I-95 Corridor Coalition's Vehicle Probe Project

A cross-vendor and cross-state analysis of the GPS-probe data latency

Transportation Research Board
97th Annual Meeting

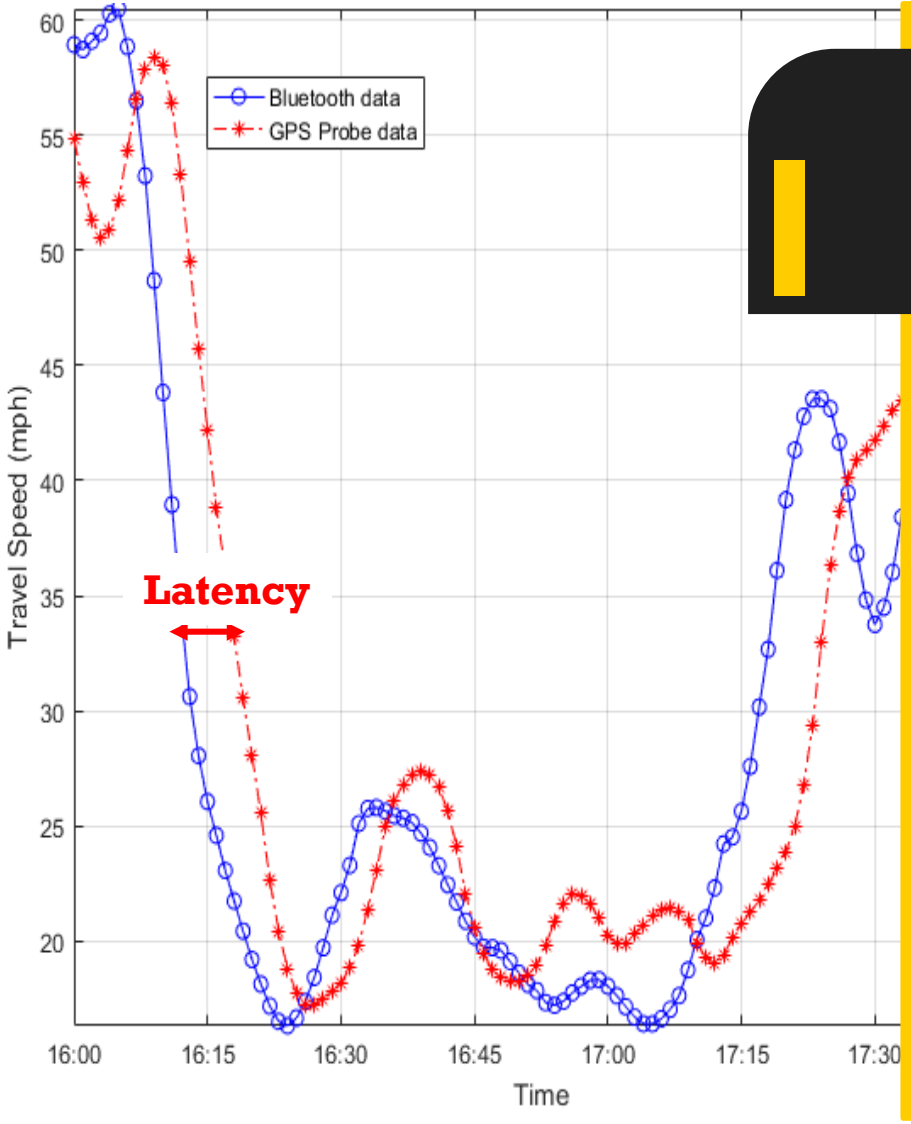
January 7–11, 2018 ▪ Washington, D.C.

I-95 CORRIDOR COALITION

CATTWORKS

Masoud Hamedi* | Zhongxiang Wang
Elham Sharifi | Stanley Young

Jan 8th 2018

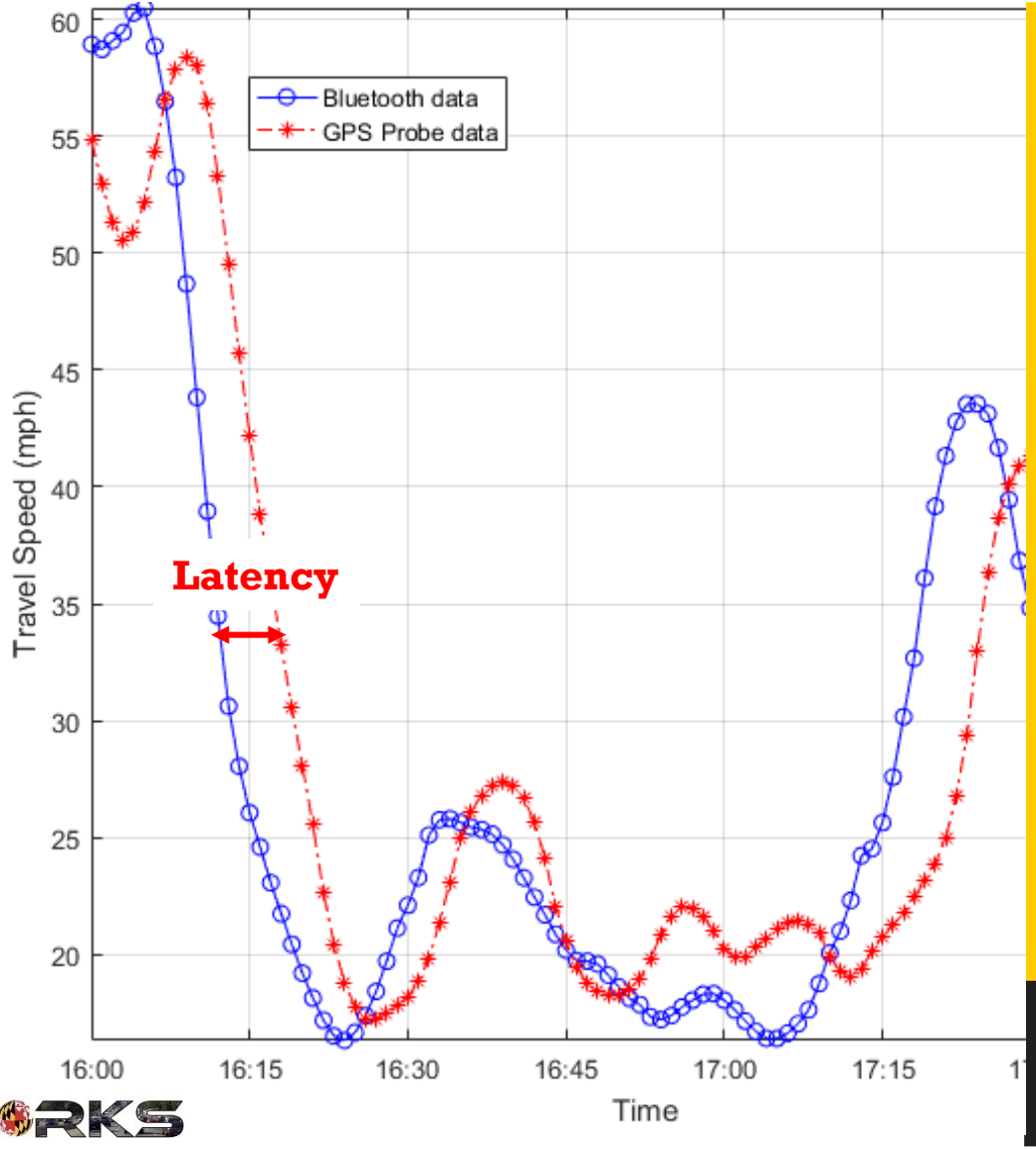


OUTLINE

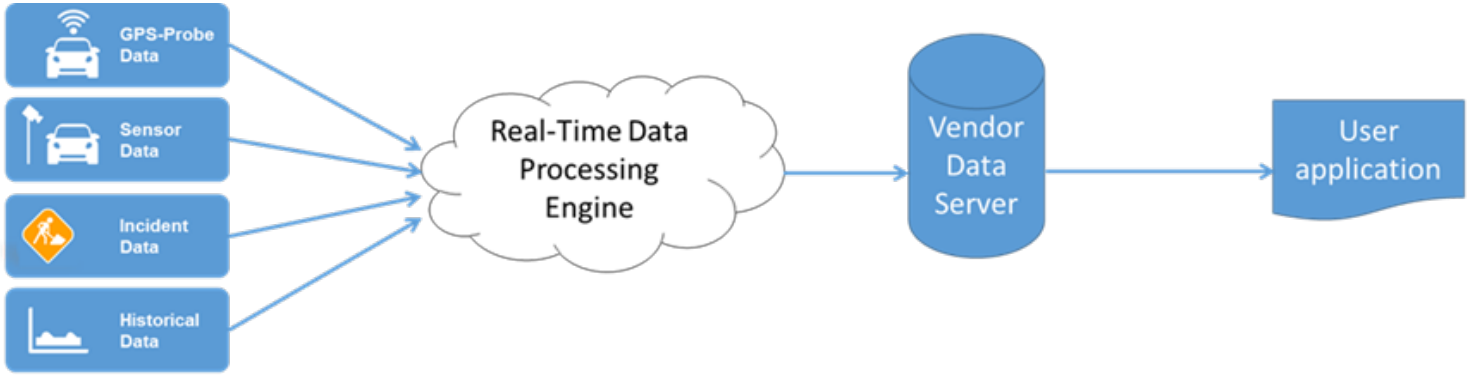
1. INTRODUCTION
2. LITERATURE REVIEW
3. METHODOLOGY
4. CASE STUDY
5. CONCLUSIONS

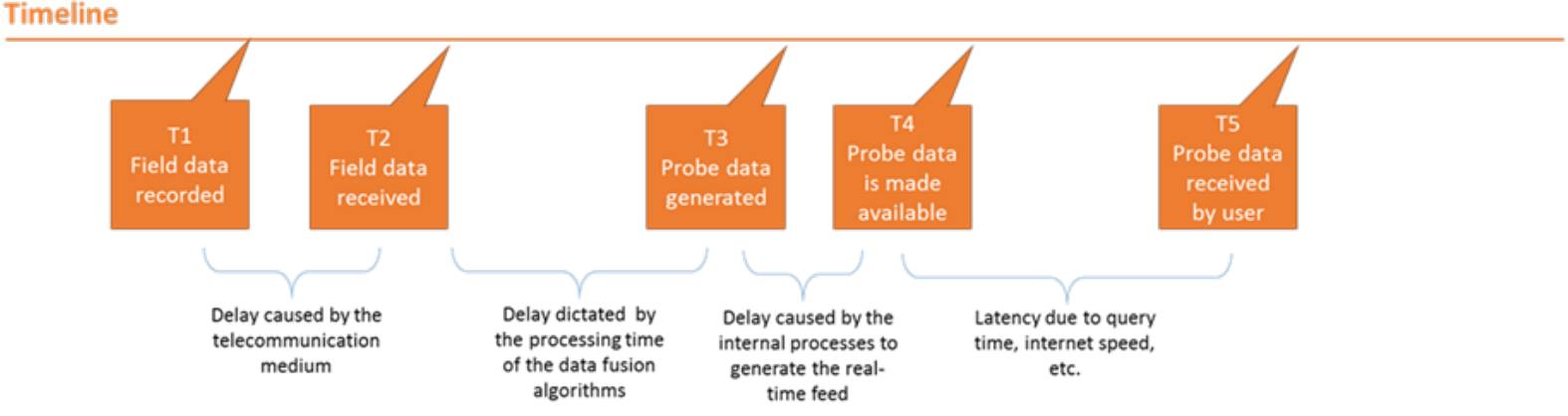
What is Latency?

- **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.



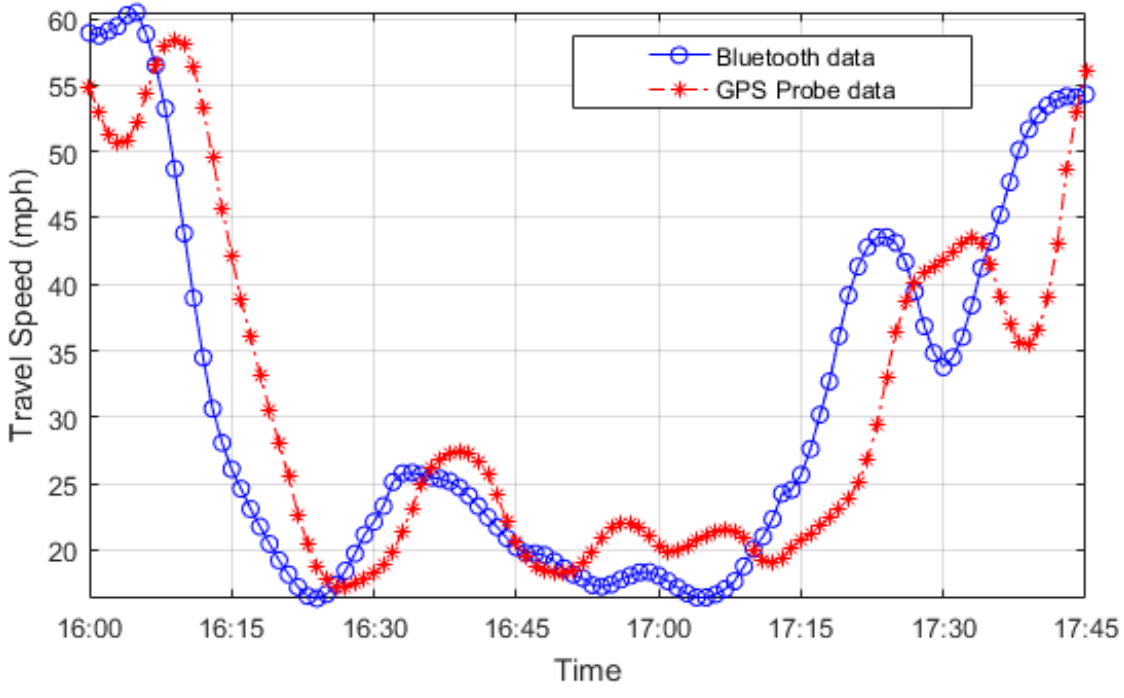
Where does the latency come from?





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.”



Related 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

Related Research

- 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.

- Wang, Z., Hamedi, M. and Young, S., 2017. A Methodology for Calculating Latency of GPS Probe Data. Transportation Research Record: Journal of the Transportation Research Board, No. 2645. DOI: 10.3141/2645-09.
  - Propose methodology to measure GPS-probe data latency in comparing to Bluetooth data. It is shown to be effective, but only on a limited dataset for one GPS-probe data vendor.

Contributions

- ✓ **Applying the latency measurement methodology (Wang et. al 2017) to a larger dataset**
  - ✓ Expanding to three states
  - ✓ Data from three major probe data vendors
- ✓ **Developing a robust algorithm for automating slowdown episode detection**
- ✓ **Revisiting conclusions of the previous research based on the extended data set;**

- ✓ **Analyzing latency for different vendors to describe:**
  - ➢ Latency distribution for different vendors;
  - ➢ Latency distribution at slowdown and recovery periods;
  - ➢ Latency distribution at different times of the day;
  - ➢ Latency distribution for segments of various lengths;
  - ➢ Statistical comparison of latency across three vendors.

Data Processing

Bluetooth/WiFi data preparation

- Matching and filtering high resolution re-identification based travel time observations to generate segment space mean speed data;
- Aggregating reference data in one minute intervals.

Data Filtering

The following observations are identified and discarded:

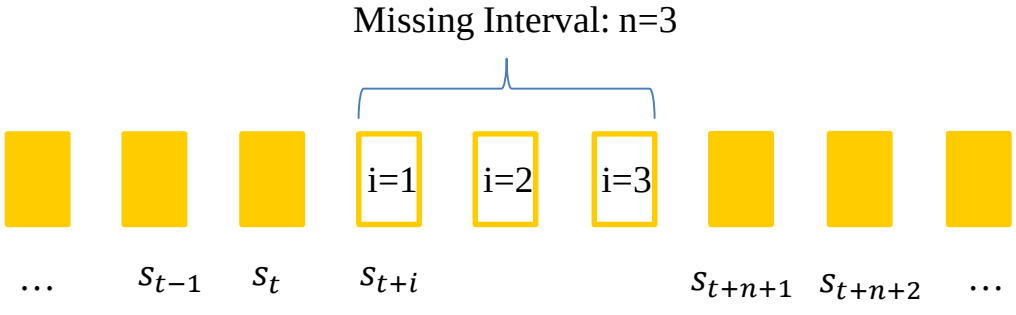
- Observations with unreasonably low or high speeds;
- Observations in a particular time interval that are far from the average of the rest of the speeds observed in the same time interval.

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).

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



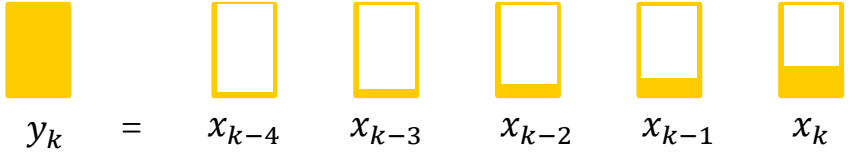
Data Processing

Data Smoothing

❖ Weighted moving average function

Arithmetic growth of the weights with the previous five time intervals

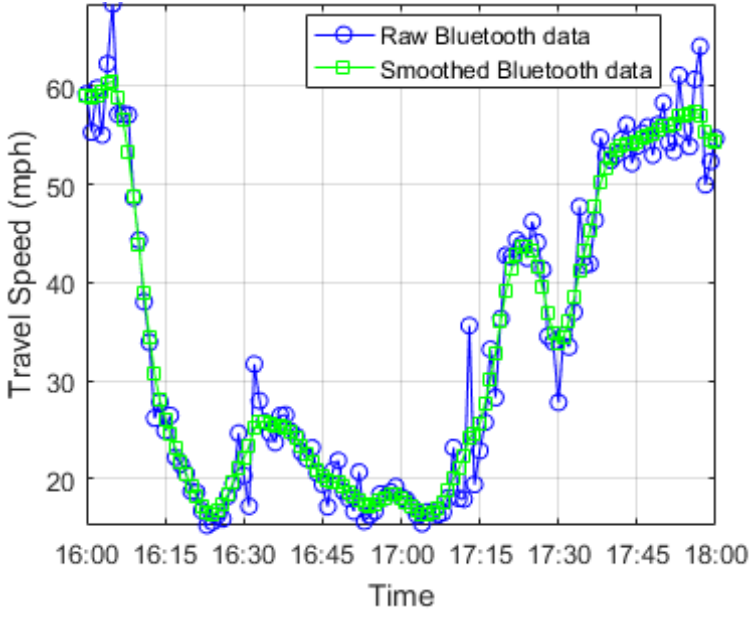
$$y_k = 0.33x_k + 0.27x_{k-1} + 0.20x_{k-2} + 0.13x_{k-3} + 0.07x_{k-4}$$



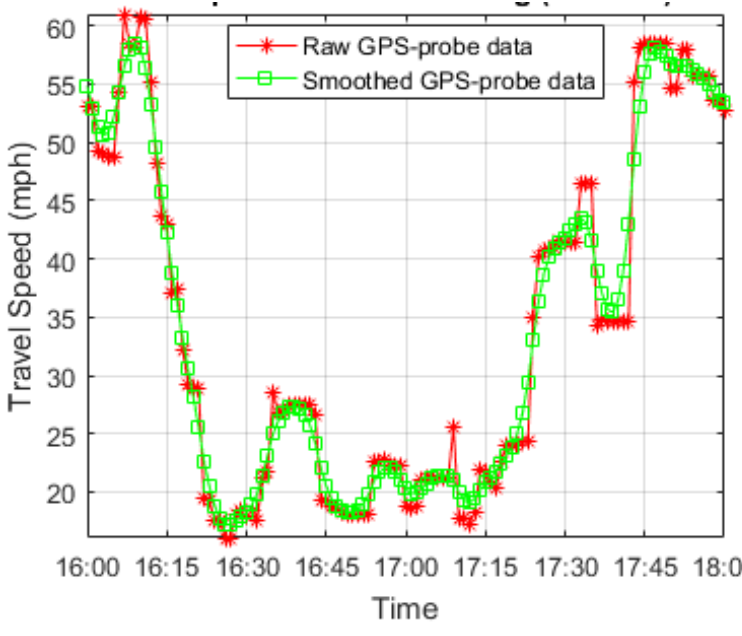
Why Zero-phase digital filtering? Smoothing may introduce undesired horizontal shift into the time series, however in *Zero-phase digital filtering,* forward shift is followed by a reverse shift, so the artificial shift is compensated.

Data Processing

Data Smoothing



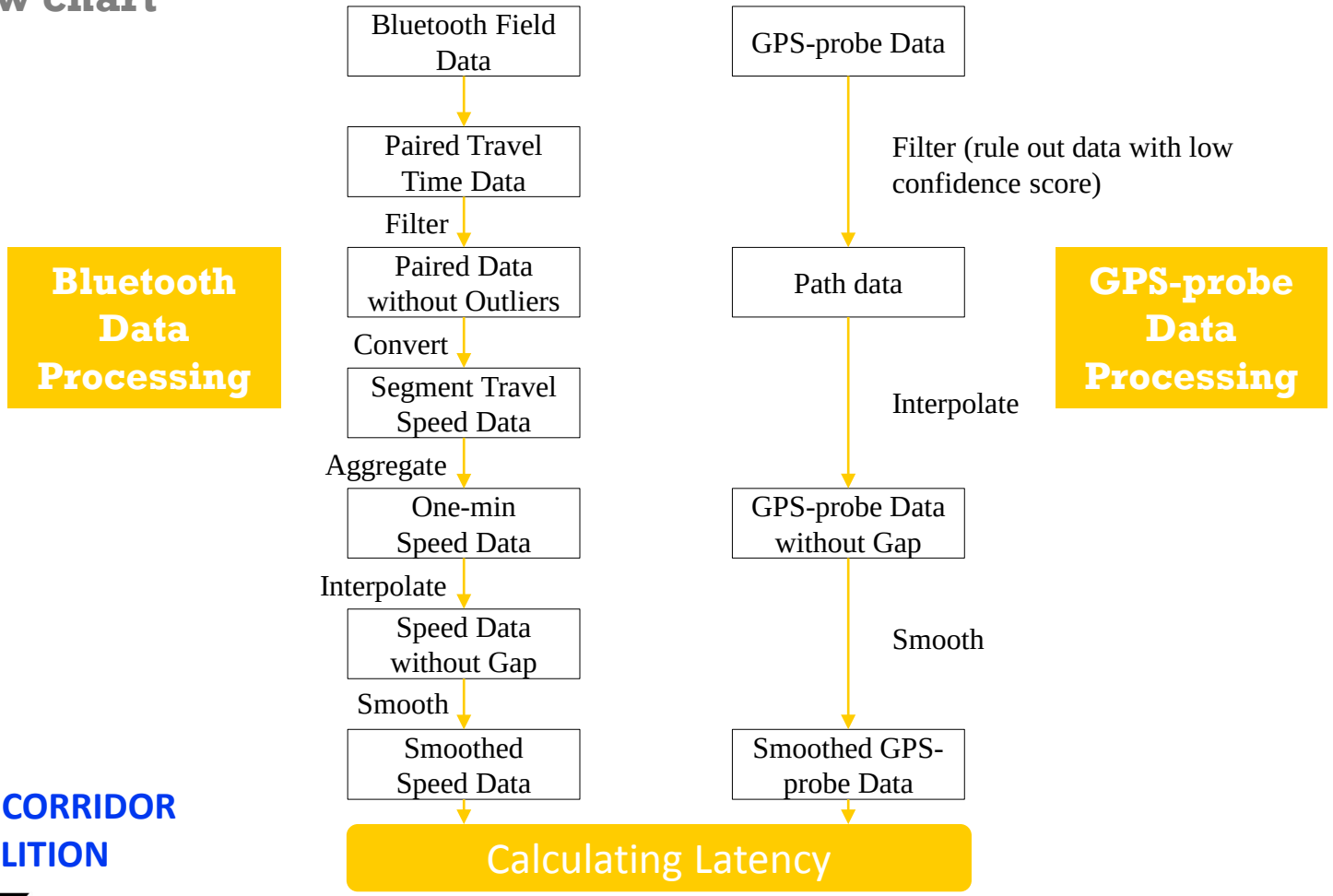
- ✓ **Less noise**
- ✓ **No shift**



METHODOLOGY

Data Processing

Flow chart



Objective

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

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

- (AVD) Absolute vertical distance between two curves

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

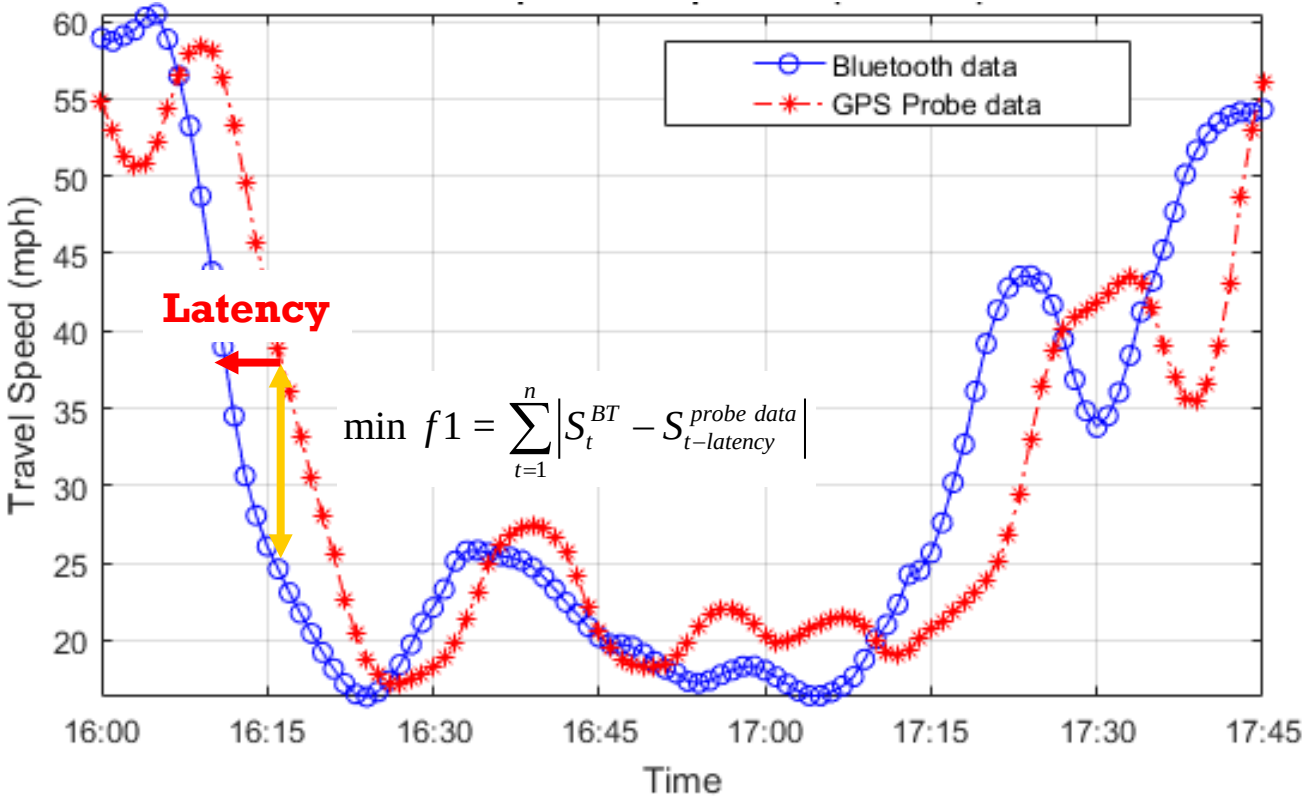
- (SVD) Squared vertical distance, which gives more weights to the points that have bigger difference

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

- (COR) Statistical representation of the linear relationship between two curves

$$lb \le latency \le ub$$

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



Data Selection

State	Road	Start Point	End Point	Length (mile)	Date
South Carolina	I-85	US-276/Exit 48	SC-14/Exit 56	7.15	12/03/2015 - 12/17/2015
	I-26	Bush River Rd/Exit 108	Harbison Blvd/Exit 103	4.47	
New Hampshire	I-89	I-93	Stickney Hill Rd/Exit 3	3.54	07/10/2016 - 07/24/2016
	I-93	I-393/US-202/US-4/Exit 15	Hackett Hill Rd/Exit 11	15.82	
	I-93	NH-28/Rockingham Rd/Exit 5	NH-102/Nashua Rd/Exit 4	3.63	
North Carolina	I-240	US-70/Charlotte St/Exit 5B	US-23/US-19/Exit 3	2.23	12/16/2016 - 12/26/2016
	I-40	NC-191/Exit 47	US-23/US-19/Exit 44	2.56	
	I-26	I-40/ Exit 46 A/US 74	I-26/Exit 37	14.43	

CASE STUDY

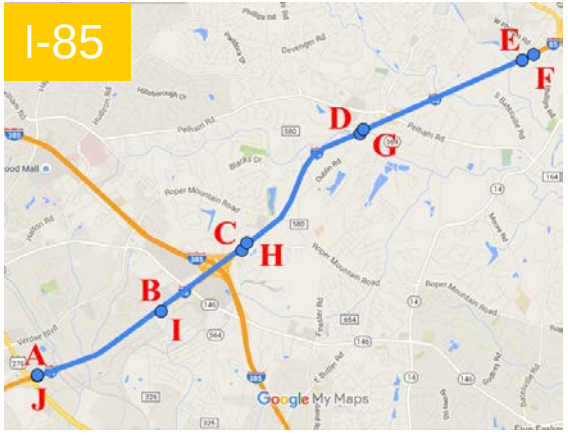
Data Selection

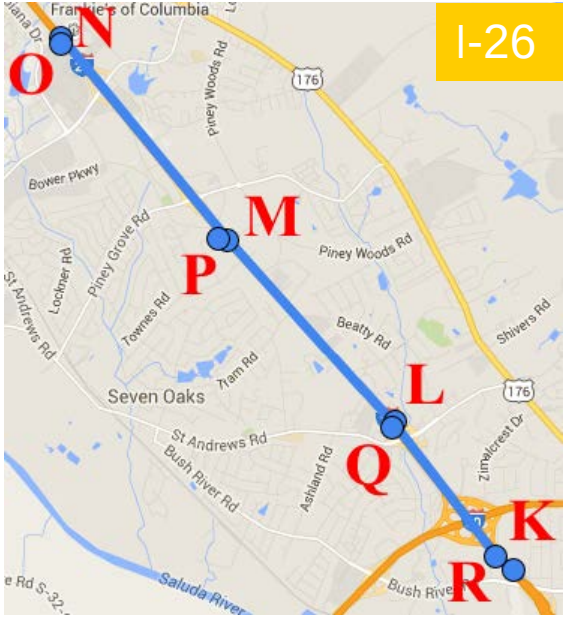
Road type: Freeway

Location: South Carolina, New Hampshire, North Carolina

Direction: both directions
Time: 2 weeks for each state
Segment length: 1~3.5 mile

South Carolina





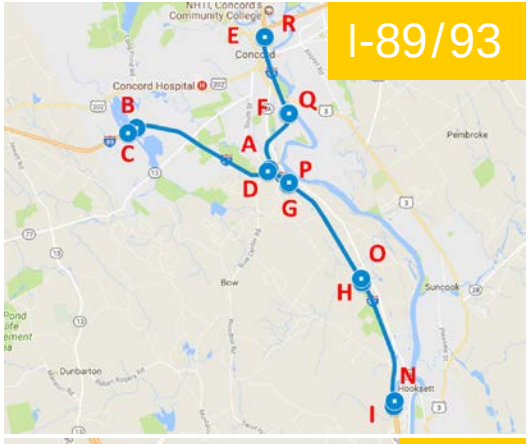
CASE STUDY

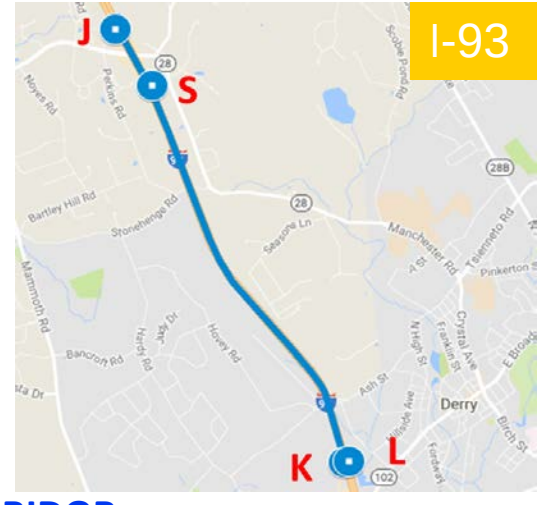
I-95 CORRIDOR COALITION

CATTWORKS

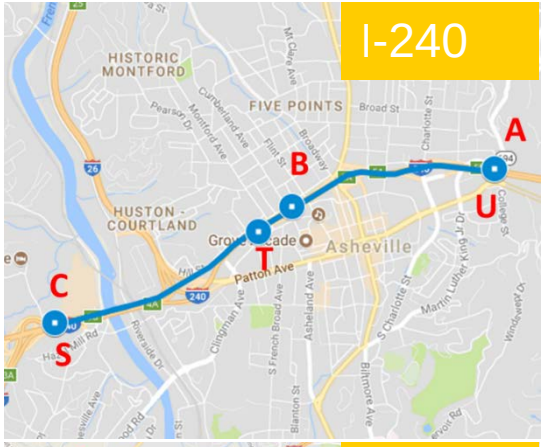
Data Selection

New Hampshire





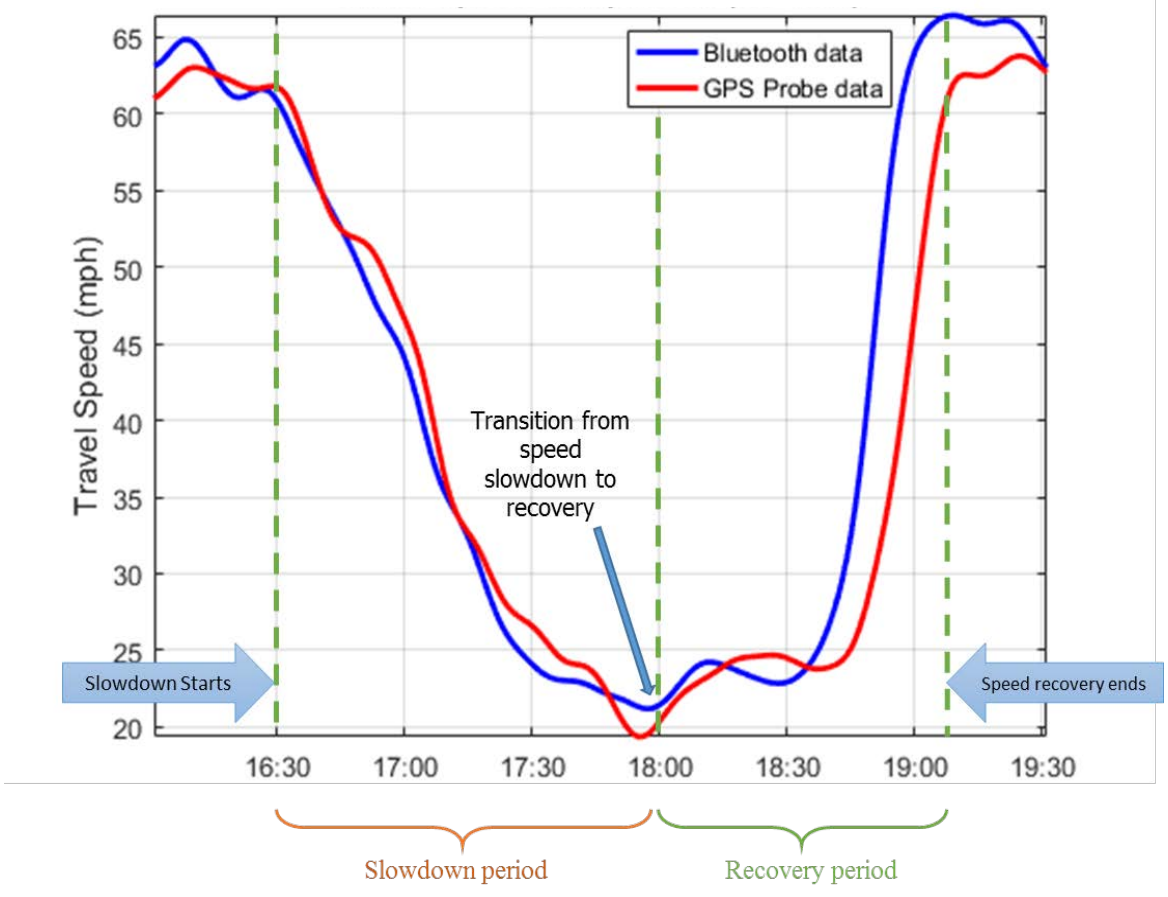
North Carolina



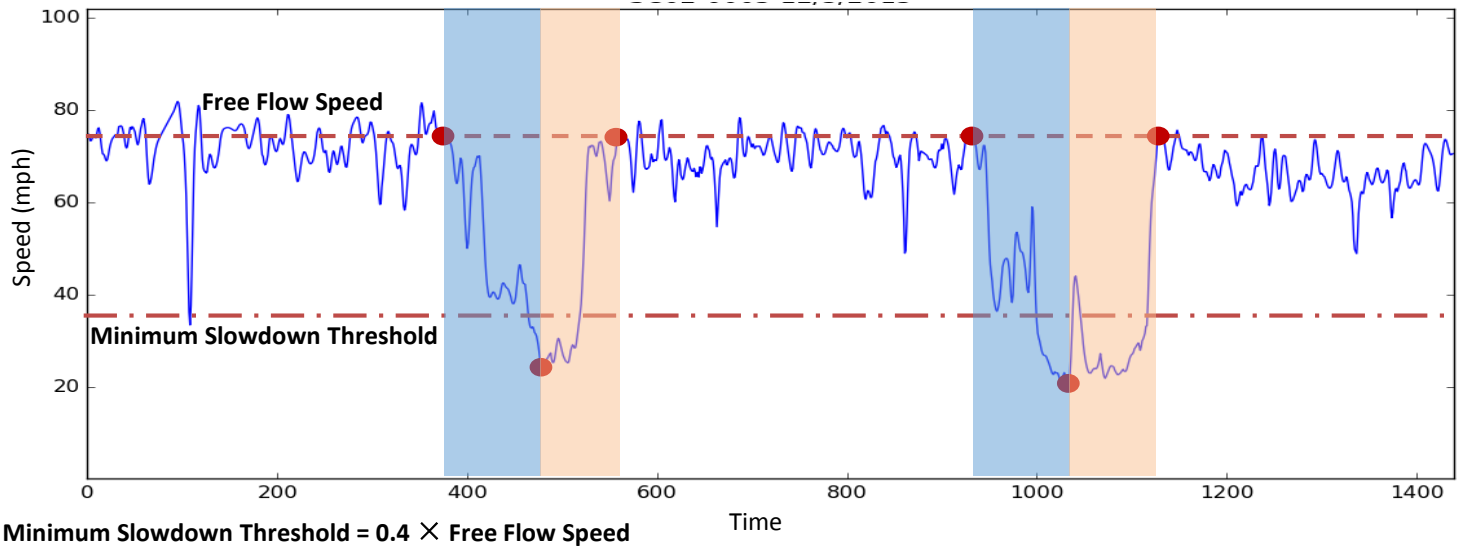
CASE STUDY

Data Preparation

Capturing Slowdown and Recovery Episodes



Data Preparation: Automated Speed Pattern Recognition Filter (ASPRF)



Minimum Slowdown Threshold = 0.4 × Free Flow Speed

State	# of paths	# of days	# of Slowdown and Recovery Episodes
South Carolina	14	15	72
New Hampshire	12	15	26
North Carolina	15	11	18
Total	41	41	116

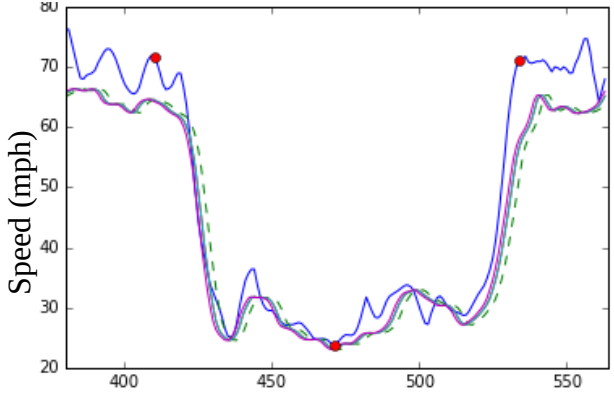
Sample Result

South Carolina

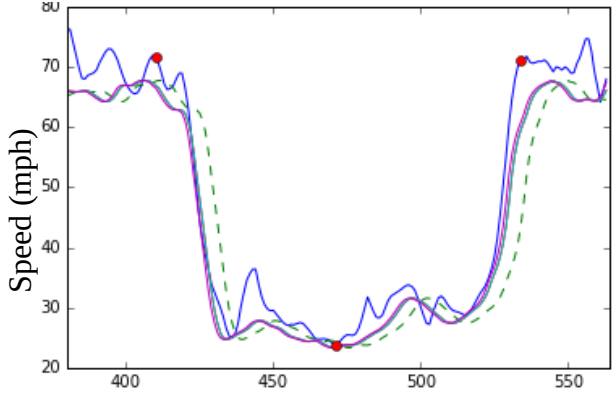
12/10/2015

Length: 2.3 miles

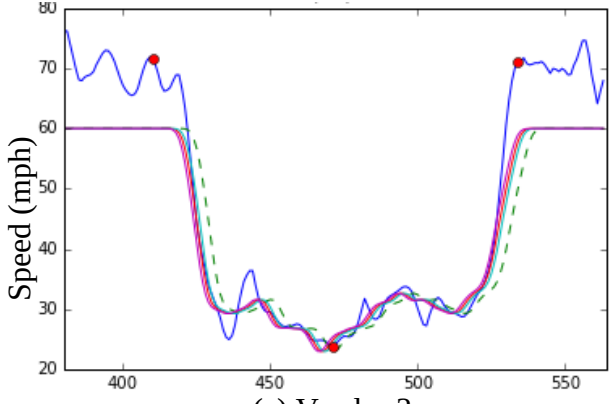
Data: 3 vendors



(a) Vendor 1



(b) Vendor 2



(c) Vendor 3

Legend

REF	Reference (Bluetooth)
ORI	Original (GPS-probe data)
AVD	Absolute Vertical Distance
SVD	Squared Vertical Distance
COR	Correlation coefficient

CASE STUDY

Results

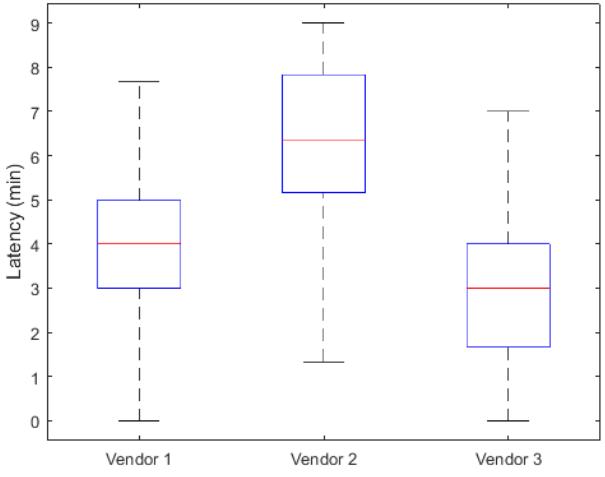
Average Latency for all segments over all periods (minute)

		Vendor 1				Vendor 2				Vendor 3			
		AVD	SVD	COR	Mean	AVD	SVD	COR	Mean	AVD	SVD	COR	Mean
SC	SRE	3.9	4.4	4.6	**4.3**	6.1	6.2	6.2	**6.2**	2.9	3.1	3.1	**3.0**
	SP	3.6	3.6	4.0	**3.7**	6.7	6.8	6.6	**6.7**	4.4	4.3	3.8	**4.2**
	RP	4.7	5.3	4.0	**4.7**	5.0	5.2	5.0	**5.1**	2.0	2.4	2.5	**2.3**
NH	SRE	3.3	3.4	3.6	**3.4**	5.6	6.0	6.4	**6.0**	2.3	2.5	2.7	**2.5**
	SP	3.3	3.2	3.2	**3.2**	7.0	7.0	6.7	**6.9**	3.5	3.0	3.2	**3.2**
	RP	4.0	4.2	3.9	**4.0**	3.6	3.7	4.7	**4.0**	2.9	3.5	2.6	**3.0**
NC	SRE	3.5	4.1	4.1	**3.9**	7.3	7.8	7.3	**7.5**	3.2	3.2	2.7	**3.0**
	SP	4.6	4.6	3.6	**4.2**	7.7	7.8	7.4	**7.6**	3.2	2.9	2.5	**2.9**
	RP	3.4	3.7	4.1	**3.7**	5.9	6.5	6.8	**6.4**	3.2	3.1	3.8	**3.4**

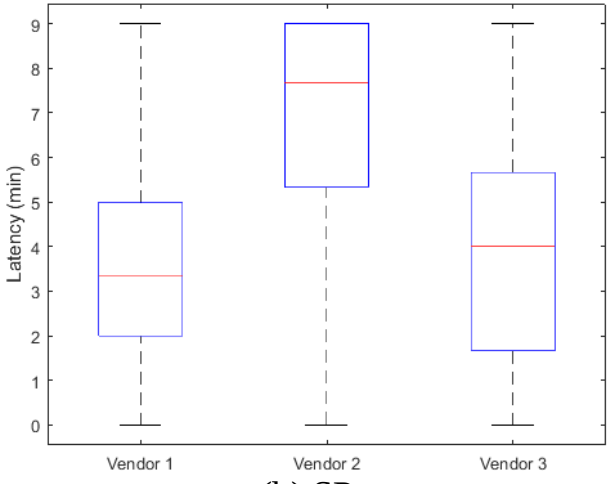
SC: South Carolina, **NH**: New Hampshire, **NC**: North Carolina; **AVD**: Absolute vertical distance; **SVD**: Squared vertical distance; **COR**: correlation coefficient; **SRE**: Slowdown and recovery episode, **SP**: Slowdown period, **RP**: recovery period

Results

Latency comparison between speed slowdown and recovery periods

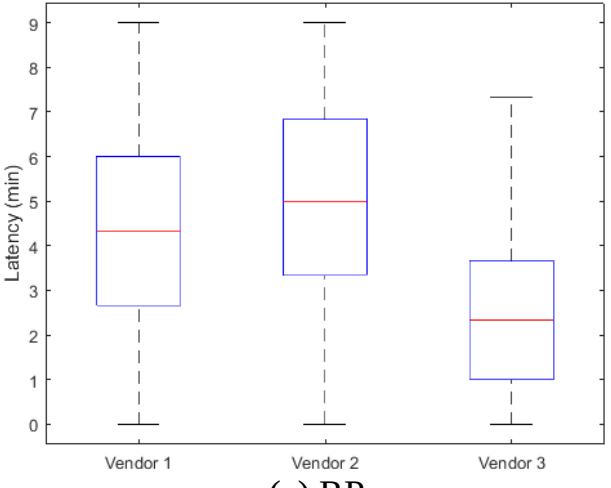


(a) SRE



(b) SP

- ✓ Comparison results vary from vendor to vendor.
- ✓ Latency is different during slowdown and recovery.

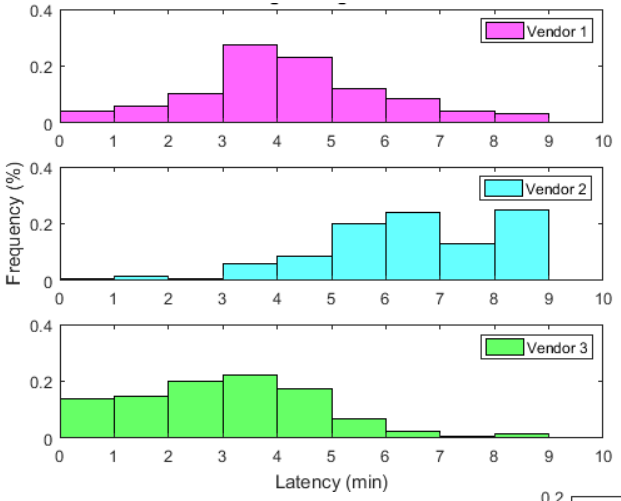


(c) RP

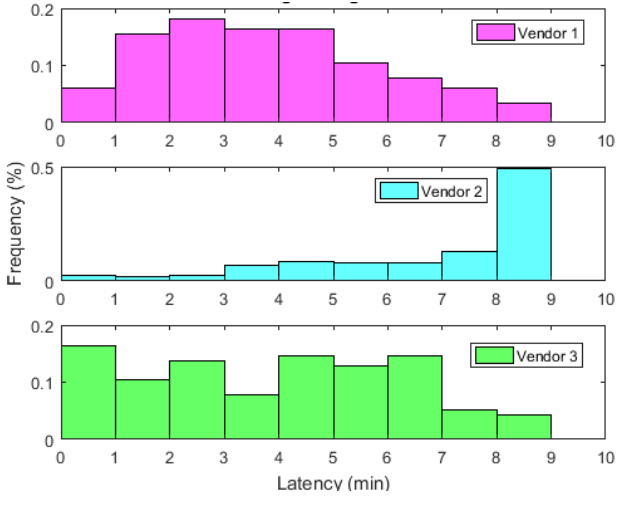
Note: whiskers represent the 1.5 interquartile range of the lower and upper quartiles, corresponding to approximately +/–2.7 times standard deviation and 99.3 percent coverage if the latency is normally distributed.

Results

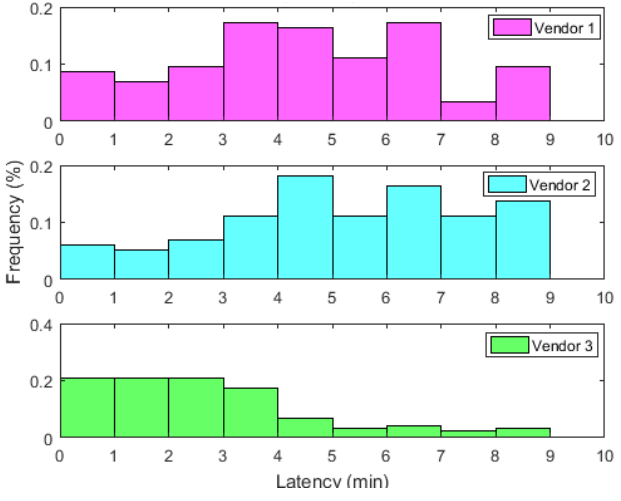
Latency histogram



(a) SRE



(b) SP



(c) RP

CASE STUDY

Result

Is difference in probe data latency statistically significant between pair of vendors?

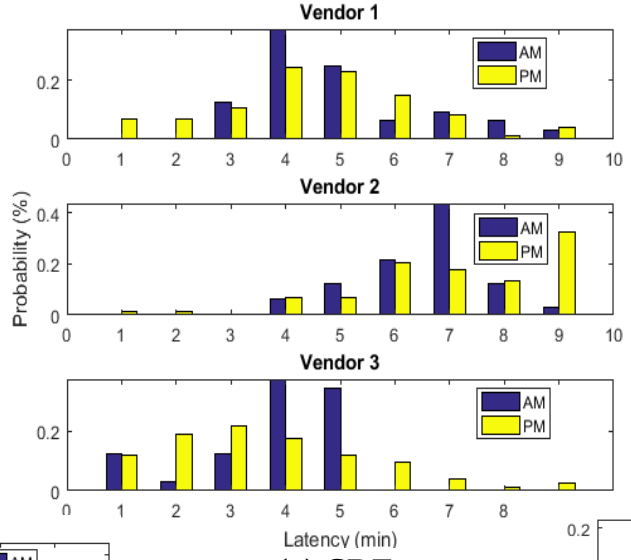
P-value	t-stat			Chi-square		
	Vendor 1	Vendor 2	Vendor 3	Vendor 1	Vendor 2	Vendor 3
Vendor 1		-9.370	4.729		2.685	0.598
Vendor 2	0.000		14.185	0.975		7.454
Vendor 3	0.000	0.000		1.000	0.590	

- ✓ Statistically speaking, average latency among vendors is different.
- ✓ The latency distribution is statistically independent among vendors.

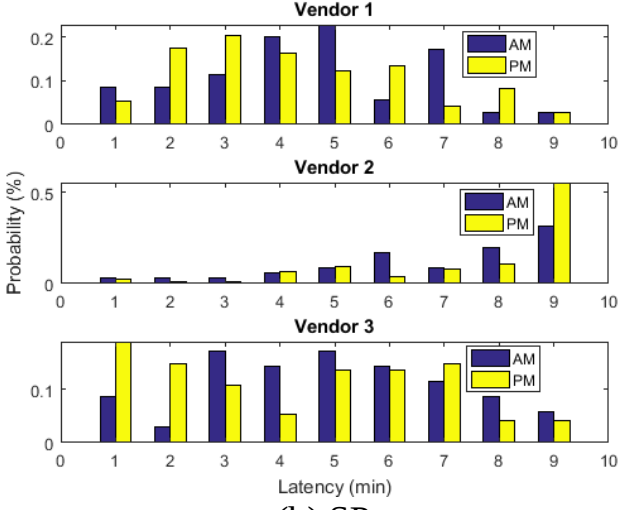
CASE STUDY

Result

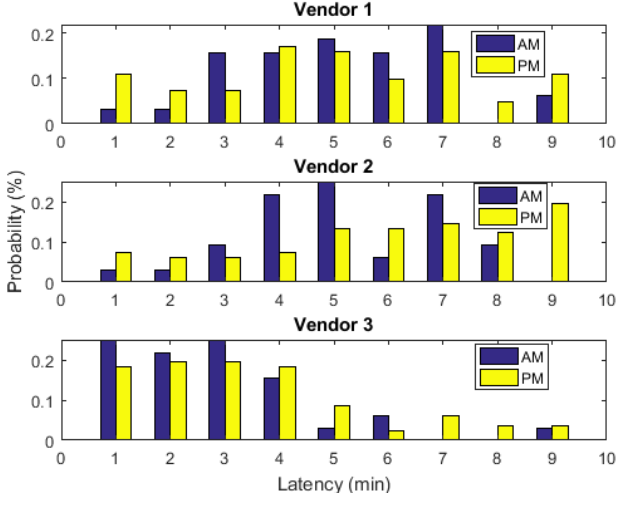
Time of the day



(a) SRE



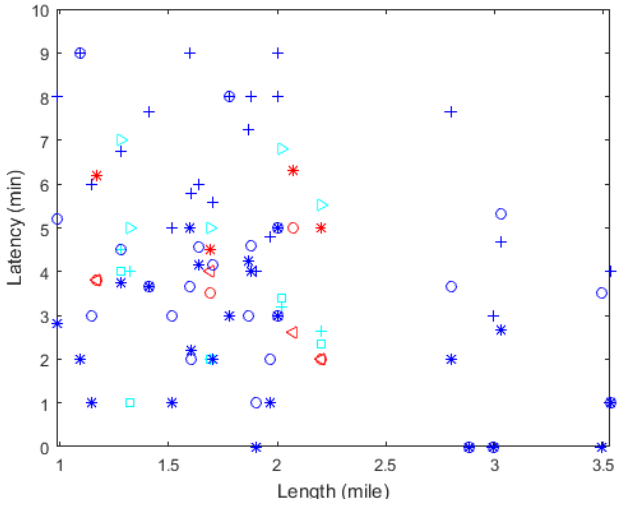
(b) SP



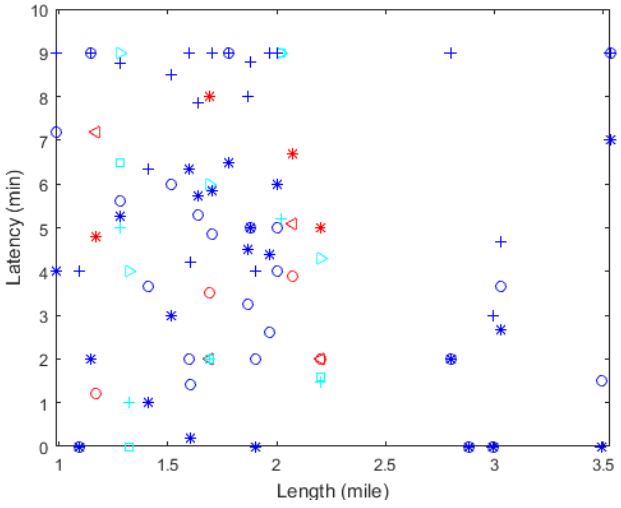
(c) RP

CASE STUDY

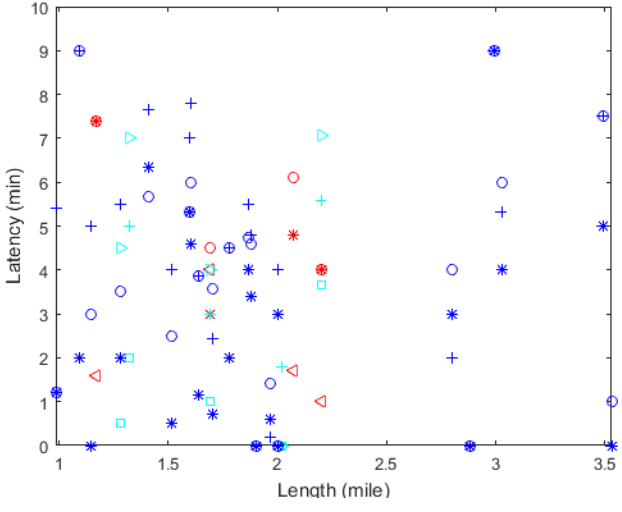
Result
Segment length



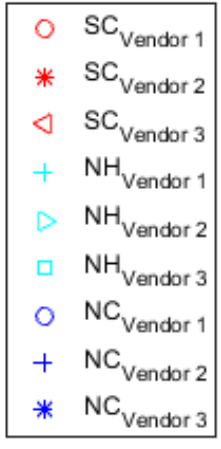
(a) SRE



(b) SP



(c) RP



- ✓ There is no strong relationship between the segment length and latency for all vendors.

CASE STUDY

ConclusionS

- Across all vendors and for all tested freeway segments, average latency is found to be 4.4 minutes with standard deviation of 2.3 minutes;
- Average latency is statistically different among vendors;
- No strong correlation between latency and time of the day, and also latency and segment length were found.
- Instead of interpreting the latency as a single number, distribution of latency should be measured and evaluated

Future Research

- Investigating the latency of GPS-probe data on signalized arterials.
- Studying potential impact of traffic volume on latency of probe data.
- Keep analyzing more cases in future to study long time trend of latency improvement for different vendors
- Understanding implications of the latency and its distribution on real-time applications and exploring solutions to compensate for latency

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

Elham Sharifi
Faculty Research Assistant
Center for Advanced Transportation Technology
University of Maryland
Email: esharifi@umd.edu

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

I-95 CORRIDOR COALITION

CATTWORKS