Using NPMRDS: Lessons Learned

Kartik Kaushik

Center for Advanced Transportation Technologies

University of Maryland

Lesson 1: Accessing

Downloading

- Need to sign data use agreement with HERE Inc. to get access
  - Coordinated through email
- Total archive size is roughly 5 to 7 GB per month (packed)
  - For the whole country, all files
- States are grouped in zones, with an archive for each zone
  - 700 MB per zone on average (packed)
- Downloading using browser is not recommended (even for zonal archives)
- Use of FTP is recommended
  - Winscp is a good FTP client for Windows
  - Can be scripted to automatically check and download updates each month

Files Overview

- Three basic types of files
- **Data files** only contain travel times for each record
  - Join with TMC definition file is necessary to obtain speeds
  - File changes each month, to reflect the month's observations
  - Data file size depends on the size of state – few 100 MB to 1-3 GB per month (unpacked)
- **TMC definition file** contains information to identify a segment
  - Contains the lengths of each TMC segment
  - Changes only when TMC table gets updated (updates released quarterly)
  - Packaged with each month's data – about 4 MB in size (unpacked)
- **Shapefile** only contains sub-TMC segments
  - Join with lookup table, TMC definition file, and Data files is required to visualize data on map
  - Has update/release cycle of 6 months
  - About 2 GB in size (unpacked)

Opening and Viewing Files

- Microsoft cannot help
  - Cannot use Excel, or Access because the files are too large
- Need core database programs or scripting programs adept to large text files
  - MySQL is good for beginning – storing, viewing, querying
  - PostgreSQL is more advanced – can perform lot more analysis, and visualizations
  - SAS, SPSS, etc. are also good commercial alternatives
  - Programming languages like Python, R, Matlab, etc. can also be used, especially if data needs to be analyzed
- To open and view the included Shapefile, ArcGIS is required
  - Can also use PostGIS (extends PostgreSQL), but is a little complicated

Analysis and Visualization

- MySQL for basic analysis – descriptive statistics and the like
- PostgreSQL and ArcGIS for advanced analysis
  - PostgreSQL has extensions to allow programming languages to connect
  - ArcGIS has embedded Python and Visual Basic, in addition to model builder
- For detailed analysis Python with Pandas, R or Matlab is recommended
  - Analysis includes computing performance measures: PTI, TTI, BTI, etc.
- For graphical visualizations Python, R or Matlab are recommended
  - One can also import the results to Excel for graphical visualizations
- Easiest to use ArcGIS to visualize on a map
  - If using PostgreSQL, download supplemental GIS package: PostGIS

Lesson 2: Data

Traffic Message Channel (TMC) codes

- TMC code references to a particular segment on the road

- However, TMC information contained within the static file is limited
  - Local name and one random coordinate point is given

- Probably use the provided shapefile to extend the geo information within the static files
  - Can get accurate start and end coordinates from the shapefile (but not easy)
  - Ordering obtained coordinate points by road direction can be a nightmare

- Also, caution is advised when comparing TMC definitions from multiple different vendors

Example of limited info about TMC (110N06799)

TMC information in Static File
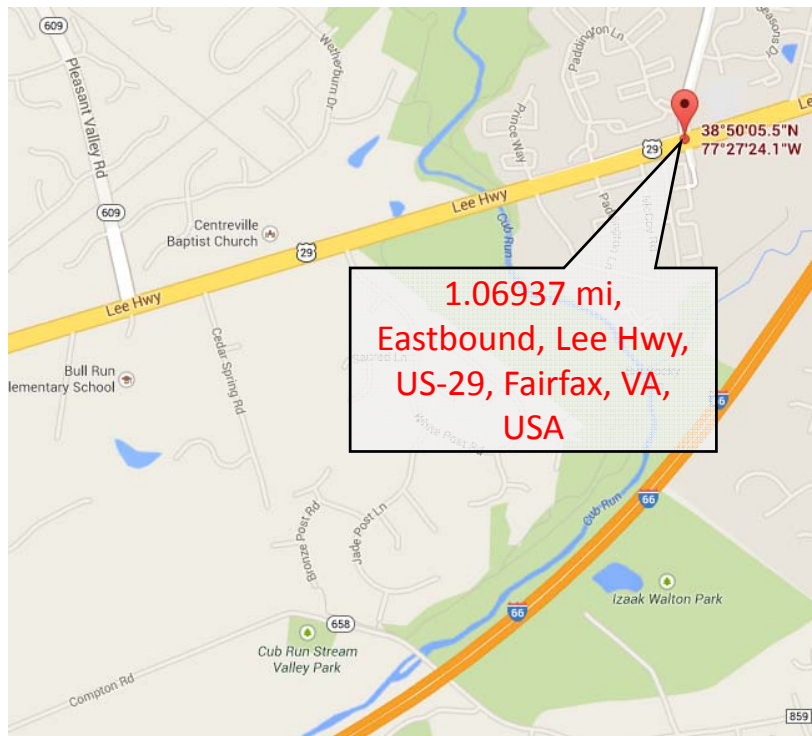
TMC information in Shapefile

38°50'05.5"N
77°27'24.1"W

1.06937 mi,
Eastbound, Lee Hwy,
US-29, Fairfax, VA,
USA

Lee Hwy, From
Reference Node, Not a
Ramp, No frontage to
property, No contra-flow
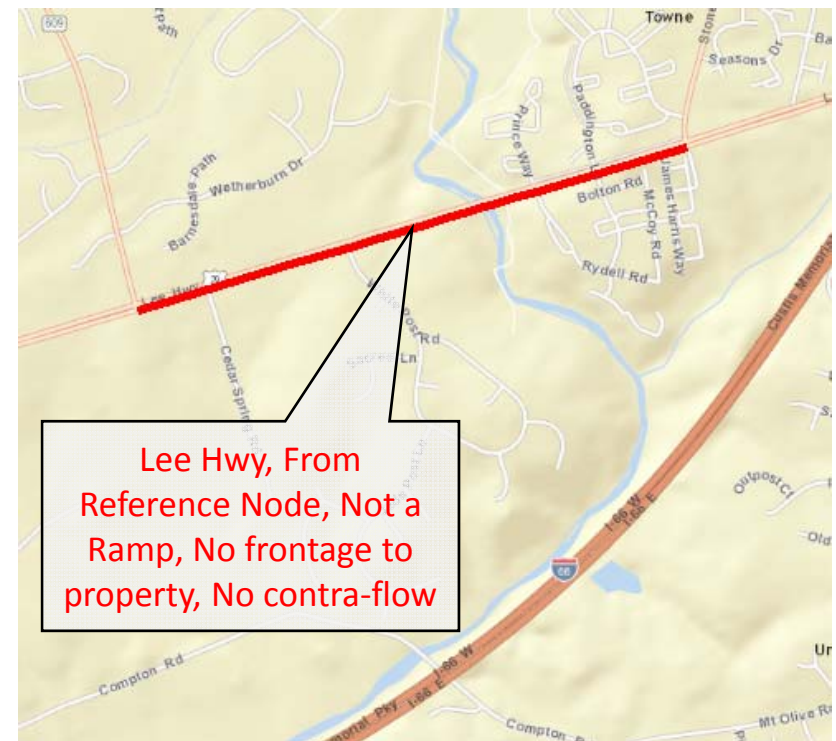
- Joining the static file to the shapefile provides a full picture
  - Joining may be problematic due to many-to-many relationship between the shapefile and TMC static file
- Lat/Long point in the static file may be begin, end or somewhere in between on a segment

Travel Time Data

- Direct calculation of traditional performance measures (TTI, PTI, and BTI) is problematic
- Data is missing for some reporting intervals
  - If no base data, no traffic record is provided
  - Creates sparse datasets
  - Complicates outlier filtering and smoothing
- Data is unfiltered, and noisy
  - Causes problems when trying to impute missing data
- Other methods of processing data are recommended
  - Methods that do not require imputation of data works best
  - Methods that can do without outlier filtering enable small and easy algorithms

Lesson 3: Analysis

Overlaying Data

- Originally developed at CATT to deal with issues presented by interrupted flow facilities (arterials)
- Involves repeatability of traffic conditions and characteristics over days or months
- Overlay data from those days where repeated traffic is expected
  - Weekday data from a couple of weeks (used in this presentation)
  - Only Monday, Tuesday... data for a month or two
  - Any other similar logical grouping
- Overlay size can be adjustable, but 24 hour periods is recommended

Example of Sparse Data:
NPMRDS data for one day, one random TMC

NPMRDS 24-hour Single Day Scatter Plot
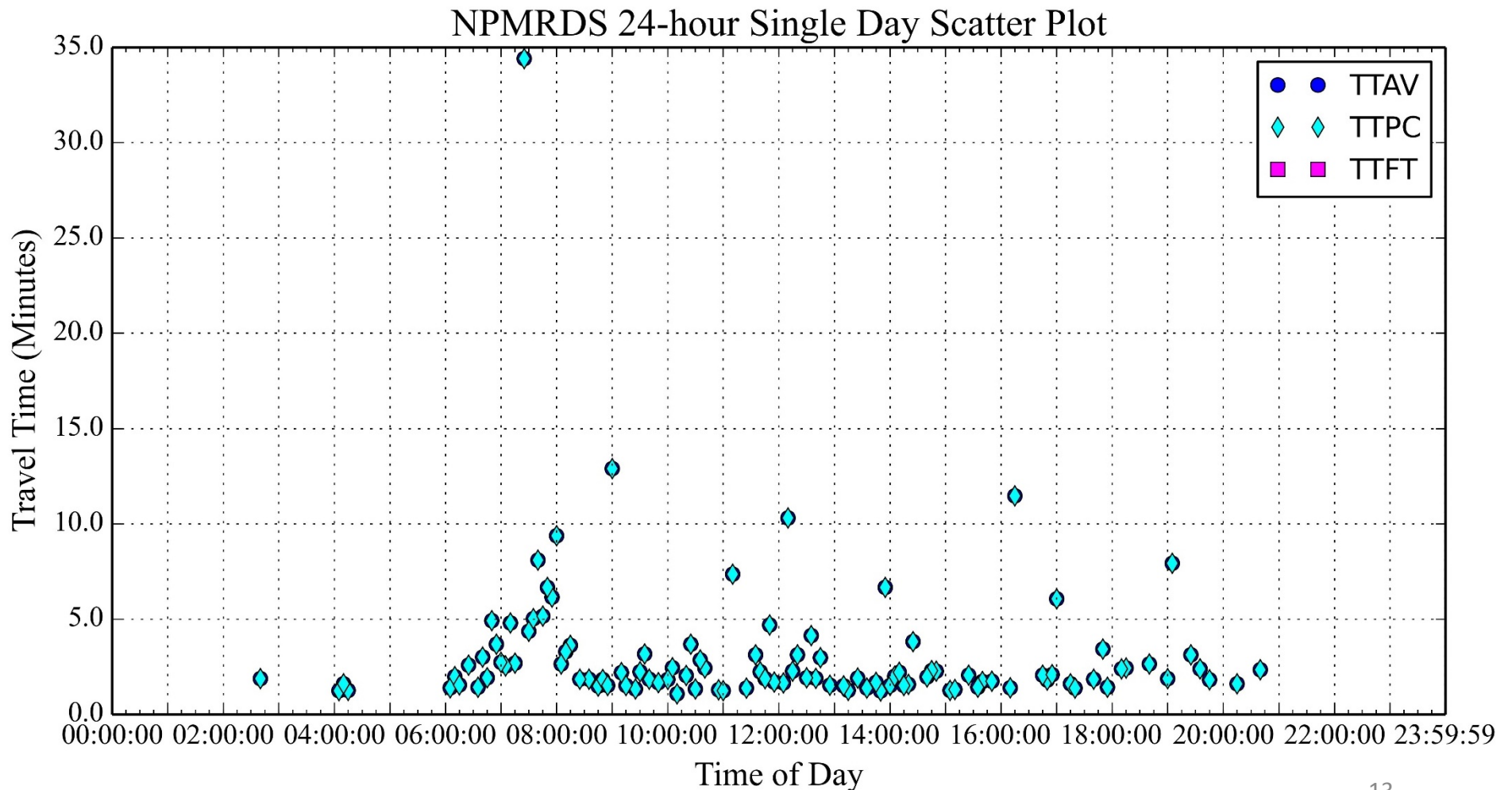
Result of Overlaying Data:
Same TMC, 2 weeks' worth of weekday data

NPMRDS 24-hour Weekday Overlay Scatter Plot
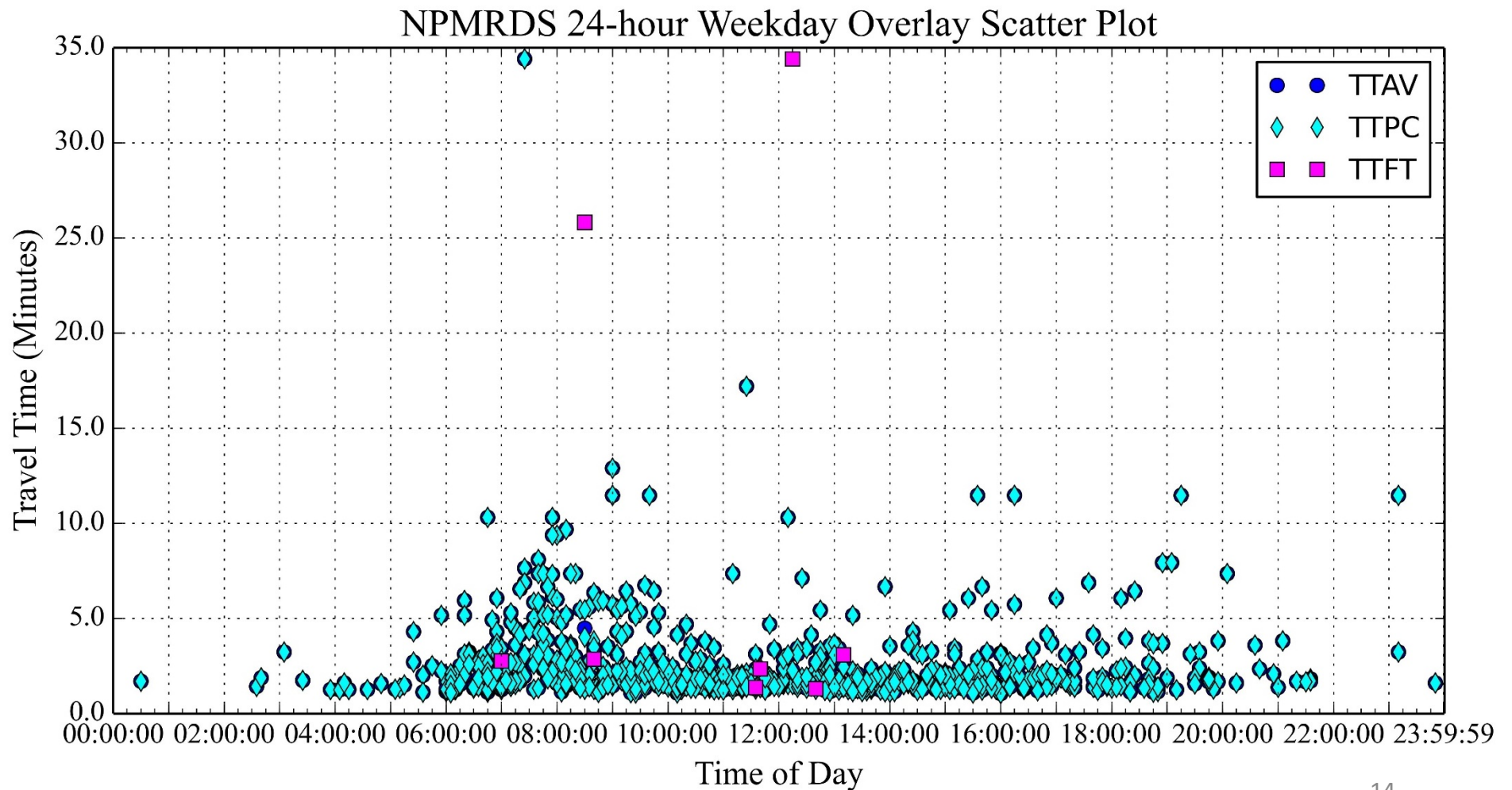
Computing Performance

- Performance is computed using 24 hour overlay scatter plots
- Slices of data from the overlay plots is taken
  - Used here is 1 hour windows, also recommended (unless overlaid data is also sparse)
- Percentiles of travel time within the slice is obtained
  - 5% increments are used here, going from 5% to 95%, (recommended)
  - Percentiles ensure that one or two outliers have minimal impact on results (because of the nature of percentile computation)
  - Percentiles are then graphed to create Cumulative Distribution Functions (CDF)
- Performance measures can be computed from CDFs

Lesson 4: Case Studies

Validation with VPP and BTM data

- Results obtained from NPMRDS was compared against Vehicle Probe Project (VPP) and Bluetooth Traffic Monitoring (BTM) data

- BTM is obtained using portable Bluetooth sensors developed by UMD
  - BTM is raw observations of actual vehicle traversals
  - Like NPMRDS, only observed vehicles are reported

- VPP is commercial probe data, available through contracts awarded by I-95 Corridor Coalition
  - VPP is probe data similar to NPMRDS, but is a real time feed at 1 minute intervals
  - VPP is smoothed, filtered and imputed

- Two case studies are presented here: a freeway-ish corridor, and an arterial with heavily interrupted flow

VA-7 in Virginia: Freeway-ish

## 4 Lanes, 1 signal, 0.88 miles

24 Hour Scatter Plots

BTM

VPP
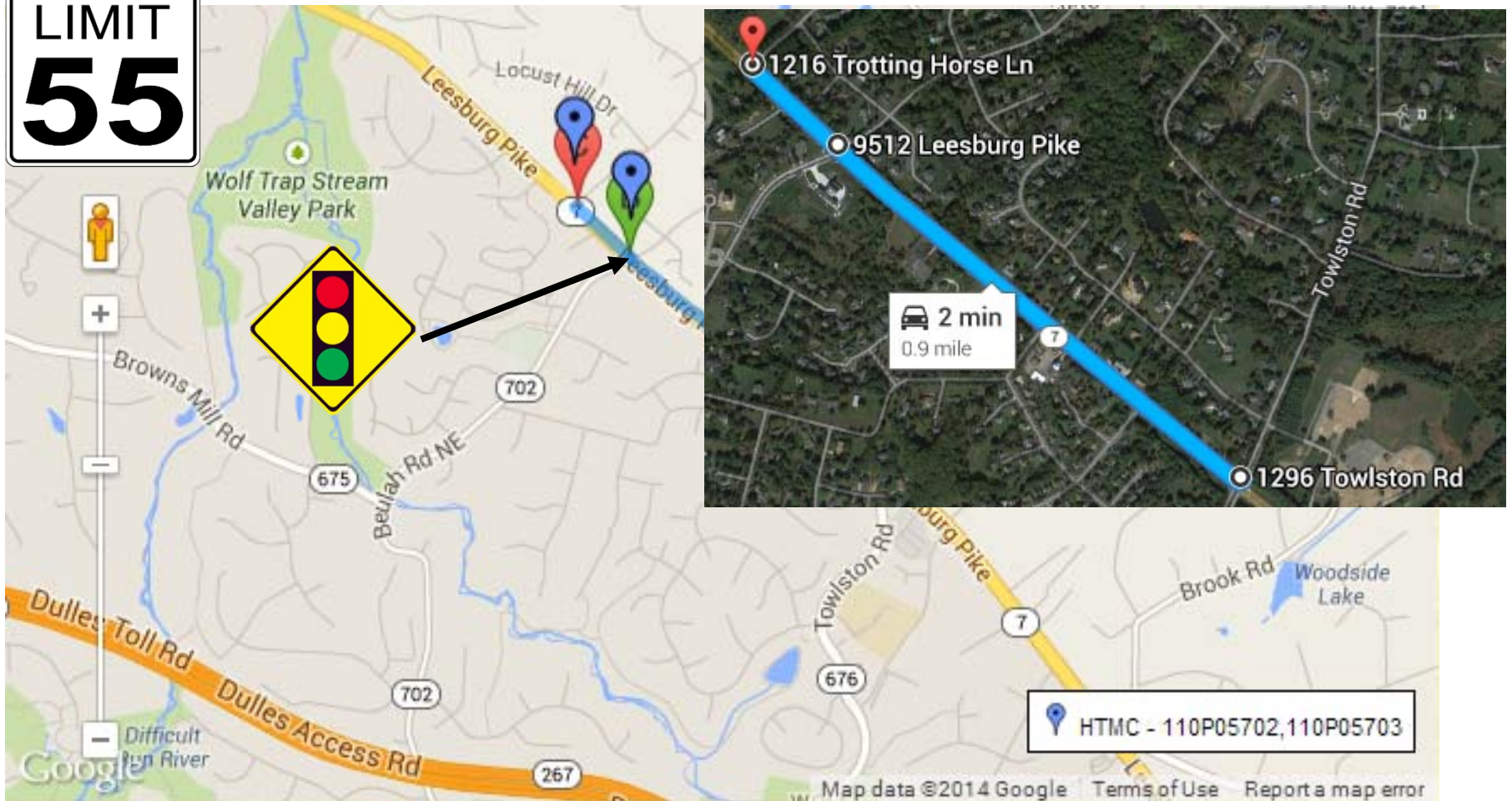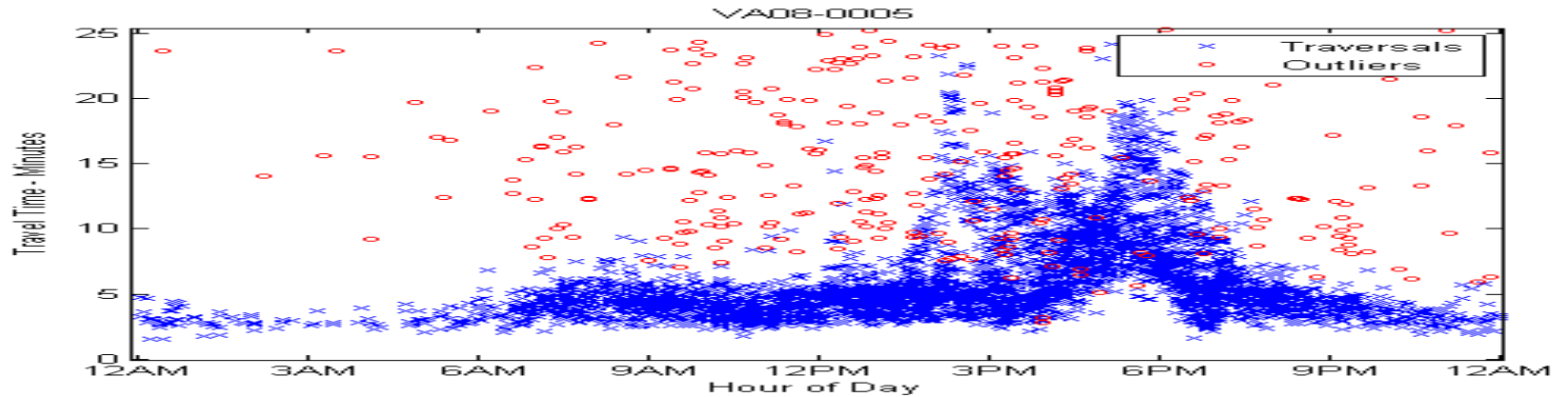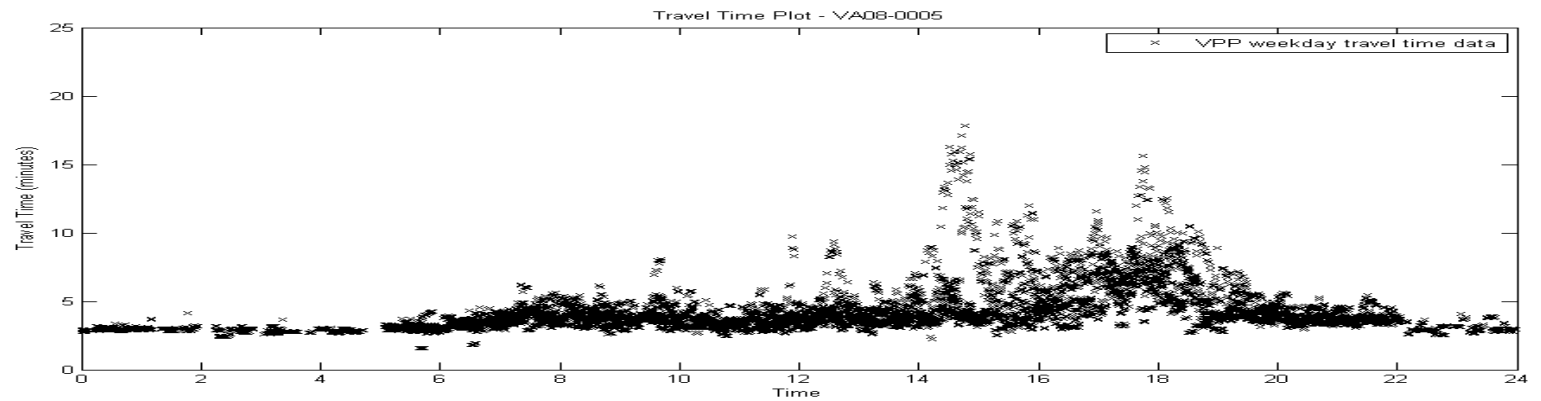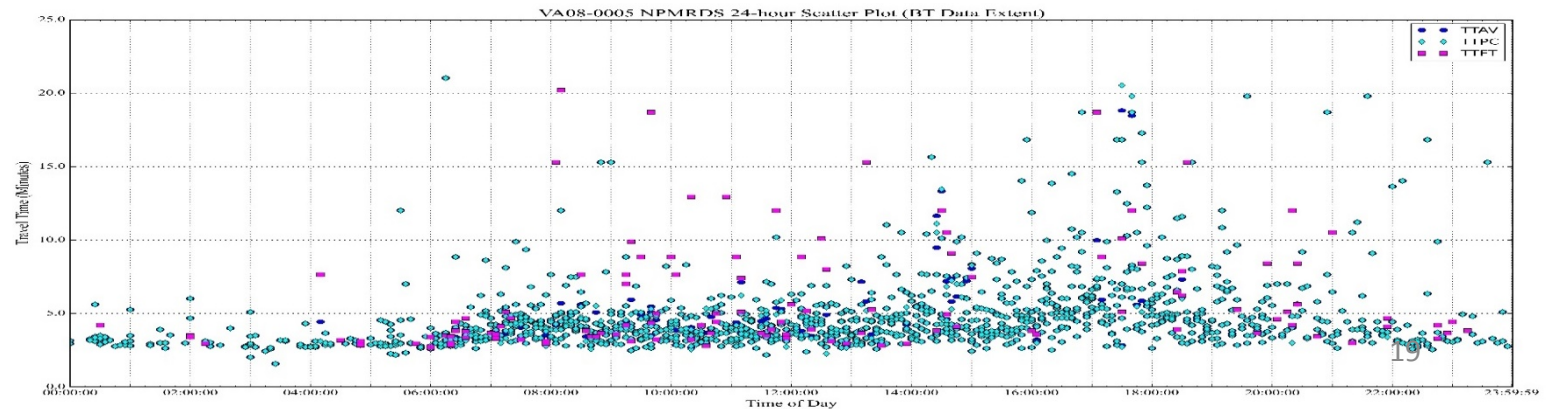
NPMRDS
(All Vehicle)

# Peak Hour: 5 PM to 6 PM (BTM, VPP, NPMRDS)

BTM

VPP
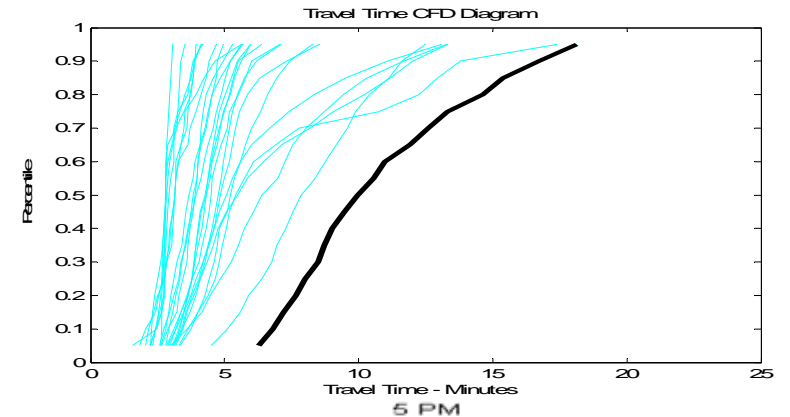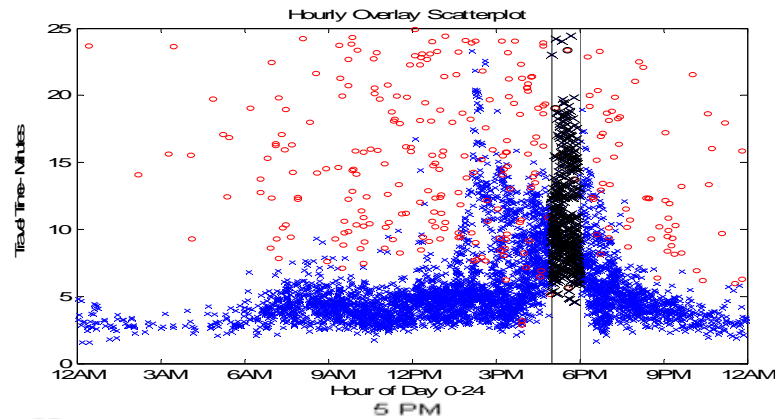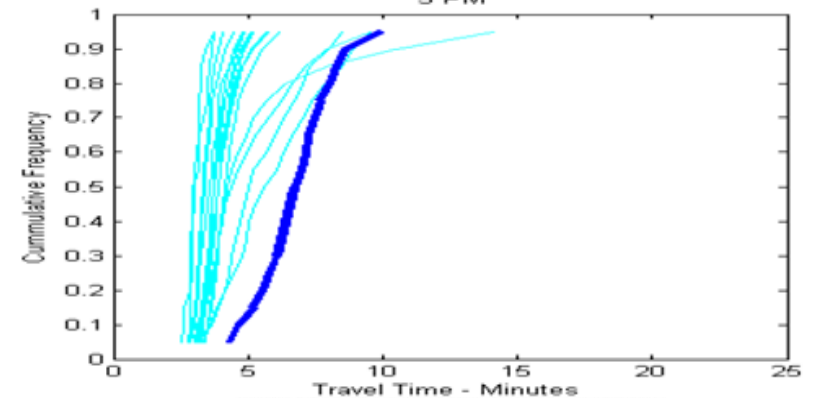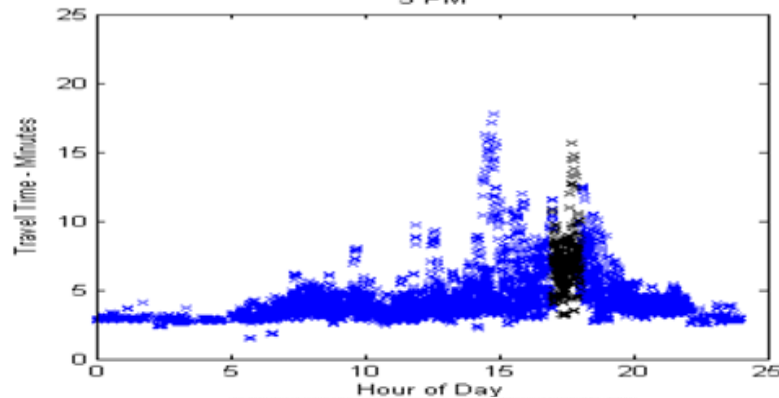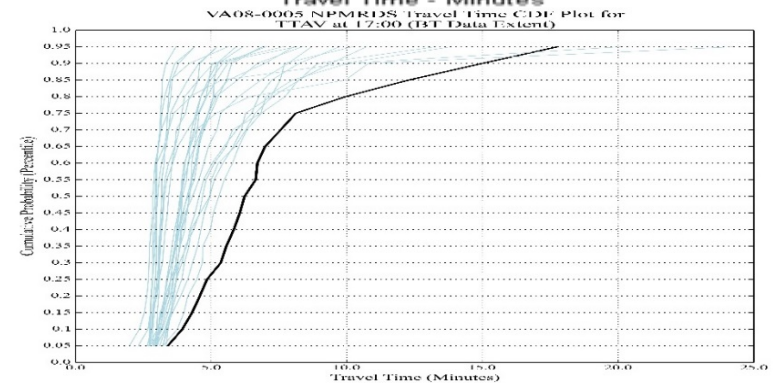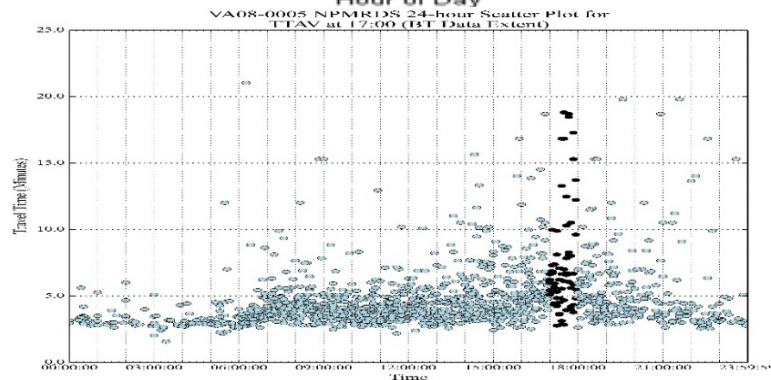
NPMRDS
(All Vehicle)

Peak Hour: 5 PM to 6 PM (NPMRDS Only)

All Vehicle

VA08-0005 NPMRDS 24-hour Scatter Plot for TTAV at 17:00 (BT Data Extent)

VA08-0005 NPMRDS Travel Time CDF Plot for TTAV at 17:00 (BT Data Extent)

Passenger Car

VA08-0005 NPMRDS 24-hour Scatter Plot for TTPC at 17:00 (BT Data Extent)

VA08-0005 NPMRDS Travel Time CDF Plot for TTPC at 17:00 (BT Data Extent)

Freight Truck

VA08-0005 NPMRDS 24-hour Scatter Plot for TTFT at 17:00 (BT Data Extent)

VA08-0005 NPMRDS Travel Time CDF Plot for TTFT at 17:00 (BT Data Extent)

21

US-1 in NJ: Interrupted Flow Arterial

4 Lanes, 6 signals, 1.37

24 Hour Scatter Plots

BTM

VPP

NPMRDS
(All Vehicle)

23

Peak Hour: 4 PM to 5 PM (BTM, VPP, NPMRDS)

BTM

VPP
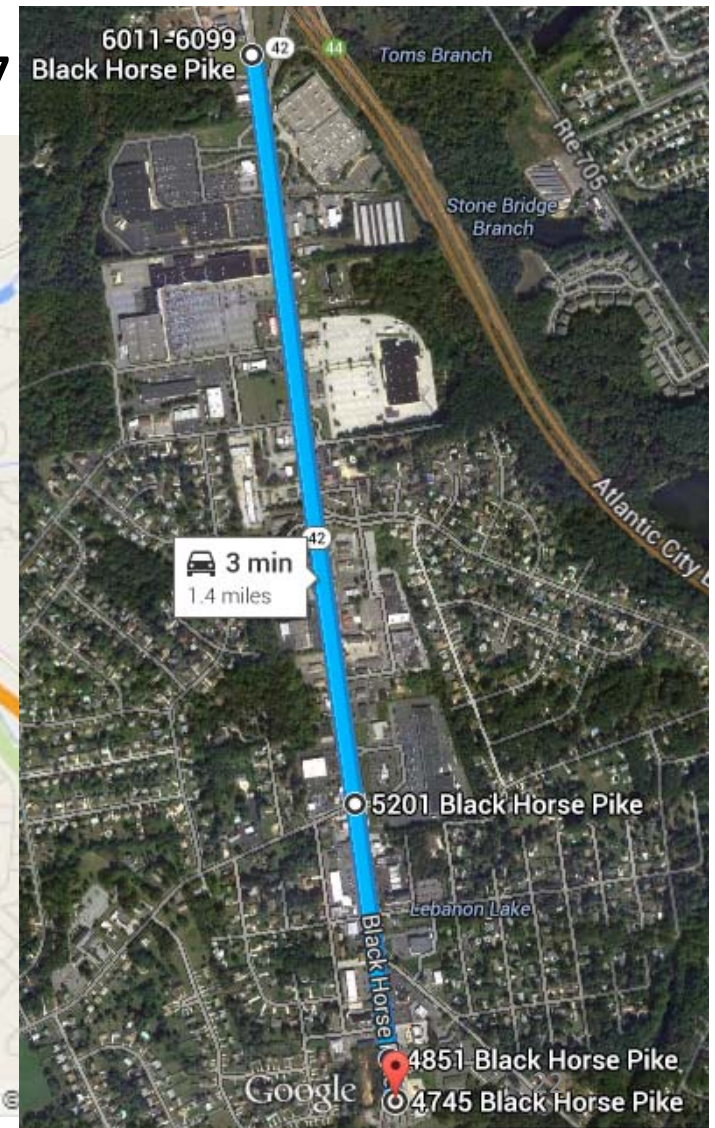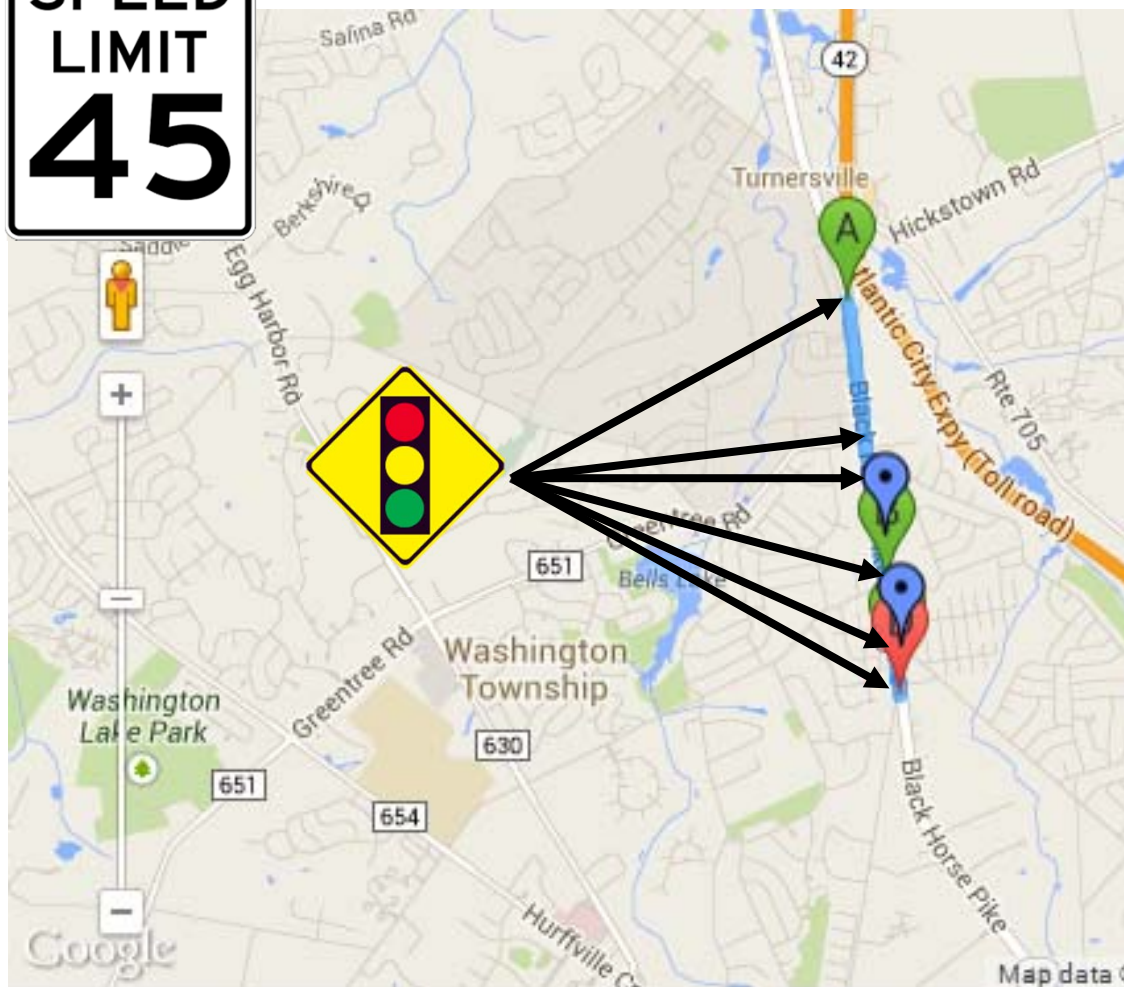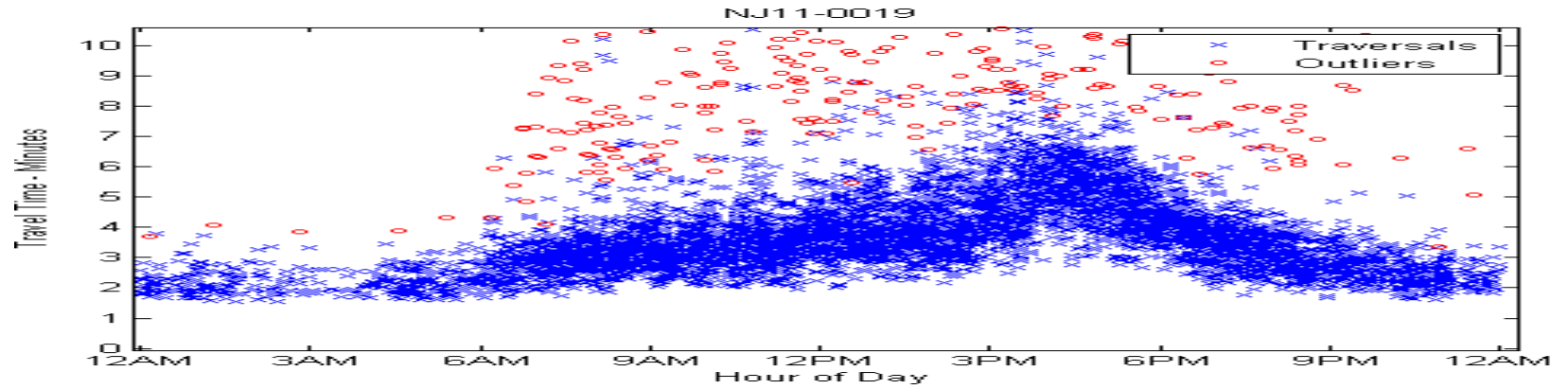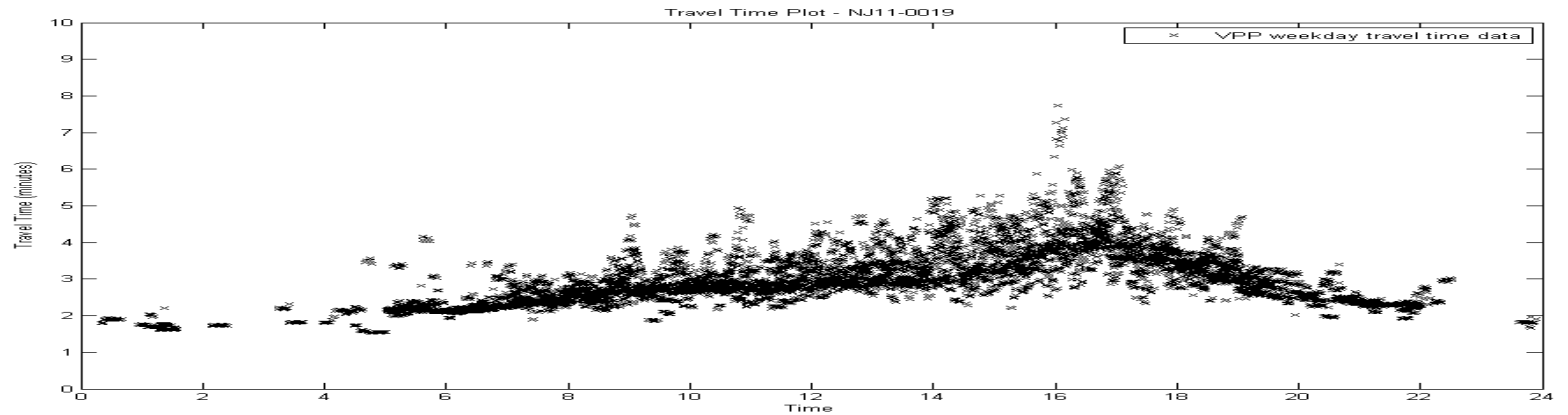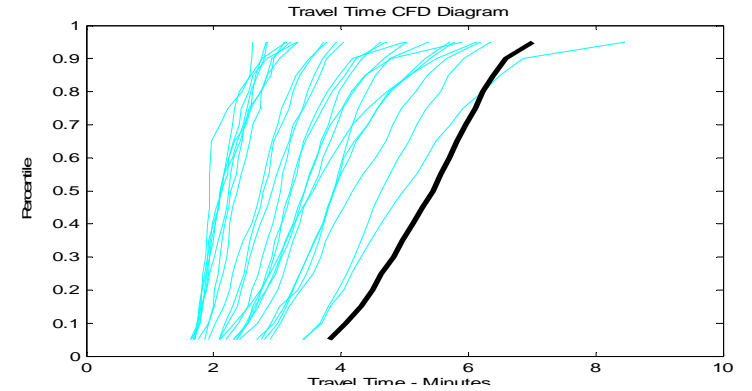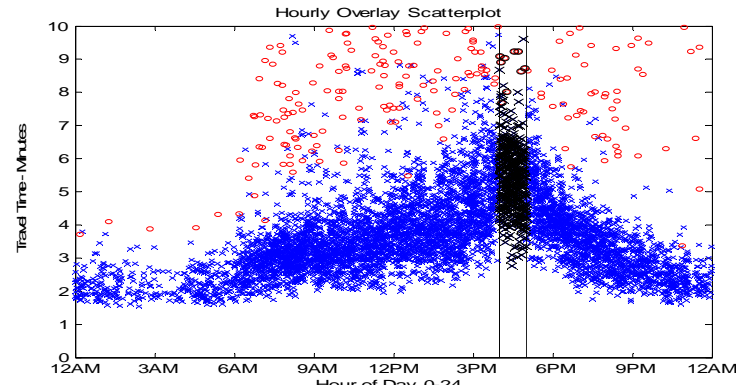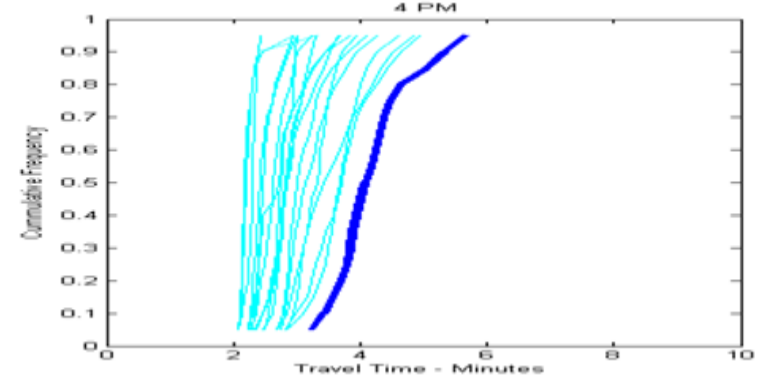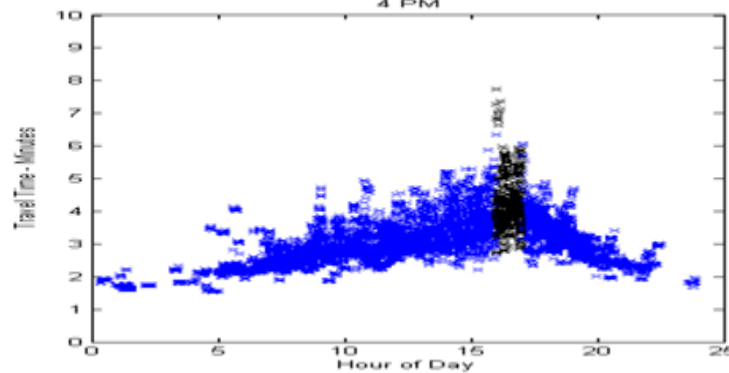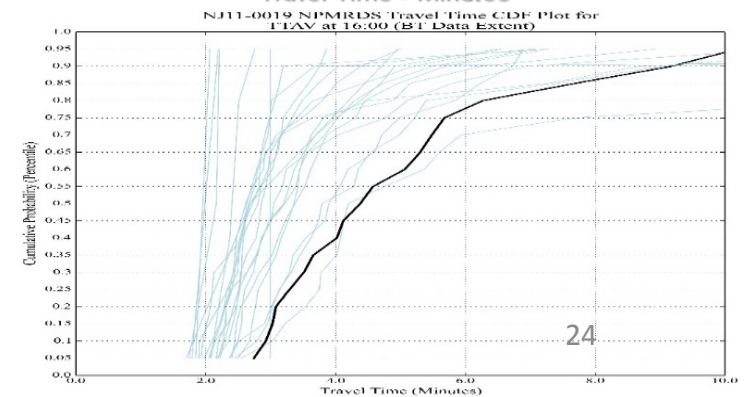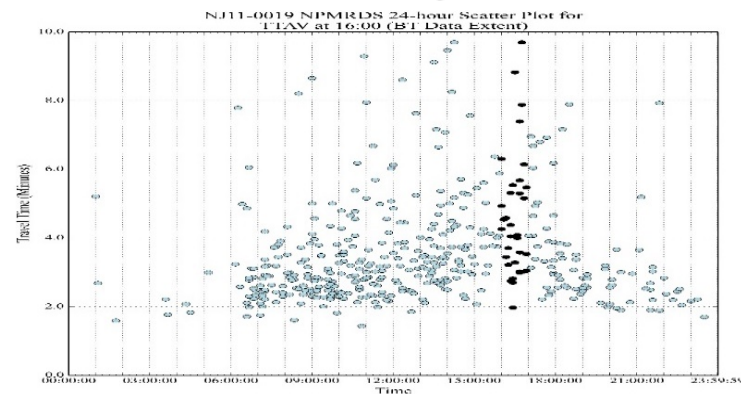
NPMRDS
(All Vehicle)

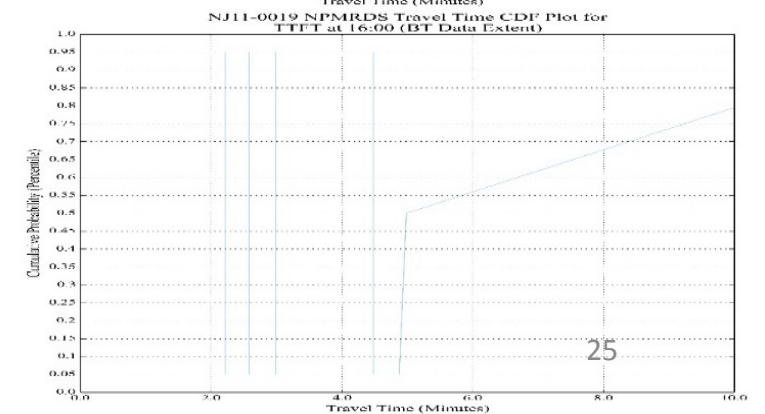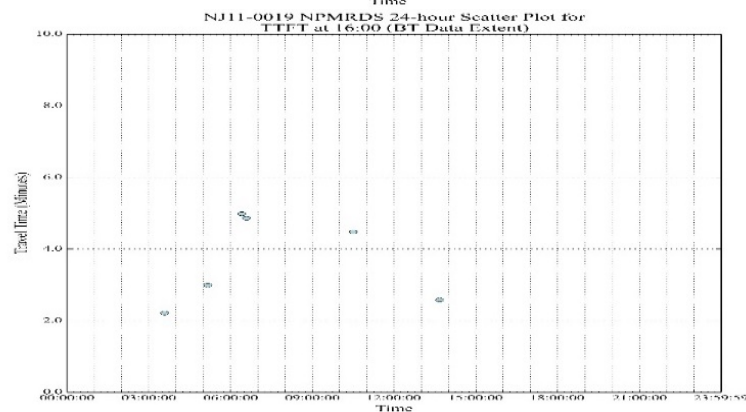# Peak Hour: 4 PM to 5 PM (NPMRDS Only)

All Vehicle
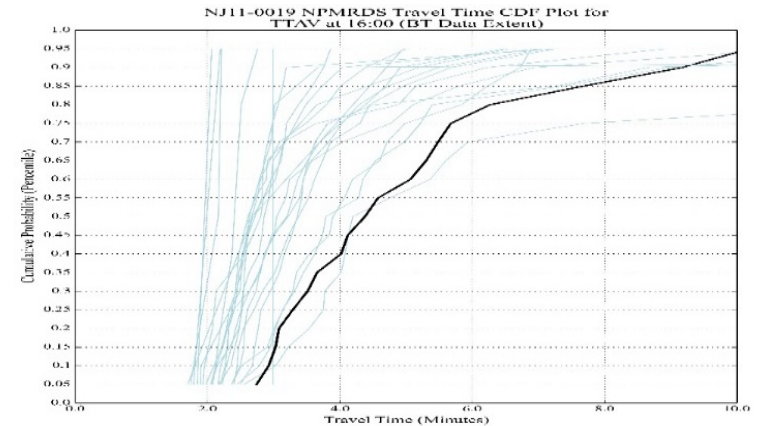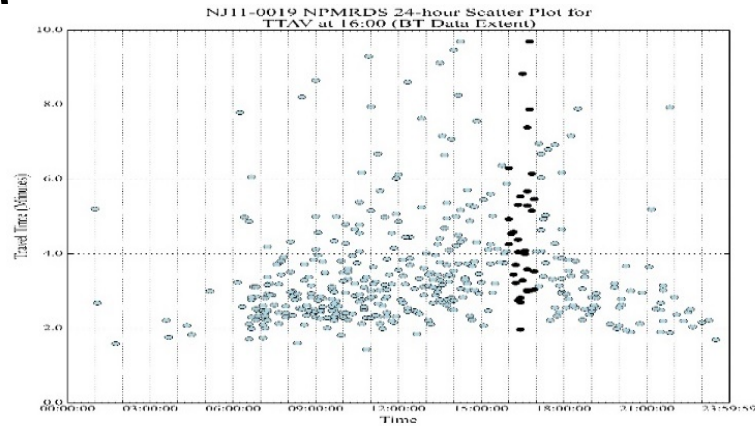
Passenger Car
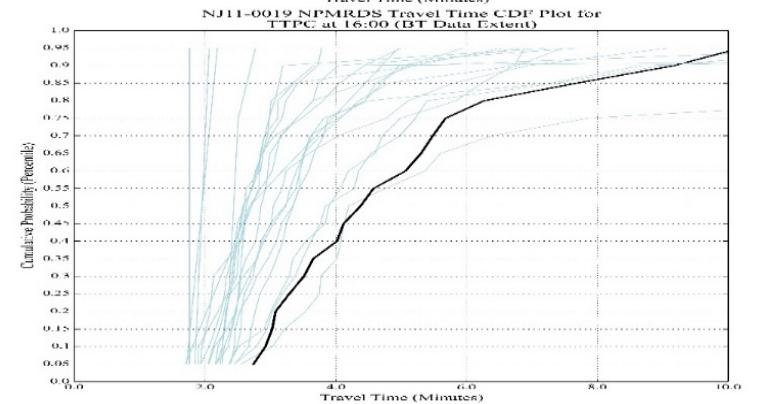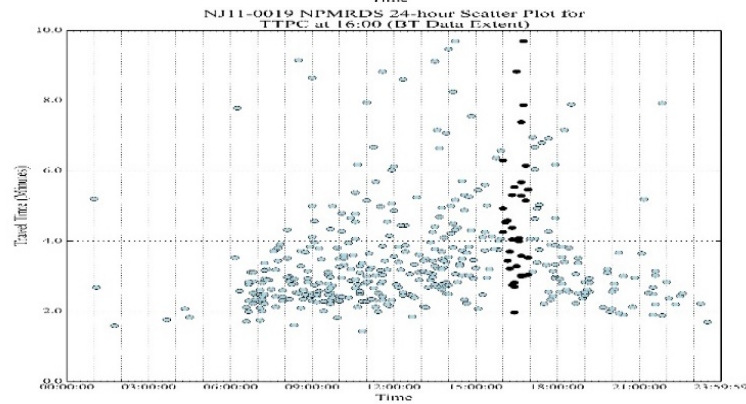
Freight Truck

Take Aways

- Accessing and Analyzing
  - Data size too large for Microsoft Office
  - Need database or other specialized analysis packages
  - GIS tools for visualizing on a map
- Comparison with probe and Bluetooth data
  - 24 hour overlay plots and corresponding CDF curves capture recurring congestion
  - NPMRDS tends to capture variance in traffic similar to Bluetooth, but still subject to same limitations of probe data
- The overlay and percentile method can be used to calculate performance measures without imputation or outlier filtering

Thank You

Questions?

Appendix

Not to be presented, unless to answer questions.

Example of TMC mismatch between vendors

- NPMRDS TMC definition vs. INRIX (TomTom) TMC definition

110+04419 INRIX
1.97 miles

110P04418 INRIX
1.09 miles

110P+04418 INRIX
0.60 miles

110P04418 NPMRDS
1.332 miles
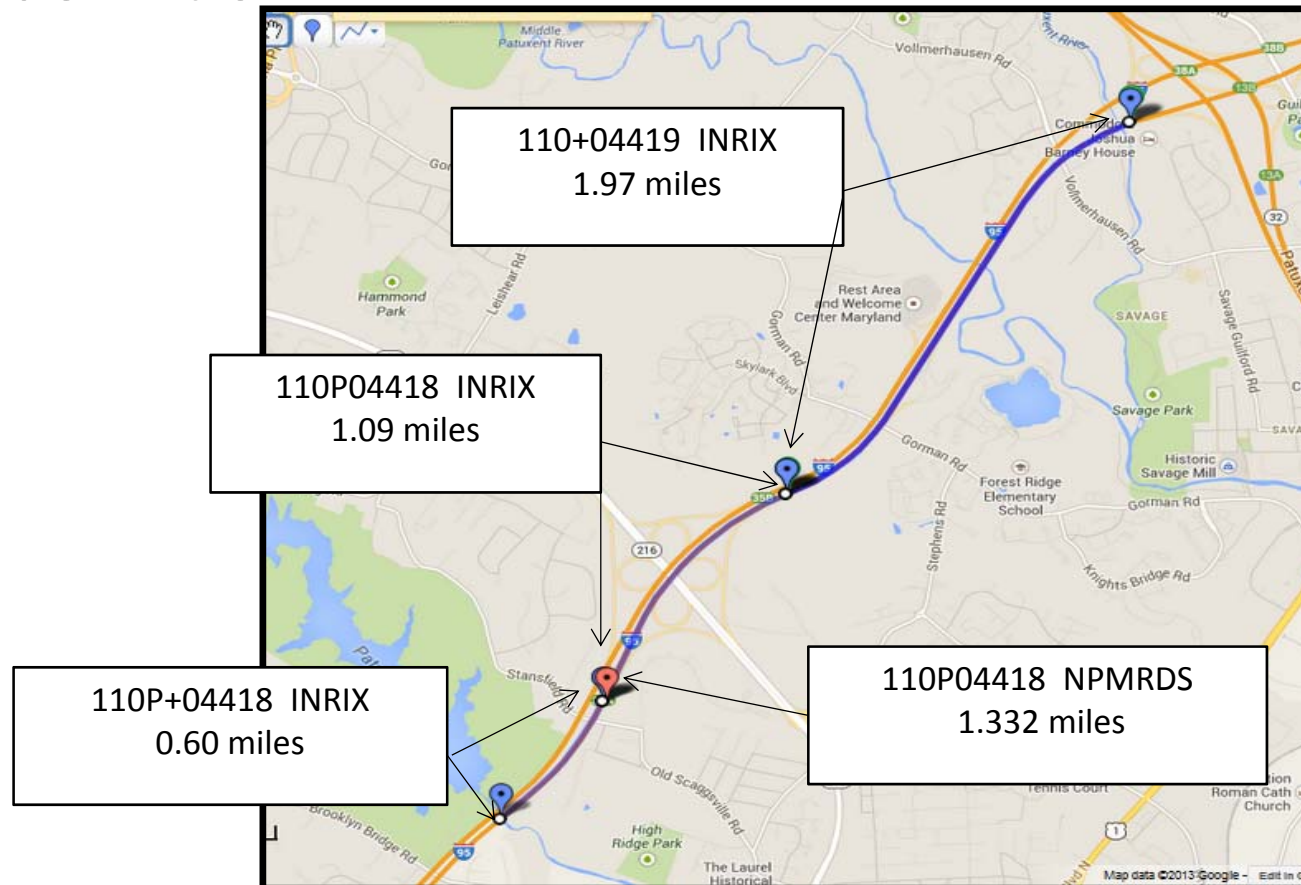
Additional Corridor

Route 29 in Virginia: Freeway-ish
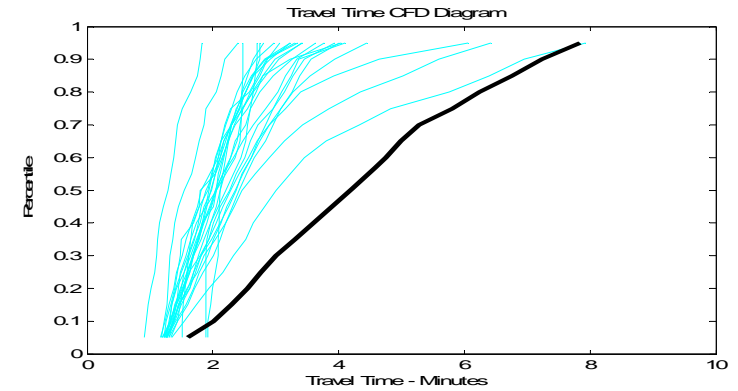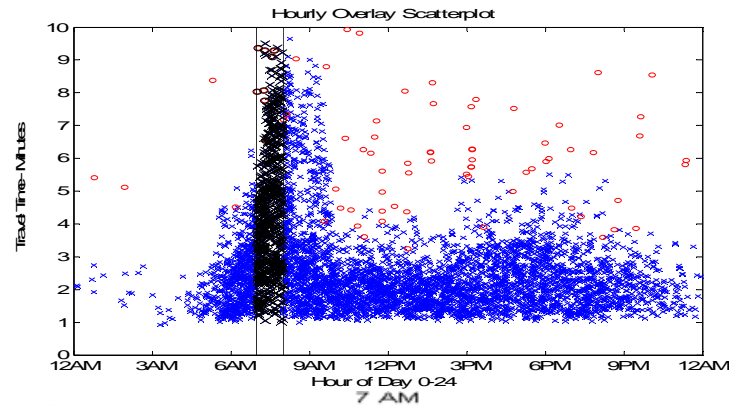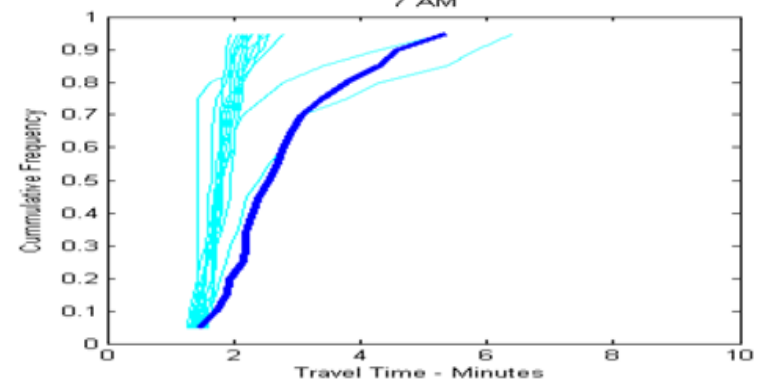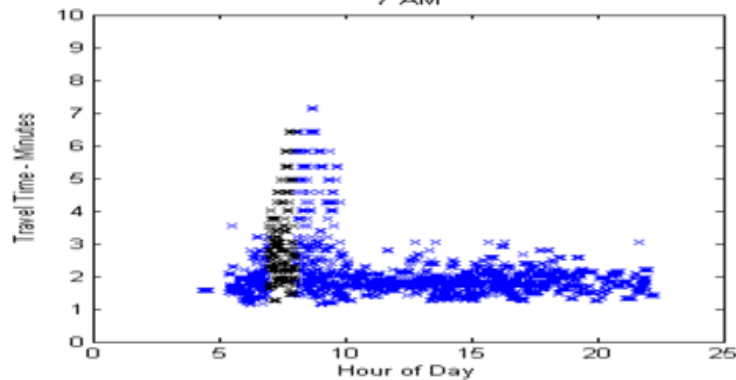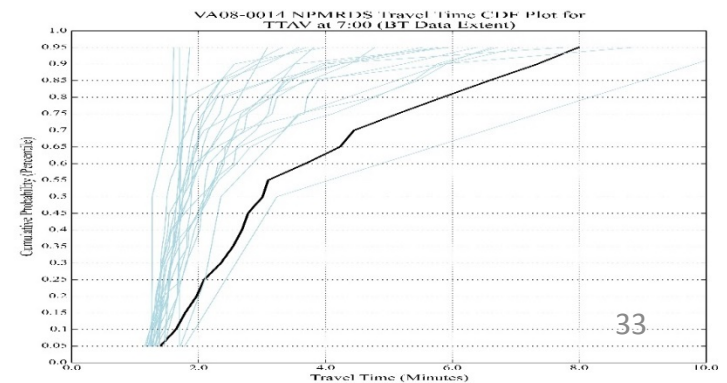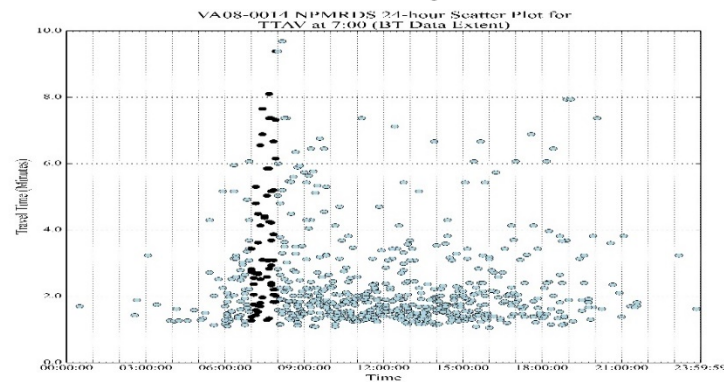## 4 Lanes, 2 signals, 1.07 miles

24 Hour Scatter Plots

BTM

VPP
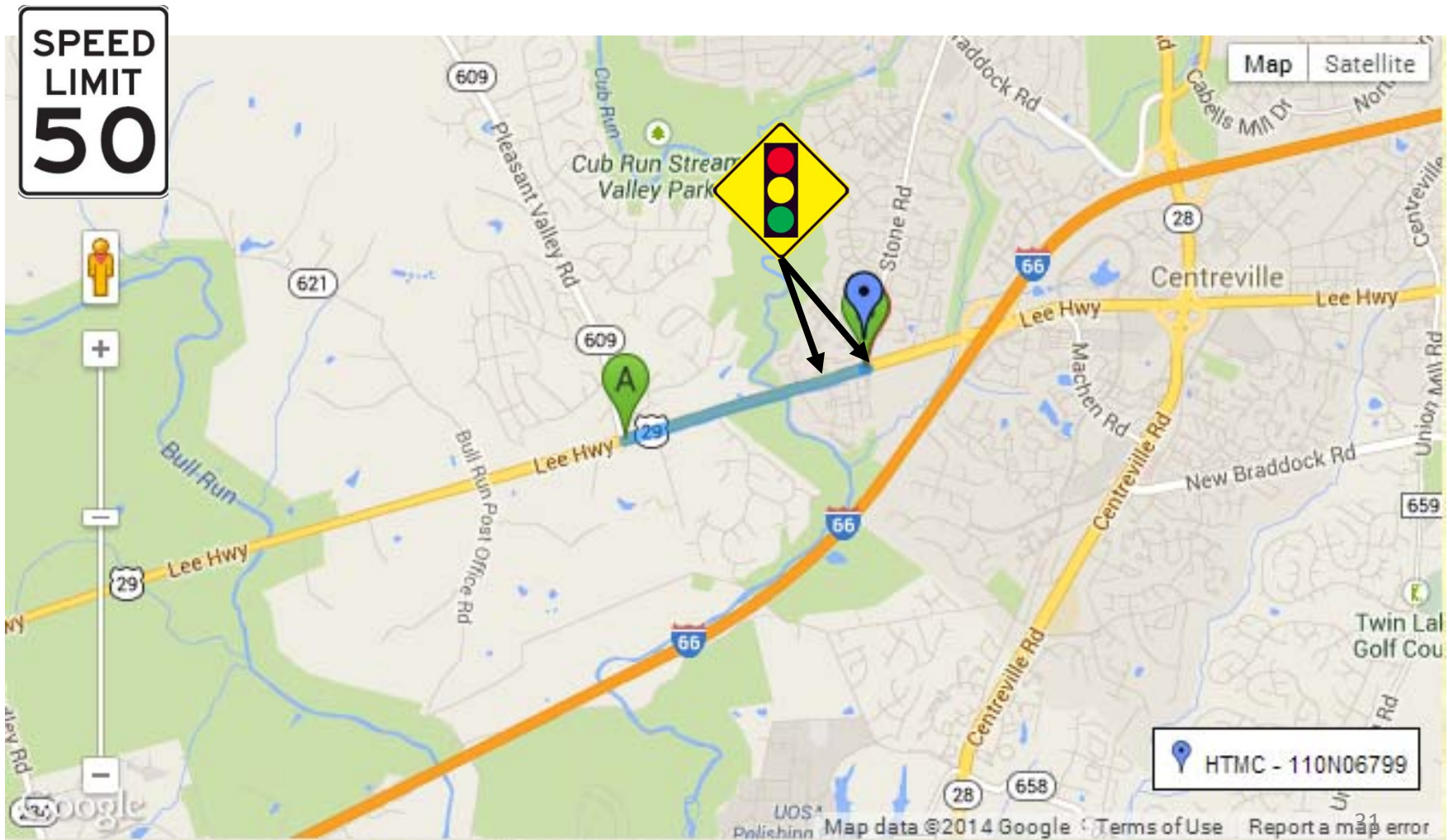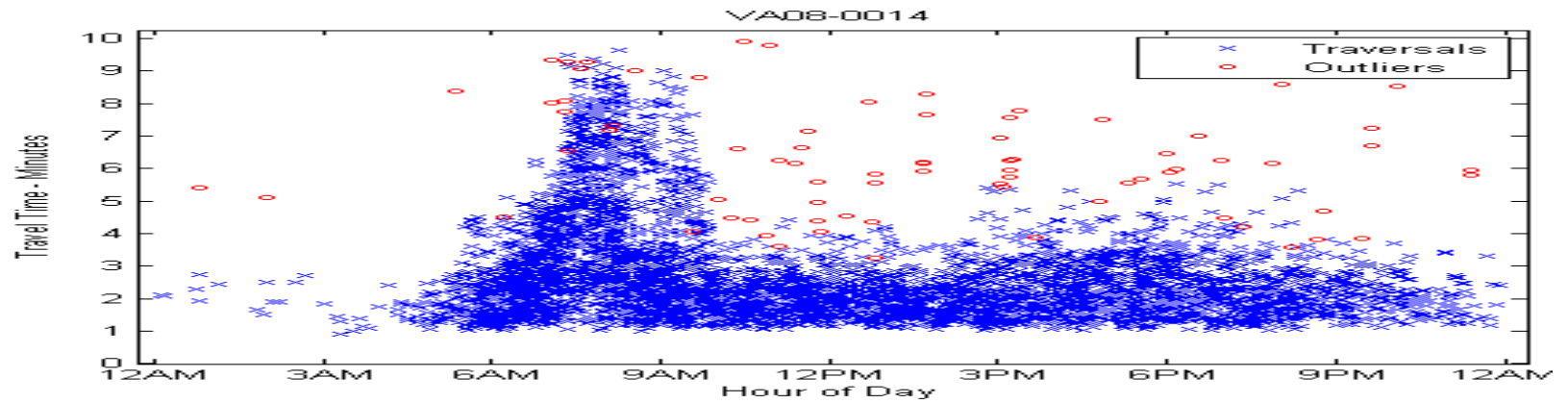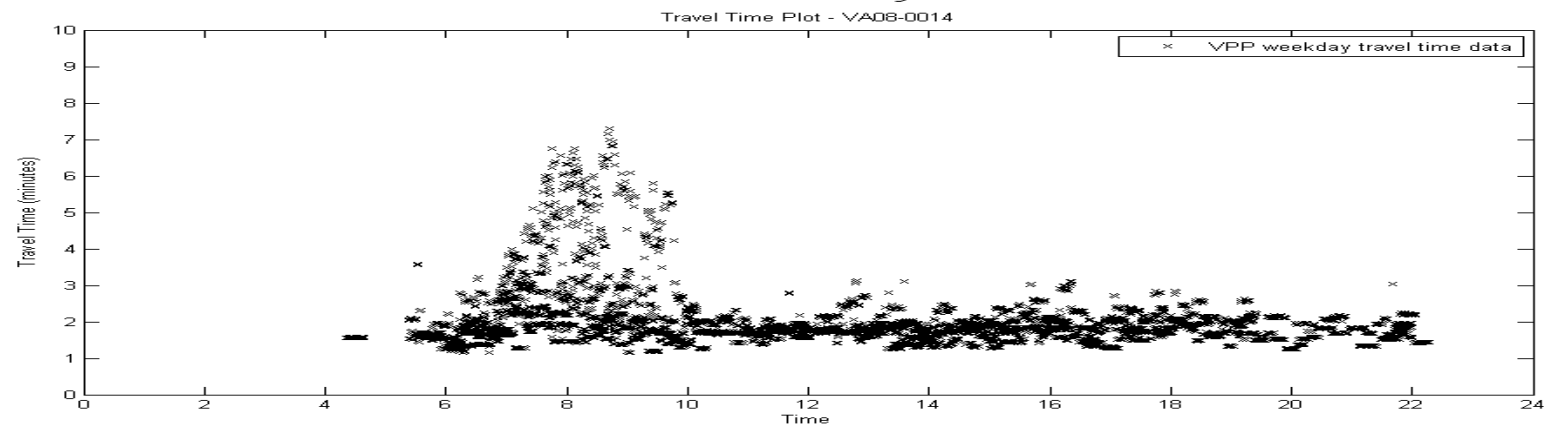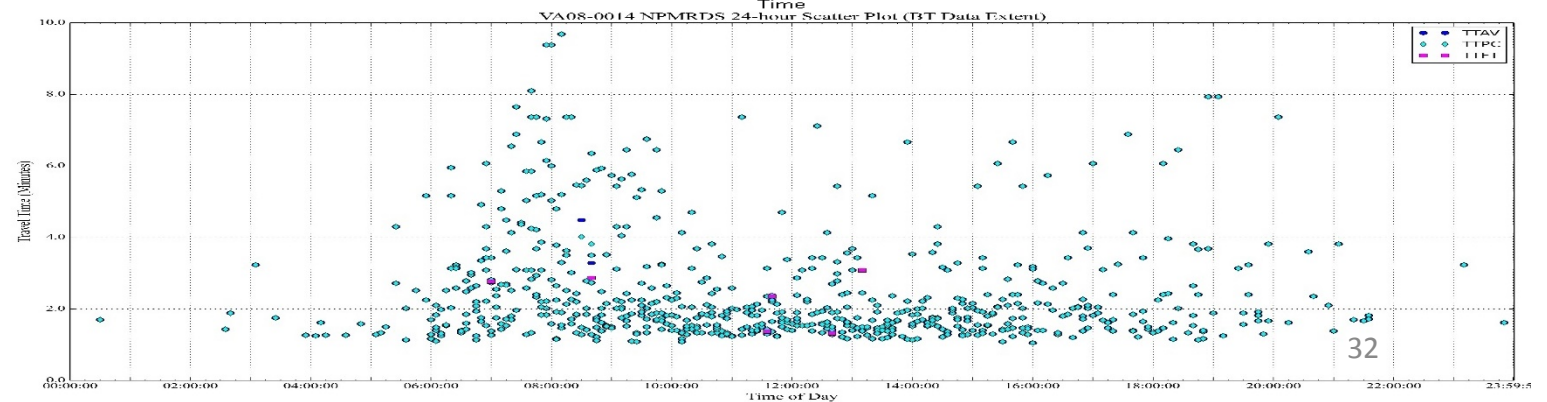
NPMRDS
(All Vehicle)

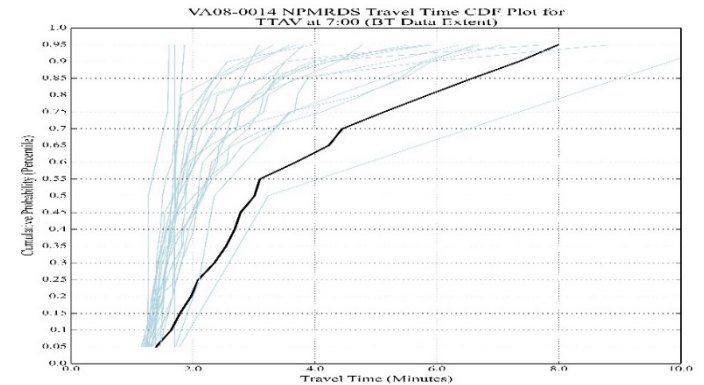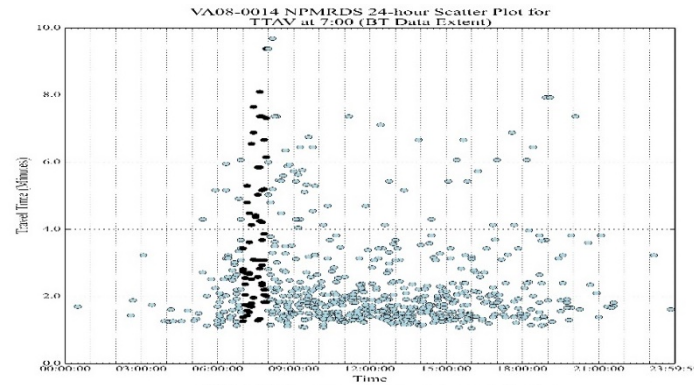# Peak Hour: 7 AM to 8 AM (BTM, VPP, NPMRDS)

BTM

VPP

NPMRDS
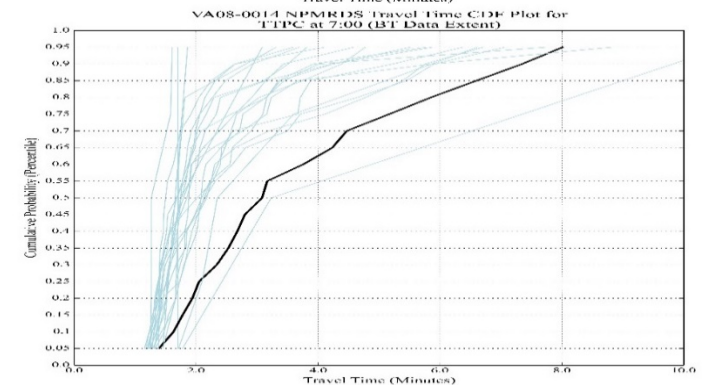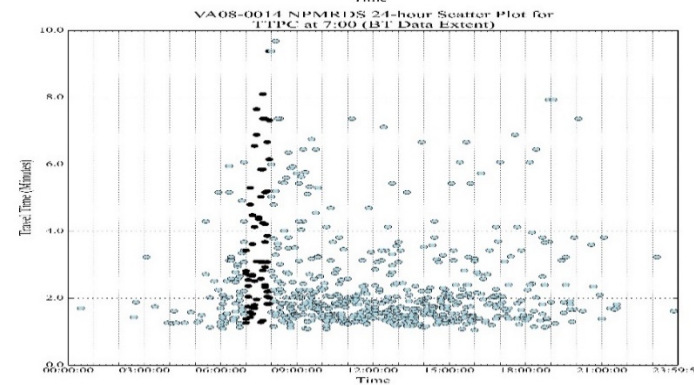**(All Vehicle)**
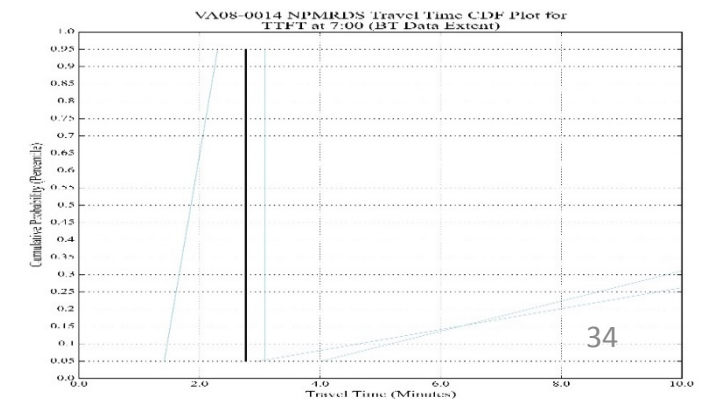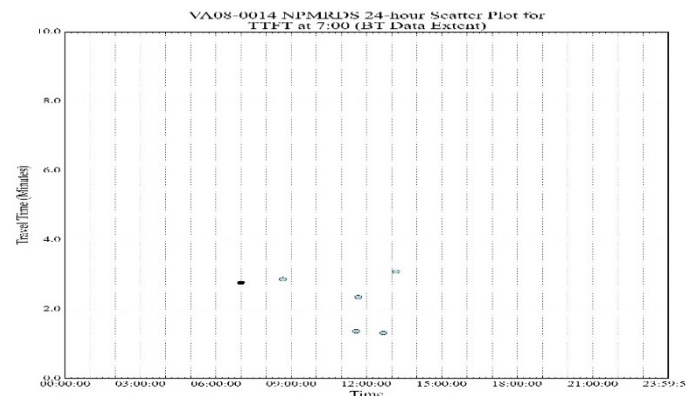

33

Peak Hour: 7 AM to 8 AM (NPMRDS Only)

All Vehicle

Passenger Car

Freight Truck

34

Compute Performance Measures from CDF

Formulae for Performance Measures

- Travel Time Index (TTI) = $\dfrac{P_{50}}{P_{15}}$
  - Or, TTI = $\dfrac{P_{50}}{\sum_{i=5}^{95}\frac{P_i}{19}}$ (because, mean = $\sum_{i=5}^{95}\dfrac{P_i}{19}$ from percentiles)

- Planning Time Index (PTI) = $\dfrac{P_{95}}{P_{15}}$

- Buffer Time Index (BTI) = $\dfrac{P_{95}-P_{50}}{P_{50}}$
  - Or, BTI = $\dfrac{P_{95}-\sum_{i=5}^{95}\frac{P_i}{19}}{P_{50}}$

- Inter-Quartile Range (IQR) = $P_{75} - P_{25}$

Method