How much GPS data do we need?

Anthony Patire
Matthew Wright
Boris Prodhomme
Alexandre Bayen

Motivation

- **Emergence of mobile internet and proliferation of smartphones**
- **Location data increasingly collected by private companies**
- **Huge opportunities for transportation organizations**
  - **Performance measures**
  - **Incident response**
  - **Situational awareness**
  - **Coordination and management**

CALIFORNIA PATH

Hybrid Data Project Background

- **Focus on** *probe data*
  - **commercially available, unaggregated, GPS-based, point-speed data**
  - **Fields include: ID, lat, long, timestamp, speed**
- **Understand practical challenges**
  - **Procurement**
  - **Real-time feeds**
  - **Filtering**
  - **Map-matching**
  - **Fusing with other data types**
- **Assess benefits**

CALIFORNIA PATH

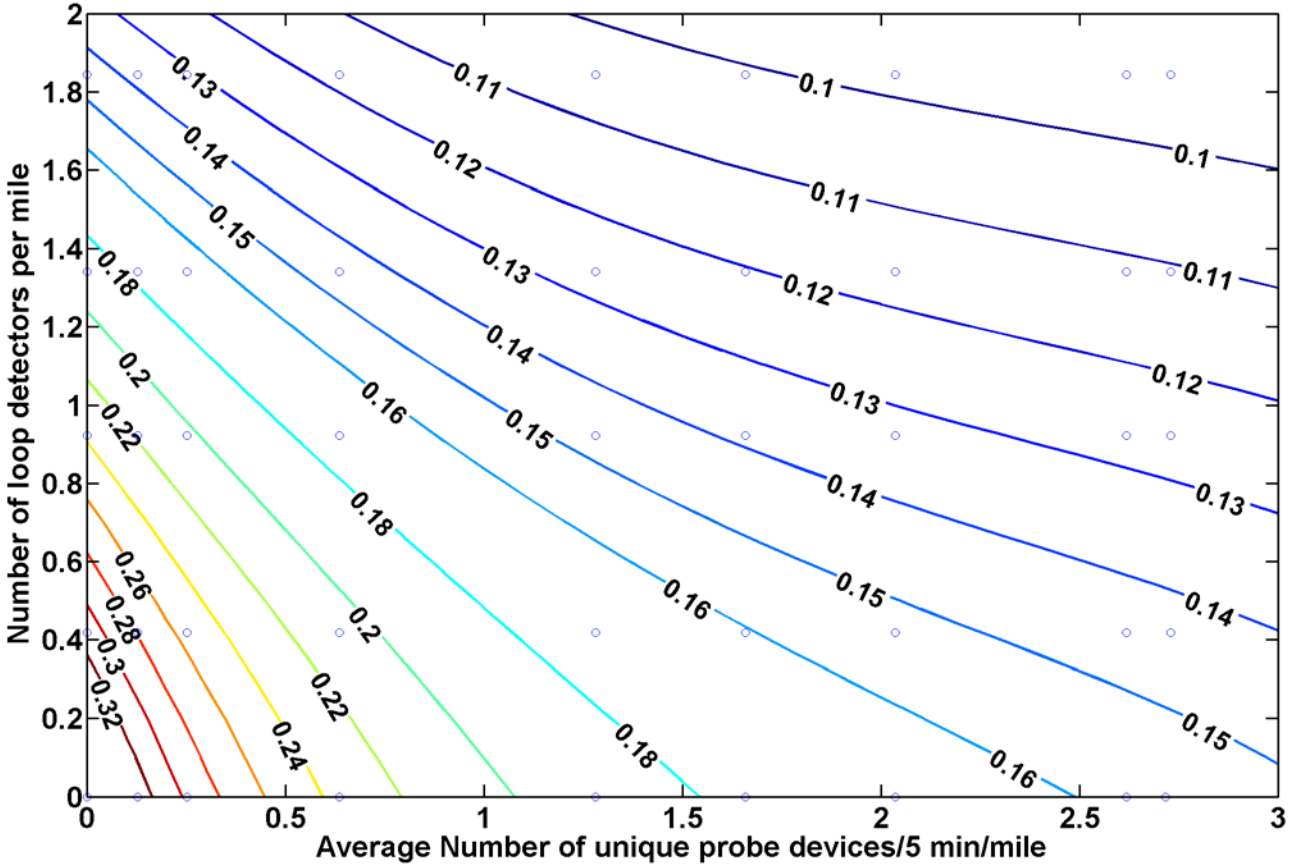
Outline

- Introduction
- One answer: how much data we need
- Overview of data and method
- Results
- Probe data quality
- Conclusion

Outline

- Introduction
- **One answer: how much data we need**
- Overview of data and method
- Results
- Probe data quality
- Conclusion

Travel Time MAPE for combinations of loop and probe data



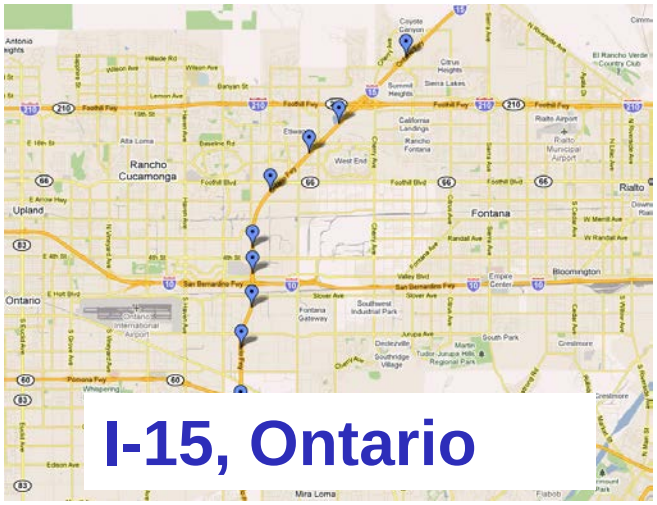
I-880 NB congested periods between 3/2 and 3/17, 2012

Outline

- Introduction
- One answer: how much data we need
- **Overview of data and method**
- Results
- Probe data quality
- Conclusion

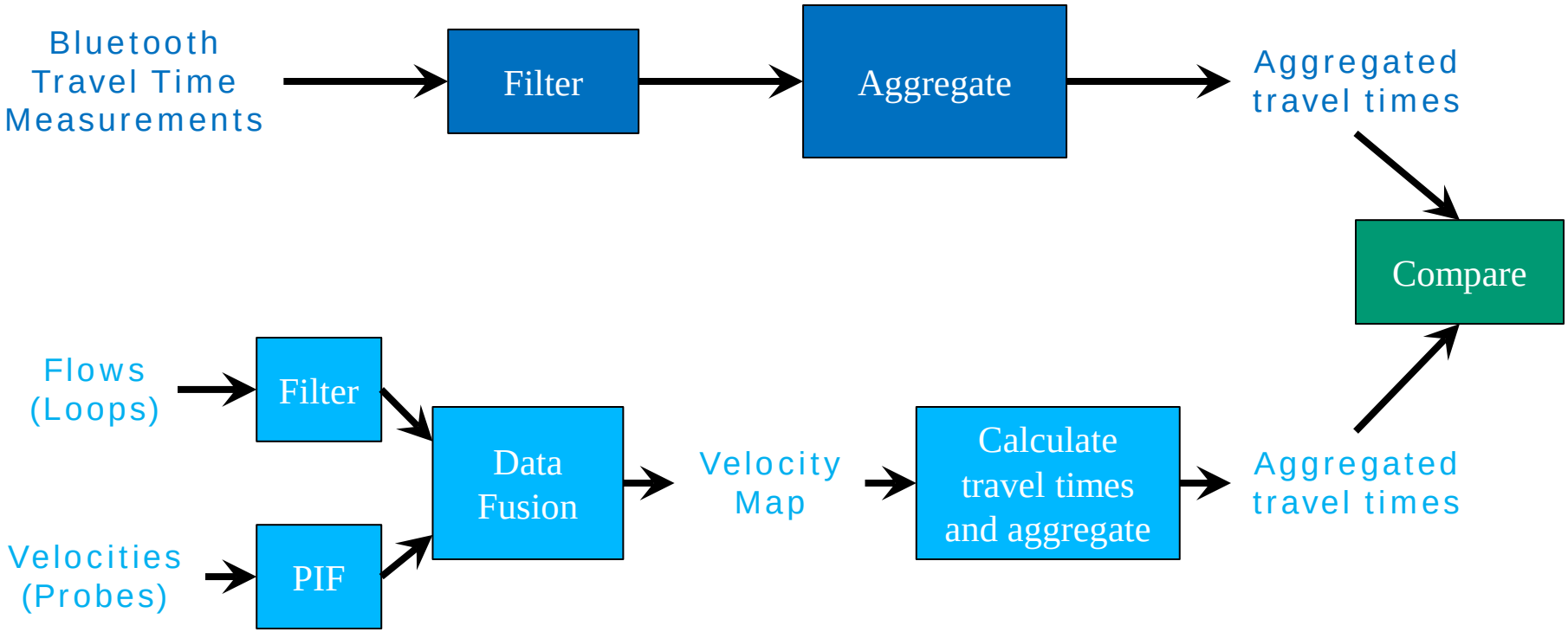
Overview of Data

- **Bluetooth**
  - Three two-week deployments of ten detection units
  - Independent measure of travel times
- **Probe data**
  - Two vendors
  - March and April of 2012
- **Loop data from PeMS**



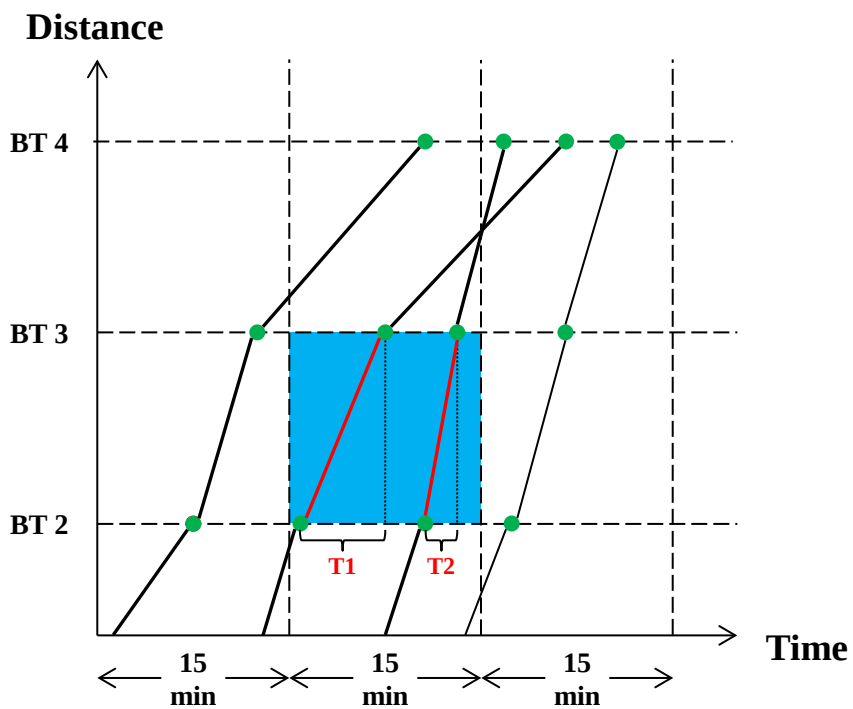
CALIFORNIA PATH

Overview of Method



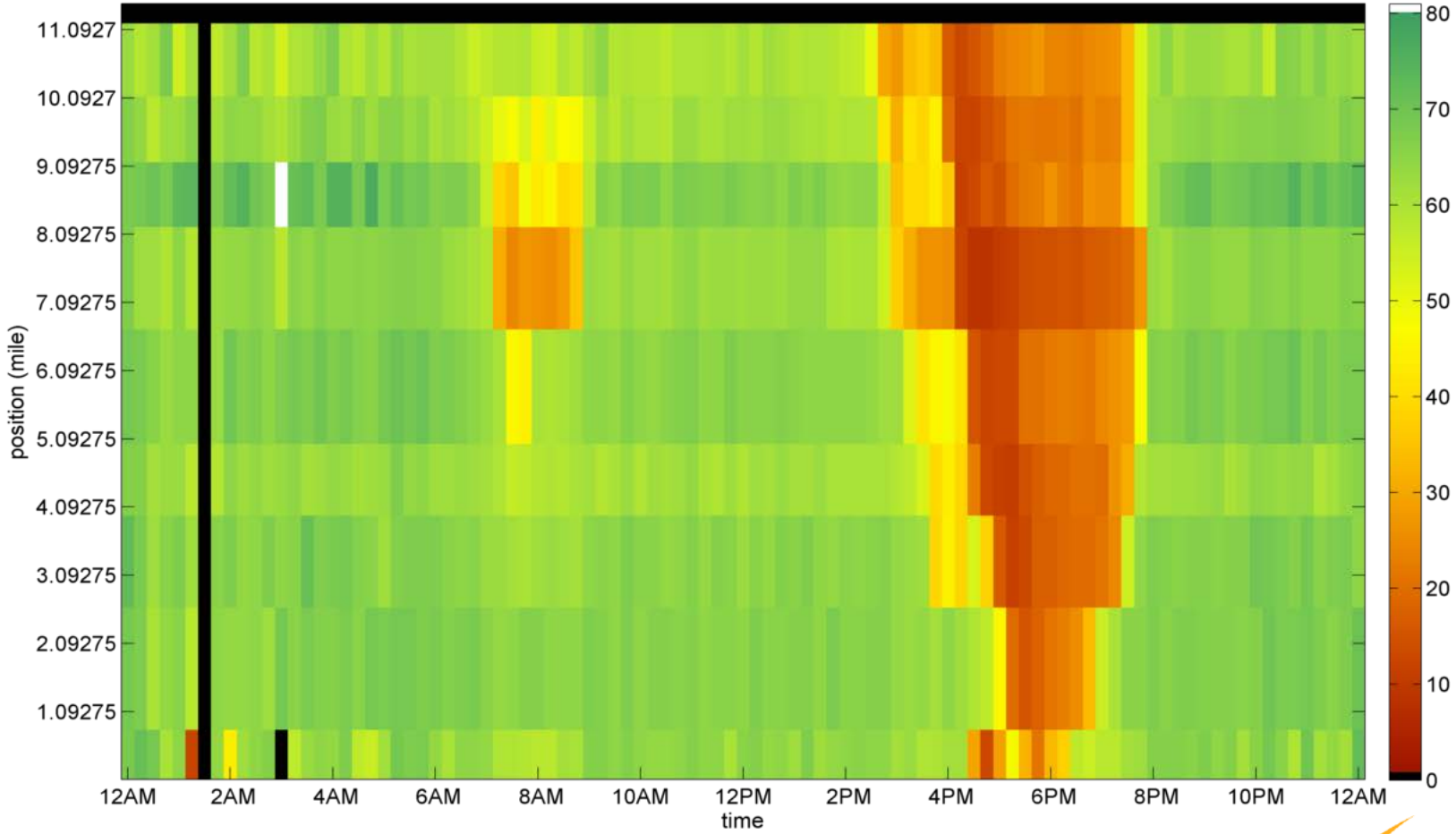
Aggregated Travel Times from Bluetooth

BT Measurements → Filter → Aggregate → Aggregated travel times



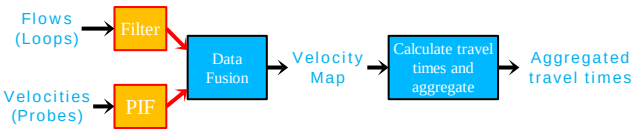
- For each 15-minute period, travel times between successive Bluetooth sensors are measured:
- $Ave\ TT = \frac{T1+T2}{2}$
- Average travel times taken as ground truth and compared with outputs of data fusion
- For visualization purposes we calculate average velocities:
- $Ave\ Velocity = \frac{section\ length}{Ave\ TT}$

Bluetooth Average Velocities

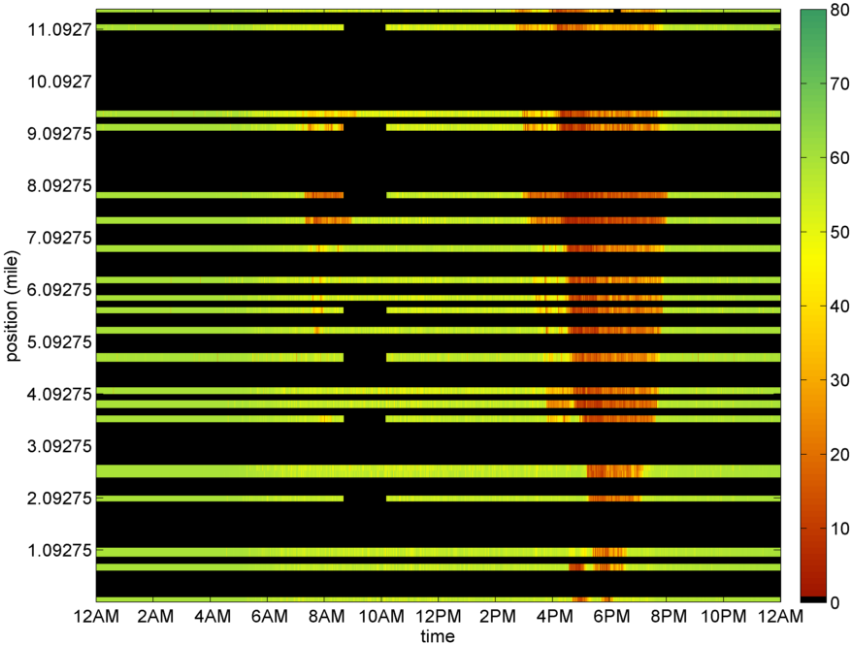


880 Northbound – Wed. March 7, 2012

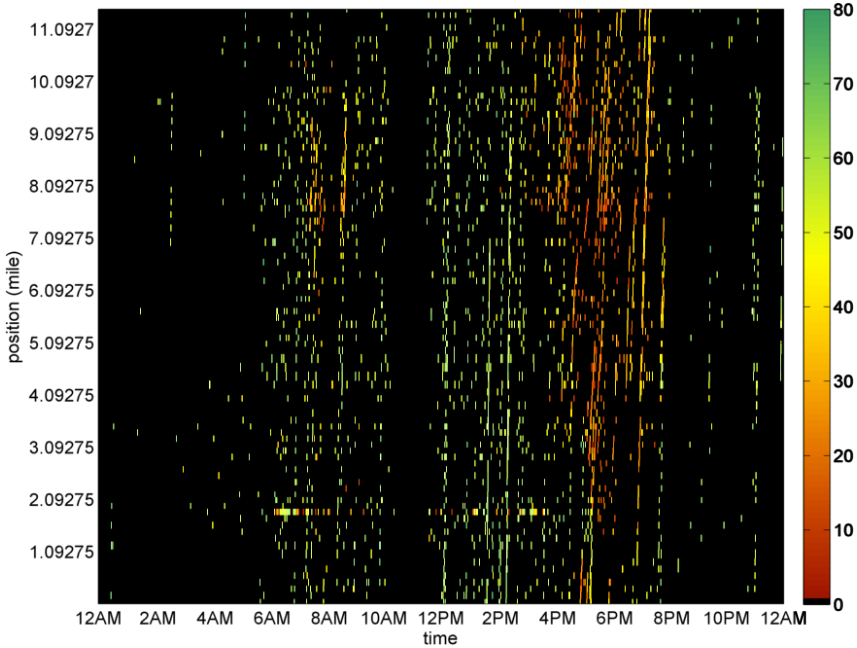
Filtered and Map-Matched Data



Filtered Loop Data



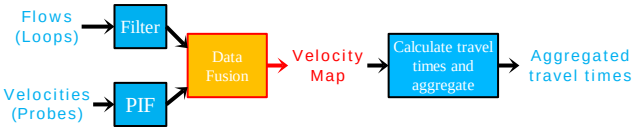
Map-matched Probe Data



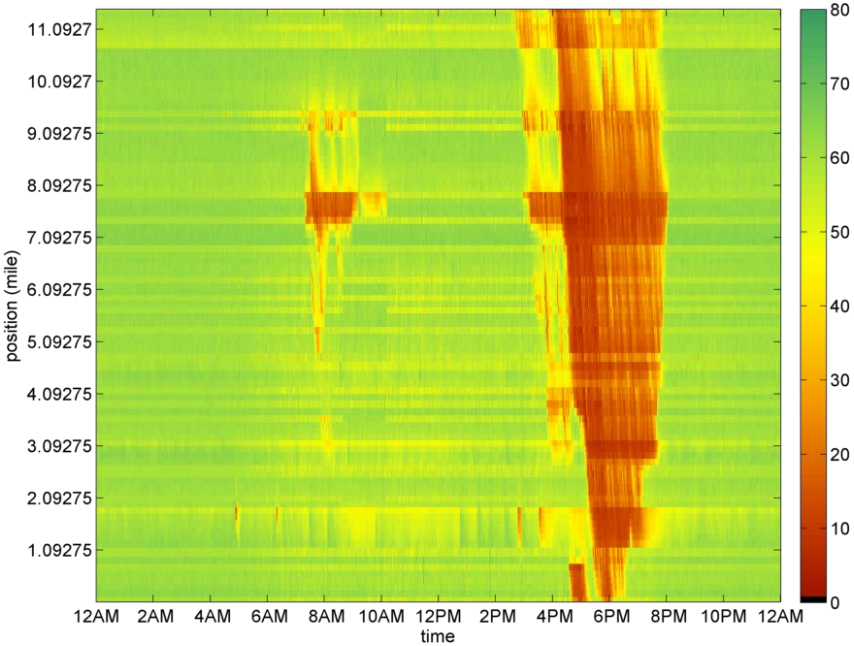
880 Northbound – Wed. March 7, 2012

CALIFORNIA PATH

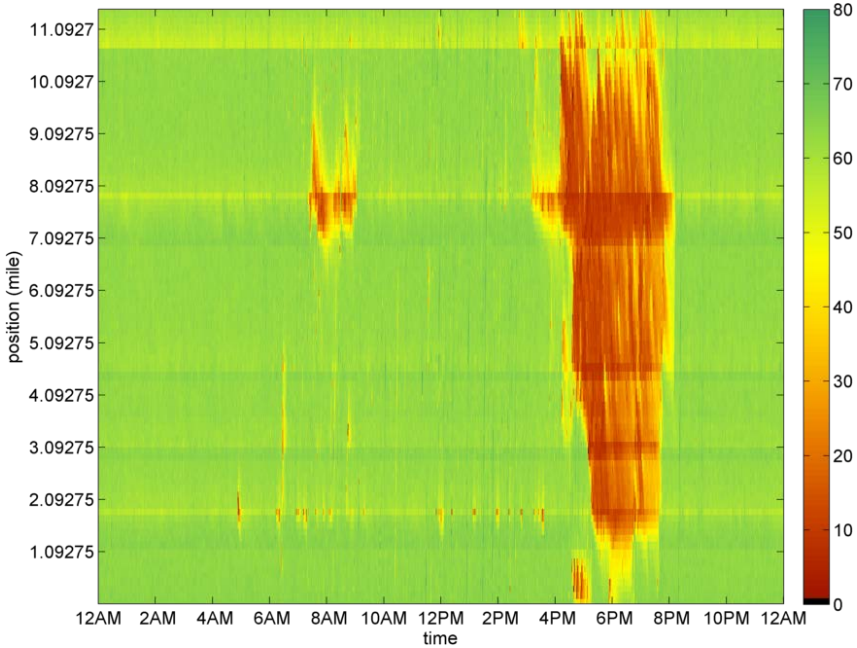
Velocity Maps



Assimilated Loop Data

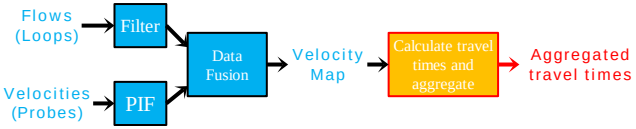


Assimilated Probe Data

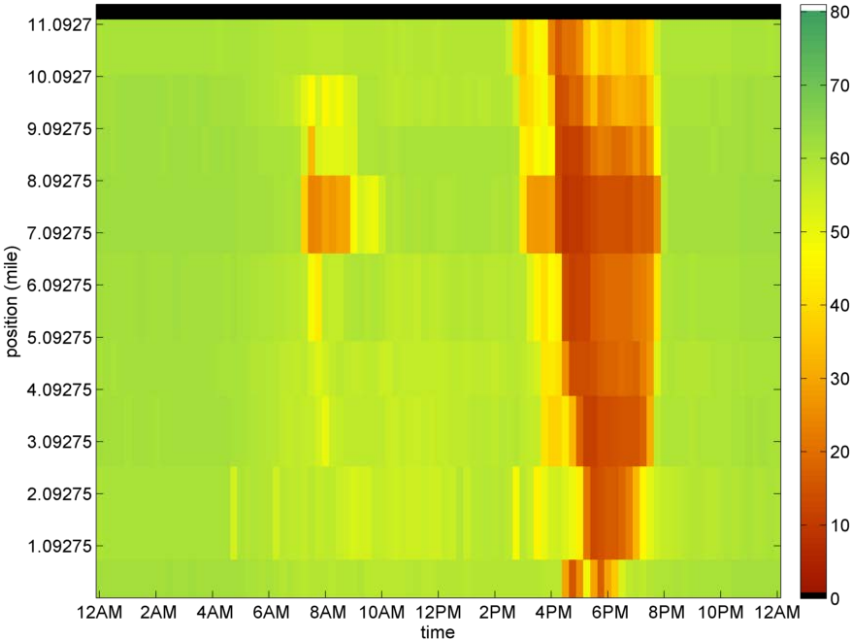


880 Northbound – Wed. March 7, 2012

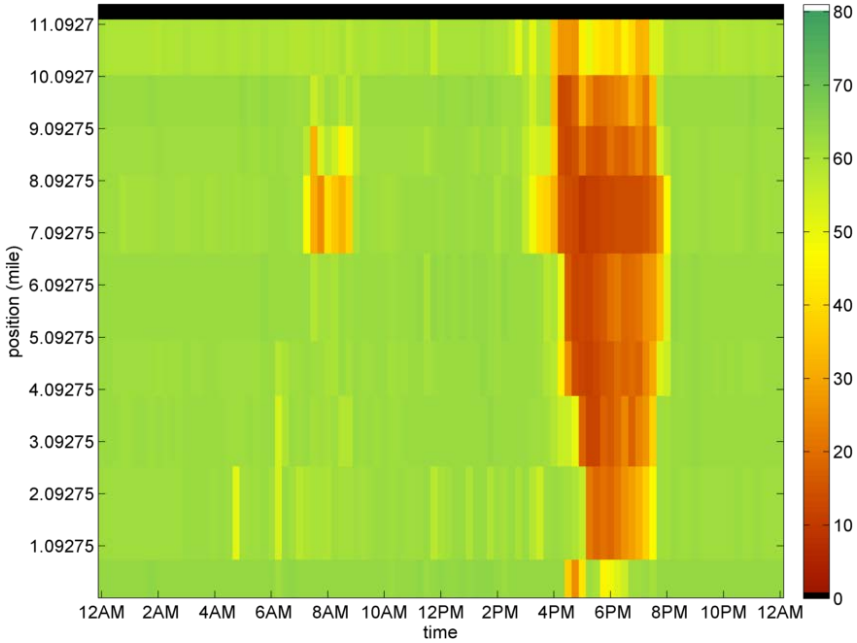
Average Velocities



Average Velocities from Loop Data



Average Velocities from Probe Data



880 Northbound – Wed. March 7, 2012

Outline

- Introduction
- One answer: how much data we need
- Overview of data and method
- **Results**
- Probe data quality
- Conclusion

Definition: Mean Absolute Percentage Error (MAPE)

- $MAPE = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{|T_{model}(i,j)-T_{bt}(i,j)|}{T_{bt}(i,j)}$
  - $T_{model}(i,j) = Modeled\ average\ travel\ times\ for\ time\ i, road\ section\ j$
  - $T_{bt}(i,j) = Bluetooth\ average\ travel\ times\ for\ time\ i, road\ section\ j$
  - $N = number\ of\ 15 - minute\ time\ periods$
  - $M = number\ of\ inter - BT\ road\ sections$

- **A measure of the percentage error between travel time from Bluetooth sensors and the model's result using available data, averaged over time and space**

Velocity State Estimation: I-880

880 Northbound – Wed. March 7, 2012

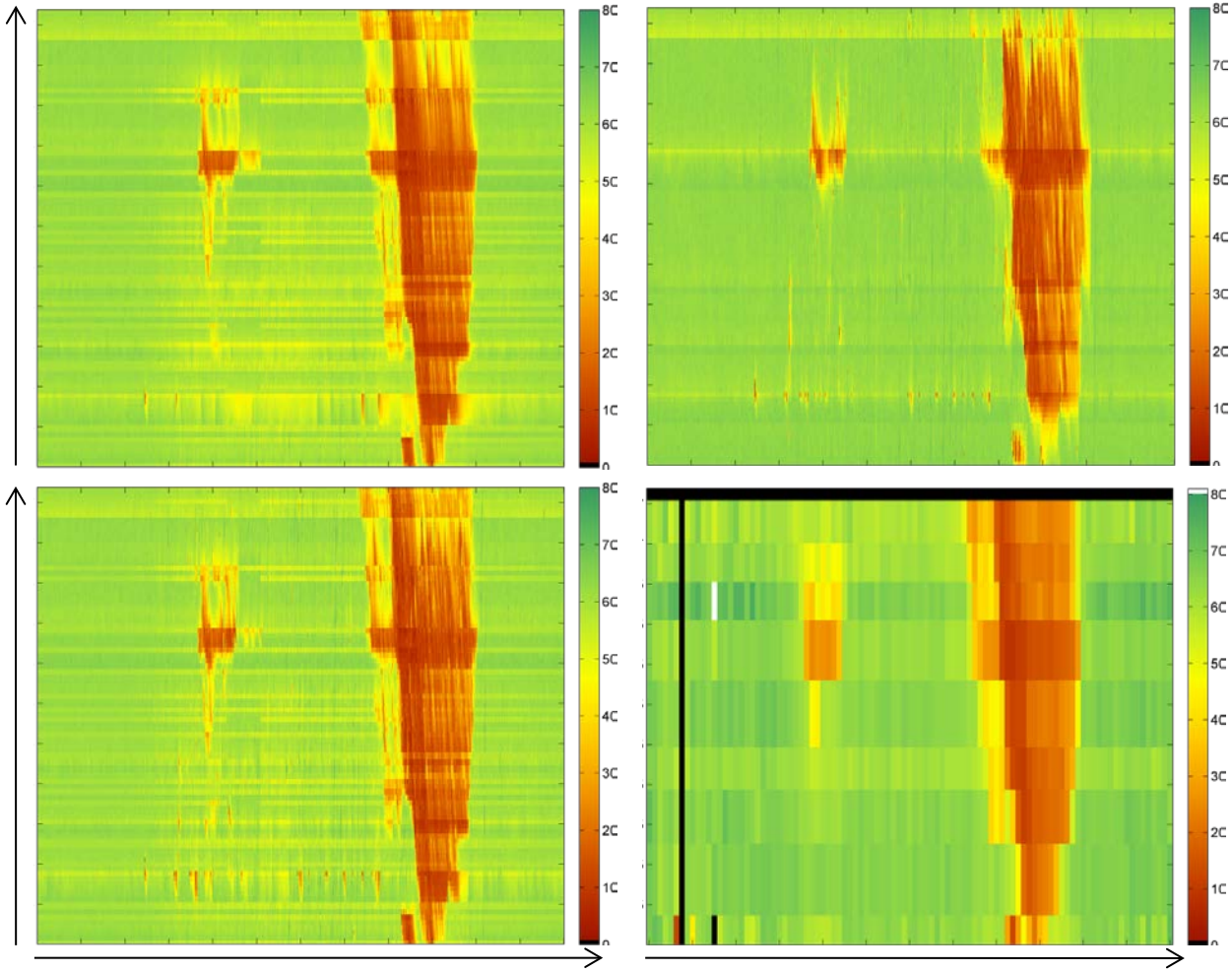
Loop data only

Probe data only

Position (12 mi)

Loop data and Probe data

Bluetooth only



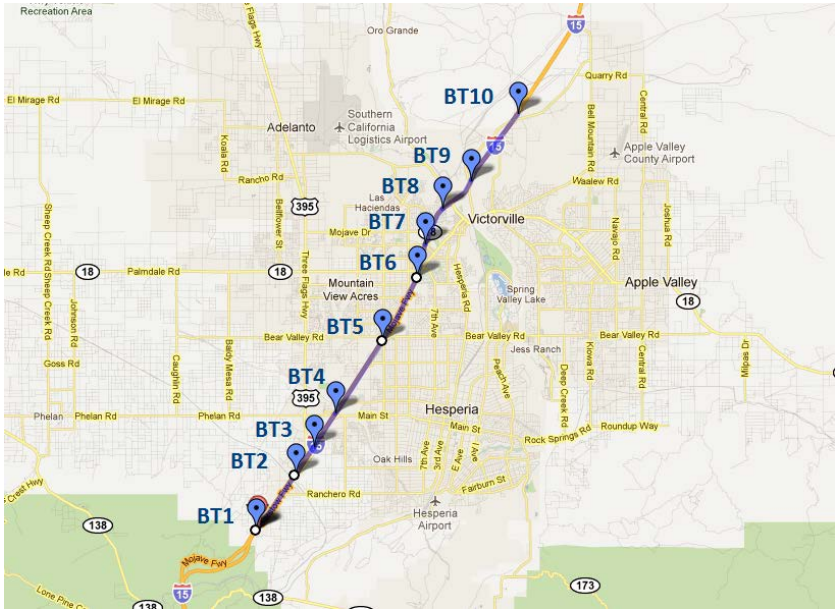
Velocity State Estimation: I-880

I-880 NB congested periods between 3/2 and 3/17, 2012

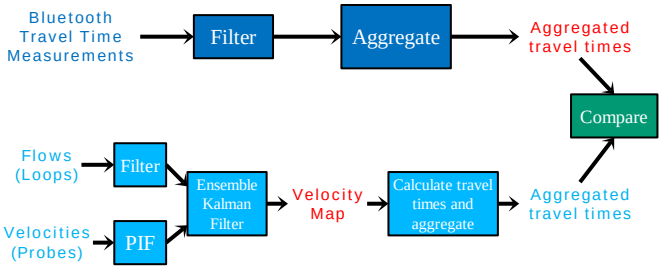
Data	MAPE or BTMAPE
Loop	21%
Probe	15%
Loop+Probe	10%
One standard deviation "noise floor"	19%

Site Description: I-15 Victorville SB

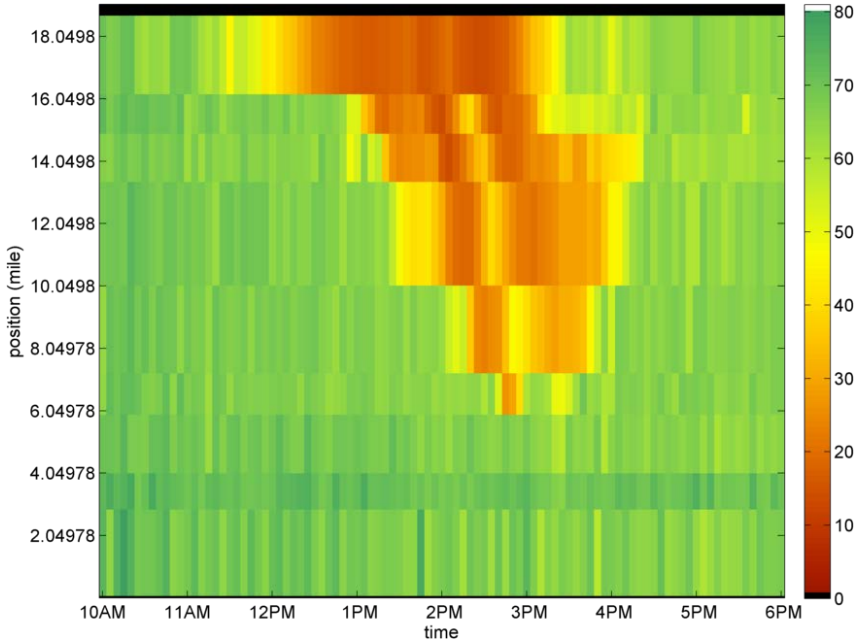
- **Ten Bluetooth sensors placed at roughly 2 mile spacing**
- **No active loop detectors in place**
- **Extremely encouraging results**



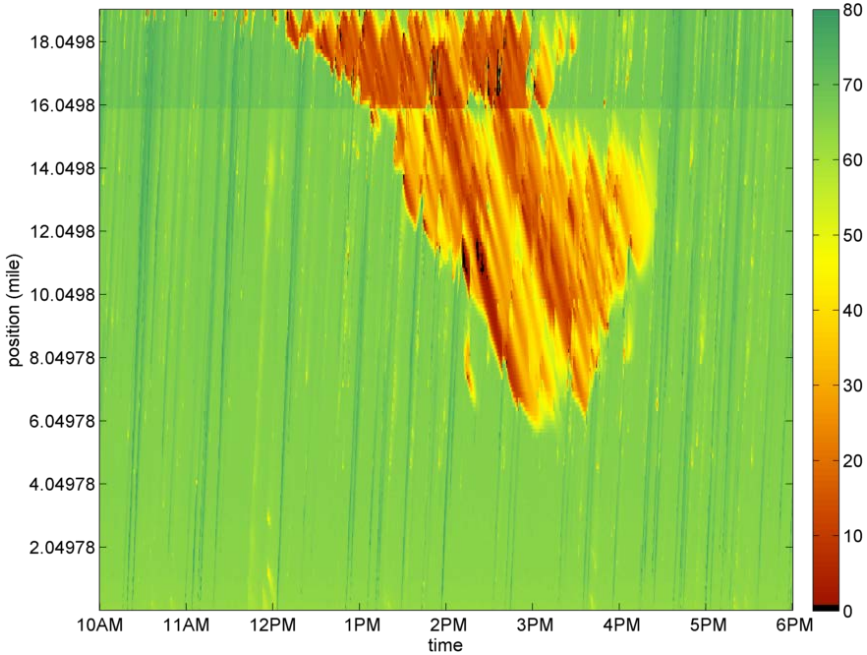
Results: I-15 Victorville



Bluetooth measured average velocities



Velocity Map from probe data



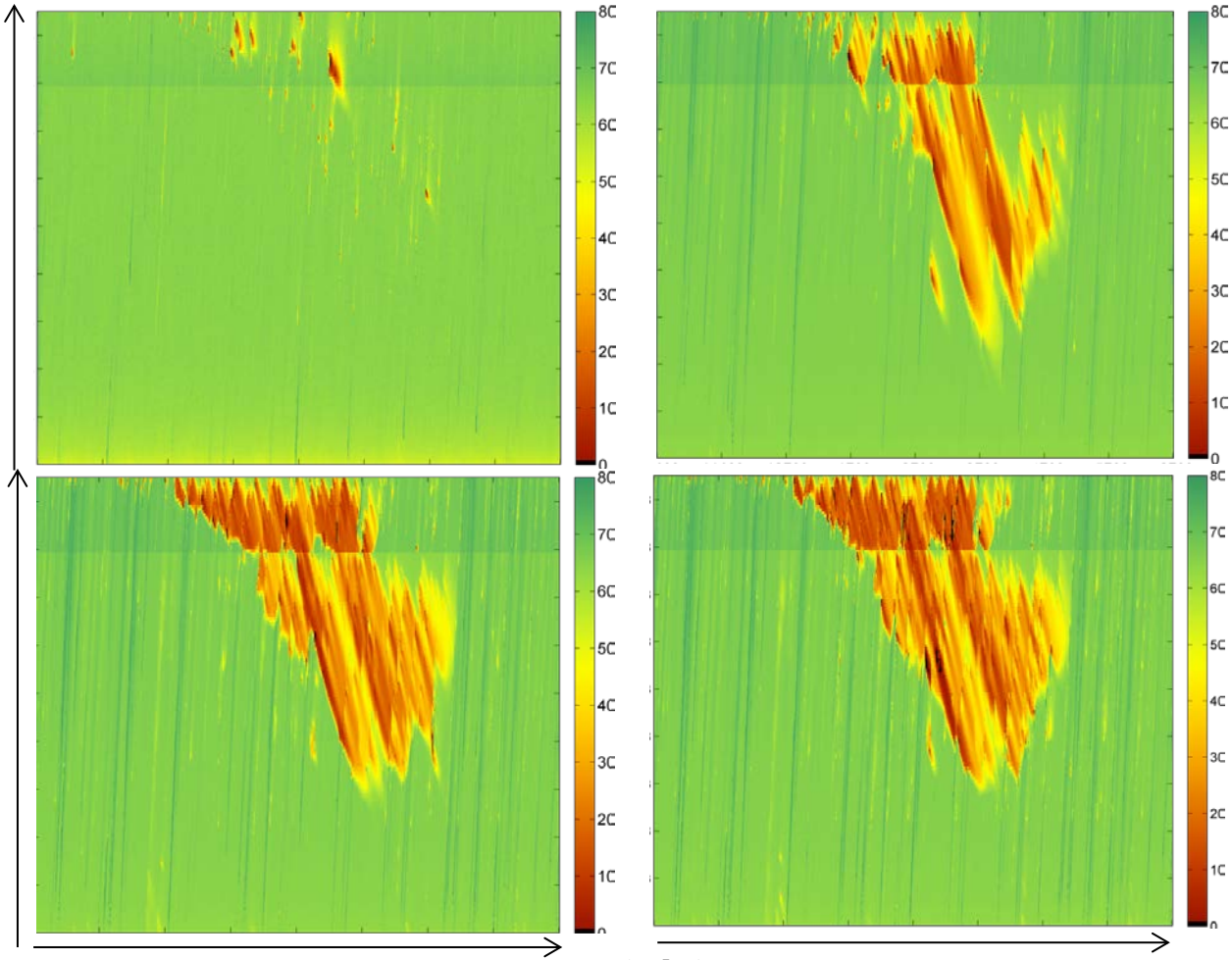
I-15 Southbound – Sunday April 22, 2012

Velocity Maps resulting from different subsets of available data

Use 9% of probe data yields 36% MAPE

Use 18% of probe data yields 24% MAPE

Position (18 mi)



Use 49% of probe data yields 15% MAPE

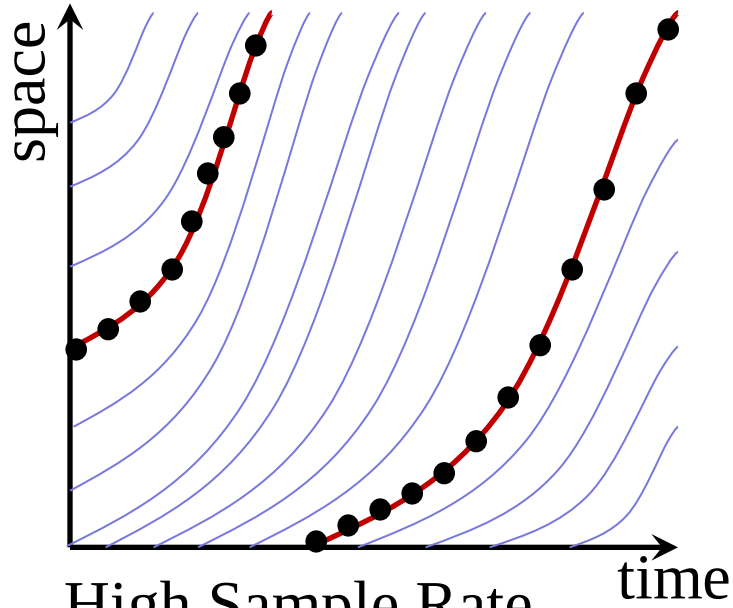
Use 100% of probe data yields 13% MAPE

Outline

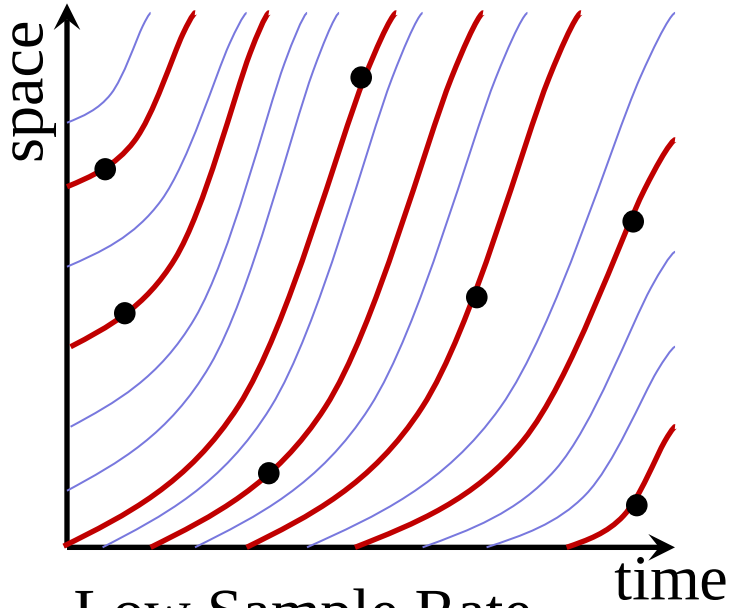
- Introduction
- One answer: how much data we need
- Overview of data and method
- Results
- **Probe data quality**
- Conclusion

Discussion of Data Quality

- **Example vehicle trajectories illustrating sample rate and penetration rate**



High Sample Rate
Low Penetration Rate



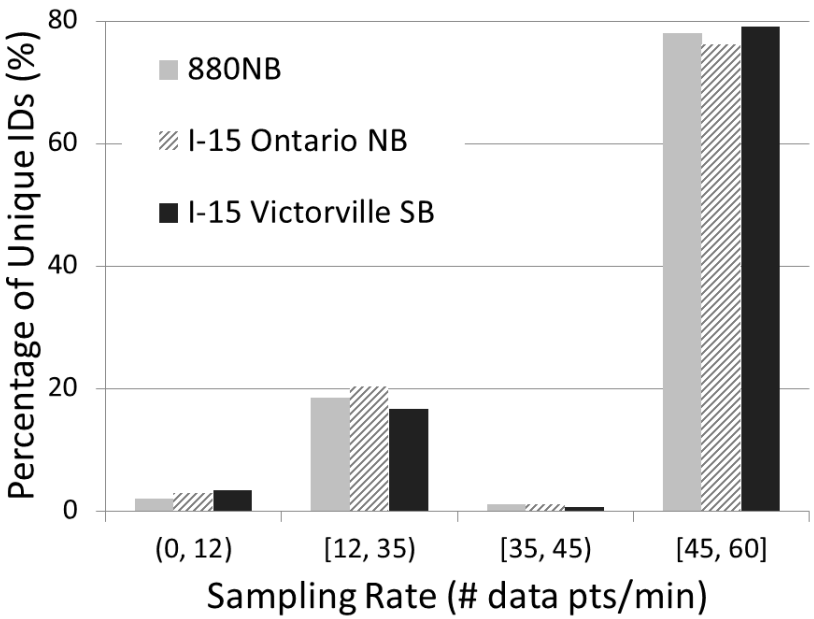
Low Sample Rate
High Penetration Rate

Penetration Rate vs. Sample Rate (1)

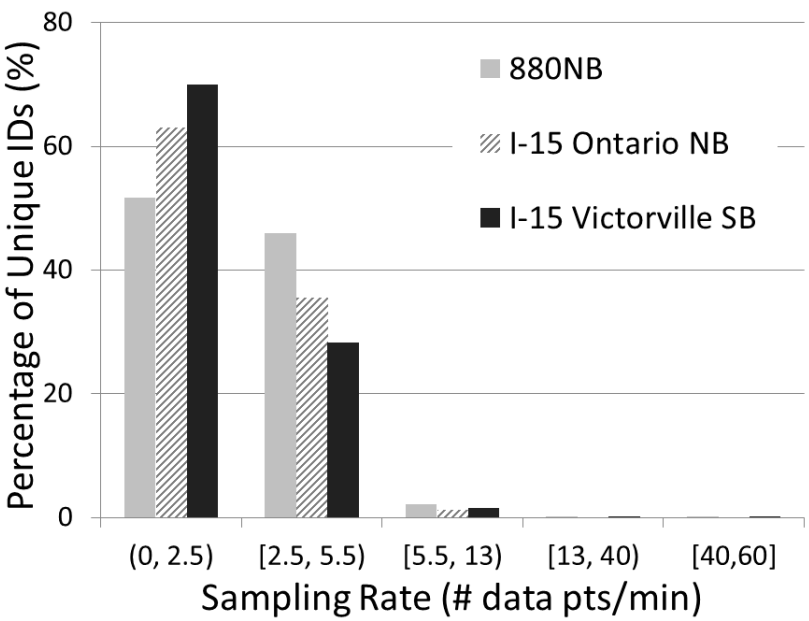
- **Probe Set A data**
  - Higher sample rate
  - Lower penetration rate: 0.1%
  - 85th percentile flows: 7 veh/hour

- **Probe Set B data**
  - Lower sample rate
  - Higher penetration rate: 0.6%
  - 85th percentile flows: 35 veh/hour

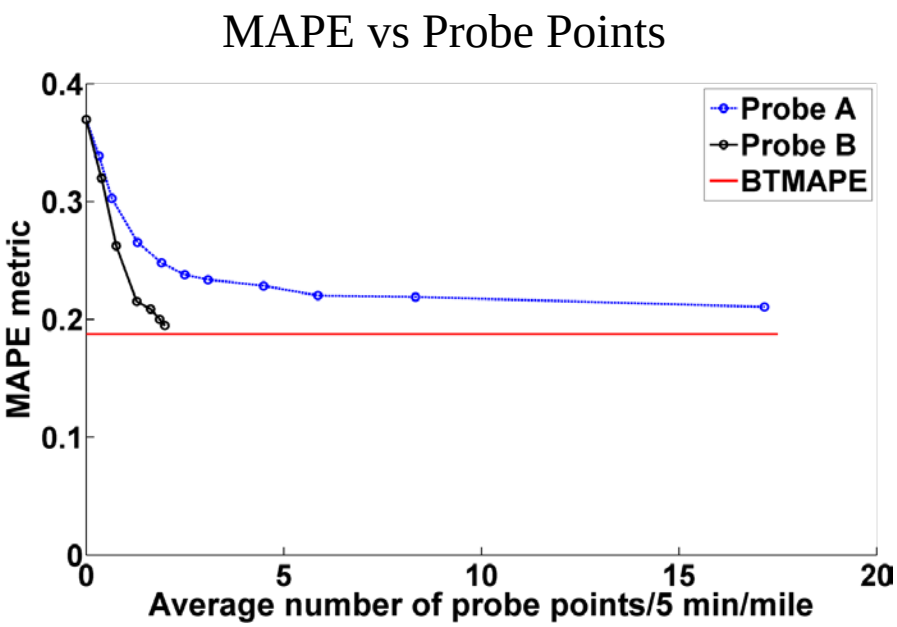
Set A Sample Rates

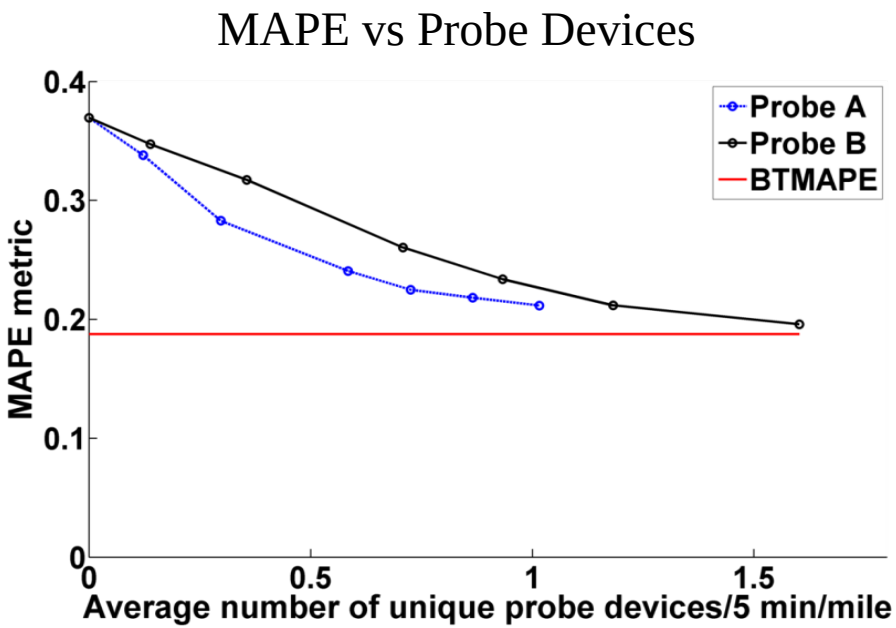


Set B Sample Rates



Penetration Rate vs. Sample Rate (2)





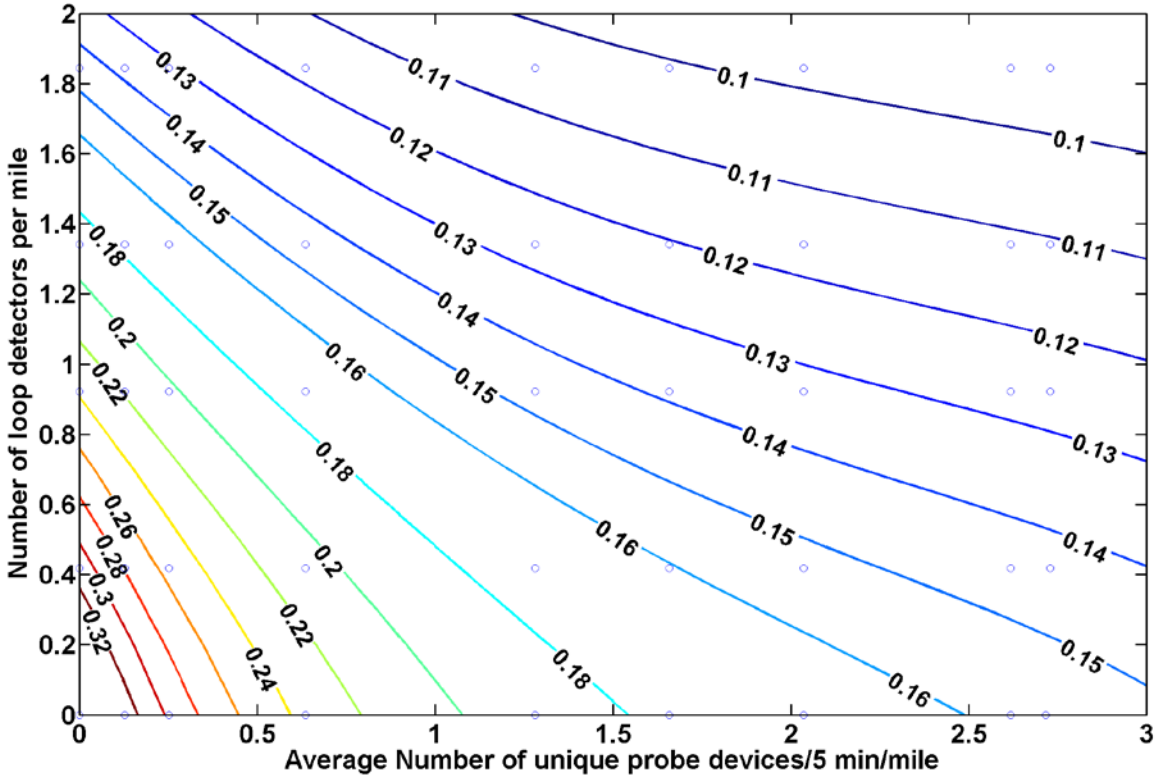
On freeways, a high penetration rate of probes is better than having a high sample rate. For this particular application, probe A data is oversampled.

Outline

- Introduction
- One answer: how much data we need
- Overview of data and method
- Results
- Probe data quality
- **Conclusion**

How much GPS data do **YOU** need?

For travel time estimation on freeways, here is a menu of data solutions for a given MAPE requirement:



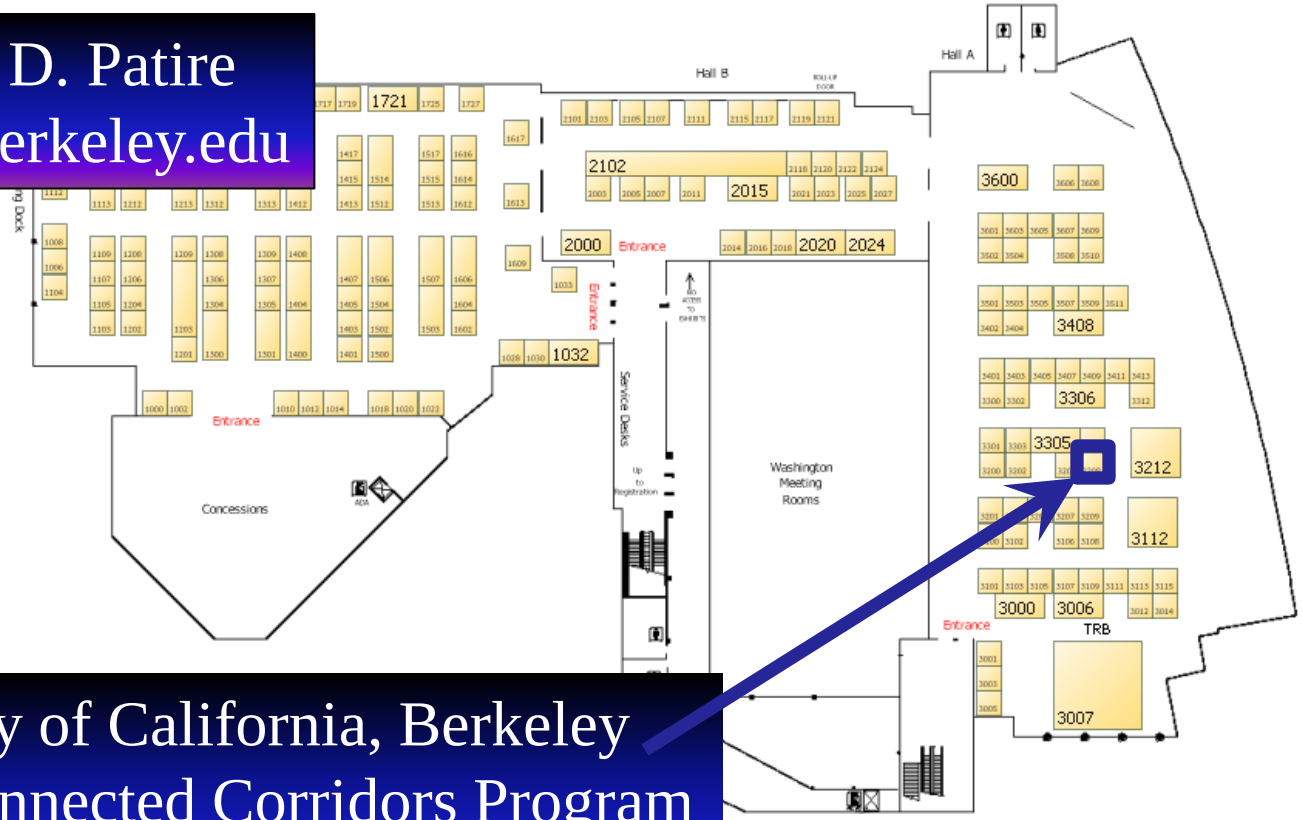
Future Work

- **Consider arterials and freeways in a coordinated way**

- **Decision support tools for a TMC**
  - **Estimation of current traffic conditions**
  - **Prediction of future traffic conditions**
  - **Assess impacts of potential control interventions**

Please visit our exhibition booth

Anthony D. Patire
adpat@berkeley.edu



University of California, Berkeley
PATH Connected Corridors Program
Booth 3208