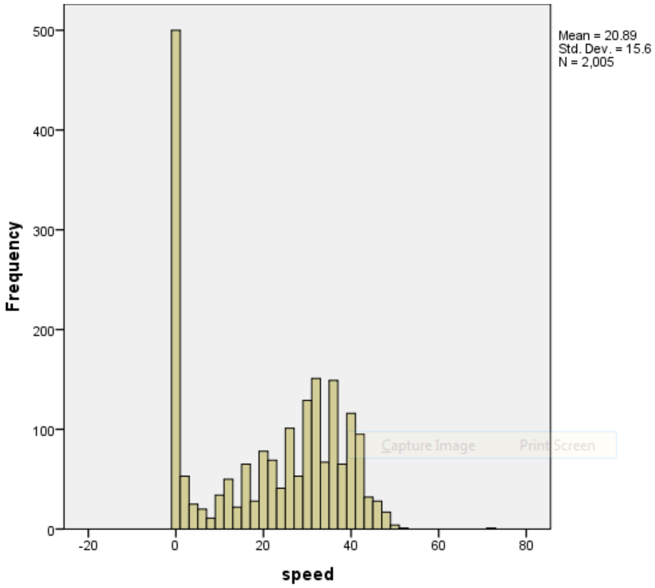
Figure 12 and Figure 13 show histograms of speeds on this link for vehicles travelling straight and turning left. Clearly, vehicles turning left are more likely to be stopped at this intersection, implying that their average travel time for that link will be longer than the average travel time for that link among vehicles continuing straight. Thus, incorporating the travel time vehicles turning left at Salem Church Road into the average of route times will add a positive bias to that average. By a similar argument, Figure 14 and Figure 15 indicate that a negative bias will be added to the average travel time for westbound trips on Old Baltimore Pike due to right turns onto Salem Church Road.

Figure 12 Histogram of Speed Estimates for Eastbound Trips on Old Baltimore Pike Travelling Straight Through the Intersection of Salem Church Road



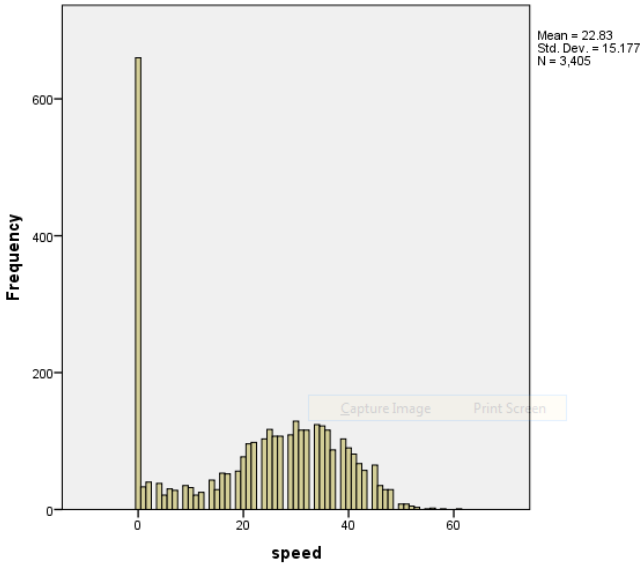
- Link id = '0000420482005080S'

Figure 13 Histogram of Speed Estimates for Eastbound Trips on Old Baltimore Pike Turning left at the Intersection of Salem Church Road



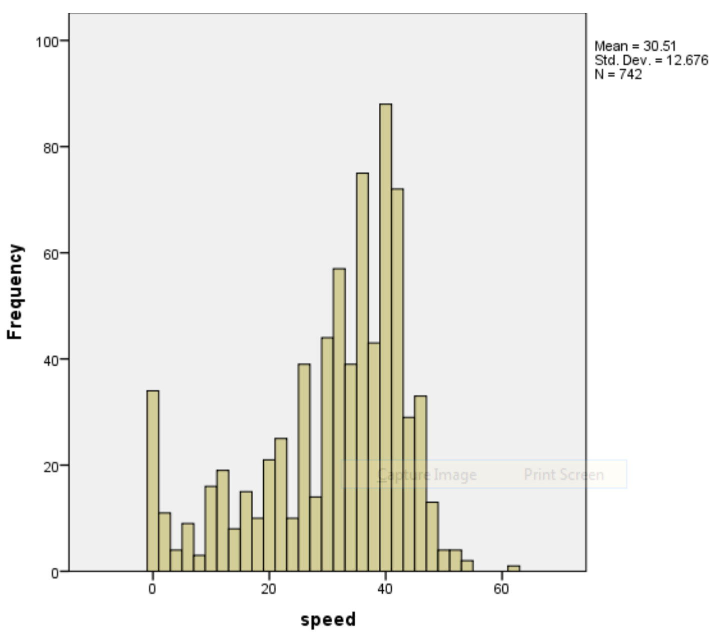
- Link id = '0000420482005080L'

Figure 14 Histogram of Speed Estimates for Westbound Trips on Old Baltimore Pike Travelling Straight Through the Intersection of Salem Church Road



- Link id = '0000420534005080S'

Figure 15 Histogram of Speed Estimates for Westbound Trips on Old Baltimore Pike Turning Right at the Intersection of Salem Church Road



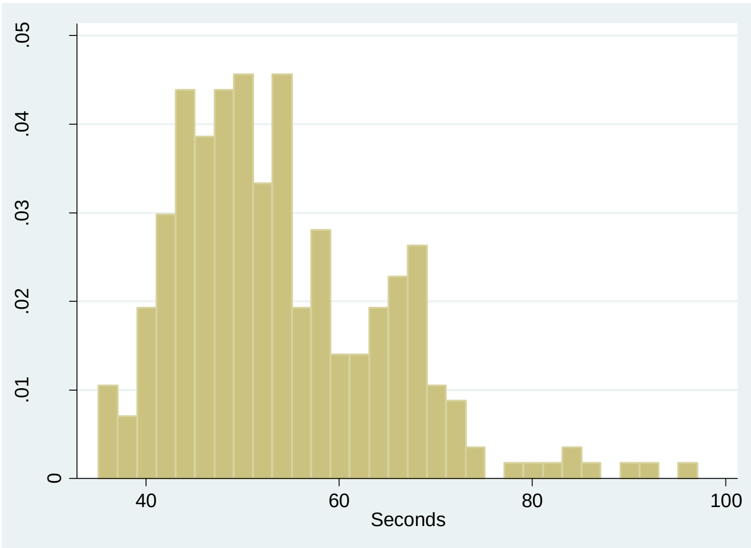
- Link id = = '0000420534005080R'

Until now, the discussion in this section has demonstrated that one cannot simply sum across the average travel time of each link in a route and expect to obtain an unbiased average of total route time. The reason is that one could be mixing travel times for the same link that have substantially different speed/travel time characteristics. Direction at the end of the link, for example, clearly matters.

Even if we obtained unbiased estimates of the average route time by summing across link specific travel time averages, another problem is that information says nothing about reliability. Unbiased information of the distributional aspects of route travel times is needed, and an average says nothing of the distribution. The remainder of this section discusses an approach to estimate the distributional parameters while taking full advantage of the information in the data. This will be particularly useful in cases where there are few observations for a particular route. The basic idea is the same: disaggregate actual trip times to a series of link times and then re-aggregate those times to estimate total travel time for a route.

In an ideal world, travel times for every link on a route could be assumed to come from a series of independent, normal distributions. In this ideal situation, hypothetical trips could easily be simulated after finding each link's mean and standard deviation from the sample data. Once defined, the normal distributions could be sampled a very large number of times, and the distributional properties of that sample would be asymptotically correct.

Figure 16 Example Histogram of the Average Travel Times for One Link



However, real world, link-based travel times are not independent or normal. Instead, these distributions have short left tails owing to maximum speeds which vehicles travel and long right tails owing to occasional congestion. Figure 16 shows an example of travel times for a particular link that is just more than half a mile in length. The figure indicates that no vehicles traversed that link in under 30 seconds, and some vehicles took nearly 100 seconds to pass that route.

Table 11 Correlations of Travel Speeds for Neighboring Links

	Lagged Speed	Actual Speed	Future Speed
Lagged Speed	1	0.5353	0.3639
Actual Speed	0.5353	1	0.5081
Future Speed	0.3639	0.5081	1

In addition to non-normality, travel times are correlated across links for any given trip. Table 11 presents the correlation matrix of travel times for vehicles traversing three subsequent westbound links on Old Baltimore Pike travelling from Route 273 to Route 72. There is a high level of correlation between lagged, actual, and future speeds.

Table 12 Trip Characteristics Example

Link Segment ID	Route Distance (meters)	Estimated Time (seconds)	Estimated Speed (miles per hour)
1000521398014160S	274.8	32	19
1000521416014400S	382.8	53	16
1000521440014620R	364.4	37	22
0000770199002270S	465.7	42	25
0000770227002380S	183.8	16	25
0000770238002760S	639.6	55	26
0000770276002930S	269.9	22	27
0000770293003210S	460.2	38	27
0000770321003440S	367.6	30	27
0000770344003760S	522.3	45	26
0000770376003890S	216.6	24	20
0000770389004210L	517.6	58	20
0000420482004740S	131.8	15	20
0000420474004700S	67.4	6	25
0000420470004550S	246.1	18	30
0000420455004470S	131.2	10	30

Table 12 shows a particular trip's composition of links, the distance of each link, the estimated time it took to traverse that link, and the corresponding speed. One can clearly see that many links have the same speed due to the process of interpolating speeds across multiple links. Even if GPS data was frequent enough such that the effects of any interpolation added only negligible correlation, roadway congestion and driver-specific characteristics would still be a source of correlation.

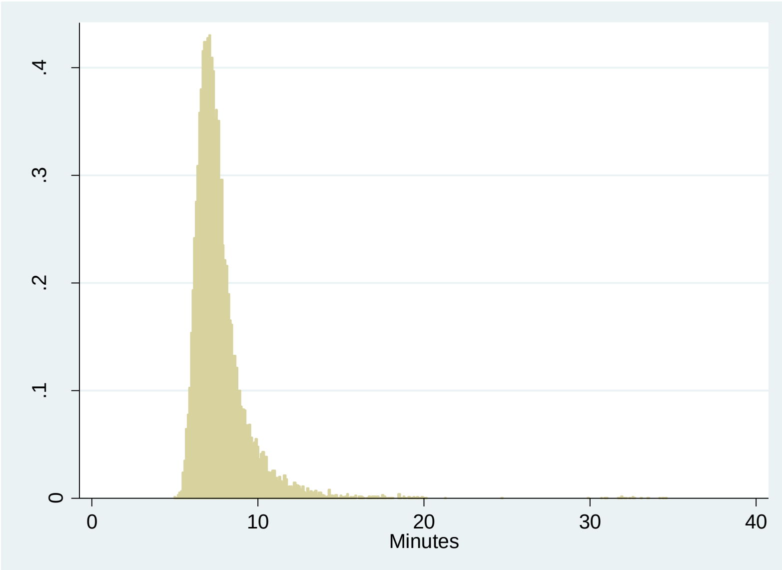
In order to simulate a distribution of travel times for a particular route, any simulation should account for both non-normality and correlated travel times across links. This implies that random draws would need to be made from a multivariate, non-normal distribution that effectively generates a series of link-based travel times.

Ruscio and Kaczetow (2008) propose an algorithm to draw random samples from any multivariate dataset from which the researcher wishes to sample.[9] Their algorithm updates an intermediate multivariate normal distribution, such that when random draws are taken from this intermediate matrix and mapped to the existing dataset, the resulting sample is random, preserves the original dataset's correlation among the variables, and maintains the non-normal properties of each variable's distribution. The advantage of implementing this algorithm is that random travel times can be simulated for multiple links at once while preserving the key, non-normal characteristics of the underlying data generating process.

In order to implement this randomization process, we created one observation for a particular link from the GPS data, as in Table 12. For each observation, we noted the previous and future link that each vehicle travelled and the corresponding speeds on those links. Each observation was therefore a series of three links with three separate speeds, to which we refer to as a triplet. The dataset was then filtered to include only those triplets whose link id's and directions matched a subset of link ids and directions that were on the westbound route of Old Baltimore Pike between Route 273 and Route 72.

[9] Ruscio, John and Kaczetow, Walter. (2008) "Simulating Multivariate Nonnormal Data Using an Iterative Algorithm." *Multivariate Behavioral Research.* Vol. 43(3), 355-381.

Figure 17 Distribution of Synthetic Trip Times



The Ruscio and Kaczetow randomization process was then performed for each unique triplet until 5,000 link speeds were generated for each triplet.[10] Again, the link speeds in each triplet are random, non-normal, and have correlations that approximate actual observations. Link speeds were then converted to travel times. Next, triplets were matched randomly such that when the total travel times were summed across all matched triplets, a total route time would be created.[11] In this fashion, 15,000 synthetic trips were created on Old Baltimore Pike from Rt 273 to Rt 72. The resulting distribution is shown in Figure 17.

[10] Link speeds at the beginning and end of this route were weighted to reflect the observed proportion of state vehicles on the road.

[11] Because this matching process was random, we are implicitly assuming independence of measured speeds across neighboring triplets. Of course, if future research indicates more correlation exists amongst link speeds, then one could simply alter the process to create a series of four or more links instead of three. Initial evidence seems to suggest that any correlation after five links is negligible.

Table 13 Distributional Characteristics of Actual and Synthetic Travel Times, in minutes

	5th pct	25th pct	50th pct	Mean	75th pct	95th pct
Actual	6.06	6.57	7.52	7.79	8.42	10.50
Correlated Simulation	6.07	6.73	7.32	7.73	8.18	10.67
Uncorrelated Simulation	6.31	6.83	7.29	7.64	7.96	10.12

The 5th percentile, 25th percentile, mean, median, 75th percentile and 95th percentile of the actual and simulated distributions are reported in Table 13. The table shows that the synthetic distribution which corrects for correlated link times with triplets matched the actual distribution quite well. In particular, the 5th and 95th percentiles were quite close to the actual sample.

To highlight the importance of correlation across neighboring links, we ignored the correlation across link speeds and simply matched random link times for each link across the entire route. The table clearly shows that ignoring correlation will cause the distribution of trip times to converge towards the median, falsely giving the impression of improved reliability. Intuitively, this is because the second simulation ignores the positive correlation among travel times. And when summing any random variables, positive correlation adds to the variance of that sum.

There are two main advantages of taking a link based approach to generate measures of reliability. First, if one was interested in the reliability of a particular route, synthetic routes will provide trip time distributions using data from all observations that travelled even partial sections of that route. It would not limit the sample to just those drivers who travelled the entire route and throw away good information. This implies that synthetic routes will be the best alternative to creating reliability measures for routes that are infrequently travelled.

Second, by creating travel time distributions for small sections of roads (triplets, in this example) that can be assumed to be independent from one another, one is not limited to measuring reliability of a specific route. In particular, traffic volume data can be integrated into the design of synthetic trips such that reliability measures could potentially be devised for roadway networks, instead of a particular route. Moreover, volume adjustments could also control for the nonrandom routing inherent in the state fleet vehicles. Such implications seem to be an area rife for future research.

Discussion

In this section we discuss several topics that are related the analysis of previous sections.

Specification of the Route Origin and Destination

When a particular route is being defined, the arbitrary choice of where to begin and end that route could make a significant impact on the estimated travel times. For example, if the first link in a route is accessed by a left hand turn, then the speed on that first like will likely be low due to interpolation. If the final link on a route precedes a major intersection that is typically congested, that congestion will spill over to the route's estimated travel time. Such effects on a route's travel times will be proportionally larger as the route's total distance falls.

Thus, rules of thumb might be in order when defining routes of interest. For example, perhaps routes should exclude end links that immediately precede a major intersection. Perhaps the beginning of route times should be a weighted average of multiple roads each leading onto the main route of interest. Further research may want to explore this issue.

Table 14 Comparison of the Average Speeds during Turning Movements - Old Baltimore Pike and Salem Church Road

	Road/Turn	2010 am	2011 am	am Vol.	2010 pm	2011 pm	pm Vol.
EB Old Baltimore Left	0000420482005080L	18.3	21.3	760	8.3	15.3	503
EB Old Baltimore Thru	0000420482005080S	31.0	32.5	667	26.0	28.2	964
WB Old Baltimore Right	0000420534005080R	31.7	29.8	287	29.0	27.3	161
WB Old Baltimore Thru	0000420534005080S	23.5	21.0	705	20.0	12.5	955
SB Salem Church Left	0002760015000000L	17.1	14.5	198	15.0	14.4	497
SB Salem Church Right	0002760015000000R	31.0	26.9	165	16.1	19.5	577

Using GPS Data to Monitor Delays at Specific Intersections

Link level estimates of travel time could be particular useful in planning and routing applications. For instance, knowing what speed a vehicle would likely travel when it continues straight through an intersection, when it turns right, or when it turns left could be used to design more efficient routing for emergency vehicles.

GPS data could also indicate which intersections experience lengthy delays during different times of the day and week. Table 14 below compares the average speeds at the intersection of Salem Church Road and Old Baltimore Pike by trip characteristic. The table indicates, for example, that the average speed of PM peak trips had improved between 2010 and 2011 for east bound vehicles on Old Baltimore Pike which turning left onto Salem Church Road. Though this data cannot be causally tied to differences in signal timing, future research might be able to explore these possible connections.

However, the analysis does not need to be limited to intersections. Figure 18, for example, shows the average speed for each link for those trips travelling straight, and Figure 19 maps the difference in the average speed for each link between 2010 and 2011.

Figure 18 GIS Mapping of Average Travel Speeds (mph)

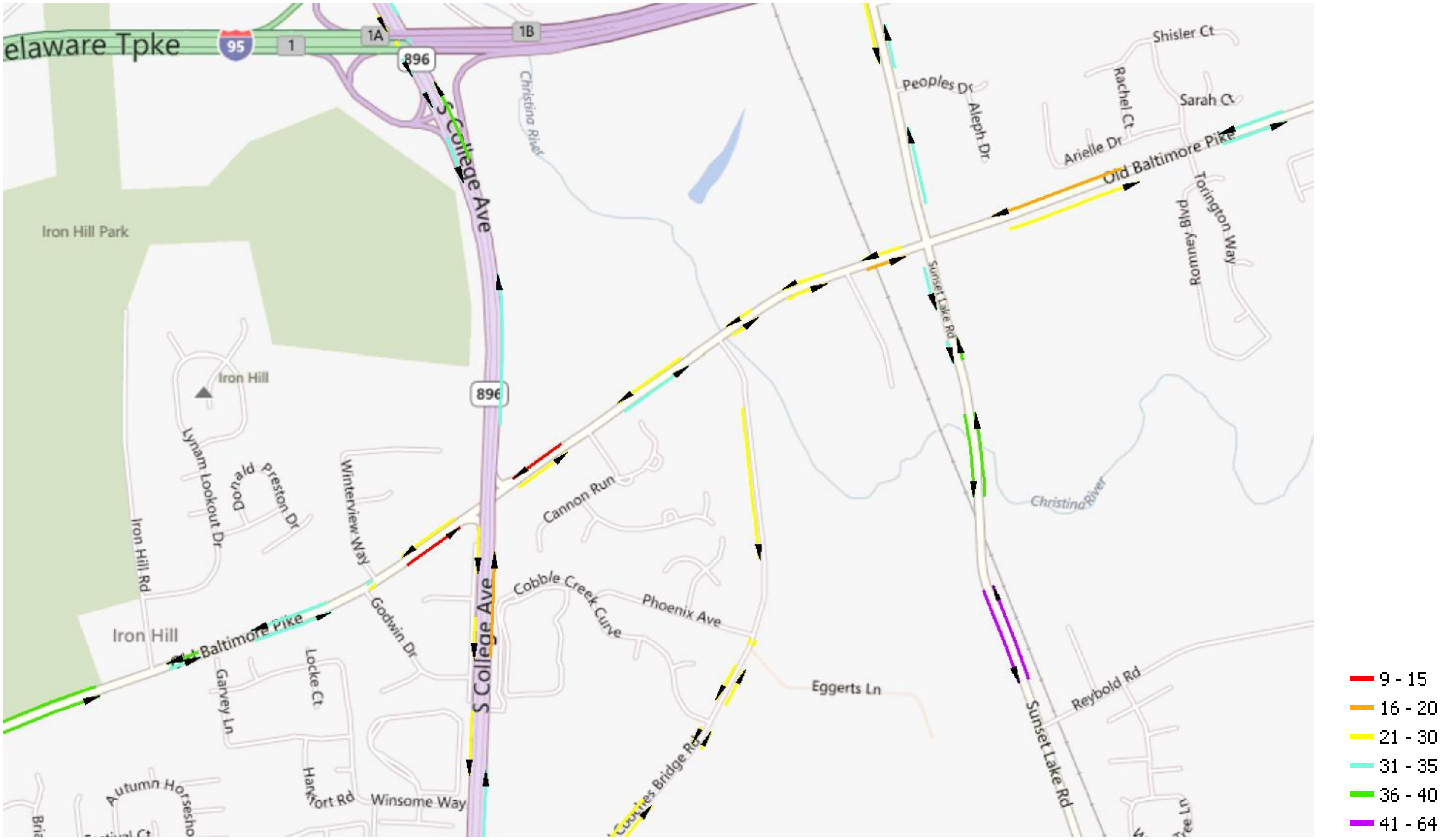
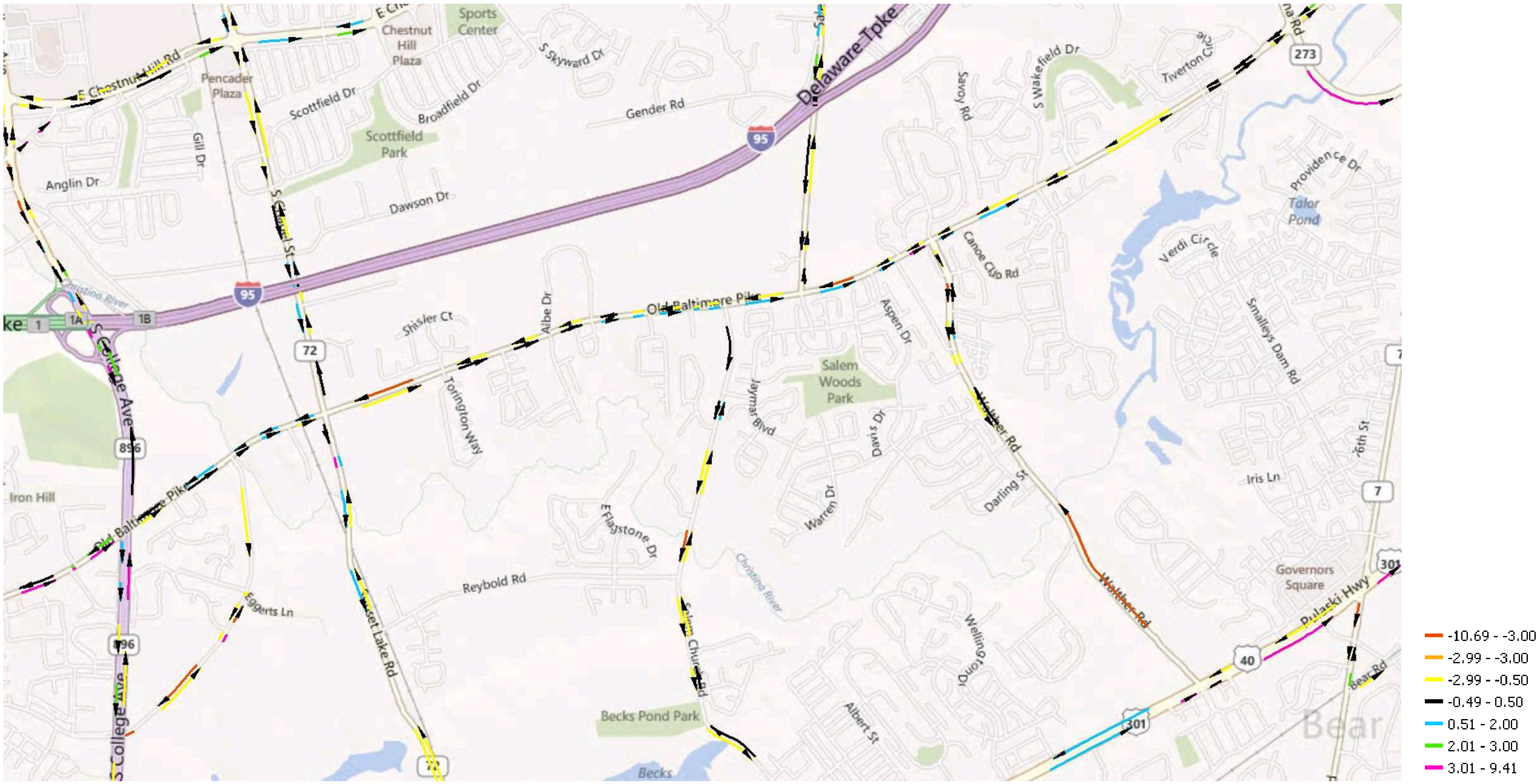


Figure 19 Peak Average Speeds (mph) in 2011 minus Peak Average Speeds in 2010 for Old Baltimore Pike Dataset



Suitability of State Fleet GPS Vehicle Data for Planning and Operations Applications

There are two possible problems mentioned regarding the state fleet vehicle dataset. First, the fact that these are state vehicles means that they are not a random sample of vehicles on the road. The second issue is that the frequency of the GPS is two minutes, so interpolation bias would be larger than more frequent readings. We discuss each of these issues in turn.

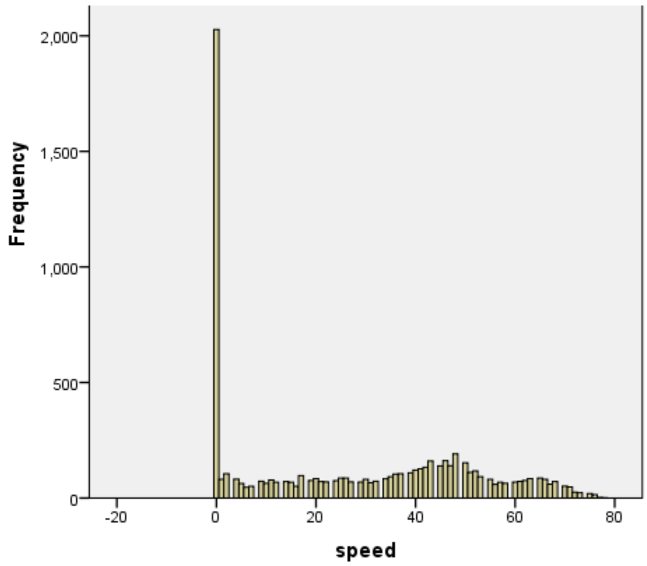
One consequence of using GPS readings from state fleet vehicles is that they likely have different routing patterns than the typical vehicle on the road. This means that they may oversample routes that are rarely used by the public and undersample routes that are heavily used. Of course, since traffic counts of actual data are available and trips can be derived synthetically, any bias from nonrandom differences in routing patterns should be correctable.

Another concern is that the drivers of state fleet vehicles may travel at slower speeds than the general public.[12] The flow of traffic on many roads is often above posted speed limits. But since drivers operating state owned vehicles know they are being monitored, they seem less likely to drive faster than the speed limit. Intuition tells us that this bias will be more pronounced on major highways where actual speeds are often substantially greater than the posted speed limit.

Figure 20 and Figure 21 are histograms of instantaneous speed readings on our two case study corridors. The speed limit ranges between 35 mph and 45 mph on Old Baltimore Pike and between 40 mph and 50 mph on Route 40. Clearly, state fleet vehicles can be travelling faster than the posted speed limits at any given time. Moreover, any bias due to driving slower will likely be most pervasive during free-flow conditions, and less of an issue when measuring congestion and reliability.

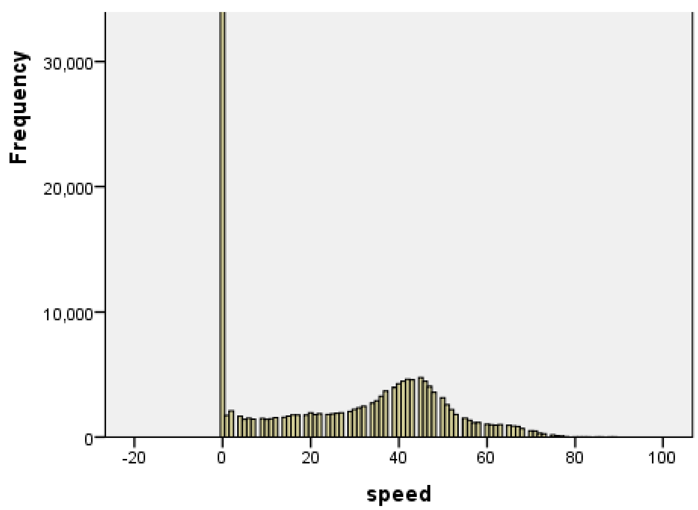
[12] Public safety vehicles (e.g. police cars and ambulances) are categorized differently and were removed from all primary data sets.

Figure 20 Speed Histogram for Route 40 in New Castle County, from State Fleet Vehicle GPS Data



- N = 7,127

Figure 21 Speed Histogram for Old Baltimore Pike in New Castle County, from State Fleet Vehicle GPS Data



- N = 165,812

Though we do not expect that nonrandomness in the State Fleet Vehicle dataset will substantially bias our measures of reliability, we can only test for this by comparing it to outside data. Historical speed studies are one source which could be used for comparison purposes. For limited access roads like Interstate I-95 and Route 1, GPS speed readings may be compared to existing measurement devices already in the field to quantify any bias. Distributions from the State fleet data could also be compared with distributions of measures from other sources and from that additional adjustments may be made.

Another issue is the two minute lag between subsequent GPS measures. This two minute lag required that we interpolate speeds over multiple links, which undoubtedly raised the correlation in link speeds across certain links. However, we were able to control for this correlation in the derivation of synthetic trips. Moreover, since we are interested in the performance of *route* times, instead of *link* times, this bias is less important in the final measure. Longer gaps between GPS readings also made the adjustments in travel times at the beginning and end links of each trip more sensitive.

Further Development of a State Fleet GPS Based Product

Most of the processing of the State Fleet vehicle GPS data could be automated. To streamline and improve this part of the process, the following steps are needed:

- Refine programming codes to increase speed of processing. Currently processing one month of data can take 2 to 3 weeks of processing time using 2 personal computers. By using more computing power and a strategy to break the job down into smaller pieces, we expect this time to be significantly reduced.
- Refine post processing steps to improve error checking and data preparation.
- Develop programs to process segment/chain based processing.
- Development of additional statistics such as those related to delay.
- Develop post processing steps to calculate averages and confidence intervals.
- Develop a rudimentary user interface to filter data and automate creation and display of performance measures.
- Maintain and update a routable road network file in a GIS format that can reference travel time and speed data in a standard manner. Such a file has been created by CADSR and was used in this research. The file is generally necessary to support the processing of many types of traffic related data.
- Compare speed estimates from other sources to identify any sources of bias or error.

Each of the steps above would build onto current progress. We expect that many of the steps above could be completed relatively quickly given current progress. Advanced interfaces that incorporate a high level of automation and integration with other traffic data would take considerably longer.

Free Flow Travel Time

Another issue that arose during the calculations of reliability was estimating each route's free flow time.[13] The FHWA describes free flow time as the ideal travel time, but the literature has not agreed on any single method of estimating that ideal. Deardoff, Wiesner, and Fazio (2011) modeled free flow speed and found it was usually close to the posted speed limit.[14] However, basing free flow conditions on the posted speed limit may actually create too ideal of a measure. Some routes might involve so many left hand turns, for example, that few vehicles ever observe such conditions in reality.

To address this criticism, the FHWA (2012) suggested that free flow travel time be calculated as the 15th percentile of travel times during traditional off-peak times, so long as that value does not imply a travel speed above the posted speed limit.[15,16] The main disadvantage of using this definition is that one must have enough observations of a specific route to calculate the 15th percentile during traditional off peak times. Unless one used synthetic trips or acquired a substantially larger dataset, this definition would be of less practical use.

---

[13] Planning Time and the Planning Time Index, for example, subtracts the 95th percentile from the free flow time.
[14] Deardoff, Mathew D., Wiesner, Brady N., and Fazio, Joseph. (2011.) "Estimating Free-flow Speed from Posted Speed Limit Signs" *Procedia – Social and Behavioral Science*. Vol. 16, 306-316.
[15] U.S. Department of Transportation, Federal Highway Administration. *The Urban Congestion Report: Documentation and Definitions*. (Online Resource) http://www.ops.fhwa.dot.gov/perf_measurement/ucr/documentation.htm
[16] Traditional off peak travel times on weekdays are between 9 am to 5 pm and also between 7 pm to 10 pm. Traditional off peak travel times on weekends are between 6 am and 10 pm.

Data Management for Traffic Data

In order to understand how measures of reliability affect drivers on Delaware roadways, these measures must be connected to traffic volumes and road capacities. This report heavily relied upon the standard road network referencing scheme in order to make that connection. Without such a referencing scheme, merging such large datasets would be practically impossible. Any future work using this or other GIS-related traffic data will require that such referencing schemes remain up-to-date. While not the focus of this research, it must be noted that the proper management of traffic data is vital to perform comprehensive analyses.

Area Wide Measures

Travel time reliability studies typically focus solely on the performance of a particular route or corridor. This can produce an incomplete picture. For instance, if a change is made to a corridor, traffic patterns may change on that corridor and on other roads that neighbor or intersect it. Thus, route specific measures of reliability should ultimately yield to network wide measures of reliability. Additional research into area-wide measures of reliability is appropriate. It would seem that data coming from GPS devices are crucial for calculating such measures.

Other Applications That Could Be Served

A goal of this research was to use the State Fleet vehicle GPS data to calculate measures of travel time reliability. The data appears to be very promising in this regard. However, the data could also be used for other applications that were beyond the scope of this report.

First, the data could be used to better explore performance at the intersections of major roadways. Turning movement speeds in the GPS data could merged with turning movement volume counts to study the capacity of intersections to handle various volumes of traffic. Traffic flows at a particular intersection could be evaluated before and after improvements were made.

Routing programs are a second application that could use GPS enabled data. Most routing software use impedances associated with posted speed and rules of thumb for turning movements. These programs also do not account for daily and seasonal fluctuations in speed. Data from GPS enabled devices could correct each of these deficiencies.

Finally, data from GPS enabled devices could help us understand the connection between land use and congestion. For example, it may be possible to show how travel flow is affected by proximity to various kinds of land development and to access points on the roadway.

Comparison of Alternative Sources for Speed Estimates

GPS readings from the State Vehicle database are not the only available source of speed estimates. Other sources of speed data could come from a study conducted by the Delaware Transportation Center, mobile Bluetooth scanners, and commercially available GPS data. A detailed comparison of these data sources is beyond the scope of this research, but there are a few basic characteristics that distinguish each data source. Those characteristics are:

- Coverage
- Number and frequency of measurements
- Proprietary considerations
- Reliability and accuracy of measures
- Detail required and scope of interest
- Costs for data collection, processing, and analysis[17]

An ongoing study by the Delaware Transportation Center at the University of Delaware equips GPS units in a vehicle and makes runs for major corridors in Delaware. The study conducts two runs per year, per corridor, per different peak condition. The data collected during this study is not proprietary, and the Center releases the study's findings in annual reports. Although coverage for this data set is excellent, the number of measurements is extremely low.

Bluetooth scanners are a second option to collect speed data for Delaware's roadways. These devices are placed at the beginning and end of a particular route of interest and record Bluetooth signals from passing vehicles. These scanners can produce thousands of travel time estimates throughout the day for a particular route. Bluetooth sensors can be deployed and redeployed to examine any route of interest. The number of measurements for each route will be excellent, the estimated speeds will be accurate and reliable, and DelDot would own this information. However, coverage will only apply to those routes equipped with scanners, and cost would prevent every route of interest to be equipped with Bluetooth scanners.

[17] The authors are unfamiliar with the exact costs of funding or purchasing any of the alternative data sources.

A third alternative for speed data is to use a commercial source. Companies like Tom Tom and NavTeq continue to compile vast amounts of GPS data from the mobile devices that their clients willingly share. With the purchase of a subscription, it is possible to gather travel time statistics for particular routes using a web-based interface. The source data of these products are high-frequency GPS signals for vehicles throughout the entire road network. Thus, the number and frequency of measures, the accuracy, reliability, and coverage should all be excellent.

There are also disadvantages to using commercial data. First, we expect there to be a multitude of problems stemming from legal and technological safeguards that protect private and/or proprietary information. In addition, commercial products are based on independent mappings of the road network, making it difficult to reference supplementary data from a commercial database. Source trip level data that could be used to develop products to serve a number of applications is unavailable. Finally, the costs of commercial data are expected to be large, especially if coverage was requested for the entire network in the state. On the other hand, GPS-based products that are more streamlined may significantly reduce the costs of processing and analyzing the raw data.

Conclusions

The State Fleet Vehicle data is a large and continually updated dataset of GPS information for all state vehicles. This report uses that dataset to calculate performance measures of travel time reliability. We find that this dataset is particularly useful for calculating such measures. One of the best features of this dataset is that it covers most of Delaware's road network with a reasonable number of measurements. In addition, it is a state-owned resource that is already funded for reasons other than traffic management.

One potential drawback of using Delaware Fleet data is that the nonrandom nature data could be a source bias. However, there are also reasons why that bias could be negligible. It is important to compare the speeds in this dataset with other sources.

We acknowledge that the steps of processing the data and performing quality control are quite involved and require significant amounts of skill, effort, and computational time. However, much of that legwork has already been done for this report. Should there be an interest in using the State Fleet Vehicle dataset for future transportation planning and operations applications, it is crucial to streamline and possibly automate the data processing step. Of course, raw data would still need to be collected, databases would need to be maintained, and basic interfaces would need to be constructed.

Most travel time reliability studies focus on specific routes, even though specific routes are only part of the road network. In fact, how that travel times on one route interacts with the travel times on connecting or adjacent roads are also important. The availability of statewide coverage could lay the foundations for developing area wide measures of reliability.

The State Fleet Vehicle data also has many other potential applications besides measuring reliability. For example, the data could be used to evaluate level of service, perform a trend analysis, improve existing vehicle routing applications, examine the relationship between land and traffic congestion, test for the effects of specific capital improvements, and analyze traffic congestion at intersections. As research develops in the future, more applications could arise.