SUMMARY

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. One alternative to the deployment of roadside equipment to produce travel time data is to subscribe to travel time data from commercial data providers. In 2011, FDOT was approached by two commercial data providers with a desire to share their travel time data on a trial basis. This provided FDOT the unique opportunity to test commercial travel time data from two vendors simultaneously. This paper compares travel time data from the two providers against ground truth data obtained through floating car methodology. Comparison with INRIX travel time data was also possible along one of the routes. The routes examined encompassed both limited access routes and arterials, along with urban and rural segments.

KEY WORDS

Travel Times, Data Collection, Florida Department of Transportation (FDOT), Tallahassee, Data Analysis, Third-Party Data, INRIX, TrafficCast, NAVTEQ.

BACKGROUND

The Florida Department of Transportation (FDOT) desires a cost-effective method of receiving current traffic information, particularly travel time information, on major portions of the interstate and state road networks throughout the state. Deploying traditional intelligent transportation systems along the roadway to gather statewide traffic information for arterial roadways and rural areas in the state is currently impractical and cost prohibitive.

Commercial traffic and travel time data is now available from a number of private companies, which can mitigate the need for agencies to deploy and maintain communication infrastructure, vehicle detectors, and other equipment necessary to gather traffic data. Commercial traffic data is sold to agencies on a subscription basis and 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 traffic 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 paper 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 time data generated from roadside sensors by SunGuide® software (Florida’s statewide Traffic Management Center 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 travelled these roadway segments and provided floating car travel time data for data comparison.

**COLLECTING THE DATA**

As described above, the data for the test was collected along four separate routes:

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

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

Data was collected from several sources using the following methods. Each vendor provided a live XML stream for their speed and travel time data, so FDOT staff simply logged the live XML stream from each system. Each XML stream contained speeds and/or travel times and other information for each TMC location segment that composed the driven route. FDOT also requested each vendor to provide a disc of archived system data as a backup.

Ground truth data was obtained using vehicles equipped with GPS loggers. The vehicles traveled each of the four routes over a total period of four days, spending half a day on one
route before moving to another route. In this way, each route was driven in peak and off-peak conditions. These vehicles were driven along the test routes at a speeds representative of the surrounding traffic, following the floating car methodology specified in the Florida Department of Transportation’s Manual on Uniform Traffic Studies (MUTS). Four drivers traveled in a loosely-spaced group to allow the variability in travel speeds to be visually approximated. The ground tracks from each GPS logger were downloaded at the conclusion of each day and post-processed to correlate with the boundaries of each TMC location segment.

FDOT has also deployed a license plate reader (LPR) system on a small section of I-10 in the Tallahassee area to calculate their own travel times. This data provided another benchmark by which to compare vendor travel times against.

**ANALYZING THE DATA**

The goal of the data analysis was to evaluate the data based on several factors. For the data to be trustworthy, it not only needed to be accurate at a given point in time but it also needed to be consistent over time. This called for a series of qualitative and quantitative methods. Quantitative comparisons of the vendor data to floating vehicle ground truth tests identified the accuracy of each vendor’s data in both peak and off-peak conditions. Other qualitative metrics were used to identify the reliability of data over time and spot potential issues with data refresh intervals.

**Data Harmonization**

Before the data sets were analyzed and compared, a level of data harmonization was required. Vendor XML logs were imported into tables that listed speeds and travel times based on specific timestamps. The timestamps of the vendor data were typically 1-2 minutes apart. Separate tables were created for each TMC location segment across the four routes.

The evaluation team also observed that certain vendors reported their data in slightly different ways with respect to TMC location segments. TMC location segments are composed of two major types—external segments and internal segments. External TMC segments typically represent links between interchanges or other geographic boundaries. Internal TMC segments typically represent distances between the off-ramps and on-ramps within a specific interchange. As such, internal TMC segments are usually much smaller than the adjacent external segments. INRIX and TrafficCast report discreet data for internal and external TMC location segments. NAVTEQ combines the internal segment with the adjacent external segment. While this difference does not present, it does underscore that agency users may need to configure systems in slightly different ways based upon a particular vendor’s data format. In order to facilitate comparisons in this project, the team needed to make sure travel time links were equivalent. This was accomplished by taking the speed data reported by NAVTEQ and converting it to a calculated travel time based on the official length of the corresponding TMC segment.
As mentioned earlier, the GPS log data was also harmonized to correspond with each individual TMC location segment. Using mapping software, the evaluation team identified the time each vehicle entered and exited a TMC location segment and used this information to calculate the driven travel time. Using the length of the segment, this travel time was converted to an average speed.

**Floating Vehicle Data versus Vendor Data**

**Visual Comparison of Data**

After harmonizing the data, it was superimposed on plots that compared the vendor speed and travel time data with those of the ground truth data and average data obtained from the LPR travel time system (where available). Figure 1 provides an example of typical speed data reported for a typical TMC location segment on I-10. The three vendor data sources were fairly consistent with each other and with the spread of ground truth speeds that were recorded. The data set labeled “LPR” represents speeds obtained from FDOT’s existing travel time sensors on this stretch of I-10.

![Speed Data for TMC Segment 102+04895 (I-10): 9/12/11](image)

*Figure 1: Sample speed data from one TMC segment on I-10*

**Data Analysis Metrics**
The next step was to identify a set of appropriate data analysis metrics to allow quantitative comparisons between the ground truth data and the vendor-supplied speed data. The speed-based metrics that were used are adaptations of those selected 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:

\[
\text{Absolute Error (i,j)} = \text{Abs}(S_{ij}(\text{Vendor}) - S_{ij}(\text{Validation}))
\]

\[
\text{Average Absolute Error} = \text{Mean(Absolute Error(i,j))}
\]

where:

- \(S_{ij}(\text{Vendor})\) = data for segment i at time j from Vendor feed
- \(S_{ij}(\text{Validation})\) = ground truth time for segment i at time j

**Average Speed Bias**

The average speed bias indicates the tendency of the vendor data to over- or under-report speed as compared to ground-truth observations:

\[
\text{Error (i,j)} = S_{ij}(\text{Vendor}) - S_{ij}(\text{Validation})
\]

\[
\text{Average Error} = \text{Mean(Error(i,j))}
\]

where:

- \(S_{ij}(\text{Vendor})\) = data for segment i at time j from Vendor feed
- \(S_{ij}(\text{Validation})\) = ground truth time for segment i at time j

**DATA ANALYSIS**

Before proceeding with a presentation of the data analysis, it is important to understand that short TMC segments are prone to higher intrinsic error than longer TMC segments. This error is due in part to the inherent positional uncertainty of commercial GPS data. Most commercial non-differential systems do a reasonable job establishing position within ±3 meters (about 10 feet) of the true position. Stated another way, the true location of each ground truth vehicle would lie somewhere within a 20-foot diameter circle surrounding the vehicle. Because of this, there is a level of uncertainty of exactly when the ground-truth vehicle crossed a TMC segment boundary. Over a one mile segment, a ±10 foot error associated with a vehicle’s location would be negligible. However, over a 0.1 mile segment this error can be quite high, particularly if those segments are signalized. Therefore, very short TMC segments were excluded from the analysis (0.1 miles or less)

Some of the routes spanned heavily-signalized urban arterials, which created high variability on some of the associated TMC location segments. This variability was confirmed in the
corresponding ground-truth readings as well. An example of this variability can be seen in Figure 2. Note that the Department only has access to INRIX data on I-10 within the area used for this study. Therefore, it does not appear on this graph.

**Figure 2: Speed readings from a signalized TMC segment**

When consecutive TMC segments were combined into longer urban segments, the variability in the average speed was noticeably reduced, as shown in Figure 3. The average speed of an urban segment was determined using the space mean speed of the segment. That is, the travel times for the individual TMC segments were summed up. The overall length of the urban segment was then divided by the sum of the individual travel times to determine the average speed. This reduction in variability makes intuitive sense when driving along a signalized arterial, particularly during coordination when some platoons of vehicles travel within the green band and others don’t. These characteristics tend to even out the overall average speed over a long segment.
We then computed the absolute average speed error and speed error bias for each vendor over each of the four routes. The summary data error data is listed below in Table 1.

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). However, the differences between vendor accuracy did not appear to be significant when compared over the one-week period of the test.
Table 1: Summary of Vendor Data Accuracy Metrics

<table>
<thead>
<tr>
<th>Route 1</th>
<th>INRIX</th>
<th>TrafficCast</th>
<th>NAVTEQ</th>
<th>INRIX</th>
<th>TrafficCast</th>
<th>NAVTEQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>6.27</td>
<td>7.32</td>
<td>7.39</td>
<td>-5.87</td>
<td>-6.48</td>
<td>-6.76</td>
</tr>
<tr>
<td>Route 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban NB</td>
<td>-</td>
<td>3.85</td>
<td>3.93</td>
<td>-</td>
<td>1.51</td>
<td>1.94</td>
</tr>
<tr>
<td>Urban SB</td>
<td>-</td>
<td>4.71</td>
<td>3.65</td>
<td>-</td>
<td>-1.83</td>
<td>-0.04</td>
</tr>
<tr>
<td>Overall</td>
<td>-</td>
<td>6.79</td>
<td>5.64</td>
<td>-</td>
<td>-3.57</td>
<td>-2.95</td>
</tr>
<tr>
<td>Route 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>-</td>
<td>5.14</td>
<td>6.10</td>
<td>-</td>
<td>-1.84</td>
<td>-3.79</td>
</tr>
<tr>
<td>Route 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban CW</td>
<td>-</td>
<td>4.64</td>
<td>5.07</td>
<td>-</td>
<td>-2.35</td>
<td>-3.52</td>
</tr>
<tr>
<td>Urban CCW</td>
<td>-</td>
<td>4.38</td>
<td>3.58</td>
<td>-</td>
<td>-4.07</td>
<td>-1.93</td>
</tr>
<tr>
<td>Overall</td>
<td>-</td>
<td>8.70</td>
<td>8.72</td>
<td>-</td>
<td>-4.83</td>
<td>-5.17</td>
</tr>
</tbody>
</table>

**CONCLUSION**

In conclusion, when the aggregated data metrics were compared, the test observed very little difference between the travel time accuracy of NAVTEQ and TrafficCast data on the routes tested. Care should be taken before applying these results to other locations, as the quality of the underlying travel time data may differ by region, depending upon the data coverage of each provider and their sources. For this reason, it is important for agencies to evaluate third-party data before purchasing, either by performing a test similar to the one above or by utilizing Bluetooth readers over an established corridor, as others have done. Understanding how data accuracy compares between vendors in a region makes an agency a more informed customer and better able to select an appropriate vendor which provides the right balance of price and accuracy.