Technical Memorandum

Probe Data Analysis

Evaluation of NAVTEQ, TrafficCast, and INRIX® Travel Time System Data in the Tallahassee Region

Version 2.0

March 27, 2012

Prepared for:

Florida Department of Transportation
Intelligent Transportation Systems Program
605 Suwannee Street, M.S. 90
Tallahassee, Florida 32399-0450
(850) 410-5600

Table of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

FDOT .... Florida Department of Transportation

GPS .... Global Positioning System

I-10 .... Interstate 10

LPR .... License Plate Reader

TMC .... Traffic Message Channel

US .... United States

XML .... Extensible Markup Language

1 Scope of Analysis

This technical memorandum provides an overview of the analysis performed on travel time data collected on Interstate 10 (I-10), US 27, US 319, and Capital Circle in Tallahassee from separate data sources between September 12 and 15, 2011. The goal of this analysis is to compare the travel time system data provided by NAVTEQ[1], TrafficCast[2], and INRIX®[3] for travel time and speed reporting purposes.

2 General Information and Project Background

The Florida Department of Transportation (FDOT) desires a cost-effective method of receiving travel time information on major portions of the interstate and state road networks throughout the state. Typical intelligent transportation systems deployments along the roadway are cost prohibitive for statewide coverage along large numbers of arterial roadways throughout rural areas in the state.

Commercial travel time data is also available from a number of private companies, which can mitigate the need for agencies to deploy and maintain travel time sensors on their roadway networks. This travel time data, which is sold to agencies on a subscription basis, is typically collected from various sources, such as mobile global positioning system (GPS) fleet devices, roadside sensors, or wireless communication devices carried on-person by the public at large.

The potential use of commercial travel time data along rural interstates, rural and urban highways, and urban arterial roads is particularly attractive since these facilities rarely include traditional roadside sensors. The Tallahassee area in FDOT District Three provides a good environment to evaluate the effectiveness of the commercial travel time data provided by NAVTEQ and TrafficCast. This area includes both urban and rural segments of I-10, urban and rural highway segments, and arterials that can be used for the purposes of this analysis.

This analysis compares the NAVTEQ and TrafficCast data against other data sources for selected roadways in the Tallahassee area. Data for comparison includes "ground truth" data from drive tests along selected road segments from four FDOT-provided probe vehicles, travel

---

[1] NAVTEQ is a private company that anonymously collects and analyzes data from GPS probes, sensors, and other sources to provide traffic data. http://www.navteq.com

[2] TrafficCast is a private company that anonymously collects and analyzes data from GPS probes, sensors, and other sources to provide traffic data. http://www.trafficcast.com

[3] INRIX is a private company that anonymously collects and analyzes data primary through the use of GPS equipped fleet vehicles. http://www.inrix.com

time data from the SunGuide® software, travel time data from INRIX, and test data provided by NAVTEQ and TrafficCast.

NAVTEQ and TrafficCast provided system data for the tested roadway segments covering a four-day period from September 12-15, 2011. During the same four-day period, FDOT drivers traveled these roadway segments and provided floating car travel time data for data comparison. Data from the license plate reader (LPR) system on I-10 was used as a secondary comparison for the I-10 data.

2.1 Description of Test Routes

This project identified four routes in the Tallahassee area for testing of the travel time system data. These routes are:

Table 2-1: Description of Routes

ID	Roadway	Description
1	I-10	I-10, including the length of road in FDOT District 3 between Exits 192 and 209 consisting of approximately 20 miles.
2	US 27 (Monroe St.)	US 27 from downtown Tallahassee (South Pensacola Street) to the Georgia state line, consisting of approximately 18 miles.
3	US 319 (Thomasville Rd.)	US 319 from Capital Circle to the Georgia state line, consisting of approximately 13 miles.
4	Capital Circle	Capital Circle from US 319 to West Orange Avenue, consisting of approximately 16 miles.

Figures 2-1 through 2-4 represent the locations of the traffic message channel (TMC) codes for the roadway segments tested during this analysis:

Figure 2-1: Route 1 (I-10)

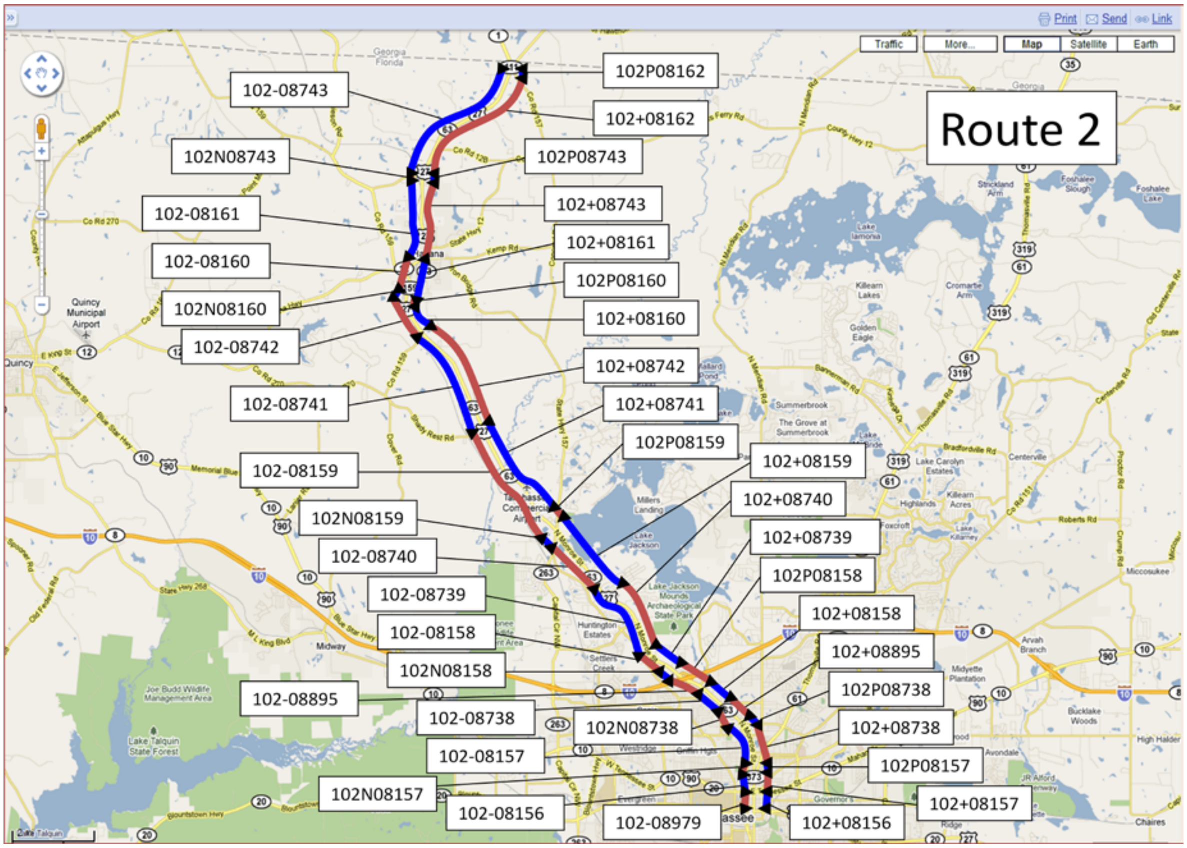

Figure 2-2: Route 2 (US 27)

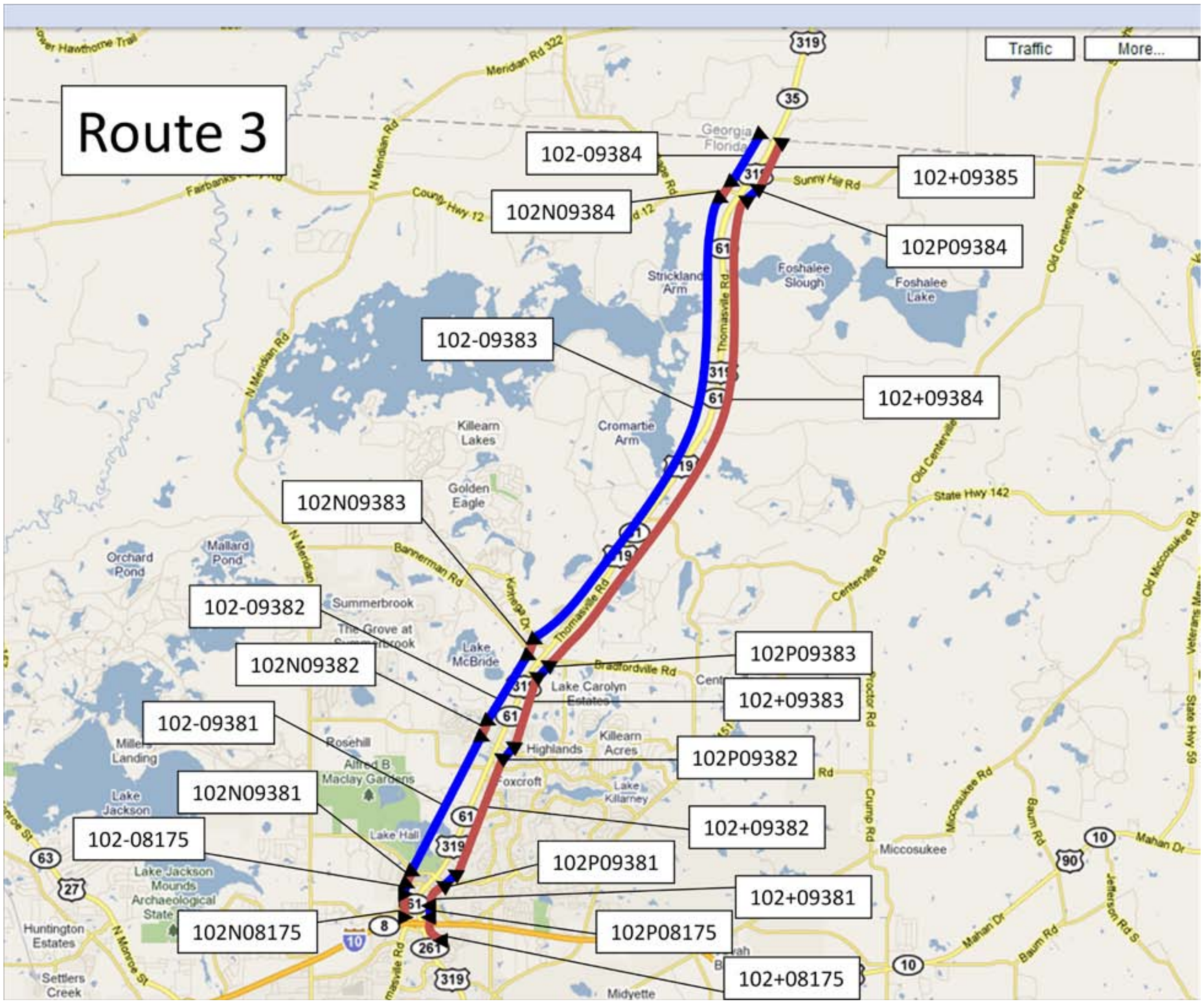

Figure 2-3: Route 3 (US 319)

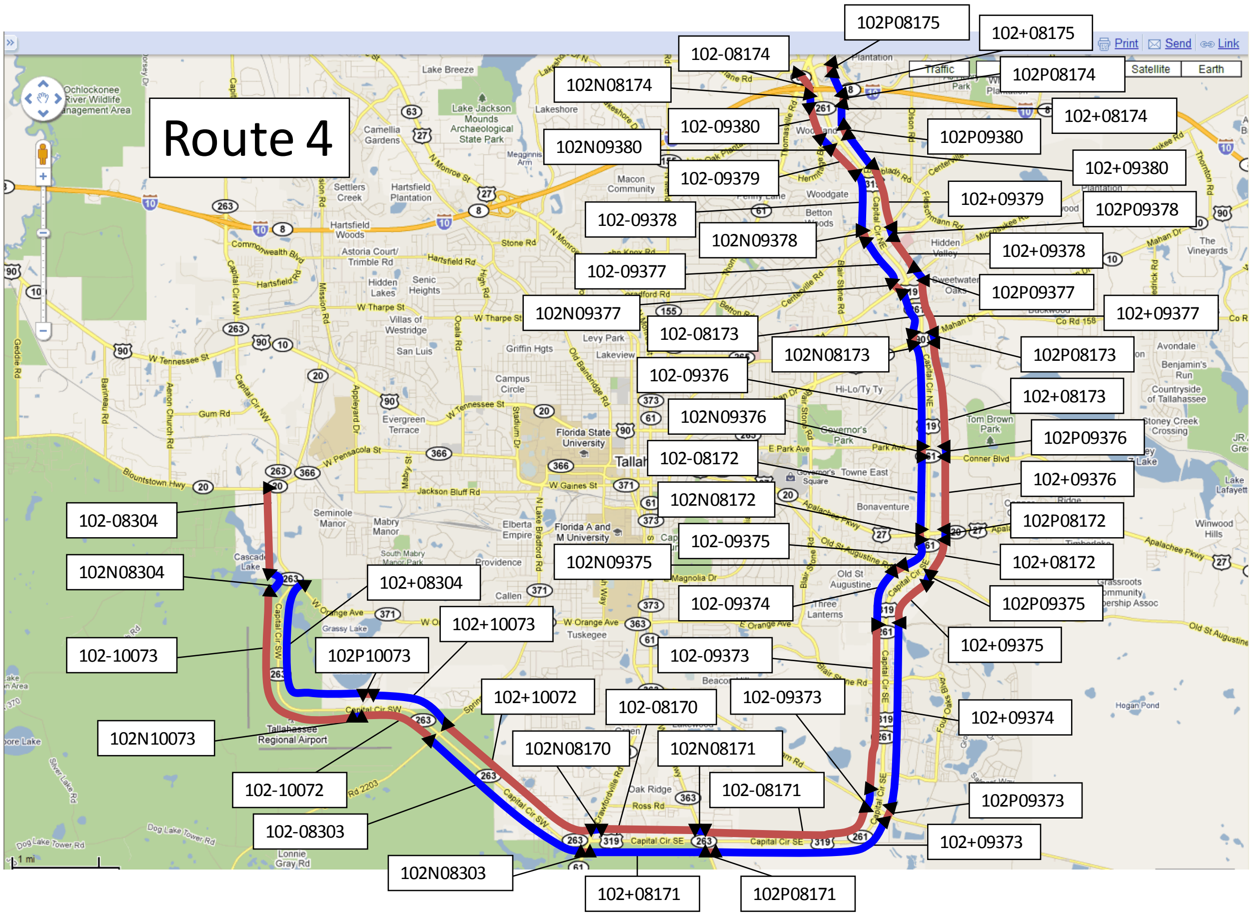

Figure 2-4: Route 4 (Capital Circle)

I-10 provides a test section of multi-lane interstate. US 27 and US 319 provide a test bed for arterials that begin in an urban area and extend to a rural area. Capital Circle provides a long section of signalized arterial within an urban area.

Adjacent TMC segments with high numbers of traffic signals were rolled up into a single "urban" route for analysis. This is because traffic signals can cause highly-variable travel times on short roadway segments. This variability becomes less pronounced over longer segments, since vehicles that stop at one intersection may receive a green light at another intersection and vice versa. Rural TMC segments were studied individually. Figure 2-5 provides an example of how TMC segments were grouped together.

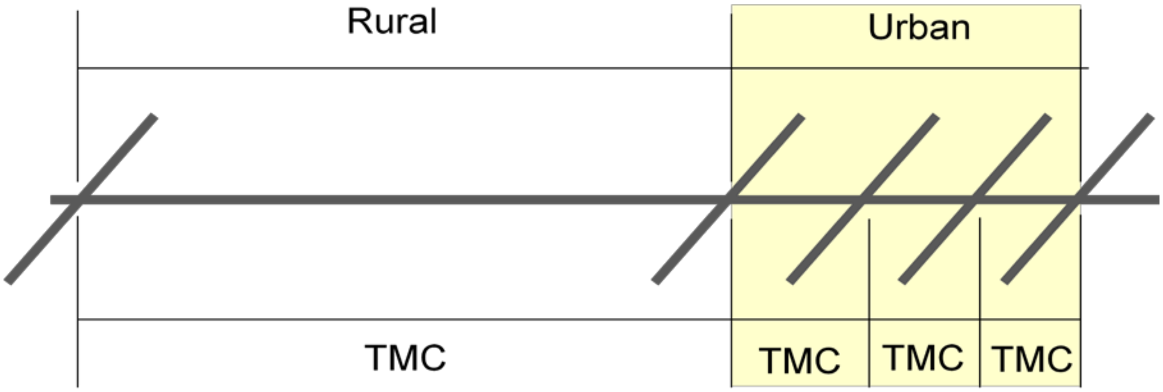

Figure 2-5: Urban and Rural TMC Code Groupings

Note: TrafficCast system data for a few of the TMC codes included in the urban routes were not available; however, the missing data only accounted for approximately two percent of the total distance for the urban segment and was deemed negligible.

2.2 Description of Data Sources

In total, five separate sources of data were examined for this analysis. INRIX data is available only on Route 1 (I-10). NAVTEQ, TrafficCast, and INRIX provided raw data in extensible mark-up language (XML) format. For processing of the data, a data ingest process was written for each vendor. In general, these processes parsed the XML files and wrote values to a data table, which was used for the analysis.

2.2.1 NAVTEQ

NAVTEQ calculates travel time information through processing of commercial and consumer GPS probe, sensor, traffic incidents, road closures and historical data into algorithms to general travel times to produce depiction of the roadway state. NAVTEQ generates travel time data by processing real-time data sources through their Smart Data Processor.

NAVTEQ advised they aggregate and weight historical time-of-day speed observations from several years to estimate the most probable speeds when real-time data is not available. If there is no data during a collection period for a particular TMC, the algorithm uses previous data weighted by time and distance to the TMC. If no data is available, for example during nights, historical data will be used.

NAVTEQ compares data inputs and uses ground truth testing to check inputs and data processing outcomes in high congestion, moderate congestion, and free flow periods. The testing determines the percent accuracy for the tested region for each congestion level. NAVTEQ investigates areas of questionable performance to determine the cause of divergent results and uses the test results to improve the algorithms and ensure accurate data.

NAVTEQ's confidence metric takes into consideration the type of data, the age of the data, and the spatial coverage of the data. Sparser and less frequently available real-time data provides for a lower confidence value in the data feed. NAVTEQ advised that a lower confidence value does not necessarily mean the data is any less accurate, it means the quantity of data and coverage calculated in the algorithm is lower than it is at a higher confidence value. NAVTEQ advised the current data will have a higher confidence weighting, but the fact that the confidence is lower does not mean the data is necessarily inaccurate. Lower confidence means there is higher possibility for error in the data, either due to less availability of real-time data or high latency of the available real-time data. An example of the NAVTEQ data used is shown in Table 2-2.

Table 2-2: Sample NAVTEQ Data

Navteq				
vTime	NTSpeed	traveltimeMinutes	confidence	Calc. TT
2011-09-12 08:30:05.000	61	2.33	0.88	1.757482
2011-09-12 08:32:04.000	62	2.33	0.88	1.729136
2011-09-12 08:34:04.000	62	2.32	0.87	1.729136
2011-09-12 08:36:04.000	62	2.31	0.86	1.729136
2011-09-12 08:38:04.000	63	2.29	0.86	1.701689
2011-09-12 08:40:04.000	66	2.19	0.86	1.62434
2011-09-12 08:42:04.000	66	2.18	0.85	1.62434
2011-09-12 08:44:04.000	66	2.16	0.84	1.62434
2011-09-12 08:46:04.000	67	2.15	0.85	1.600096
2011-09-12 08:48:04.000	67	2.15	0.84	1.600096
2011-09-12 08:50:04.000	66	2.17	0.84	1.62434
2011-09-12 08:52:04.000	66	2.17	0.84	1.62434
2011-09-12 08:54:04.000	62	2.3	0.84	1.729136
2011-09-12 08:56:04.000	65	2.19	0.84	1.649329
2011-09-12 08:58:04.000	65	2.19	0.83	1.649329
2011-09-12 09:00:04.000	65	2.19	0.83	1.649329

One of the NAVTEQ data characteristics that differs from the other providers is that NAVTEQ does not report separate data for many internal TMC codes. Rather, the data is rolled up into the adjacent external TMC segment. For example, NAVTEQ reports data for TMC code 102+04894, but does not report for TMC code 102P04894 because data for TMC code 102+04894 covers both segments. In order to compare travel times between NAVTEQ and TrafficCast, we had to calculate the travel time from NAVTEQ's reported speed data using the following formula:

*Travel Time [mins] = (60 * Segment Length [miles]) / (Average Speed [MPH])TrafficCast*

TrafficCast collects travel time data though multiple sources, such as GPS probes, sensors, historical data, weather, and incidents. The data provided by TrafficCast is reported as TMC location codes. This data includes a timestamp, speed, and severity information reported at one-minute intervals. Table 2-3 is an example of the data used to compare against FDOT ground truth.

Table 2-3: Sample TrafficCast Data

TrafficCast			
TimeSlice	Speed	Severity	Calculated Travel TIme
2011-09-12 08:40:48.000	67	0	1.597409
2011-09-12 08:41:43.000	67	0	1.597409
2011-09-12 08:41:43.000	67	0	1.597409
2011-09-12 08:43:44.000	66	0	1.621612
2011-09-12 08:44:46.000	64	0	1.672288
2011-09-12 08:45:46.000	64	0	1.672288
2011-09-12 08:46:49.000	63	0	1.698832
2011-09-12 08:46:49.000	63	0	1.698832
2011-09-12 08:48:46.000	61	0	1.754531
2011-09-12 08:49:42.000	57	0	1.877656
2011-09-12 08:51:42.000	57	0	1.877656
2011-09-12 08:51:42.000	57	0	1.877656
2011-09-12 08:53:45.000	59	0	1.814007
2011-09-12 08:54:48.000	66	0	1.621612
2011-09-12 08:55:44.000	67	0	1.597409
2011-09-12 08:56:46.000	67	0	1.597409
2011-09-12 08:57:51.000	67	0	1.597409
2011-09-12 08:58:43.000	67	0	1.597409
2011-09-12 08:59:45.000	67	0	1.597409
2011-09-12 09:00:49.000	67	0	1.597409

TrafficCast provides speed, not travel time data, for each TMC segment. In order to provide a travel time for comparison against the ground truth, the travel time for each TMC segment was calculated based on the reported TMC speed and its length.

According to TrafficCast, “Speed” is the fusion of historical speed, real-time speed, and traffic impact. Negative speed value implies special situation: “1” no information is available or sensor malfunction and “2” road closure. “Severity” is assigned based on speed slow down ratio, which is calculated by {1- [(real-time speed) / (posted speed limit or historical speed)]}

2.2.2 INRIX

INRIX collects travel time data primary through the use of GPS-equipped fleet vehicles. The data provided by INRIX is reported in terms of TMC location codes. This data includes a timestamp, speeds, and travel time information reported at one-minute intervals. The speeds and travel times reported by INRIX are based on an aggregation of data provided by GPS probes.

Table 2-4 provides an example of INRIX data.

Table 2-4: Sample INRIX Data

INRIX						
DateTime	Average Speed	Speed	Reference Speed	Confidence Score	Cvalue	Travel Time (Minutes)
2011-09-12 09:08:29.000	65	65	65	30	100	1.649
2011-09-12 09:09:31.000	65	65	65	30	100	1.649
2011-09-12 09:10:33.000	65	65	65	30	100	1.649
2011-09-12 09:11:35.000	65	65	65	30	100	1.649
2011-09-12 09:11:36.000	65	67	65	30	100	1.6
2011-09-12 09:12:38.000	65	67	65	30	100	1.6
2011-09-12 09:12:39.000	65	67	65	30	100	1.6
2011-09-12 09:13:41.000	65	67	65	30	100	1.6
2011-09-12 09:14:43.000	65	75	65	30	28	1.429
2011-09-12 09:15:45.000	65	75	65	30	31	1.429
2011-09-12 09:16:48.000	65	75	65	30	33	1.429

"Reference Speed" is analogous to the free-flow speed along the reporting link. According to INRIX, "the reference attribute is the calculated 'free flow' mean speed for the roadway segment in miles per hour (capped at 65 miles per hour). This attribute is calculated based on the 85th-percentile point of the observed speeds on that segment for all time periods, which establishes a reliable proxy for the speed of traffic at free-flow for that segment."

"Average Speed" is the historical average mean speed for the reporting segment for that time of the day and day of the week in miles per hour.

"Speed" represents the average speed for a given TMC code, calculated from live data over the most current time slice.

INRIX also reports data quality in terms of a "confidence score," which is 10, 20, or 30 (with a score of 30 denoting data of the highest quality) making use of the greatest amount of real-time traffic data). An analysis of the reported data reveals that when data is assigned a score of 10, the system reports a speed equal to the reference speed. When the data is assigned a score of 20, the system reports a speed equal to the average speed. When the data is assigned a score of 30, the system reports a speed calculated from live data over the most recent time slice.

"Cvalue" represents the confidence value (range 0-100) that is designed to help agencies determine whether the INRIX value meets their criteria for real-time data.

2.2.3 Floating Vehicle Data

Floating vehicle data was collected to provide a ground truth comparison to each vendor's system data. The floating vehicle data came from four drivers (from FDOT) who traveled the four roadway segments on September 12-15, 2011. Each driver typically made three to five passes for each route. The drive test data was collected primarily during the morning and evening peak periods (6:00 – 9:00 a.m. and 3:30 – 6:30 p.m., respectively). Mid-day, off-peak data (11:30 a.m. – 1:30 p.m.) was also collected, but made up no more than a maximum of 15 percent of all drive test data collected.

This dataset, an example of which is shown in Table 2-5, includes a timestamp, latitude/longitude, and speed of the vehicle on a second-by-second basis as the vehicle traversed each route.

Table 2-5: Sample Floating Vehicle Data

Index	Route	Driver	Trip	Time	Latitude	Longitude	Speed
51	1	1	1	2011-09-12 13:30:01.000	30.485168	-84.0242	63.936001
52	1	1	1	2011-09-12 13:30:02.000	30.485321	-84.024269	67.608002
53	1	1	1	2011-09-12 13:30:03.000	30.485481	-84.024345	71.064003
54	1	1	1	2011-09-12 13:30:04.000	30.485649	-84.024422	73.080002
55	1	1	1	2011-09-12 13:30:05.000	30.485823	-84.024506	74.879997
56	1	1	1	2011-09-12 13:30:06.000	30.485998	-84.02459	76.571999
57	1	1	1	2011-09-12 13:30:07.000	30.486177	-84.024673	77.255997
58	1	1	1	2011-09-12 13:30:08.000	30.486361	-84.024757	78.012001
59	1	1	1	2011-09-12 13:30:09.000	30.486544	-84.024849	78.695999
60	1	1	1	2011-09-12 13:30:10.000	30.486734	-84.024933	79.524002
61	1	1	1	2011-09-12 13:30:11.000	30.486925	-84.025009	79.704002
62	1	1	1	2011-09-12 13:30:12.000	30.487118	-84.025085	78.480003

Index	Route	Driver	Trip	Time	Latitude	Longitude	Speed
63	1	1	1	2011-09-12 13:30:13.000	30.487301	-84.025154	75.671997
64	1	1	1	2011-09-12 13:30:14.000	30.48748	-84.025223	72.431999

Probe vehicle data obtained from vehicle runs conducted by FDOT was provided in the following format:

- Timestamp: Updated on a per-second basis (Eastern Standard Time)
- Geographic Segment: Breadcrumb trail of GPS location coordinates
- Instantaneous Point speed (miles per hour)

The GPS data from the probe vehicles was imported into Microsoft® Streets and Trips software utilizing latitudinal and longitudinal information provided by the GPS probes. As shown in Figure 2-6, the start and end times for the drivers can be determined by importing the vehicle probe data onto the map and comparing the start and end points for each TMC code.

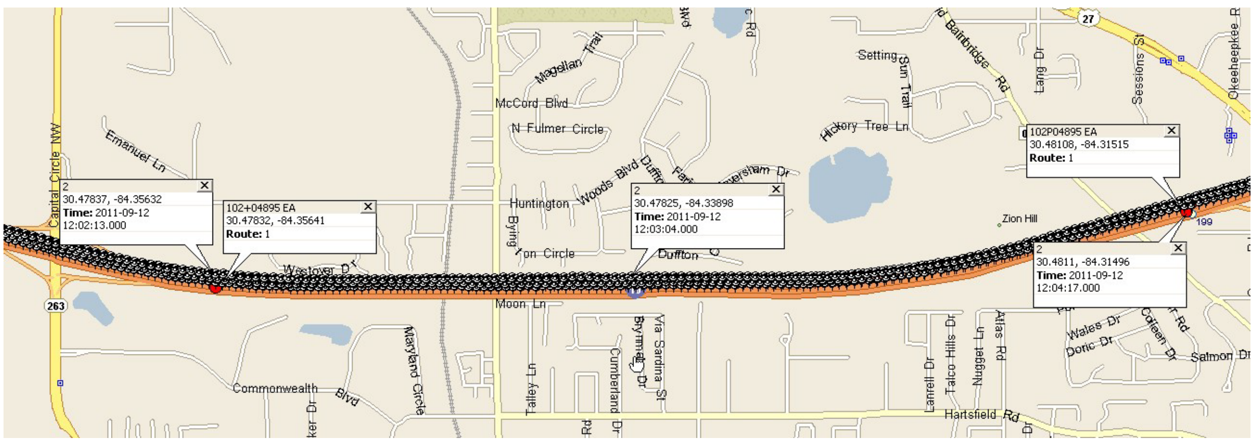

Figure 2-6: Floating Vehicle Data & TMC Codes in Street and Trips

Data from the drive testing was aggregated for analysis against TMC code-based roadway segments. This data was subsequently processed and compared with NAVTEQ, TrafficCast, and INRIX data collected along the same roadway segments during the same period of time for analysis.

After the raw location data and time stamps were mapped to their corresponding TMC segments, the travel time and distance for each segment were used to calculate the average travel speed along that segment. An example of the data spreadsheet analysis is shown in Table 2-6.

Table 2-6: Sample Floating Vehicle Data for a TMC Segment

102+08156	Date	Start	End	TT (min)	TT (sec)	dist (miles)	MPH
						0.355	
FDOT GT Driver 1							
Pass 1	9/13/2011	6:06:38	6:08:46	2.13	128	0.355	9.99
Pass 2	9/13/2011	7:49:38	7:52:41	3.05	183	0.355	6.99
Pass 3	9/13/2011	9:24:20	9:26:50	2.50	150	0.355	8.52
Pass 4	9/13/2011	10:54:05	10:56:25	2.33	140	0.355	9.13
FDOT GT Driver 2							
Pass 1	9/13/2011	6:11:45	6:13:05	1.33	80	0.355	15.98
Pass 2	9/13/2011	7:55:02	7:57:27	2.42	145	0.355	8.82
Pass 3	9/13/2011	9:28:05	9:29:58	1.88	113	0.355	11.32
Pass 4	9/13/2011	10:58:41	11:00:43	2.03	122	0.355	10.48
FDOT GT Driver 3							
Pass 1	9/13/2011	6:16:16	6:17:32	1.27	76	0.355	16.82
Pass 2	9/13/2011	8:00:11	8:02:16	2.08	125	0.355	10.23
Pass 3	9/13/2011	9:33:42	9:35:42	2.00	120	0.355	10.66
Pass 4	9/13/2011	11:03:31	11:05:20	1.82	109	0.355	11.73
FDOT GT Driver 4							
Pass 1	9/13/2011	6:20:56	6:23:20	2.40	144	0.355	8.88
Pass 2	9/13/2011	8:05:32	8:07:06	1.57	94	0.355	13.60
Pass 3	9/13/2011	9:37:56	9:40:02	2.10	126	0.355	10.15
Pass 4	9/13/2011	11:07:19	11:09:31	2.20	132	0.355	9.69

2.2.4 License Plate Reader System

License plate reader (LPR) data was compared to the NAVTEQ, TrafficCast, and INRIX system data on I-10 for multiple TMC codes. The LPR data was reported in terms of LPR segments, so the speed and travel time was linearly recalculated to estimate the portion that overlays the TMC codes.

The data provided by SunGuide software includes a timestamp, average speed, total volume, and travel time for each LPR segment, reported at 15-minute intervals. The travel time and speeds reported are based on an average of the past 15 minutes. Table 2-7 provides an example of the recalculated LPR speed and travel time for TMC code 102-04896.

Table 2-7: Sample LPR Data

102-04896 is 68.4 Percent of LPR Segment 210WB			
Date Time	Average Speed	Total Volume	Travel Time
09/12/2011 08:00	73	14	6.2000
09/12/2011 08:15	73	7	6.1667
09/12/2011 08:30	69	13	8.1333
09/12/2011 08:45	72	11	6.6500
09/12/2011 09:00	74	14	6.1167
09/12/2011 09:15	68	10	7.6000
09/12/2011 09:30	76	12	5.9500
09/12/2011 09:45	74	16	6.0833
09/12/2011 10:00	76	8	5.9833
09/12/2011 10:15	71	19	6.5000
09/12/2011 10:30	73	10	6.1833
09/12/2011 10:45	75	14	6.0500
09/12/2011 11:00	75	14	6.0333
09/12/2011 11:15	65	17	8.6667
09/12/2011 11:30	71	18	6.7167
09/12/2011 11:45	72	23	6.2833
09/12/2011 12:00	66	13	7.0167

3 Methods of Analysis

The goal of this analysis was to make a direct comparison of floating car data with system data reported by NAVTEQ, TrafficCast, and INRIX. A secondary comparison (for multiple I-10 segments) was conducted to compare the NAVTEQ, TrafficCast, and INRIX system data with LPR data obtained from the SunGuide software.

3.1 Floating Vehicle versus Vendor Data

The floating vehicle data provided by FDOT was compared with data from the three vendor's system data.

Examples of typical travel time comparisons and speed comparisons for Route 1 (I-10) are shown in the figures below. The FDOT ground truth results generally match the vendor's data well for this particular TMC code from graphical inspection. Complete results for each TMC segment are provided in Appendix A.

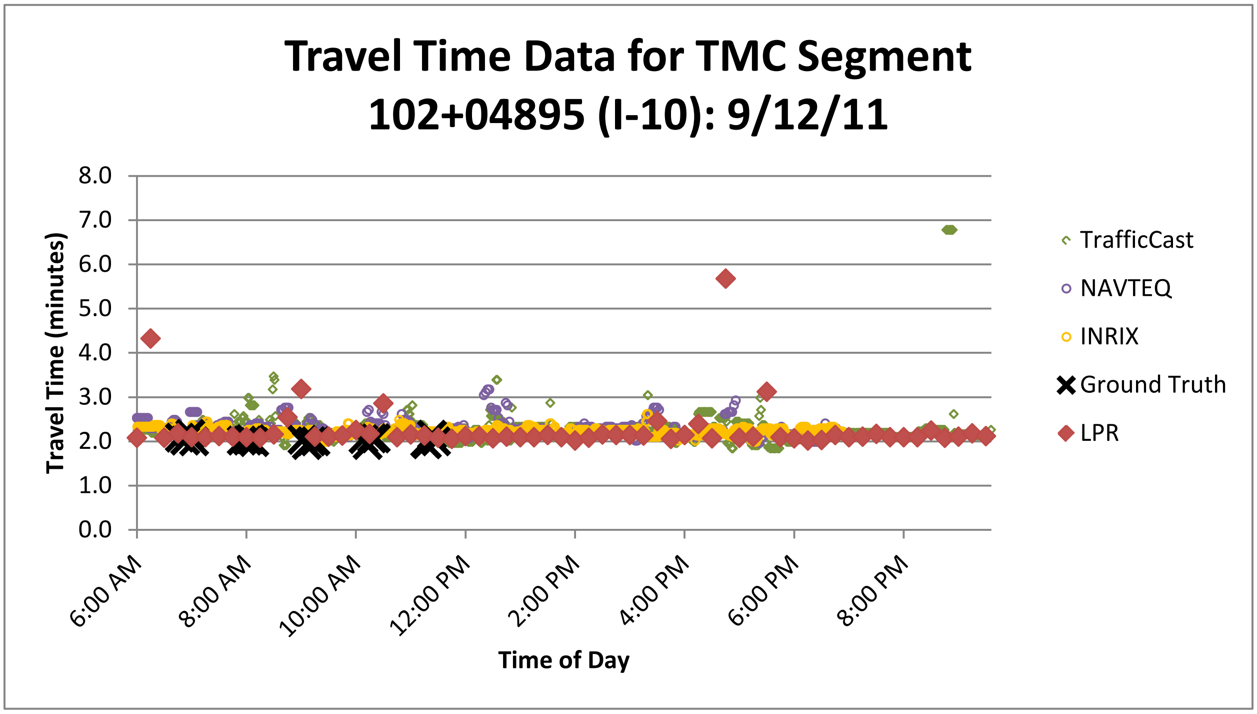

Figure 3-1: Typical Travel Time Comparison for Route 1

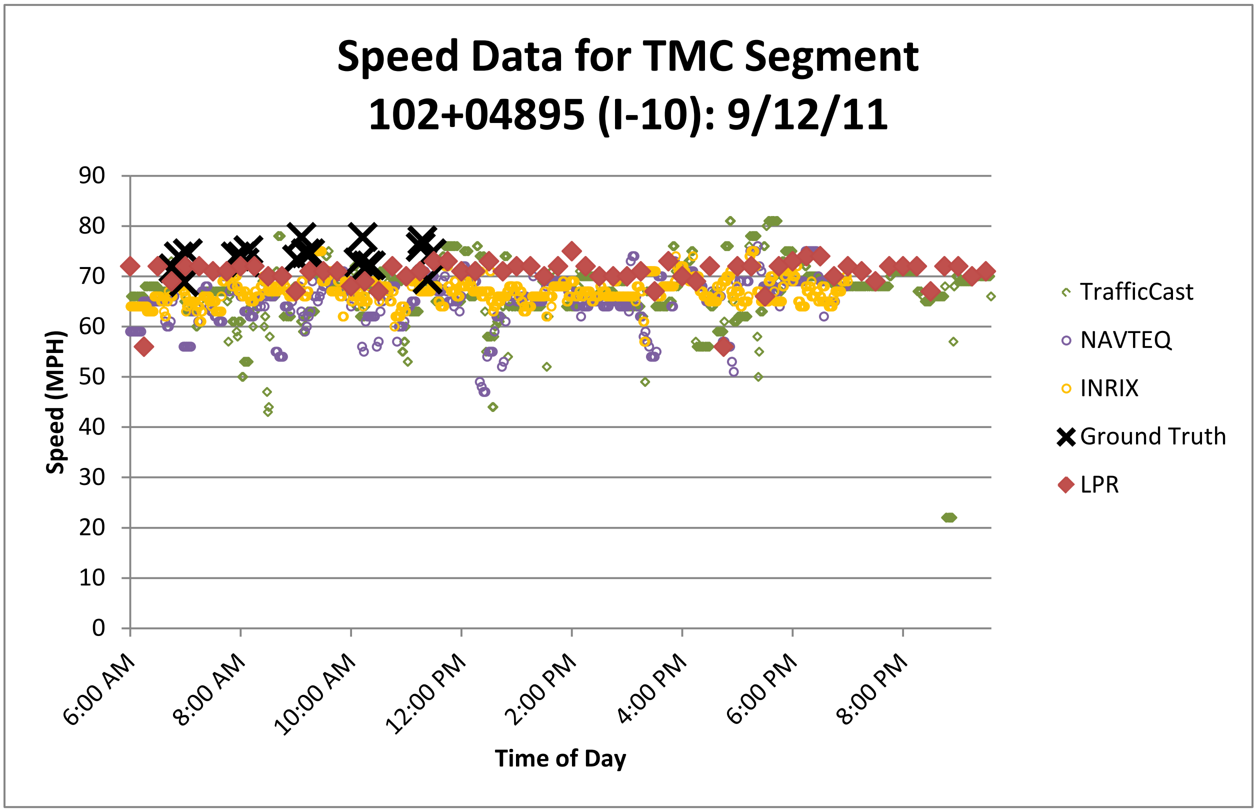

Figure 3-2: Typical Speed Comparison for Route 1

3.1.1 Data Analysis Metrics

Several metrics were used to determine the accuracy of the vendor's system data. The speed-based metrics are derived from metrics used in recent work by the I-95 Corridor Coalition to evaluate the accuracy of commercial travel time data. These metrics are described below.

Absolute Average Speed Error

- The absolute average speed error indicates the difference in ground truth speed from the vendor system speed and is calculated by:

    1. Calculate the difference between each ground truth speed and the corresponding vendor speed. (Corresponding speed is determined by when the vehicle exits the TMC segment.)

    2. Take the absolute value of each difference calculated in Step 1.

    3. Take the average of the values calculated in Step 2.

$$\text{Absolute Error (i,j)} = \text{Abs}(S_{ij}(\text{Vendor}) - S_{ij}(\text{Validation}))$$
$$\text{Average Absolute Error} = \text{Mean}(\text{Absolute Error(i,j)})$$
where:
S_{ij}(Vendor) = data for segment i at time j from Vendor feed
S_{ij}(Validation) = ground truth time for segment i at time j

Average Speed Bias

- The average speed bias indicates the tendency to over- or under-report speed and is calculated by:

  1. Calculate the difference between each ground truth speed and the corresponding vendor speed. (Corresponding speed is determined by when the vehicle exits the TMC segment.)

  2. Take the average of the values calculated in Step 1.

$$\text{Error (i,j)} = S_{ij}(\text{Vendor}) - S_{ij}(\text{Validation})$$
$$\text{Average Error} = \text{Mean}(\text{Error(i,j)})$$
where:
S_{ij}(Vendor) = data for segment i at time j from Vendor feed
S_{ij}(Validation) = ground truth time for segment i at time j

For comparison, we also chose to calculate metrics in terms of absolute average travel time error and travel time bias.

Absolute Average Travel Time Error

- Used to determine difference in ground truth travel time from vendor system, this metric is calculated by:

  1. Calculate the difference between each ground truth travel time and the corresponding vendor travel time. (Corresponding time is determined by when the vehicle exits the TMC segment.)

  2. Take the absolute value of each difference calculated in Step 1.

  3. Take the average of the values calculated in Step 2.

$$\text{Absolute Error (i,j)} = \text{Abs}(S_{ij}(\text{Vendor}) - S_{ij}(\text{Validation}))$$
$$\text{Average Absolute Error} = \text{Mean}(\text{Absolute Error(i,j)})$$
where:

S_{ij}(Vendor) = data for segment i at time j from Vendor feed

S_{ij}(Validation) = ground truth time for segment i at time j

Travel Time Bias

- Used to determine the tendency to over- or under-report travel time, the travel time bias is calculated by:

  1. Calculate the difference between each ground truth travel time and the corresponding vendor travel time. (Corresponding time is determined by when the vehicle exits the TMC segment.)

  2. Take the average of the values calculated in Step 1.

Error (i,j) = S_{ij}(Vendor) – S_{ij}(Validation)

Average Error = Mean(Error(i,j))

where:

S_{ij}(Vendor) = data for segment i at time j from Vendor feed

S_{ij}(Validation) = ground truth time for segment i at time j

Other metrics utilized in the analysis of data included the following.

Percentage of Time that Data is Changing

- Demonstrates the percentage of time the vendor speed data changed minute-by-minute.

Percent Change (t) = Sum(If([S_{t+1}]-$S_t \neq 0$ then “1” else “0” for t = 1 to n-1)) / n

Where:

S_t = Speed at time t

n= total number of speed samples

3.1.2 Analysis of the Data

Part of the data analysis included charting the percent of time the reported speeds from a vendor changed from the previous speed reported. If the same speed is reported by a vendor for an extended time (hours or even days) without changing, it may suggest that the vendor does not have penetration of probe data in that TMC segment. Figure 3-3 shows the percent of data change for Route 1 where the percent of data change was fairly consistent along the route.

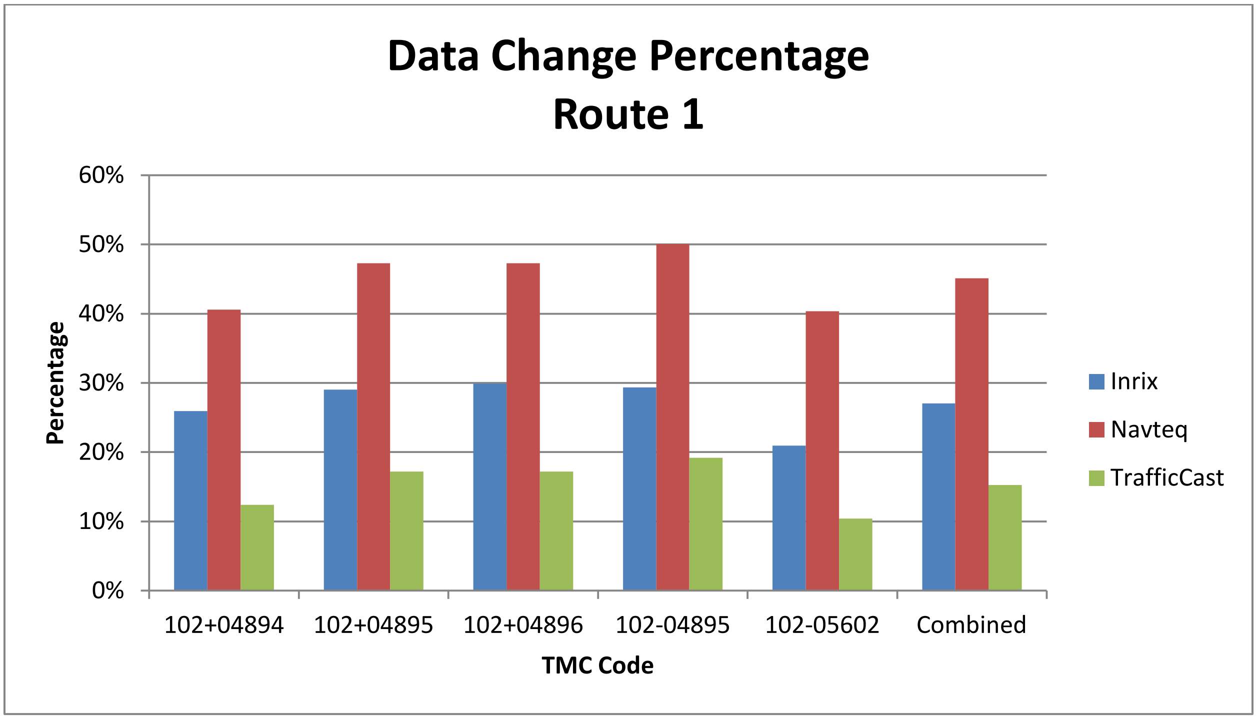

Figure 3-3: Percentage of Time Data is Changing (Route 1)

Figure 3-4 shows an example where the percent of data change results were low. By comparison of the map in Figure 2-3 it is reasonable to suggest that the vendors may have less penetration in these rural areas since there is potentially less traffic.

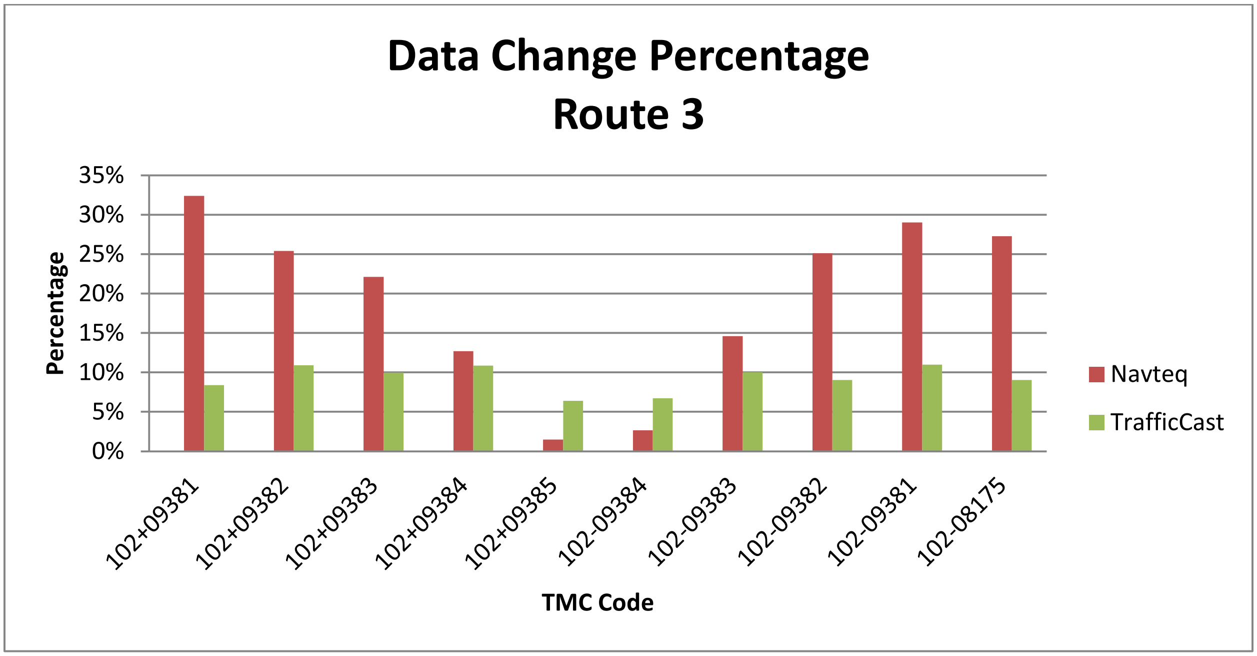

Figure 3-4: Percentage of Time Data is Changing (Route 3)

Figure 3-5 provides the average absolute aggregated speed error for the TMC segments above. Interestingly, the error reported in these rural segments are some of the lowest errors for the route.

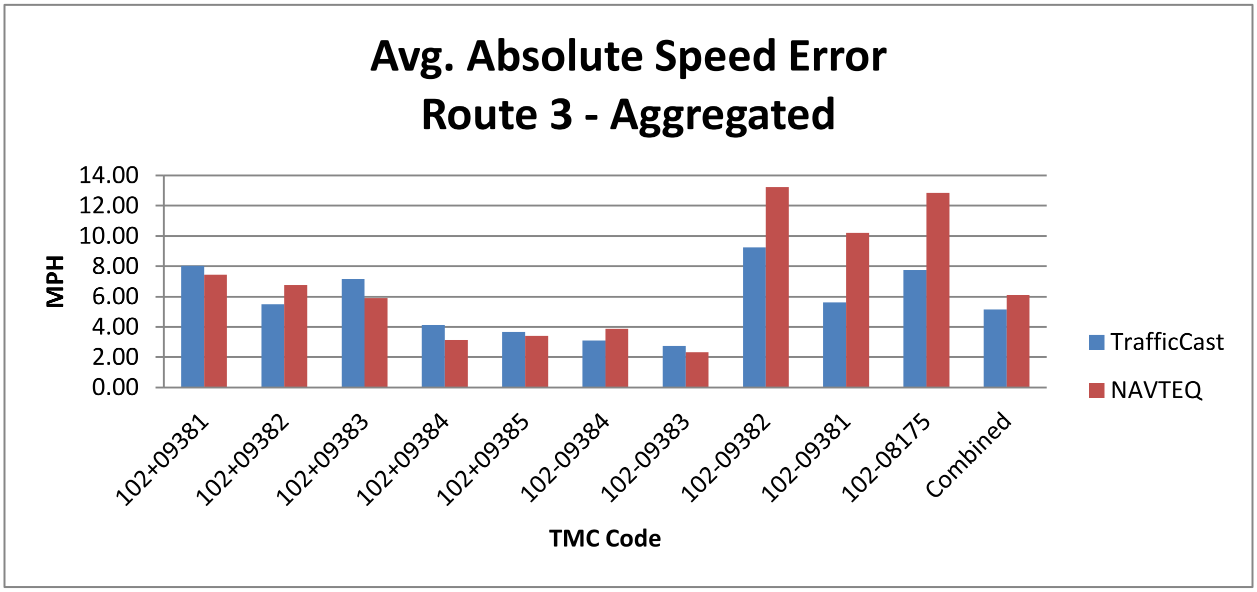

Figure 3-5: Average Absolute Speed Error (Route 3)

In urban areas, particularly Capital Circle (Route 4), the TMC codes are generally shorter and there are traffic signals within many of the segments. This resulted in large variability in the observed speeds and travel times from all data sources. Figure 3-6 shows a typical example of a TMC code with these characteristics.

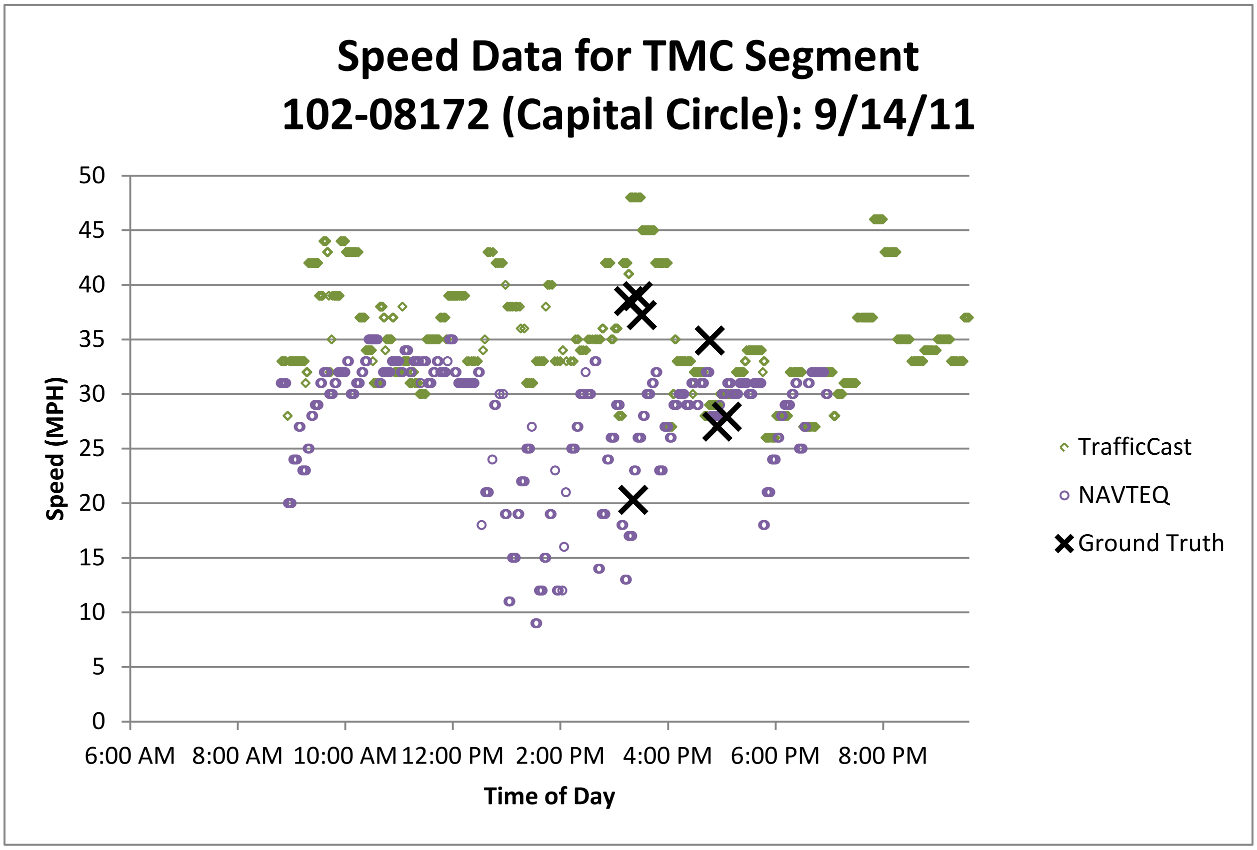

Figure 3-6: Typical Speed Results for an Urban Segment

By combining several individual segments into one "Urban" route the variability was much less pronounced. Drivers that were "caught" at one red light may get a "green" light at the next providing as average of the travel times across a larger region. Figure 3-7 shows speed data for the clockwise urban Route 4.

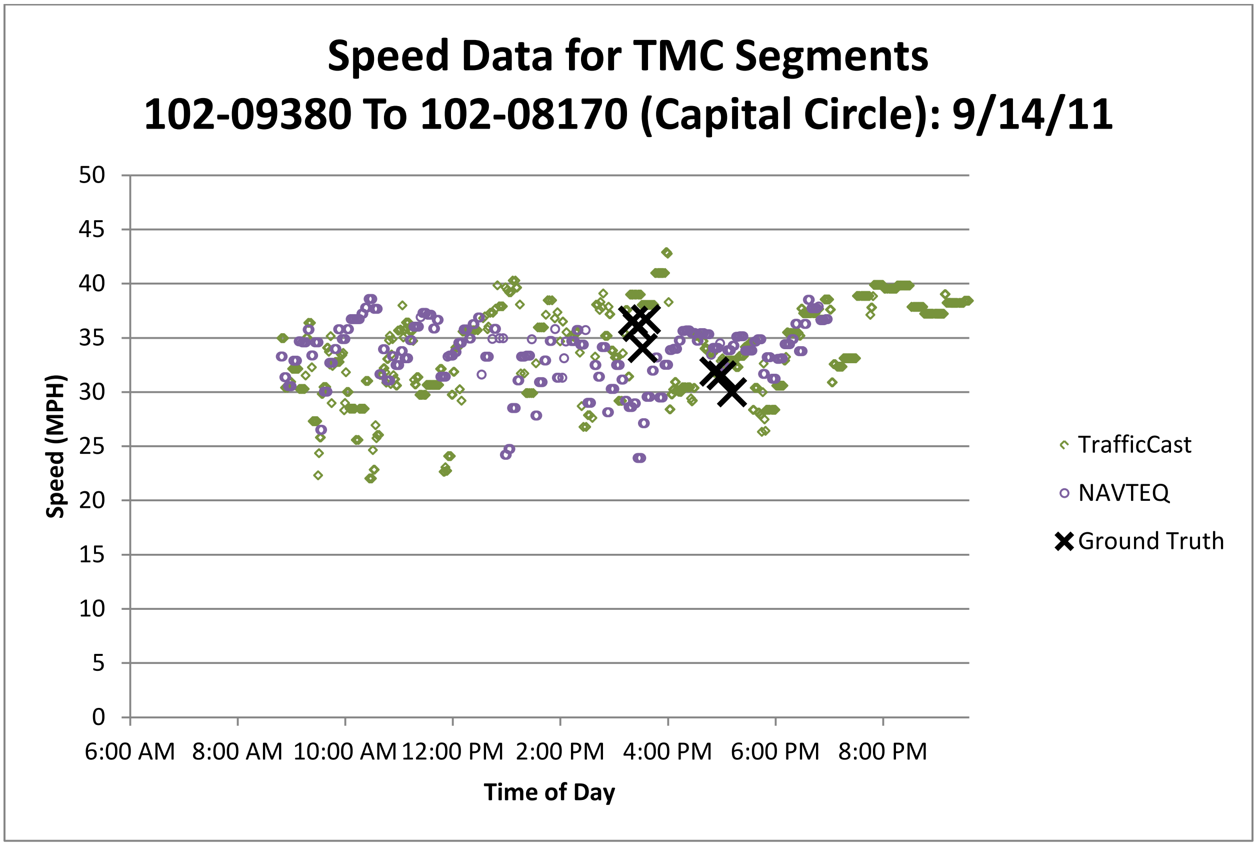

Figure 3-7: Speed Data for Clockwise Urban Route 4

Time of day was also considered for the percent of time the data changed. During peak times, a slight increase in the percent of time data change was observed; however, no strong differences or trends were suggested in the results. During peak times, there is greater traffic and hence, increased GPS probe data available.

Some of the TMC codes analyzed were deemed too short to be considered for further analysis. TMC codes, particularly TMC codes containing a “P” or “N” were often less than one mile and many were less than one-tenth of a mile in length. The GPS probe data provided by FDOT drivers is reported each second. With a resolution of one second for the ground truth, TMC codes that were extremely short experienced very high error. This can be attributed to the fact that drivers would pass through these short TMC codes within one to two seconds. Therefore, TMC codes less than one-tenth of a mile were flagged and not included in the results.

The data provided by the vendors appears to remain stable throughout the testing period. No “holes” or “missing” data were observed when analyzing the results. It was noted during the analysis that NAVTEQ’s file sizes are much larger than the other vendors. This resulted in a lengthier processing time.

Vendor schemas varied in fields and size. Table 3-1 shows vendor-specific data regarding file content and size.

Table 3-1: Vendor-specific Data

Vendor	Fields Per/observation	Average File Size (KB)
Traffic Cast	3	58
INRIX (core)	7	45
INRIX (extended)	7	204
NAVTEQ	14	24586

In the context of real-time or near real-time use of the data, processing of TrafficCast and INRIX requires reasonable resources. Changing the amount of observations considered would change resource costs of processing the data. The NAVTEQ data contains significant overhead per observation; for a real-time or near real-time system, redundant details are given with each observation. It is suggested that if NAVTEQ data is used in such a system, NAVTEQ should provide a leaner schema to reduce observation size significantly.

4 Conclusions

Overall, the data looks fairly consistent by observation of the graphs (see appendix for the complete results). NAVTEQ, TrafficCast, and INRIX are all generally consistent with the ground truth and the LPR data and no significant differences in data accuracy between the three vendors were observed, as shown in Table 4-1. The percent of speed bias and speed error appear within a reasonable range for the majority of TMC segments. For comparison, the I-95 Corridor Coalition (which subscribes to INRIX data) considers an absolute average speed error of 10 MPH or less and a speed error bias of ± 5MPH to be acceptable. In the majority of cases, the data observed during this test passed these criteria. As discussed in Methods of Analysis, TMC segments in urban areas with traffic signals experienced a larger variability in the results.

Table 4-1: Vendor Data Accuracy Summary

	Abs. Avg. Speed Error (mph)			Speed Error Bias (mph)		
	INRIX	TrafficCast	NAVTEQ	INRIX	TrafficCast	NAVTEQ
Route 1						
Overall	6.27	7.32	7.39	-5.87	-6.48	-6.76
Route 2						
Urban NB	-	3.85	3.93	-	1.51	1.94
Urban SB	-	4.71	3.65	-	-1.83	-0.04
Overall	-	6.79	5.64	-	-3.57	-2.95
Route 3						
Overall	-	5.14	6.10	-	-1.84	-3.79
Route 4						
Urban CW	-	4.64	5.07	-	-2.35	-3.52
Urban CCW	-	4.38	3.58	-	-4.07	-1.93
Overall	-	8.70	8.72	-	-4.83	-5.17

The percent of data change analysis did not suggest that any of the vendors were lacking obvious data penetration for any extended time periods. It appears that if "live" probe data or senor data was not available the systems would use other sources of data such as historical speeds to report.

With regards to specific vendors, the INRIX data on Route 1 appeared to have a slight advantage in accuracy; however, we were not able to evaluate INRIX data in parallel with TrafficCast and NAVTEQ on Routes 2-4. In comparing the summary data for the remaining vendors, the errors in the TrafficCast data appear to be slightly lower on average than the NAVTEQ data and also more consistent (less variable when compared route-to-route). The TrafficCast XML feed is also more streamlined and closer to the current format provided by INRIX to FDOT; therefore, the TrafficCast data may be easier to work with and integrate into current systems. However, it should be noted that NAVTEQ has some extensive developer resources and their feeds have been successfully integrated into numerous large-scale travel time systems, so their more complicated XML format should not automatically preclude them from consideration.

Appendix A:

Floating Vehicle Data
Versus
NAVTEQ, TrafficCast, and INRIX System Data

September 12th, 2011 to September 15th, 2011

Appendix B:

Average Time Error
And
Average Speed Error
Summaries

September 12th, 2011 to September 15th, 2011

Appendix C:

Percent of Data Change

September 12th, 2011 to September 15th, 2011