Stuck in Traffic: Analyzing Real Time Traffic Capabilities of Personal Navigation Devices and Traffic Phone Applications

by

Bruce M. Belzowski
Research Area Specialist
Automotive Analysis

Andrew Ekstrom
Research Associate
Automotive Analysis

UNIVERSITY OF MICHIGAN
TRANSPORTATION RESEARCH INSTITUTE

University of Michigan Transportation Research Institute

for

TomTom Group

Revised January, 2014

Contents

Abstract

The global positioning system (GPS) market is a fast changing, highly competitive market. Products change frequently as they try to provide the best customer experience for a service that is based on the need for real-time data. Two major functions of the GPS unit are to correctly report traffic jams on a driver's route and provide an accurate and timely estimated time of arrival (ETA) for the driver whether he/she is in a traffic jam or just following driving directions from a GPS unit. This study measures the accuracy of traffic jam reporting by having Personal Navigational Devices (PNDs) from TomTom and Garmin and phone apps from TomTom, INRIX, and Google in the same vehicle programmed to arrive at the same destination. We found significant differences between the units in terms of their ability to recognize an upcoming traffic jam. We also found differences in how well the devices responded to jams when driving on surface streets versus highways, and whether the jams were shorter or longer in length. We see potential for auto manufacturers to employ real-time traffic in their new vehicles, providing potential growth for real-time traffic providers through access to new vehicles as well as the aftermarket.

Keywords: global positioning system, GPS, PND, traffic app, jam hunt

Acknowledgements

We would like to thank our sponsor, TomTom Group, for their support for this project, particularly Harriet Chen-Pyle, Tomas Coomans, Jeff McClung, and Nick Cohn. Our own UMTRI support came from our drivers Ewan Compton and Daniel Hershberger, and our Engineering Group: Dave LeBlanc, Mark Gilbert, John Koch, Dan Huddleston, Doug Mechan, John Drew, and John Hall. Without their mobile electronic data collection insight and technical capability, our project would not have been possible.

Stuck in Traffic: Analyzing Real Time Traffic Capabilities of Personal Navigation Devices and Traffic Phone Applications

Executive Summary

Our research focus for this report is on measuring the real time traffic capability and accuracy of Personal Navigation Devices (PND) and Smartphone applications (apps). Our emphasis is on how well these devices accurately report traffic jams. We examined PNDs from TomTom and Garmin, and Smartphone apps from TomTom, Google, and INRIX, using a unique field-test process for measuring the accuracy of each device that entailed taking simultaneous videos of each unit on the same vehicle.

Our analyses include comparisons of how well all the units reported:

- All traffic jams
- All traffic jams on surface streets
- Traffic jams on surface streets lasting less than or equal to five minutes
- Traffic jams on surface streets lasting longer than five minutes
- All traffic jams on highways
- Traffic jams on highways lasting less than or equal to ten minutes
- Traffic jams on highway lasting longer than ten minutes

For all traffic jams, the TomTom PND and App accurately reported 67 percent and 66 percent of the jams, respectively. Using logistic regression, we found that there are statistically significant differences between the TomTom PND and the Google App (52 percent), the INRIX App (38 percent), the Garmin HD PND (22 percent), and the Garmin SIM PND (4 percent).

It was difficult for the tested devices to accurately report surface street jams. The Google App (48 percent), the TomTom PND (43 percent), and the TomTom App (38 percent) reported jams more accurately than the other units. There are statistically significant differences between the TomTom PND and the INRIX App (12 percent), the Garmin HD PND (7 percent), and the Garmin SIM PND (3 percent). There are not statistically significant differences between the TomTom PND and the TomTom App and the Google App.

We divided traffic jams on surface streets into those that last less than or equal to five minutes and those longer than five minutes. The tested devices had difficulty accurately reporting both types of surface street jams, and the shorter the jam the harder it was to report accurately. For the less than or equal to 5 minutes jams, the TomTom PND (43 percent), Google App (39 percent), and the TomTom App (35 percent) were best at accurately reporting jams. For the over five minute jams on surface streets, the Google App (58 percent), the TomTom PND (42 percent) and the TomTom App (42 percent) reported these types of jams accurately, though there is not statistical difference between the TomTom PND and the other units.

The tested devices were better at reporting traffic jams on highways than on surface streets. For accurately reporting traffic jams on highways, the TomTom PND (76 percent) and the TomTom App (76 percent) outperformed the other devices. There is a statistically significant difference between the TomTom PND and the Google App (54 percent), the INRIX App (46 percent), the Garmin HD PND (28 percent), and the Garmin SIM PND (4 percent).

Traffic jams on highways were divided into those less than or equal to 10 minutes, and those more than 10 minutes in length. For the highway jams less than or equal to 10 minutes, the TomTom App (74 percent) and TomTom PND (73 percent) reported more jams accurately. The Google App (49 percent), the INRIX App (44 percent), the Garmin HD PND (30 percent), and the Garmin SIM PND (5 percent) are all statistically different from the TomTom PND for this analysis. For highway jams greater than 10 minutes in length, the TomTom PND (86 percent) and App (83 percent) and the Google App (70 percent) recorded the highest readings for accurately reporting traffic jams in this study. There is no statistically significant difference between the TomTom PND and the TomTom and Google Apps.

Three specific issues affected the generalizability of our results: the choice of the Detroit Metropolitan Statistical Area (MSA) for our area of study, device operation, and our decision to define a traffic jam as delaying drivers 90 seconds in getting to their destination while driving half the speed limit. Because the systems in our study may provide different traffic coverage throughout the U.S., our results are only applicable to the Detroit MSA (though it is the 14^{th} largest MSA of 381 MSAs, by population, in the U.S.).

All of our devices required some type of intervention on our part to optimize their operation, but our pre-testing and continual monitoring during testing mitigated these disruptions and offered each device the opportunity to function properly.

Finally, our choice of a 90 second destination delay time while driving half the speed limit as the definition of a traffic jam, based on the results of the study, proved a challenging metric for most of the systems in the study because it rewarded systems that updated their traffic feeds faster and better than other systems. Though more restrictive, a 90 second destination delay provides drivers with more timely information about their traffic situation.

Based on our results, providing dynamic traffic information varies by unit and where the traffic occurs and continues to be a challenge for all PNDs and apps. We are encouraged by the ability of these companies to manage and manipulate the vast amounts of data that is needed to provide real-time traffic data, and we are interested to see what improvements companies will develop for these products that will keep us from being “stuck in traffic.”

Introduction

The global positioning system (GPS) market is a fast changing, highly competitive market. The market is divided into the installed GPS units manufacturers provided to new vehicle buyers, Personal Navigation Device (PND) devices sold as independent units and Smartphone applications (apps). Our research focus for this report is on measuring the traffic capability and accuracy of PNDs and apps.

Especially in the case of PNDs, these products change frequently as they try to provide the best customer experience for a service that is based on the need for real-time data. In order to be more competitive, PNDs try to provide additional functions related to their mapping functions such as local information on restaurants, airports, entertainment venues, hardware stores, and museums to name a few. These are static addresses that tend to remain in one location, yet even these locations need continual updating as businesses close and new businesses open. Keeping abreast of these changes can be a daunting task for companies focused on providing driving maps and traffic information, but all the major PNDs and traffic apps offer these functions. How well each product performs these tasks, as well as how well they handle dynamic tasks such as parking, table seating in restaurants, and current local event information can be a way of measuring the performance of a device or system.

For this report, we will primarily focus on a dynamic core function of the PND/app unit: correctly reporting traffic jams on a driver's route. This study measures the accuracy of traffic jam reporting by examining the stand alone PND and app on a Smartphone. For our study we equipped two vehicles with each of the following six units:

- TomTom PND and App
- Two different Garmin PNDs
- Google App
- INRIX App

Our testing procedure for measuring a PND's response to traffic jams is unique because we actually look for traffic jams, drive into them instead of avoiding them, and measure how the PND responds before, during, and after entering the jam. Because not all of our trips incur a jam, we call non-jam trips "staging trips" that put the driver in place so he can set a new destination and enter a jam during his next trip.

Method

For this study, we used the most recent versions of the PNDs and apps available 'off the shelf' in retail outlets and app stores during early May, 2013. The following are the details for each of the devices.

- TomTom PNDs: GO LIVE 2535M in each vehicle used map version 4/2013 (from 3/2013 quarterly map release) and software version app 12.065.1252068.84 (0) (2081, 4/24/2013)
- Garmin PNDs: one vehicle had a NUVI 1690 (SIM) and NUVI 3590 (HD). The other vehicle was equipped with a NUVI 1695 (SIM) and NUVI 3490 (HD). Maps and traffic information provided by Nokia/NAVTEQ. The NUVI 1690 used map version CN North America NT 2010.20 and software version: 3.2. The Garmin NUVI 1695 used map version CN North America NT 2011.20 and software version: 3.20. The Garmin NUVI 3490 used map version CN North America NT 2013.10 3D and CN North America NT 2013.10 and software version: 7.90. The Garmin NUVI 3590 used map version CN North America NT 2013.10 3D and CN North America NT 2013.10 and software version: 7.90
- TomTom App: Version 1.14 (downloaded 5/24/13 on an iPhone 4)
- Google App: Version 1.1.6 (downloaded 5/22/13 on an Android-based Samsung Galaxy 4G LTE, moved to an iPhone 4 on 6/21/13 and then moved to a Motorola Droid RAZR M phone on 6/24/13
- INRIX App: Version 4.5.1 (downloaded on 5/13/13 on an iPhone 4 and then moved to an Android-based Samsung Galaxy 4G LTE phone on 6/21/13)

Traffic information providers use complex historical and real-time traffic sources to generate traffic updates that are sent to the devices in a vehicle at different intervals, depending on the traffic information provider. Thus the devices in our study rely on obtaining their traffic "feed" regularly in order to provide drivers with the most up-to-date traffic information.

The TomTom and Garmin 1690 and 1695 PNDs in the study received their updated traffic information via an internal SIM card, while the Garmin 3490 and 3590 received their traffic information via an HD Digital Radio signal. All of the phone apps used the Virgin Mobile network for receiving their updated traffic information from their respective provider (i.e. TomTom, Google, and INRIX).

The performance of the PNDs and apps tested depends on the proprietary means of combining traffic data from various sources and across various time scales, as well as the instantaneous quality of network connections. Both the traffic data streams used in companies' algorithms and the network connections may vary across geographical areas, and from city to city across the country. We also need to consider the fact that a wide range of test methodologies would be desirable, because some devices may, by the nature of their algorithms, perform better with one test method than another. Because the subject algorithms are proprietary, the researchers were not able to take all of these factors into account.

Our method for measuring the accuracy of each device is tied to our measurement process that is based on designing and installing a customized shelf that is placed in the passenger side air bag section of two research vehicles. On the shelf, cameras record the screens of all PNDs and apps simultaneously, providing us with a common video array for all six devices in our study. Figure 1 shows one of the vehicles used in the study. The second vehicle is identical.

Figure 1. UMTRI Research Vehicle

Figure 2 shows the customized shelf that contains five PNDs with cameras that are focused exclusively on each device. A sixth device is mounted in the dashboard next to the customized shelf on the center console. A seventh camera is mounted to the rear view mirror and faces the road, providing a view of what the driver sees on the road. A black shield was placed above the PNDs to limit glare on the PND screens.

Figure 2. The Customized Shelf Containing PNDs, Phones, and Cameras.

Video from the cameras is stored on-board each vehicle in a digital video recorder, as shown in Figure 3.

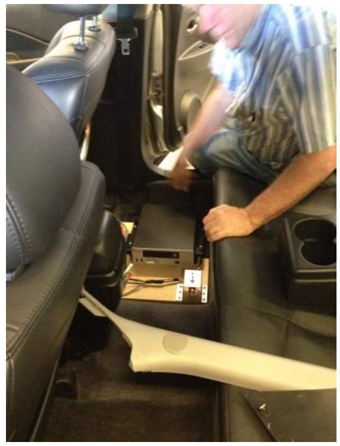

Figure 3: Digital Video Recorder

Video from the digital video recorder is downloaded to our UMTRI computer and the simultaneous videos from each PND and app are shown on one computer screen as shown in Figure 4. This process allows us to code the responses of each PND and app to traffic jams..

Figure 4: View of Simultaneous Videos on UMTRI Computer

Our preparation for data collection took place during May of 2013. During this time we

- Purchased and tested the PNDs and apps to be used in the study
- Designed and installed the customized shelf that included the PNDs, apps, and cameras
- Installed the digital video recorder
- Hired and trained the drivers
- Tested all the equipment during a “dress rehearsal” prior to actual data collection

Each of our drivers was sent on “trips” by our researcher at UMTRI who designated where each driver would drive throughout the Detroit metro area. He chose destinations for jams based primarily on our knowledge of the local area and a combination of local television and radio reports and websites dedicated to providing traffic information about the area. We used these external sources primarily to locate any accidents that created unexpected traffic jams. These traffic jams were primarily during rush hours, but there were also instances of traffic jams at different times of the day and early evening. Drivers began each trip by entering destinations into each of the PNDs and apps, completing a log sheet that included the:

- driver's initials
- date and time of the initial PND entry for the trip
- starting location for the trip
- ending location for the trip
- vehicle's location when it entered the jam
- vehicle's location when it exited the jam
- cause of the jam (if known)
- comments the driver had about this trip.

Though we relied on video provided by the camera (showing the view of the road from the driver's perspective), the log data provided valuable information about the starting and ending location as well as verification of when a vehicle entered a jam, the cause of the jam, and especially any comments by the driver that described anything unusual about trip. This information was particularly important when a PND or app acted erratically.

Ann Arbor was the base from which we drove throughout southeastern Michigan, which is part of the Detroit Metropolitan Statistical Area (MSA) as defined by the U.S. Census. The northern driving boundary was Flint, Michigan, which is about 58 miles from Ann Arbor. The southern driving boundary was Sylvania, Ohio, which is about 45 miles from Ann Arbor. The eastern driving boundary was Chesterfield Township, Michigan, which is about 66 miles from Ann Arbor. And the western driving boundary was Ann Arbor, Michigan. Analyzing data gathered in only one region, such as the Detroit MSA, allows us to make sound generalizations about how these devices perform in this area, but does not consider variations in the quality of the mapping and traffic applications across different geographies.

Some of these trips were considered "staging trips" because we did not expect to drive into jams during these trips. The staging trips set the driver up to enter a new destination and drive into a jam during his next trip. During the month of June, 2013 and the first two weeks of July, 2013, our drivers were organized to drive throughout the Detroit metro area in staggered shifts with the morning driver arriving at UMTRI at six-thirty A.M., picking up his log sheet where he entered his first trip, entering his first destination into the six PNDs, and driving to his first destination based on where the UMTRI researcher told him to go.

Each driver was equipped with a cell phone provided by the project to be used only for talking to or texting with the UMTRI researcher. (Drivers were not allowed to text or call while driving.) Most communications between the UMTRI researcher and the drivers was done via text messages. Most of these messages focused on providing destinations for the drivers to enter into the PNDs and apps. Using text messages made it easier for drivers to see exactly what destinations they were to enter into the units, compared to receiving verbal destinations that had to be written down in order to remember all the details of an address.

The A.M. driver's shift included the morning rush hour traffic and any jams that occurred during the morning and early afternoon. The UMTRI researcher continually checked for traffic jams throughout the day, sending the driver to areas where jams were reported. The A.M. driver ended his shift at two thirty P.M. The P.M. driver began his shift at eleven A.M., entering and driving to destinations based on where the UMTRI researcher told him to go. The P.M. driver's shift included early and late afternoon and early evening traffic. His shift ended at seven P.M.

Methodological Challenges

Our two main challenges in data collection for this study were in the areas of

1) finding jams and routing drivers to the jams
2) PNDs' sensitivity to the environment and the quality of network connectivity

Finding jams and routing drivers to jams

We define a jam as an event that

1) delays the driver 90 seconds or more while en route to his destination **and**
2) the vehicle's speed is reduced to about half of the posted speed limit.

There are certain areas and times in a metro area where one can be pretty much assured that a jam will occur. But most other areas are very unpredictable. One would think that road construction would be an easy place to find a jam, but construction teams are adept at routing traffic and working during times when traffic is light; accordingly, traffic may not back up and the speed of traffic, while slower, may not meet our requirement of about half the posted speed. Accidents and weather-related jams were also not a major contributor to our jam data. There are some accident and weather-related jams in our data, but the majority of our jams are based on heavy traffic during rush hour in the AM and PM.

Because of our distance from the major Detroit metro area (20 to 60 miles depending on what part of the metro area), we sometimes found it difficult to get drivers to the area of a reported jam in time to experience the effects of the jam. Staging trips of many miles were sometimes used to get a driver into position to experience a jam that we expected to occur based on past experience.

Our analyses divide our jams into those on highways and surface streets, which include streets that link to highways as well as city streets. These different types of roads provide good tests for the ability of the PNDs to find jams. Some surface street jams are difficult to characterize as jams because they are dependent on the timing of traffic lights as much as on the volume of traffic. One might see a quarter mile backup at a traffic light as a jam, but once the light turns green, all the backed up traffic passes through the intersection. We did not count these backups as jams.

In general, the question of whether or not a jam met our quantitative definition of delay and speed did not require detailed quantitative analysis. It was readily apparent which jams met our definition.

PND display sensitivity and the effects of environment and connectivity

We found that the mapping and traffic apps on Smartphones are not designed to manage the environmental stress of their position on the instrument panel of a vehicle. Automotive systems are designed to function in extreme temperatures, but because apps on Smartphones are portable consumer devices, they do not have system insulation and air flow in the way of automotive systems. They cannot remain on the instrument panel during the summer without air conditioning. Consequently, we found that leaving our apps on Smartphones in the sun without air conditioning or air flow, even with a shield covering the devices, caused them to not function at times until they cooled. In extreme cases, it took an hour for the devices to cool off enough to function. Our Android-based phones by Samsung were the most sensitive to the heat. The PNDs, cameras, or DVRs in our study were not as sensitive to the heat.

This is an important issue concerning the future of mapping and traffic apps on Smartphones. Because these apps are on phones, consumers will be taking them in and out of dashboard mounts. This will make the units less likely to overheat while positioned on the dashboard, but consumers will have to be sensitive to the amount of heat the units are exposed to. Dashboard mounts will probably be required for these phones when they are used as GPS devices, because some states are treating the use of phones for directions while driving as illegal, like texting.

For one of our Smartphone apps we had a problem with the phone dimming the display after driving for three to five minutes if the vehicle was driving on a straight road. (The display would brighten when the vehicle made a turn.) This occurred only on the Google app on iPhone, not on the Android-based phone. On the Android-based phone the Google app had a button to click that kept the screen from dimming. Because we wanted to record everything that occurred on the screen, we had to switch the Google Map app from the iPhone to the Android-based phone. The screen with the TomTom app on the iPhone did not dim.

Early in our data collection, we noticed that the INRIX app on the iPhone occasionally would not connect to the server on the phone system or to the network connection to the INRIX server. This situation worsened to the point where it would stop working in a middle of a route, go back to the home screen, and the driver would have to re-enter the destination. When this occurred more than once on a trip, we moved the INRIX app to an Android-based phone, and the INRIX app performance improved. We noticed and corrected these problems quickly, so the effects on the results of traffic reporting on INRIX phones were negligible.

These methodological challenges thus represent the main areas of concern for research such as this:

1) Device coverage
2) Device connectivity
3) Device operation

In terms of device coverage, though the Detroit Metropolitan Statistical Area (MSA) does not rank as one of the most congested areas in the country, it does rank 18th of 395 MSAs in population in the US according to the U.S. 2011 census estimates. [1] This places the Detroit MSA, which is our basic area of coverage in this study, in the same category of about 4 million people with other major metro areas.[2] Companies that choose not to include Detroit in their traffic algorithms are missing, population-wise, one of the major metro areas in the U.S.

Device connectivity may be affected by the cellular carrier used for GPS phone apps, though for this study, all the phone apps (Google, TomTom, and INRIX) used the same carrier. Also, for the INRIX and Google apps, we tested both Android-based and Apple-based I-Phones to find the best operating system for each particular app. For the TomTom and Garmin standalone units, these units rely on their own proprietary communication systems, so device connectivity is based on each particular unit's ability to link to its traffic data updates.

Our previous discussion of the operational challenges for each of the units tested shows our willingness, in our pre-test phase, to carefully understand how each unit operates, as well as how each unit reports traffic information to the driver. Though the devices may use different ways of alerting drivers to traffic situations, our testing of the devices before going into the field and our subsequent reviewing of videos of each traffic jam provided us with a very clear understanding of their traffic reporting mechanisms. Using videos to code the actual traffic jams encountered by our drivers allowed us to review operations by the devices in terms of their traffic jam alerts. In the early stages of our coding we reviewed how each device responded to traffic jams, and re-designed our coding process if a unit operated differently than we expected based on our pre-tests.

---

[1] U.S. Census Bureau: http://www.census.gov/popest/data/metro/totals/2011/tables/CBSA-EST2011-01.xls. Referenced on 12/30/13

[2] The MSAs with 4 million residents include Boston-Cambridge-Newton, Massachusetts/New Hampshire; Washington-Arlington-Alexandria, District of Columbia/Maryland/Virginia/West Virginia; San Francisco-Oakland-Fremont, California; Dallas-Plano-Irving, Texas; Riverside-San Bernadino-Ontario, California; Phoenix-Scottsdale-Mesa, Arizona; and Philadelphia, Pennsylvania.

Data Coding

Our gathering of data simultaneously from all six devices via our camera/video recorder system allowed us to accurately code the responses to each GPS device to real traffic conditions (as viewed from our forward facing camera).

For coding the jam for a particular trip, we used the following form, as shown in Figure 5, that lists the type of road (highway or surface street) and the number of jams on a trip and the number of hits, misses, ghosts, and inaccurate readings.

Jam Coding					
Unit	Jams	Hits	Misses	Ghosts	Inaccurates
TT PND					
Garmin HD					
Garmin SIM					
INRIX					
Google App					
TT App					

Figure 5: Jam Coding Form

The Jam Hunt Analysis

Our jam hunt analysis is based on entering the same destination on each of the PNDs and apps, and tracking their performance on trips where traffic jams are caused by general congestion, accidents, road closures, road work, or lane closures. As noted earlier, we define a jam as an event that

1) delays the driver 90 seconds or more while en route to his destination **and**
2) the vehicle's speed is reduced to about half of the posted speed limit.

Our total number of jams tested, as shown in Figure 6, for this analysis varies by device because of device failures and mistakes made by drivers in entering destinations or occasionally getting lost. This metric represents the total number of jams that each device was involved in. It does not represent the device's performance or accuracy. We use the following coding scheme to display the results of our study:

- TT PND: TomTom GO LIVE 2535M PND unit
- Garmin HD: Garmin NUVI 3490 (HD) and Garmin NUVI 3590 (HD)
- Garmin SIM: Garmin NUVI 1690 (SIM) and Garmin NUVI 1695 (SIM)
- Google App: App Version 1.1.6
- INRIX App: App Version 4.5.1
- TT App: TomTom App Version 1.14

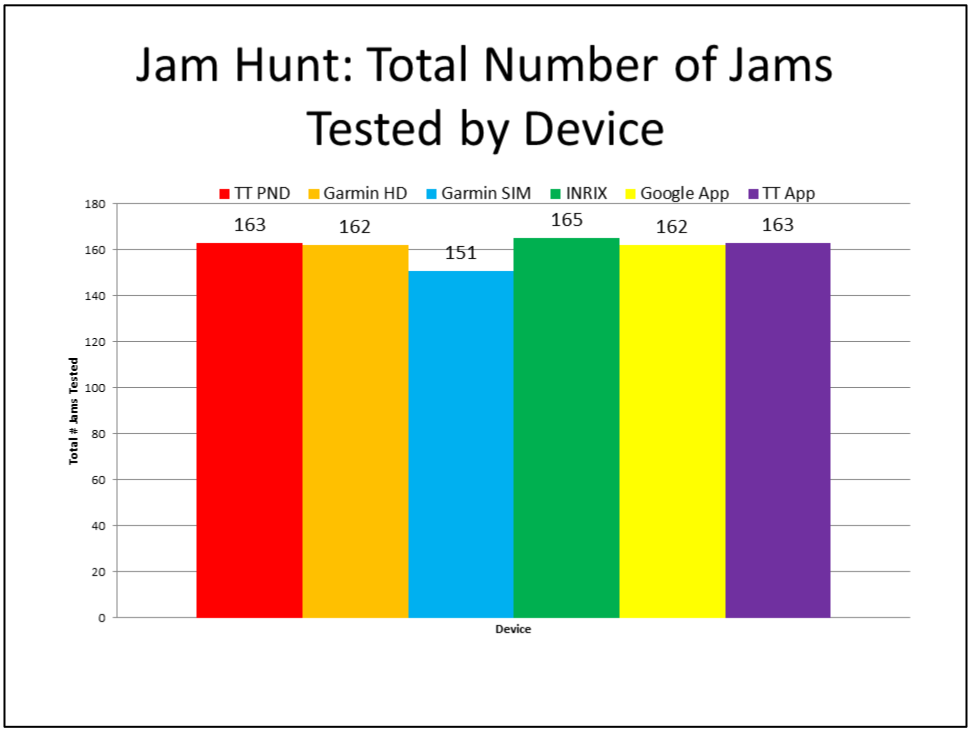

Figure 6: Total Number of Traffic Jams Tested by Each Device

We used the following rules to code the accuracy of each device in recognizing a jam:

- Hit = visually displaying a jam by warning the driver about the upcoming jam within two minutes of the jam's start and end
- Miss = not displaying a jam that exists based on our view of the road
- Ghost = displaying a jam that does not exist based on our view of the road
- Inaccurate = displaying a jam that exists but is more than two minutes off in displaying the beginning (head) or end (tail) of the jam. For example, a device does not report a jam until the driver is five minutes into the jam; it reports the jam as ended three minutes before the driver exits the jam; or it continues to report the jam after the driver has exited the jam.

For our jam analyses, the sum of all the hits, misses, and inaccurate readings should equal the total number of jams for a specific PND or app. We did not include ghosts into this calculation because they can occur multiple times during a trip.

Because our drivers had to correctly enter the same destination on each PND and app as quickly as possible, there were some errors that occurred during this process. We also created some rules for coding the data based on the non-performance of the devices:

- If a particular PND or app was not working on a trip because the driver failed to program it properly or the unit was too hot from being in a car for too long, we did not count any hits, misses or other categories for that device.
- If a particular device was not working on a trip because it failed to update its data, or it did not display sufficient data, we count this as a miss for that device if a jam occurred.
- If a device could not find the address or the destinations were not available on the device, especially after we entered a destination we visited before, we skipped programming the unit and carried on without it. If that unit happened to miss a jam because it did not accept or could not find the address we were going to, we counted this as a "miss" by the device, only if a jam occurred. This happened more often with the INRIX app and sporadically across all the other units.
- If a device shows a jam and it is under one minute in length, it will not be coded. All of the units tend to have "ghosts" that last under 1 minute. Since our threshold is 90 seconds, these short ghosts were not counted.

Ghosts were a particular issue for the Google app. When traffic occurred on the route in certain display modes, the device reported traffic, but it did not always show where the jam was located on the route because the app had a tendency to zoom in during driving. When viewing the video of the trip, we could not tell where the jam was. We could only tell from the color of the time indicator that changed from green to red that there was traffic ahead. We treated this as the app's indication of traffic, and if it did not occur we coded that part of the trip as a ghost. This led to us to count many ghosts on the Google app because of this issue.

Each of the devices has unique ways of displaying traffic.

- Garmin units that track traffic have a button on the screen that changes color from green to yellow or red. They also show yellow or red boxes on the route. We used these red and yellow boxes on the route to determine if the unit found a jam and where it thought the jam started and ended. A red button requires hitting the button to see what made it change color. A red button that stays on for more than 90 seconds was considered a ghost if there was no traffic.
- The INRIX devices sometimes warned drivers about traffic jams after they had driven past them, during them, or even warned them about traffic in the opposite lanes of traffic. These were not coded as "Misses". When INRIX devices displayed traffic information on the route, they showed a 'red' route. If it took more than 90 seconds to travel through

this area, it was considered a jam. The INRIX units also had a time display that changed color. There were many times when the time turned red for no observable reason. Sometimes the time turned red when in a jam. Random red time displays were coded as "ghosts" if there was no jam.

- When the Google app was on the iPhone, occasionally it would turn the screen dark. The drivers would have to tap the screen to brighten it again. Sometimes this would cause the unit to switch perspectives. In one perspective, Google showed the entire route as a series of colors. If the color of the route was red for a long distance, we considered this a jam. In another perspective, the route would be displayed as a shortened view (rather than the view of the entire route). In this perspective a blue line would show where to go and not show any traffic information. In this perspective, we relied on the color of the time on the screen to determine if there was traffic ahead. If the time is red, we consider this a jam. Figure 7 shows the Google app screen with this shortened perspective and the time displayed in green.

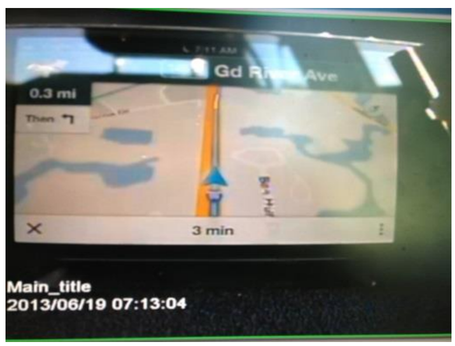

Figure 7: The Google App Screen with the Time Shown in Green

TomTom devices have the traffic display on the far right side of the screen. These traffic flags can be useful in showing how long it takes until the driver will encounter the jam.

Our analyses of hits, misses, ghosts, and inaccurates across all devices are shown in Figure 8. These numbers do not take into account the total number of jams measured for each device.

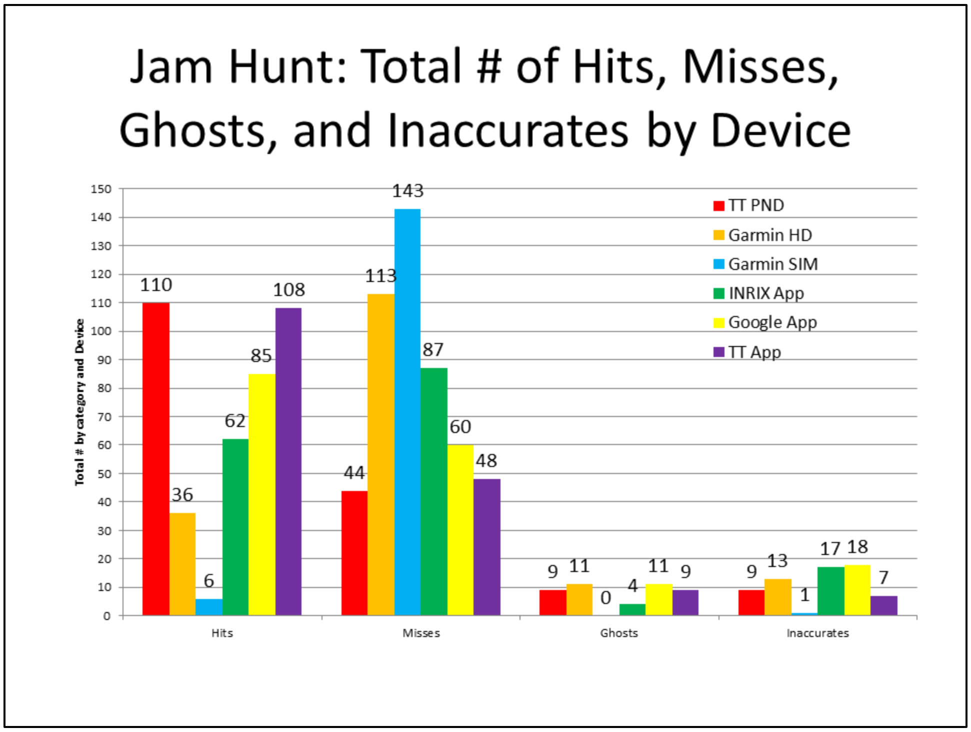

Figure 8: Jam Hunt: Total Number of Hits, Misses, Ghosts, and Inaccurates

Figure 9 adjusts for the different number of jams measured by each device: For example, the percentage of hits for a unit equals the total number of hits for that unit divided by the total number of jams recorded for that unit. The TomTom devices, both PND and App, provided more hits, by correctly identifying 67 percent and 66 percent of the traffic jams, respectively. The Google App correctly identified 52 percent of the traffic jams, and the INRIX App correctly identified 38 percent of the jams it was involved in. The Garmin HD units correctly identified 22 percent of the jams and the Garmin SIM units correctly identified 4 percent of the jams.

There are relatively few inaccurate readings with the Google App reporting the most inaccurate readings at 11 percent of the jams it encountered. The percentage of missed jams varied by device with the Garmin SIM units missing 95 percent of the jams it encountered, the Garmin HD units missing 70 percent of the jams, the INRIX App missing 53 percent of the jams, the Google App missing 37 percent of the jams, and the two TomTom devices missing 29 percent (TomTom App) and 27 percent (TomTom PND).

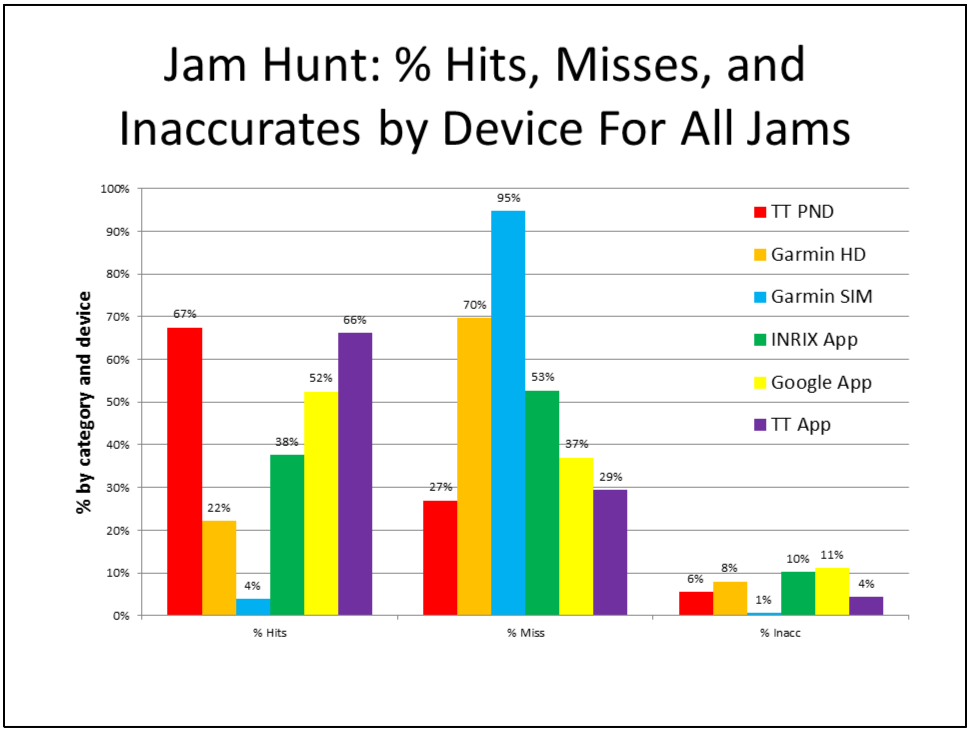

Figure 9: Jam Hunt: Percent of Hits, Misses, and Inaccurates

To determine if the differences we see among the units are statistically significant, we performed a logistic regression to determine the probability of getting a "hit" on a unit compared to the TomTom PND. We chose the TomTom PND because it had the highest hit rate in identifying traffic jams. We cannot say anything certain about the differences among the other units, such as the INRIX App compared to the Google App or the Garmin HD to the TomTom App.

Logistic regression allows us to determine if the other units are statistically different from the TomTom PND and the statistical probability that the difference does not occur by chance. In Table 1 below, we show the number of jams, hits, misses, and the percentage of hits and misses. We also show the coefficient and the probability of a difference between the TomTom PND and the other devices.[3]

The probability column displays the statistical probability that the unit is different from the TomTom PND. When we compare the TomTom PND against the Garmin HD, the Garmin SIM, and INRIX App the probability is over 99.99 percent that there is a difference. The comparison between the TomTom PND and the Google App shows that there is a 98 percent probability that there is a difference. There is no statistical difference between the TomTom PND and the TomTom App.

[3] The coefficient column tells us if a unit is more or less accurate than the TomTom PND in correctly identifying traffic jams. If the value in the coefficient column is negative, this means the device is less accurate than the TomTom PND. If it is positive, it is more accurate than the TomTom PND. In this case, the TomTom PND unit is more accurate in identifying traffic jams than all the other devices.

In determining the probability that there is a difference between the TomTom PND and the other devices, two issues are important to note. The first has to do with the sample size (in this case the number of jams). The second issue is the difference in the percentage of hits and misses between the TomTom PND and the other devices. These two issues are the key components in determining whether there is a statistically significant difference between two devices. Table 1 shows that there is a statistically significant difference between the TomTom PND and all the other devices except the TomTom App, where the percentage of hits is too similar, based on the number of jams.

UNIT	Jams	Hits	Misses	% Hits	% Miss	Coefficient	Probability
TT PND	163	110	44	67%	27%	0.00	
Garmin HD	162	36	113	22%	70%	-2.24	99+%
Garmin SIM	151	6	143	4%	95%	-4.45	99+%
INRIX App	165	62	87	38%	53%	-1.40	99+%
Google App	162	85	60	52%	37%	-0.63	98%
TT App	163	108	48	66%	29%	-0.12	ND

Table 1: Direction and Probability of a Difference Between the TomTom PND and Other Devices for All Jams

These results show the difficulty of providing accurate information about dynamic events such as traffic jams. The TomTom PND and App, which had the best performance in this particular test, correctly identified traffic jams at only modest levels of accuracy (67 percent and 66 percent of the time, respectively).

It is important for the units to accurately report jams because the underlying processes in most units are designed to route the driver away from jams to faster routes to a destination. Without the ability to receive accurate information about a jam, drivers will drive into a jam, causing unexpected delays. Consequently drivers will lose confidence in the device and not rely on it to help them navigate to their destinations. Consumers understand the challenge in providing this type of information, but they also are likely to be quick to discard products that do not meet their needs or do not provide the information they expected.

Surface Street Jams

We also tested to see if there are differences between the different systems in relation to the type of roads where the jams occurred. For these analyses, we look at the difference between jams that occurred on surface streets and those that occurred on highways. Figure 10 shows the percentage of hits, misses, and inaccuractes for all surface street jams. The TomTom PND, the Google App, and the TomTom App were the most accurate in reporting traffic jams on surface streets with 43 percent, 48 percent, and 38 percent correct reports (hits), respectively. The other units reported less than 15 percent of the surface street jams.

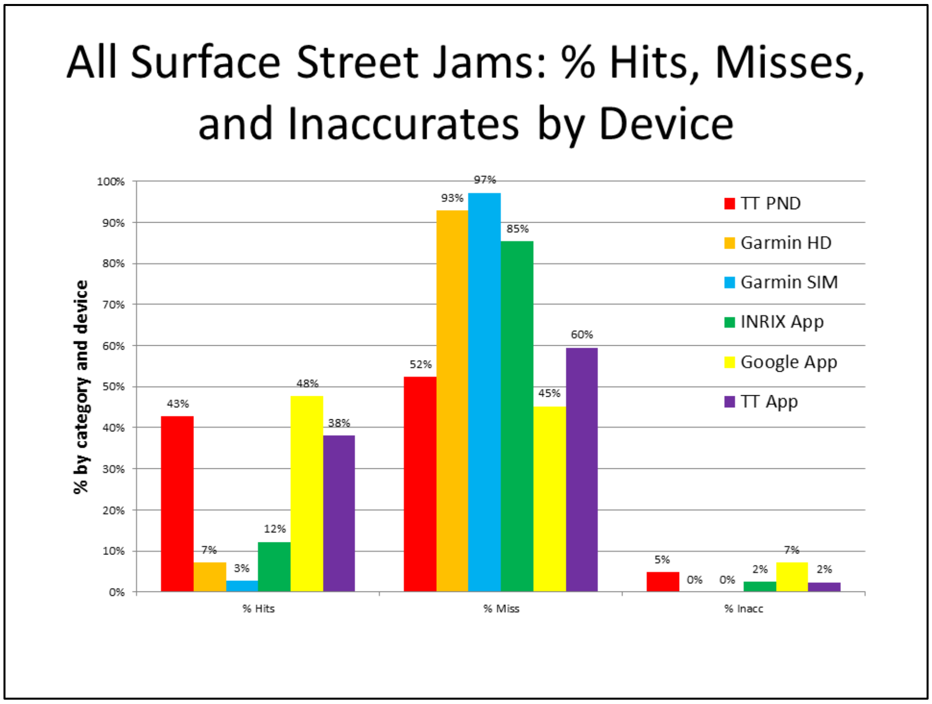

Figure 10: Percent of Hits, Misses, and Inaccurates for Surface Street Jams

We also looked at the surface street jams that were less than or equal to 5 minutes and those that were greater than 5 minutes in length. Figure 11 displays, similar to the overall surface street jam data, the TomTom PND, the Google App, and the TomTom App were the most accurate in reporting these traffic jams with 43 percent, 39 percent, and 35 percent correct reports (hits), respectively. The other units reported less than 10 percent of the surface street jams less than or equal to 5 minutes. Please note that for short jams, it could be that some services filter these out in what they provide to drivers, for example they leave out queues at traffic lights.

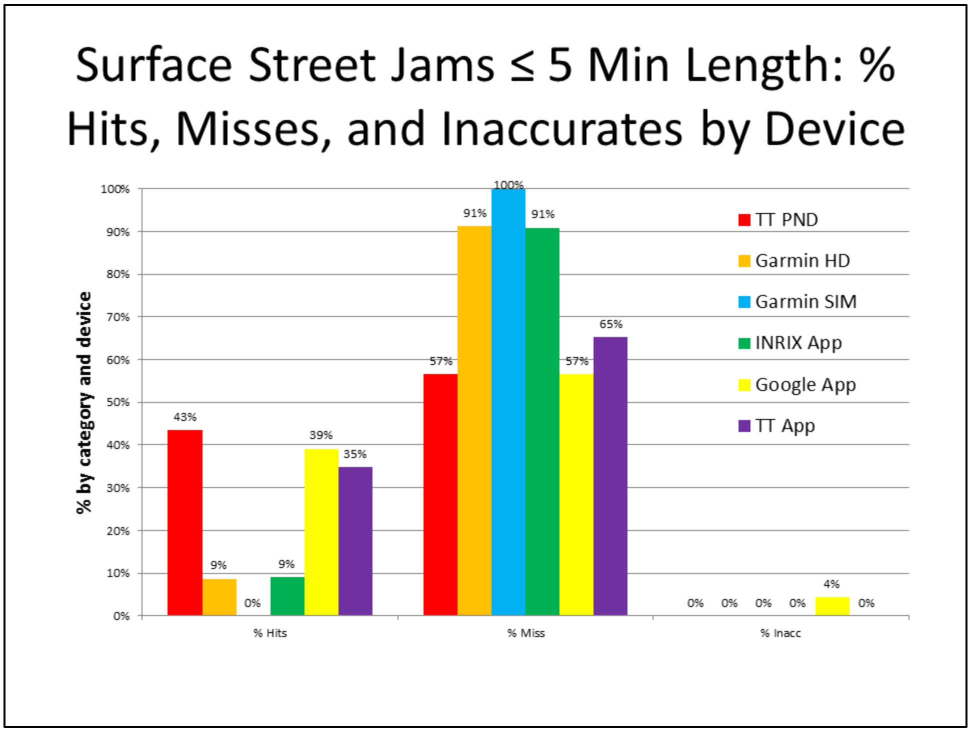

Figure 11: Percent of Hits, Misses, and Inaccurates for Surface Street Jams Less than or Equal to 5 Minutes in Length

Surface street jams that were greater than 5 minutes in length showed a slightly different pattern, as shown in Figure 12. The Google App, the TomTom App, and the TomTom PND were the most accurate in reporting traffic jams on surface streets with 42 percent, 58 percent, and 42 percent correct reports (hits), respectively. The other units reported less than 20 percent of the surface street jams greater than 5 minutes.

These analyses show that the traffic systems all need to improve how they identify surface street jams. They tend to identify the longer surface street jams better than the shorter jams, and the TomTom PND, App, and the Google App provide the best jam information.

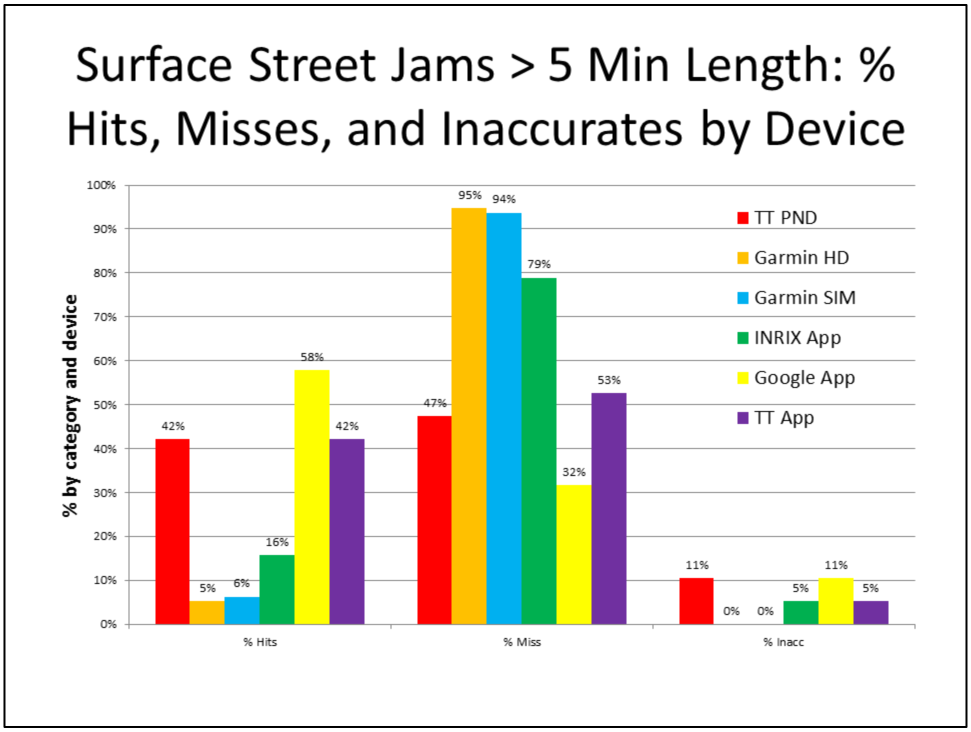

Figure 12: Percent of Hits, Misses, and Inaccurates for Surface Street Jams Greater than 5 Minutes in Length

Highway Jams

Finally, we examine the highway traffic jam to see how well the devices report these jams. We begin by looking at all highway traffic jams. Compared to surface street jams, all the units report highway jams more accurately than surface street jams. As Figure 13 shows, the TomTom PND and App accurately report almost 76 percent of the highway jams correctly. The Google App correctly reports 54 percent of the highway jams, while the INRIX App accurately reports 46 percent, the Garmin HD reports 28 percent, and the Garmin SIM reports 4 percent.

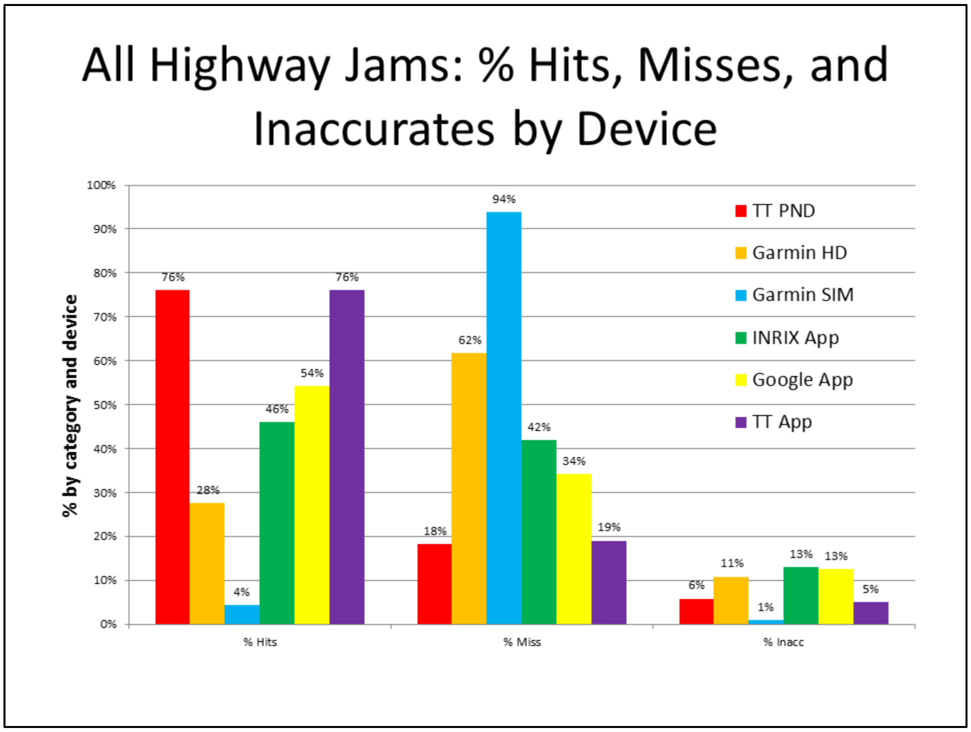

Figure 13: Percent of Hits, Misses, and Inaccurates for Highway Jams

Table 2 shows the results of our logistic regression that determines the probability of getting a "hit" on a unit compared to the TomTom PND for highway jams. From these results we see that the TomTom PND unit differs significantly from the Garmin HD, the Garmin SIM, the INRIX App, and the Google App, but there are no significant differences between the TomTom PND and the TomTom App.

UNIT	Jams	Hits	Misses	% Hits	% Miss	Coefficient	Probability
TT PND	121	92	22	76%	18%	0.00	
Garmin HD	120	33	74	28%	62%	-2.24	99+%
Garmin SIM	115	5	108	4%	94%	-4.50	99+%
INRIX App	124	57	52	46%	42%	-1.34	99+%
Google App	120	65	41	54%	34%	-0.97	99%
TT App	121	92	23	76%	19%	-0.04	ND

Table 2: Direction and Probability of a Difference Between the TomTom PND and Other Devices for Highway Jams

Because highway jams tend to last longer than surface street jams, we looked at highway jams less than or equal to 10 minutes in length and those longer than 10 minutes in length. In Figure 14, for highway jams less than or equal to 10 minutes in length, the TomTom App reports 74 percent of the jams accurately, the TomTom PND reports 73 percent of its jams accurately, the

Google App reports 49 percent, the INRIX App reports 44 percent, the Garmin HD reports 30 percent, and the Garmin SIM reports 5 percent accurately.

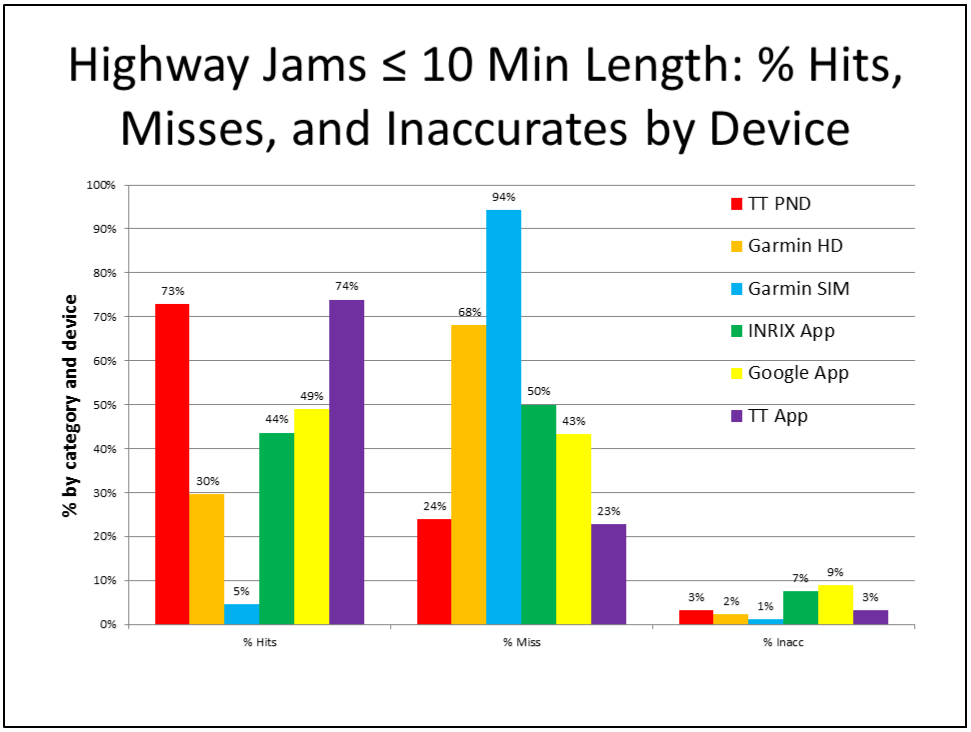

Figure 14: Percent of Hits, Misses, and Inaccurates for Highway Jams that are Less Than or Equal to 10 Minutes in Length

The results of our logistic regression that determines the probability of getting a "hit" on a unit compared to the TomTom PND for highway jams that are less than or equal to ten minutes are shown in Table 3. From these results we see that the TomTom PND unit differs significantly from the Garmin HD , the Garmin SIM, the INRIX App, and the Google App, but there are no significant differences between the TomTom PND and the TomTom App.

UNIT	Jams	Hits	Misses	% Hits	% Miss	Coefficient	Probability
TT PND	92	67	22	73%	24%	0.00	
Garmin HD	91	27	62	30%	68%	-1.94	99+%
Garmin SIM	88	4	83	5%	94%	-4.15	99+%
INRIX App	94	41	47	44%	50%	-1.25	99+%
Google App	90	44	39	49%	43%	-0.99	99%
TT App	92	68	21	74%	23%	0.06	ND

Table 3: Direction and Probability of a Difference Between the TomTom PND and Other Devices for Highway Jams Less Than or Equal to Ten Minutes

Our final jam analysis compares the six devices when reporting highway jams greater than 10 minutes in length. Figure 15 shows the highest levels of accuracy for all of our jam analyses for the TomTom PND (86 percent), the TomTom App (83 percent), the Google App (70 percent), and the INRIX App (53 percent). Both the Garmin PNDs decreased their accuracy: the Garmin HD (21 percent), and the Garmin SIM (4 percent).

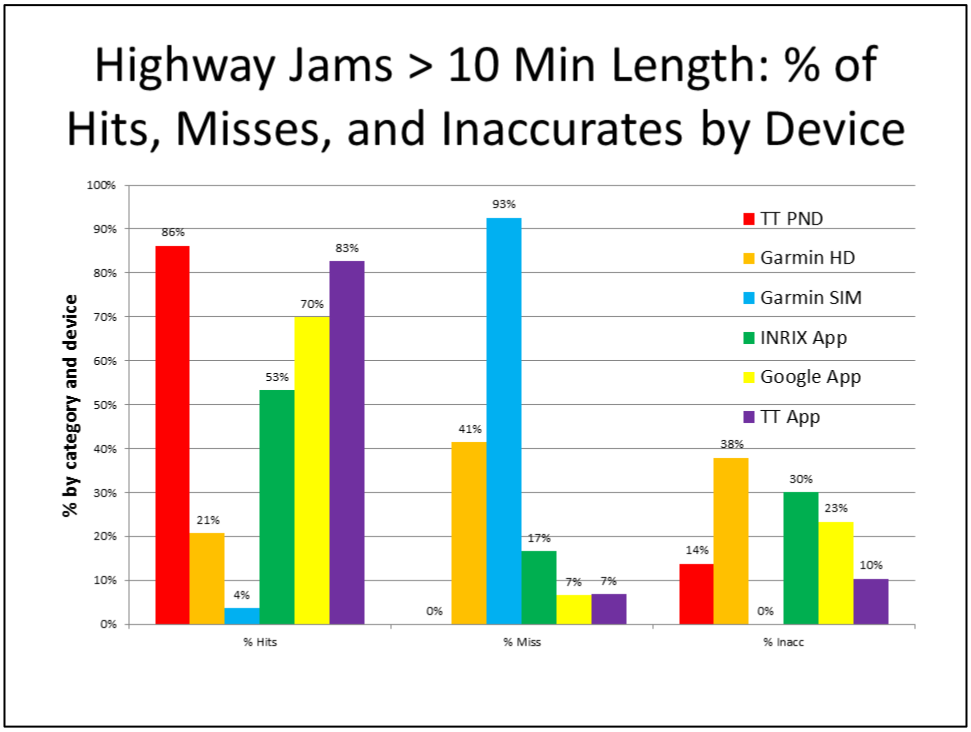

Figure 15: Percent of Hits, Misses, and Inaccurates for Highway Jams that are Greater Than 10 Minutes in Length

Table 4 shows the results of our logistic regression that determines the probability of getting a "hit" on a unit compared to the TomTom PND for highway jams greater than 10 minutes in length. From these results we see that the TomTom PND unit differs significantly from the Garmin HD , the Garmin SIM, and the INRIX App, but there are no significant differences between the TomTom PND and the Google and TomTom Apps.

UNIT	Jams	Hits	Misses	% Hits	% Miss	Coefficient	Probability
TT PND	29	25	0	86%	0%	0.00	
Garmin HD	29	6	12	21%	41%	-2.24	99+%
Garmin SIM	27	1	25	4%	93%	-4.50	99+%
INRIX App	30	16	5	53%	17%	-1.34	92%
Google App	30	21	2	70%	7%	-0.97	ND
TT App	29	24	2	83%	7%	-0.04	ND

Table 4: Direction and Probability of a Difference Between the TomTom PND and Other Devices for Highway Jams Greater Than Ten Minutes in Length

Discussion

By drilling down into the different types of jams we find some interesting patterns for the current traffic PNDs and apps in our study:

- In terms of accurately reporting traffic jams, only the TomTom PND (67 percent), TomTom App (66 percent), and the Google App (52 percent) are better than a coin flip when it comes to knowing if a traffic jam is occurring on one's route.

- All the systems are significantly better at reporting traffic jams on highways than surface streets with the TomTom devices correctly reporting almost 76 percent of the highway jams and the Google App correctly reporting 54 percent of the highway jams.)

- The longer the jam, the better chance the system will accurately report it.

- The systems tend to do a better job of reporting surface street jams that are greater than 5 minutes in length compared to those less than or equal to 5 minutes in length.

- The systems also do an even better job of reporting highway jams greater than 10 minutes in length. (TomTom PND: 86 percent accurate, TomTom App: 83 percent accurate, Google App: 70 percent accurate, and the INRIX App: 53 percent accurate.)

In this study, we devised a field test method to relate the traffic jams reported by various commercial GPS traffic devices to the traffic conditions recorded using independent means in our test vehicles. It was then possible to compare the "hit" and "miss" rates of the various commercial devices. The hit rates we determined were modest at best, and we also found significant differences among the hit rates of individual devices. We therefore need to consider the major challenges of performing research such as this.

When it comes to comparing the performance of individual commercial systems, what are we actually trying to compare? In this case, we had a very narrow focus: we measured whether a device provided adequate warning about traffic before the driver entered a traffic jam. We did not measure the effects of alternate routes systems provide when sensing traffic in terms of time saved, mapping quality, or display quality/effectiveness. Having said that, there are potential differences in the geographic coverage of particular systems, the design of individual systems, the methods of communication, and the definitions of traffic jams, that can affect results of a study such as this.

This class of devices depends on a proprietary means of combining traffic data from various sources and across various time scales, as well as the instantaneous quality of network connections. Both the traffic data streams used in companies' algorithms and the network connections may vary across geographical areas across the country. As noted earlier in the report, this analysis allows us to make sound generalizations about how these devices perform in this area, but does not consider variations in the quality of the mapping and traffic applications across different geographies.

We have no way of knowing how well a device "hit rate" measured in the Detroit Metropolitan Statistical Area (MSA) would represent other major metropolitan areas. Other traffic information markets, which experience much higher congestion levels than the Detroit MSA, may be the prime targets of these GPS traffic devices. But the Detroit MSA is the 18^{th} largest MSA by population in the U.S out of 359 other MSAs. So companies deciding not to provide traffic information for the Detroit MSA are making business decisions about which areas they cover because they cannot cover all the major MSAs in the U.S. Such decisions would impact the performance of their devices in this region.

The overall relevance of our method also depends on the "ground truth" we were able to establish in terms of traffic jams occurring in our selected region. Our definition of "traffic jam", while quantitative, is not universal. Rigorous qualification of individual jams against our criteria is both data-intensive and labor-intensive; on the other hand, we do not feel that many of the jams would have been "close call" decisions against our criteria. We also had to search for and locate traffic jams, based on available information and previous experience. It is possible – though very unlikely - that there was some kind of systematic bias in the traffic jams we found (and those we missed). If we had adopted a "heavier" criterion for traffic jam identification (particularly longer duration) we would have limited our ability to find sufficient traffic jams in a reasonable time frame in the Detroit MSA.

Because vendors may all be using similar data streams, the proprietary means by which algorithms combine the data and provide drivers information may well be regarded as the true essence of an individual product. Because this is commercially-sensitive information, how does the researcher know that all devices are being exercised in a representative manner? In our case, we thoroughly tested the operation and performance of each device before going into the field, and each device was monitored to optimize its performance during the testing phase. One might say that we represented how a consumer would interact with a new device: we learned how it worked and used it in real driving situations to see how it responded.

For devices which rely on cellular communication, how does our choice of carrier – interacting with our choice of geographic area – affect the performance of each device. "Misses" may be caused by lack of communication rather than an inferior algorithm. Our devices were divided into two classes: personal navigation devices (PNDs) and phone applications (apps). PNDs use their own proprietary communication systems, while all the apps used the same cellular provider. PND performance (TomTom and Garmin) was based on the strength of its communication system in the Detroit MSA, while the apps (Google, TomTom, and INRIX) were tested, monitored, and installed on the phone where each worked best. Because all the apps used the same cellular provider and were matched with the phone where it worked best, we feel confident that we provided each app the best opportunity to perform.

Finally, each device tends to have a unique way of displaying traffic jams. It is therefore difficult to know how well our definition of traffic jam correlated with the approaches taken by individual vendors. As noted earlier, our definition of a traffic jam defined our approach to whether a certain amount of traffic was considered a jam. One might argue that being delayed 90 seconds and driving less than half the speed limit is too restrictive, and that a 120 second or 180 second delay would have yielded different results. Changing the delay time and speed would affect the number of cases in our analysis. It might also change our coding of how soon a device warned a driver of a jam. In this case more devices could generate more "hits" because they would have more time to identify the traffic jam.

But a more restrictive delay time truly tests the capability of a device by testing how well it updates a dynamic event such as traffic. Lengthening the delay time makes it easier for devices that do not update their traffic information as often or as well to perform as well as devices that update more often and more accurately. A 90 second delay time, while more restrictive, provides drivers with timelier information about their driving situation. It also provides a good test of the capability of these systems to respond to traffic. Taking all of these considerations into account, one can see the challenges of performing this type of research on the current state of development of real-time traffic devices and in trying to compare device performance.

Conclusions

In terms of alerting users to the presence of traffic jams in the Detroit MSA, none of the devices showed better than moderate performance. All of the devices tested performed better in reporting traffic jams that were of longer duration and that occurred on highways rather than surface streets.

The TomTom PND (67 percent), TomTom App (66 percent), and the Google App (52 percent) accurately reported more than half of the traffic jams in our study, while the INRIX units accurately reported 38 percent, the Garmin HD units accurately reported 22 percent, and the Garmin SIM units accurately reported 4 percent of the traffic jams.

All the systems were significantly better at reporting traffic jams on highways than surface streets with both TomTom devices correctly reporting 76 percent of the highway jams, the Google App correctly reporting 54 percent of the highway jams, the INRIX app correctly reporting 46 percent, the Garmin HD device correctly reporting 28 percent, and the Garmin SIM device correctly reporting 4 percent of the highway jams. The devices correctly reported surface street traffic jams less than half the time with the Google App reporting 48 percent accurately, the TomTom PND reporting 43 percent accurately, the TomTom App reporting 38 percent accurately, the INRIX app reporting 12 percent accurately, the Garmin HD unit reporting 7 percent accurately, and the Garmin SIM device reporting 3 percent of the surface street jams accurately.

We found that the longer the jam, the better the chance the system will accurately report it. The systems tend to do a better job of reporting surface street jams that are greater than 5 minutes in length (Google App: 58 percent accurate reporting of jams, both TomTom PND and App: 42 percent, INRIX app: 16 percent, Garmin SIM unit: 6 percent, and the Garmin HD unit: 5 percent). The devices are less likely to accurately report surface street traffic jams less than or equal to 5 minutes in length (TomTom PND: 43 percent accurate reporting of jams, Google App: 39 percent, TomTom App: 35 percent. INRIX App and Garmin HD PND: 9 percent, and Garmin SIM: 0 percent).

The systems do a better job of reporting highway jams greater than 10 minutes in length (TomTom PND: 86 percent accurate, TomTom App: 83 percent accurate, Google App: 70 percent accurate, the INRIX App: 53 percent accurate, and the Garmin HD unit: 21 percent, and the Garmin SIM unit: 4 percent). The units are less likely to accurately report highway jams less than or equal to 10 minutes in length (TomTom App: 74 percent accurate, TomTom PND: 73 percent accurate, Google App: 49 percent accurate, INRIX App: 44 percent accurate, Garmin HD unit: 30 percent accurate, and the Garmin SIM unit: 5 percent accurate).

Finally, three major challenges may have affected the overall performance of the traffic devices: the choice of geographic area where some devices may have chosen not to provide traffic coverage, device operation, and our chosen definition of a "traffic jam".

Because of how different traffic systems choose to cover areas of the U.S., our results provide a good analysis of how these systems function only in the Detroit Metropolitan Statistical Area (MSA), the 18th largest MSA, by population, in the U.S. Because the Detroit MSA is not noted nationally for its levels of traffic congestion and delay, we believe it is important to extend such research into the major national traffic data markets.

All the systems in the study experienced some type of operational failure or disruption during our study. We mitigated these disruptions by thoroughly pre-testing the devices and continually monitoring and optimizing each system's performance throughout the two month testing period. Also, system connectivity issues occurred occasionally throughout the study for all the devices. Because Detroit is a major metropolitan area that provides good cellular service and because all the apps used the same cell service and the PNDs used their own proprietary connectivity systems, we feel that our monitoring and optimizing of the systems offered each device the best opportunity possible to function properly.

Utilizing a 90 second destination delay and half the posted speed criteria for determining a traffic jam provided a good sample of jams to analyze, but a longer destination delay time would have reduced the number of jams to analyze. A longer destination delay time may have generated a higher "hit" rate for all the devices by allowing systems that are slower in updating their traffic feeds to "catch up" with the systems that are updating their traffic feeds faster. A 90 second delay, while more restrictive, provides drivers with more timely information about their traffic situation.

We live in an age of innovation based on sensors, connectivity, and the manipulation of big data. Useful information is being created at a relatively low cost, where no such information existed before. Real-time information provided to drivers is a significant innovation, but the information must be timely and accurate in order to be trusted and valued. Despite the challenges of this experiment, it is important that the industry has access to independent testing and analysis. Such analysis is challenged by the rapid pace of development, vendor reliance on multiple, evolving data streams, and a lack of independently-acknowledged architectures and test methods.

We are encouraged by the ability of these companies to manage and manipulate the vast amounts of data that is needed to provide real-time traffic data, and we believe that more broadly-based field operational tests would help contribute to products that will truly keep us from being "stuck in traffic."

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UMTRI-2014-4		

4. Title and Subtitle	5. Report Date
Stuck in Traffic: Analyzing Real Time Traffic Capabilities of Personal Navigation Devices and Traffic Phone Applications	Revised 1/16/14
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Bruce M. Belzowski and Andrew Ekstrom	UMTRI-2014-4
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
UM Transportation Research Institute 2901 Baxter Road Ann Arbor, MI 48109	11. Contract or Grant No.
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
TomTom Group 26261 Evergreen Rd. Ste. 425 Southfield, MI 48076	Field Operational Test: May, 2013 to July, 2013
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The global positioning system (GPS) market is a fast changing, highly competitive market. Products change frequently as they try to provide the best customer experience for a service that is based on the need for real-time data. Two major functions of the GPS unit are to correctly report traffic jams on a driver's route and provide an accurate and timely estimated time of arrival (ETA) for the driver whether he/she is in a traffic jam or just following driving directions from a GPS unit. This study measures the accuracy of traffic jam reporting by having Personal Navigational Devices (PNDs) from TomTom and Garmin and phone apps from TomTom, INRIX, and Google in the same vehicle programmed to arrive at the same destination. We found significant differences among the units in terms of their ability to recognize an upcoming traffic jam. We also found differences in how well the devices responded to jams when driving on surface streets versus highways, and whether the jams were shorter or longer in length. We see potential for auto manufacturers to employ real-time traffic in their new vehicles, providing potential growth for real-time traffic providers through access to new vehicles as well as the aftermarket.

17. Key Word	18. Distribution Statement
global positioning system, GPS, PND, traffic app, jam hunt	No restrictions.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
		33	