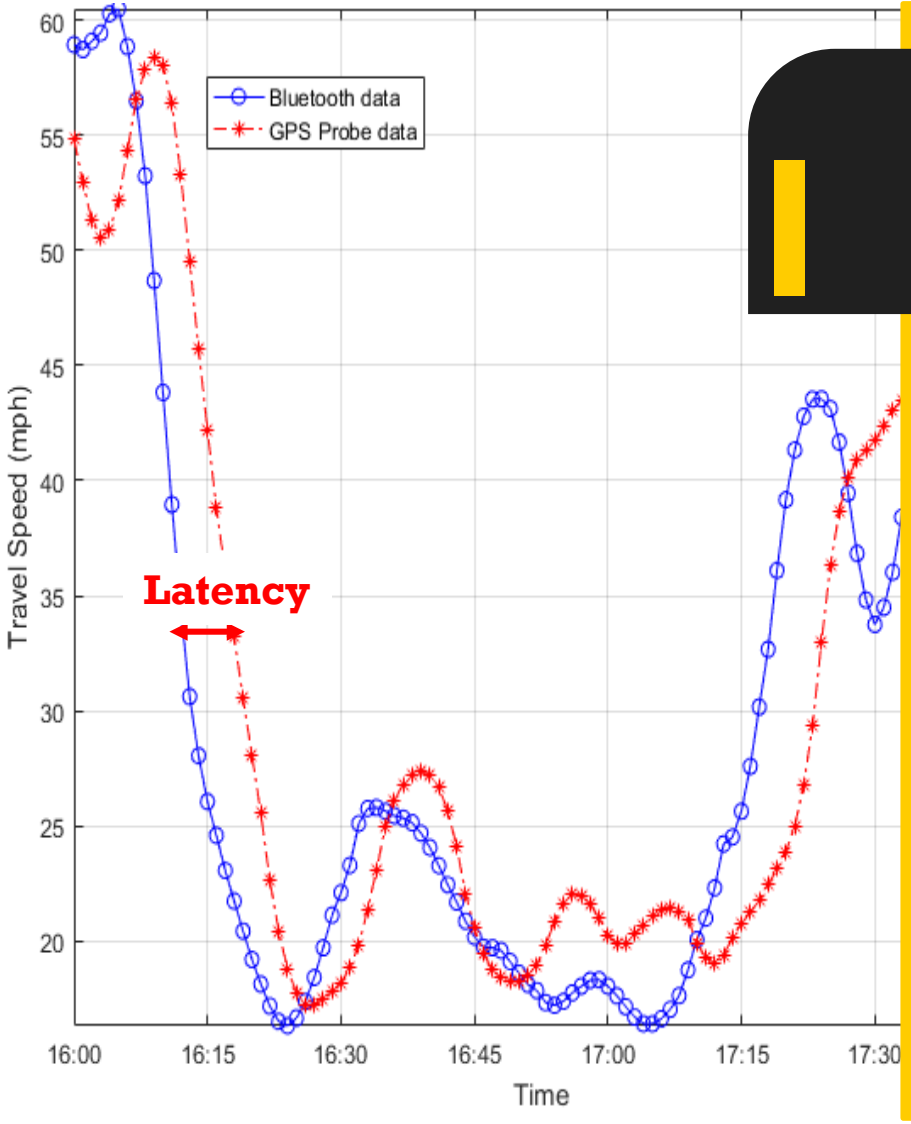
Transportation Research Board
96th Annual Meeting
January 8–12, 2017 ▪ Washington, D.C.

# Methodology for Calculating Latency of GPS Probe Data


Masoud Hamedi | Zhongxiang Wang | Stanley Young

Jan 10th 2017

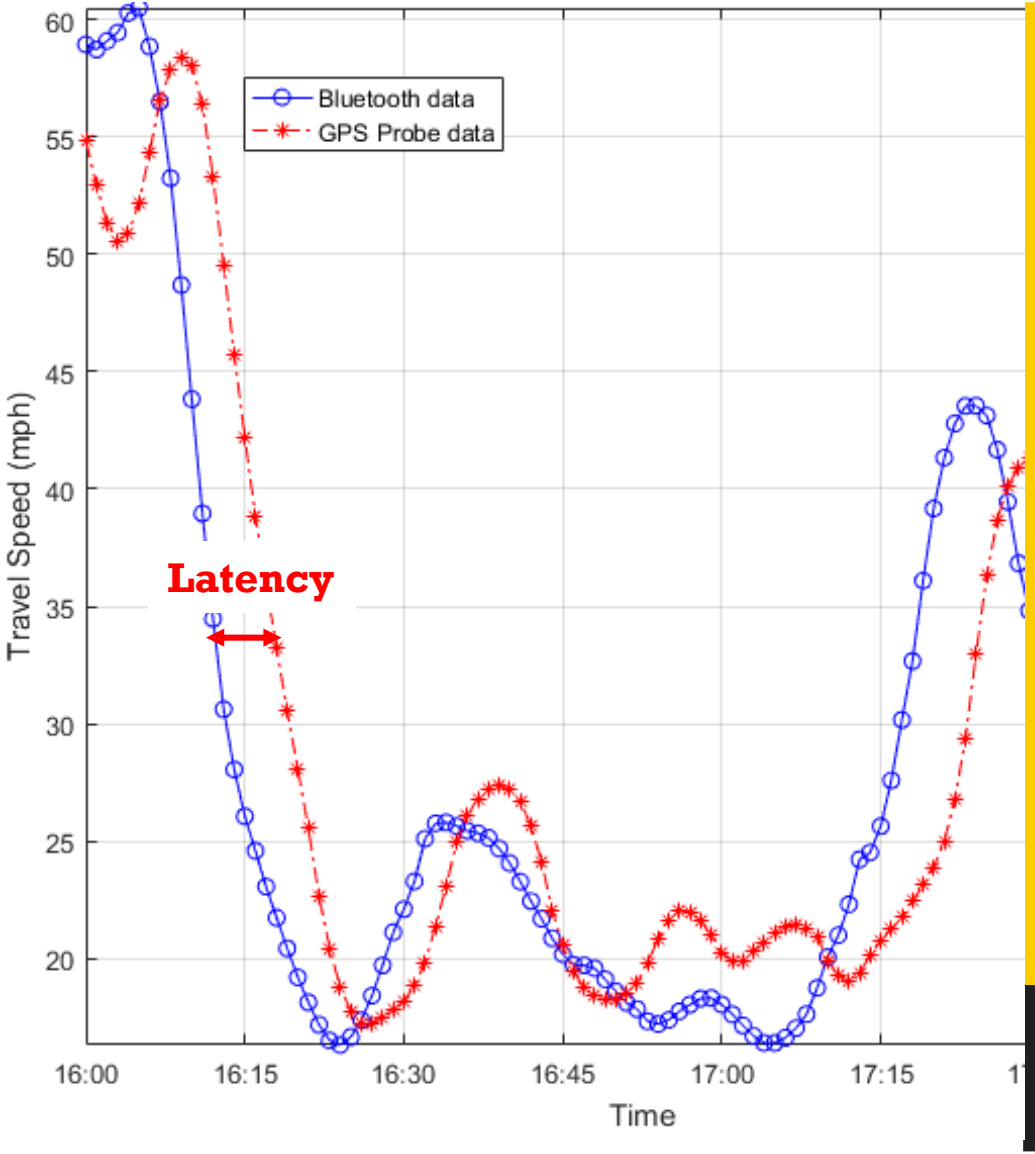



# OUTLINE

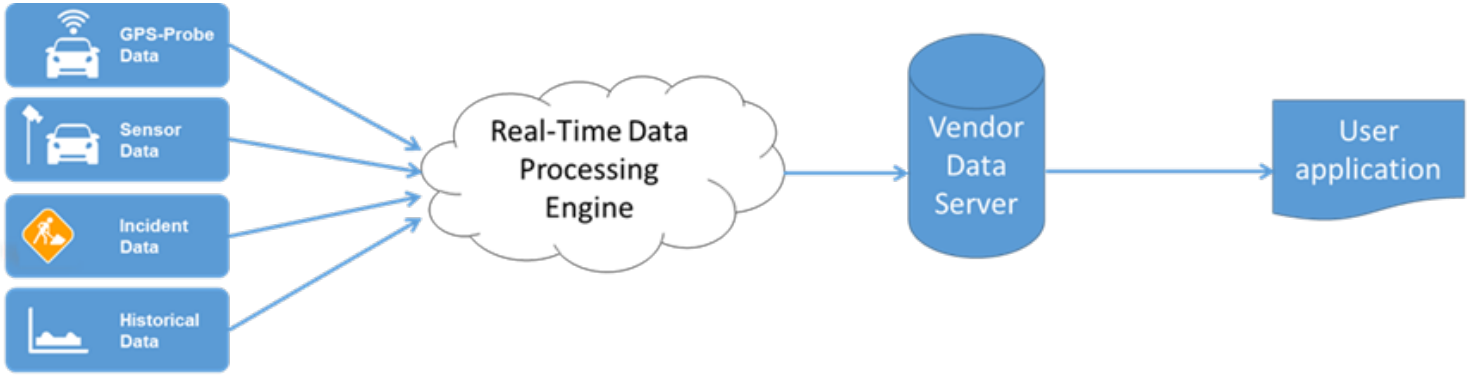
1. INTRODUCTION
2. LITERATURE REVIEW
3. METHODOLOGY
4. CASE STUDY
5. CONCLUSION

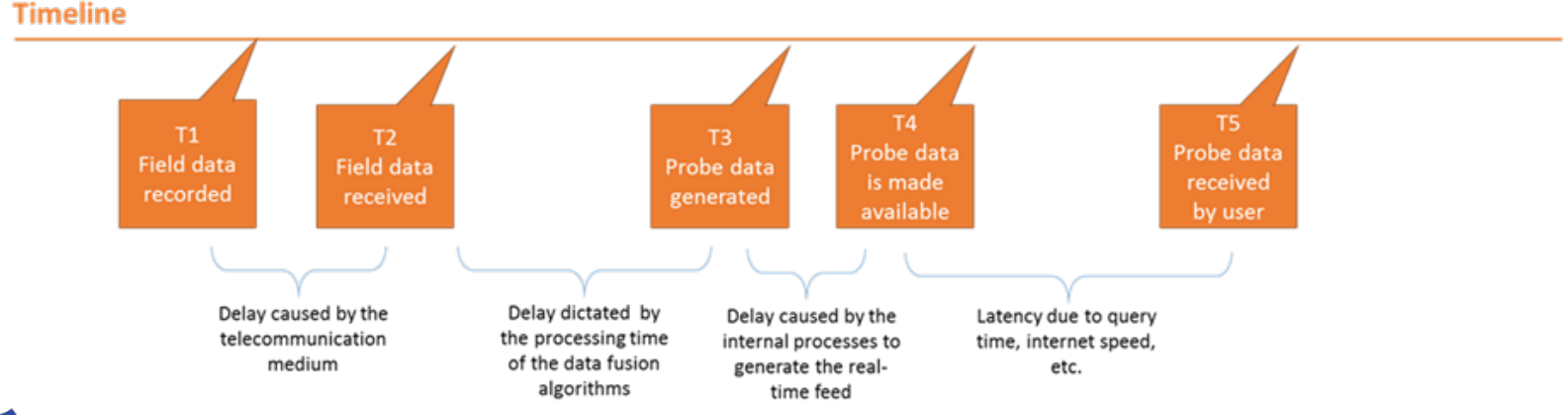
# What is Latency?

- **LATENCY** is the time difference between GPS-probe data and real traffic condition;
- It describes the punctuality of data;
- It is crucial to real-time applications.



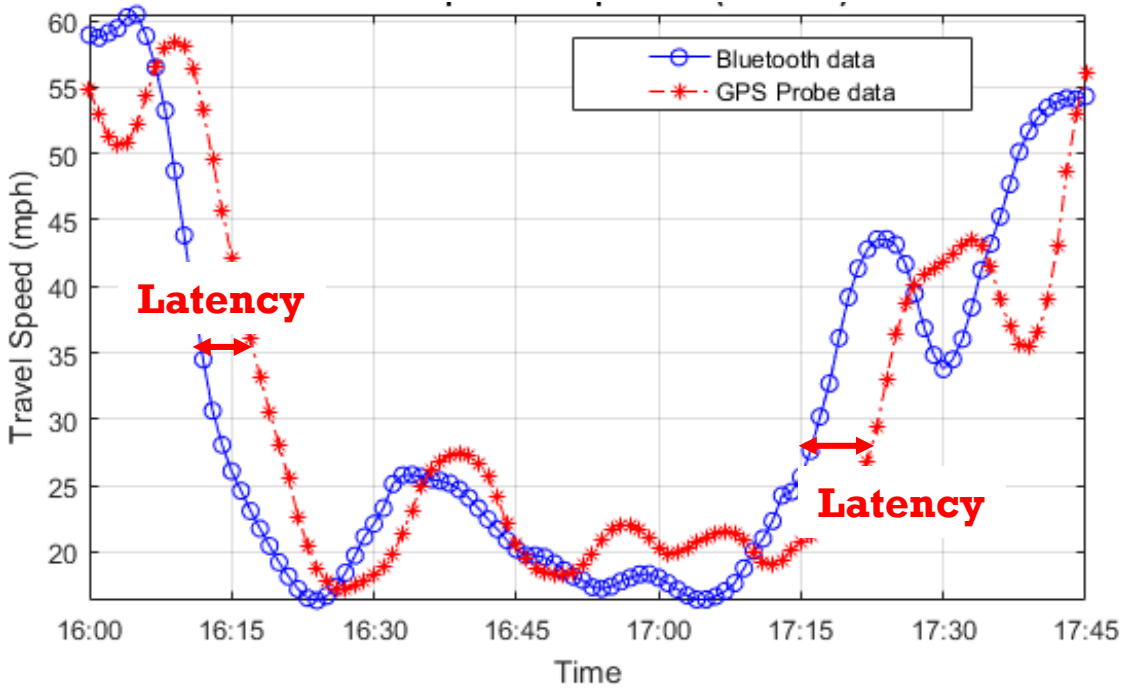
# Sources of Latency



# Definition of Latency

Latency is defined as “the difference between the time the traffic flow is perturbed and the time that the change in speed is reflected in the data”.



# Current Research

- Kim, Seoungbum, and Benjamin Coifman. "Comparing INRIX speed data against concurrent loop detector stations over several months." Transportation Research Part C: Emerging Technologies 49 (2014): 59-72.
  - Objective: Maximizing Correlation Coefficient
  - Average latency : 6.8 minutes
  - Maximum latency could exceed 10 minutes in many time periods
  - Reference data: Loop detector
  - Aggregating data into 10 second time interval

- INRIX.(2007). Traffic Data and Associated Services along the I-95 Corridor
  - INRIX will deliver current speed, travel time, average speed…with latency on average of 4.5 minutes.

# Data Processing

## Bluetooth data preparation

- Transferring travel time data into TMC segment space mean speed data;
- Aggregating Bluetooth data into one minute basis.

## Data Filtering

**The following observations are identified and discarded:**

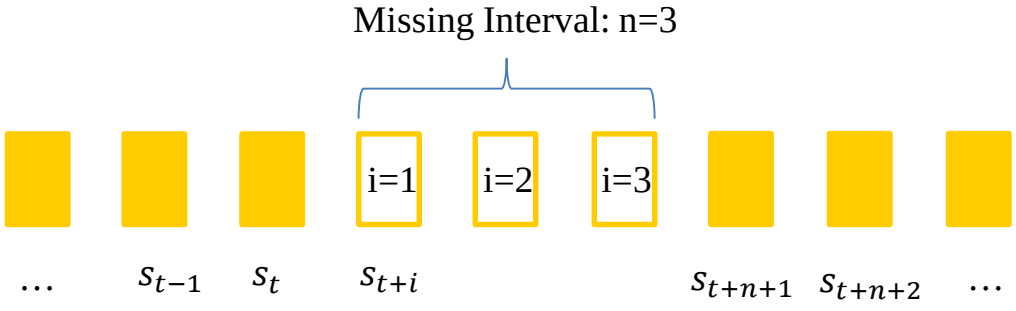
- Observations with unreasonably low or high speeds;
- Observations in a particular time interval that are far from the average of the rest of the speeds observed in the same time interval.

# Data Processing

## Data Interpolation

The average of the neighboring observations is considered as the travel speed for the missing interval (only applied to less than or equal to 5 mins).

$$s_{t+i} = s_t + \frac{i}{n+1}\left(s_{t+n+1} - s_t\right) \quad \forall i = 1,2,3$$




METHODOLOGY

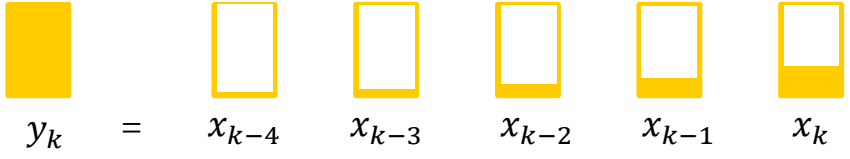
# Data Processing

## Data Smoothing

❖ Weighted moving average function

Arithmetic growth of the weights with the previous five time intervals

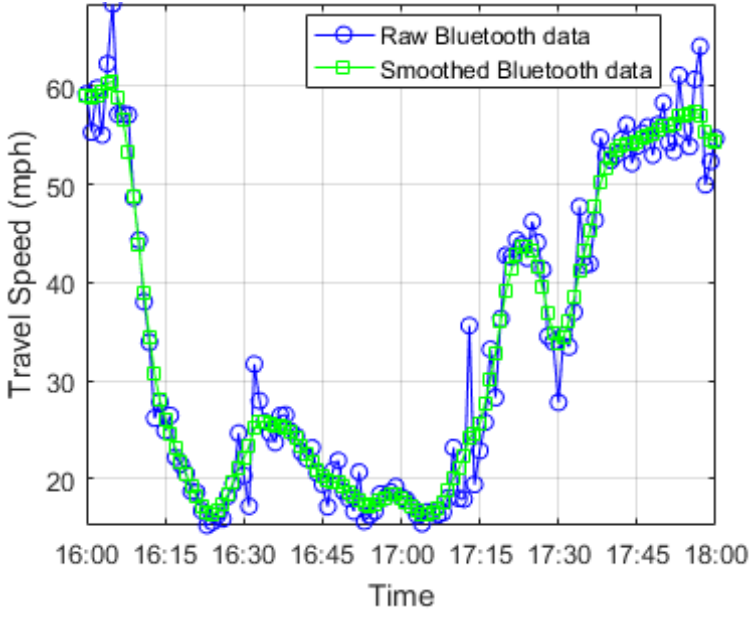
$$y_k = 0.33x_k + 0.27x_{k-1} + 0.20x_{k-2} + 0.13x_{k-3} + 0.07x_{k-4}$$



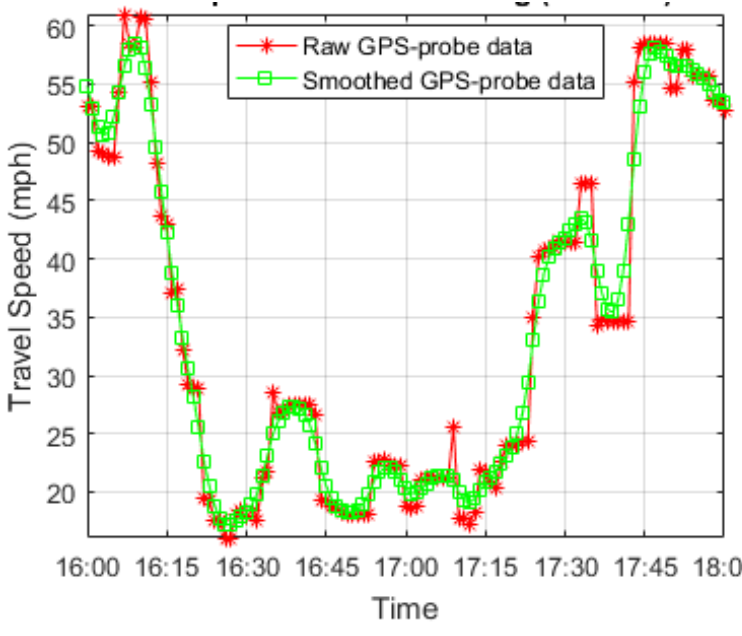

*Why filtfilt?* Smoothing may introduce undesired horizontal shift into the time series, however in *filtfilt*, forward shift is followed by a backward shift, so the artificial shift is compensated.

# Data Processing

## Data Smoothing

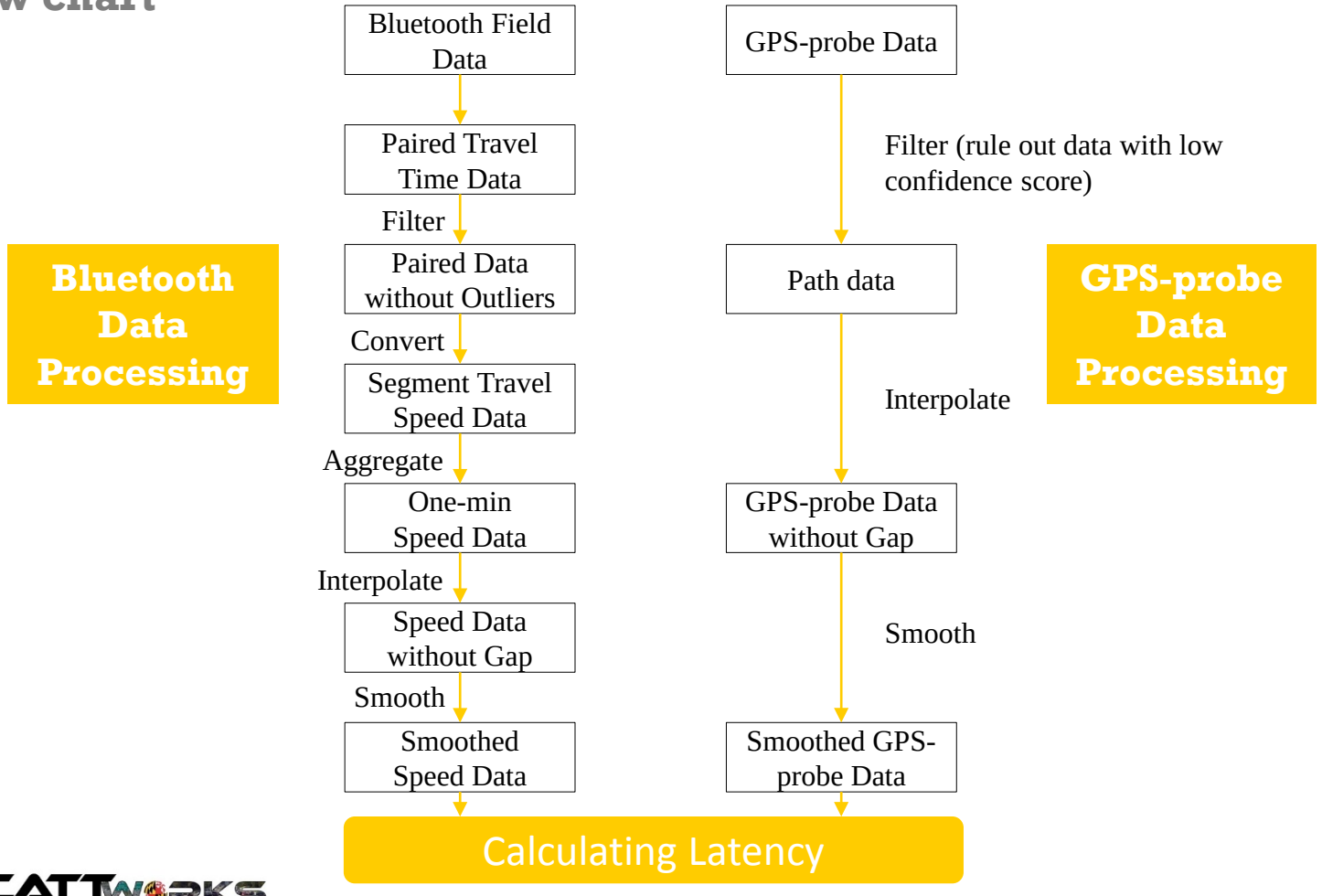



- ✓ **Less noise**
- ✓ **No shift**



# Data Processing

## Flow chart



# Objective

- Find the shift distance that maximizes the overlapping of Bluetooth data and GPS-probe data.

$$\min f1 = \sum_{t=1}^{n} \left| S_t^{BT} - S_{t-latency}^{probe\ data} \right|$$

➢ Absolute vertical distance between two curves

$$\min f2 = \sum_{t=1}^{n} \left( S_t^{BT} - S_{t-latency}^{probe\ data} \right)^2$$

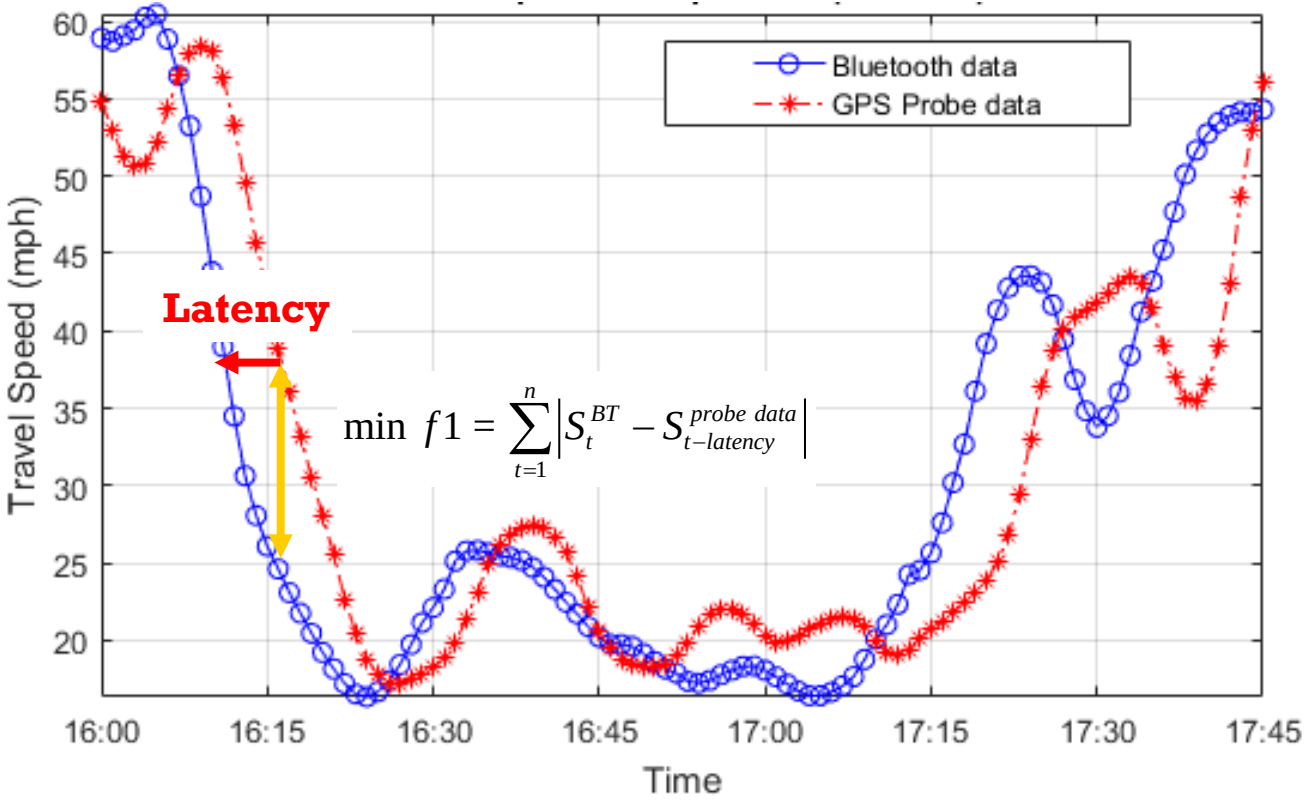
➢ Gives more weights to the points that have bigger difference

$$\min f3 = \mathrm{corr}(S_t^{BT}, S_{t-latency}^{probe\ data})$$

➢ Statistical representation of the linear relationship between two curves

$$lb \le latency \le ub$$

# Minimize Absolute Vertical Distance (*f1*) – as example




$$\min\ f1 = \sum_{t=1}^{n} \left| S_t^{BT} - S_{t-latency}^{probe\ data} \right|$$

# Data Selection

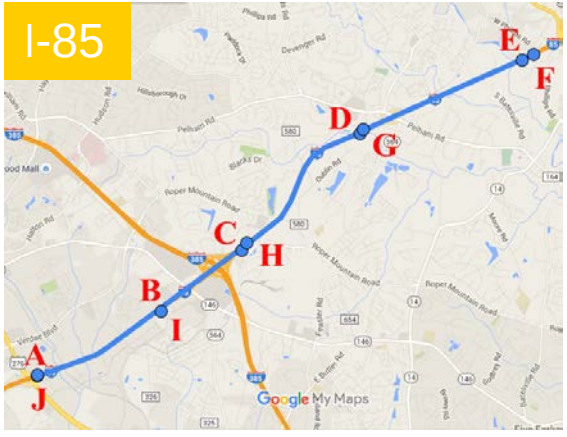
Road type: Freeway

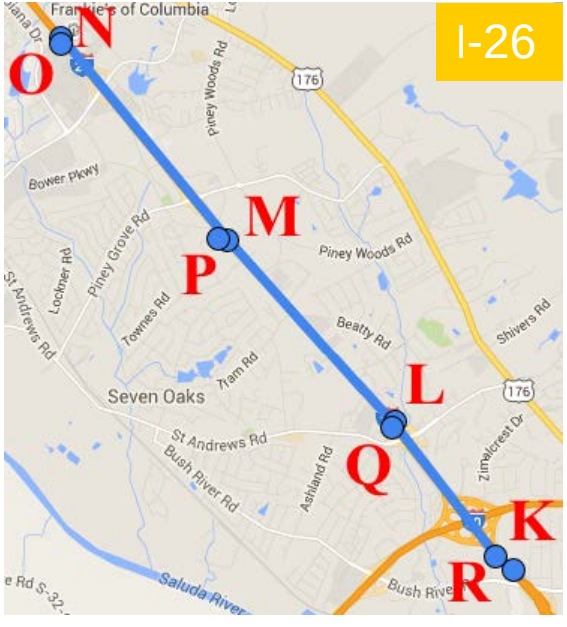
Location: South Carolina,
I-85 (Exit 48 to Exit 54, 7.15 miles) &
I-26 (Exit 103 to Exit 108 , 4.28 miles)

Direction: both directions

Time: Dec 3, 2015 to Dec15, 2015

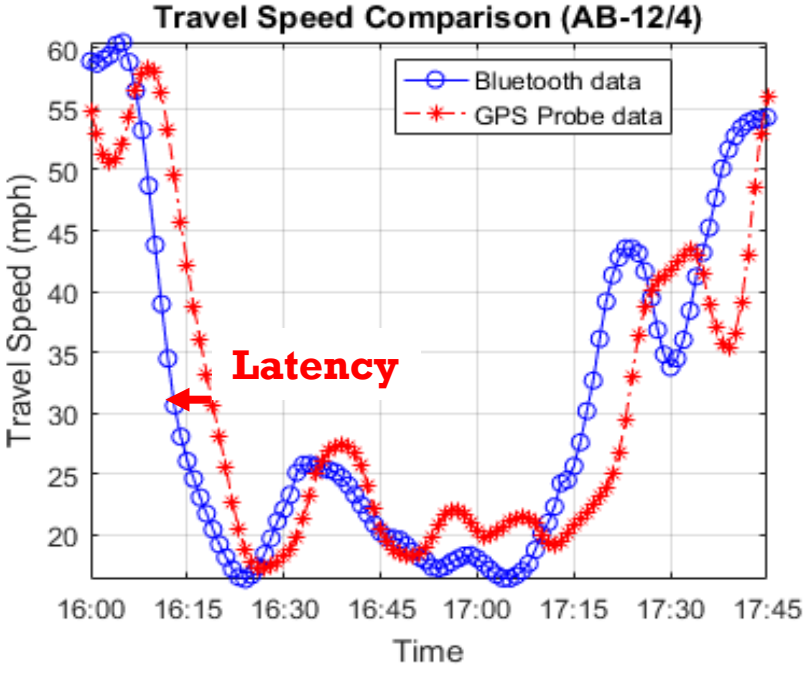
Test scenarios:
1. Morning & Afternoon Peaks;
2. Different TMC segments;
3. Slowdown & recovery.

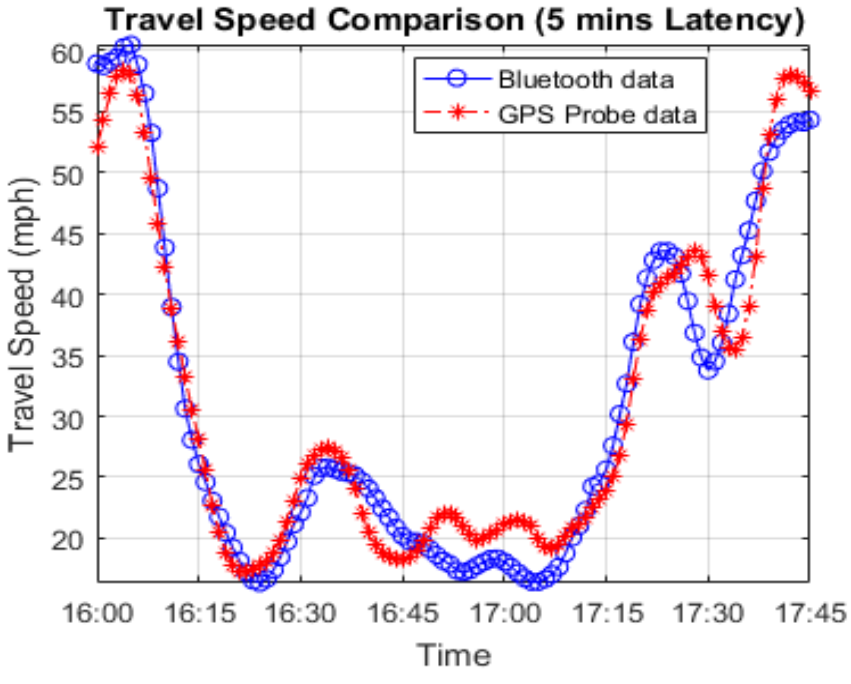



CASE STUDY

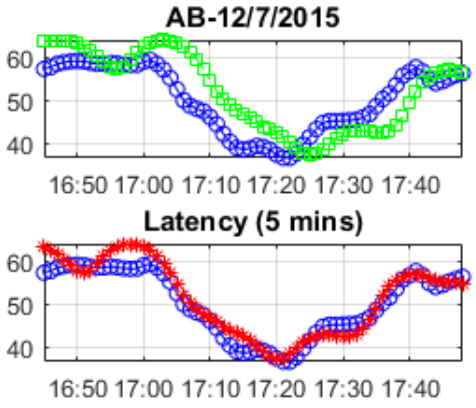
# Result
## Comparison

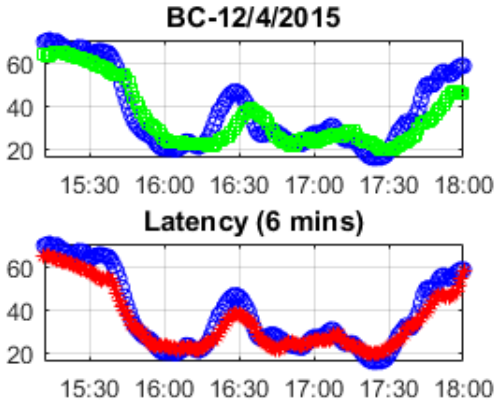


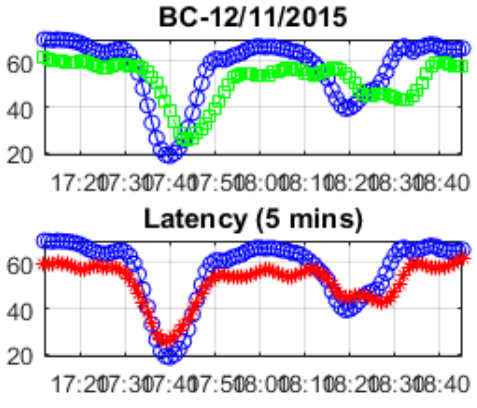



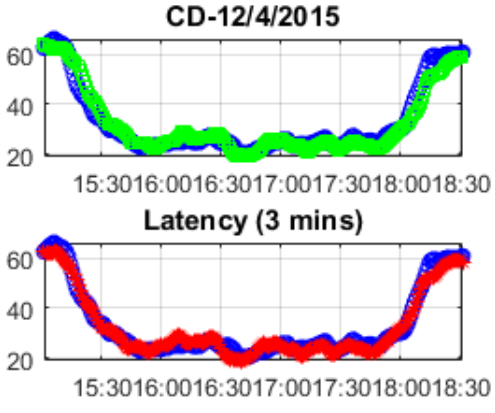
# Result

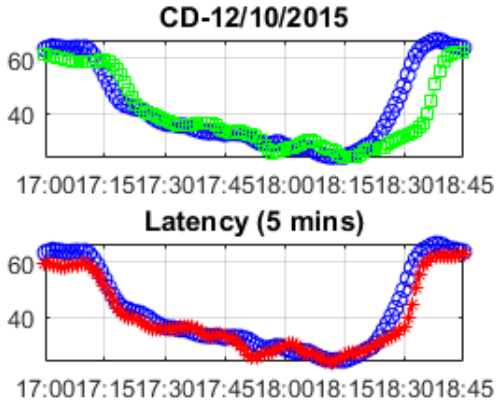
## Some Afternoon Peaks

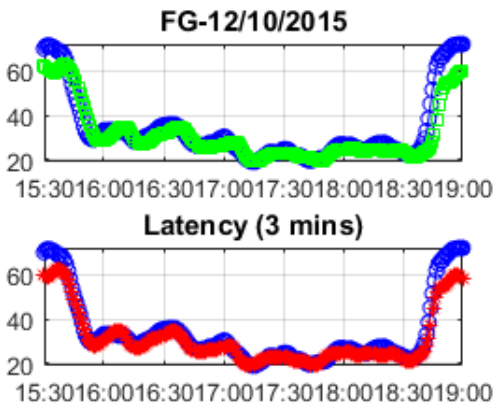



CASE STUDY

# Result

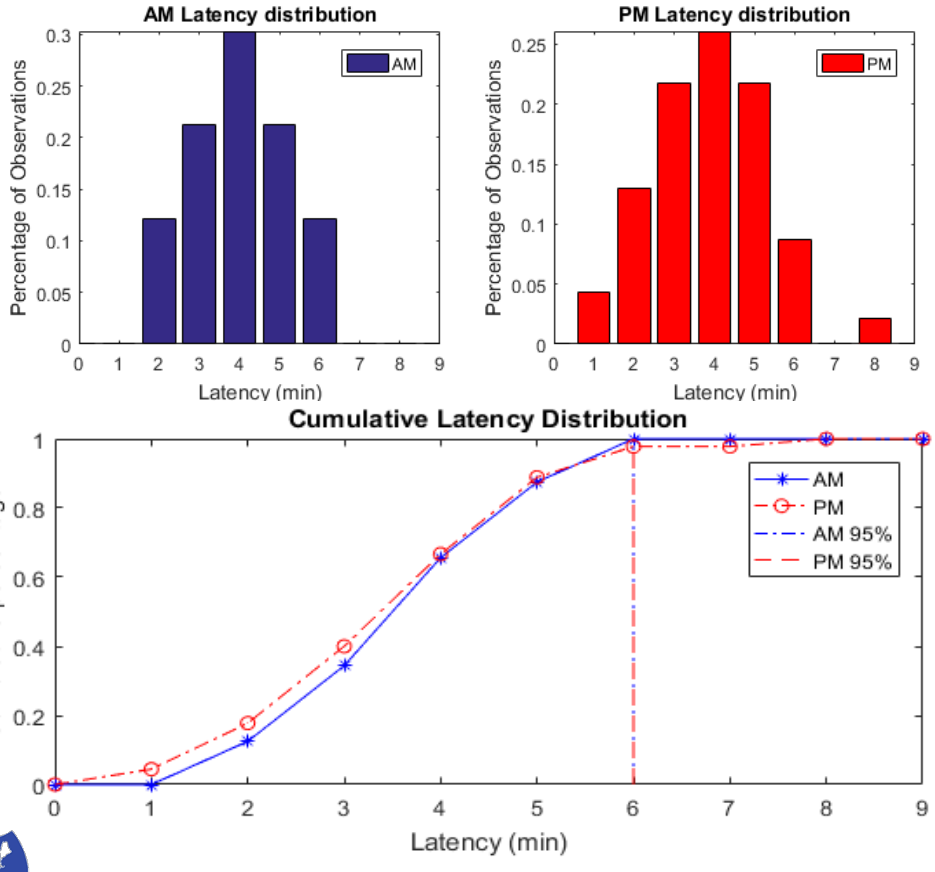
## Test One: Average Latency at Peak Periods

| Period | Number of Observations | Average Latency (minute) | | | |
|---|---|---|---|---|---|
| | | f1 (AVD) | f2 (SVD) | f3 (COR) | Average |
| Morning | 32 | 3.96 | 4.42 | 4.41 | 4.26 |
| Afternoon | 45 | 3.64 | 4.01 | 4.19 | 3.94 |

- Latency measured by three different fitness objectives “converged”;
- Latency in the morning peaks in slightly higher than that in the afternoon, but not significant;

# Result

## Test One: Average Latency at Peak Periods




- Similar distributions at morning and afternoon;
- 4 minutes latency has the highest probability/distribution;
- 95% of latency values fall within 6 minutes for both morning/afternoon peaks.

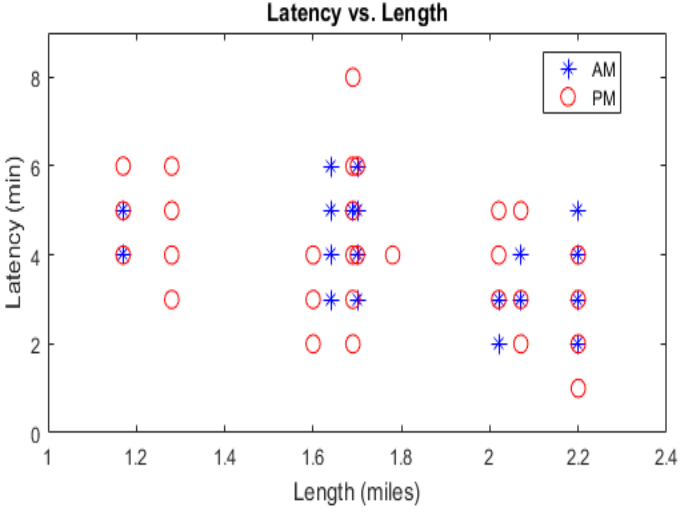
CASE STUDY

# Result

## Test Two: Average Latency at Different Segments

| Segment | Length (mile) | Average Latency (minute) | | | |
|---|---|---|---|---|---|
| | | f1 (AVD) | f2 (SVD) | f3 (COR) | Avg. |
| BC | 1.17 | 4.80 | 5.00 | 5.00 | 4.93 |
| KL | 1.28 | 4.43 | 4.86 | 5.00 | 4.76 |
| LM | 1.60 | 3.33 | 3.83 | 3.83 | 3.66 |
| OP | 1.64 | 4.67 | 5.00 | 5.00 | 4.89 |
| AB | 1.69 | 4.56 | 4.56 | 4.67 | 4.60 |
| PQ | 1.70 | 4.78 | 4.89 | 4.89 | 4.85 |
| MN | 1.78 | 4.00 | 4.18 | 3.95 | 4.04 |
| GH | 2.02 | 3.40 | 3.40 | 3.00 | 3.27 |
| CD | 2.07 | 3.92 | 4.50 | 4.50 | 4.31 |
| FG | 2.20 | 2.76 | 4.06 | 4.59 | 3.80 |
| Avg. for all segments | 1.72 | 4.06 | 4.43 | 4.44 | 4.31 |

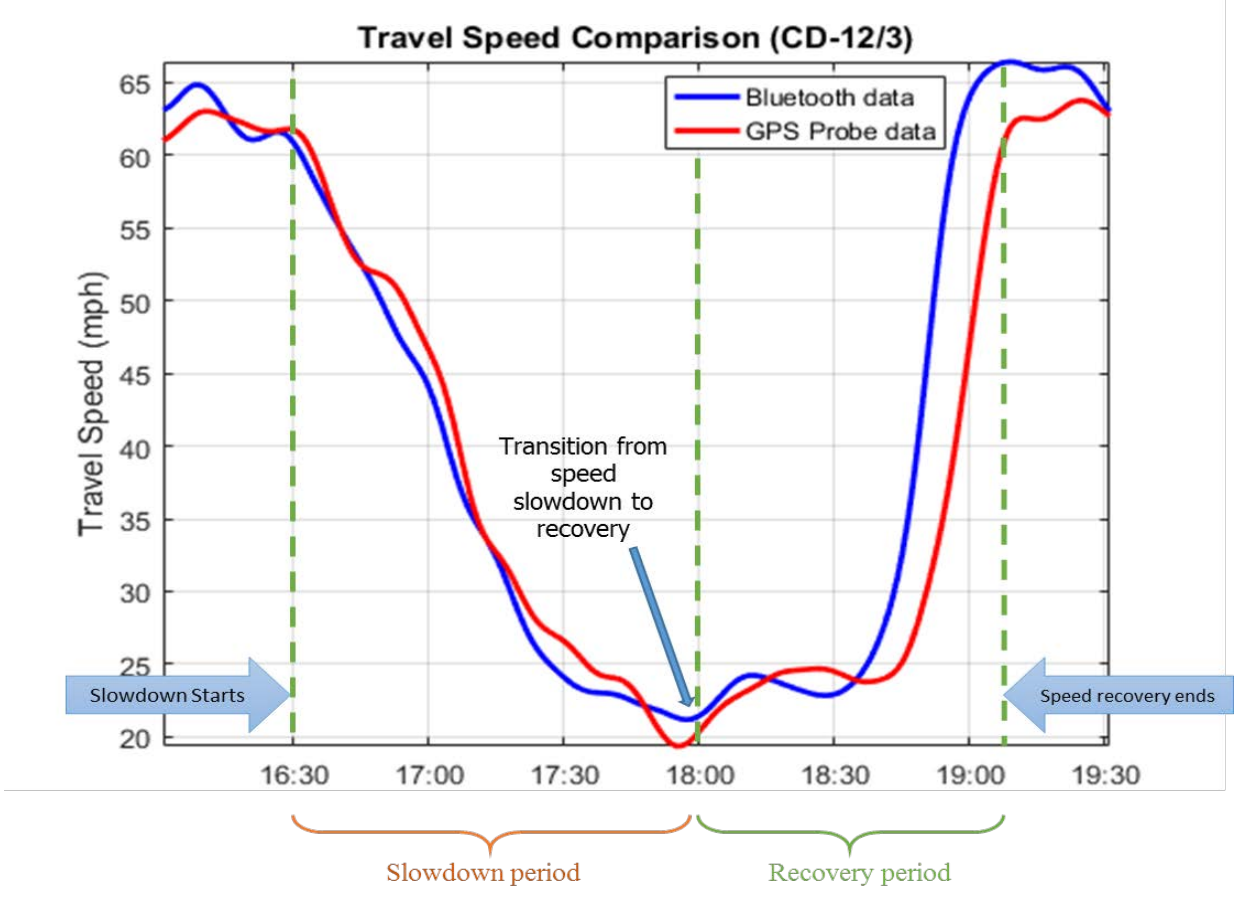
- Latency is not significantly correlated with the length of the segment;
- Latency is consistent with previous analysis.




CASE STUDY

# Result

## Test Three: Average Latency at Slowdown and Recovery




CASE STUDY

# Result

## Test Three: Average Latency at Slowdown and Recovery

| Time Period | Scenario | Number of Observations | Average Latency (minute) | | | |
|---|---|---|---|---|---|---|
| | | | f1 (AVD) | f2 (SVD) | f3 (COR) | Average |
| Morning | Slowdown | 32 | 3.55 | 3.60 | 3.90 | 3.68 |
| | Recovery | 32 | 4.76 | 5.15 | 4.45 | 4.83 |
| Afternoon | Slowdown | 45 | 3.43 | 3.45 | 3.75 | 3.54 |
| | Recovery | 45 | 4.70 | 4.94 | 4.62 | 4.76 |
| Overall | Slowdown | 77 | 3.48 | 3.51 | 3.81 | 3.60 |
| | Recovery | 77 | 4.72 | 5.03 | 4.55 | 4.79 |

**Significant reduction in traffic speed slowdown seems to be reflected in probe data with 25% less latency compared to the recovery from slowdowns.**

# Conclusion

- Analyze latency associated with GPS probe data
- Propose an iterative methodology to quantify the latency
- Conduct case study on two freeway segments at South Carolina (average latency is around 4 mins)

- **Latency:**
  - is slightly higher at morning peaks than afternoon peaks
  - has no significant difference at different segments
  - is smaller at slowdown than recovery

# Future Research

- Investigate the impact of smoothing
- Pattern matching algorithms applied to arterials
- Larger dataset from multiple probe data vendors
- Test other potential influential attributes
- If exists consistent latency under similar condition (off-line)

# Thank you
# Q & A


**Masoud Hamedi**
Senior Research Scientist
Center for Advanced Transportation Technology
University of Maryland
5000 College Ave
College Park, MD 20740
Email: masoud@umd.edu | Phone:301-405-2350

**Zhongxiang Wang**
Graduate Student
Department of Civil & Environmental Engineering
University of Maryland
Email: zxwang25@umd.edu

**Stanley Young**
Advanced Transportation and Urban Scientist
National Renewable Energy Laboratory
Email: stanley.young@nrel.gov