Methodology for Calculating Latency of GPS Probe Data
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

Timeline

- T1 Field data recorded
  - Delay caused by the telecommunication medium

- T2 Field data received
  - Delay dictated by the processing time of the data fusion algorithms

- T3 Probe data generated
  - Delay caused by the internal processes to generate the real-time feed

- T4 Probe data is made available
  - Latency due to query time, internet speed, etc.

- T5 Probe data received by user
  -
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”.

![Graph showing travel speed over time with latency highlighted](image-url)
Current Research

  - 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.

LITERATURE REVIEW
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.
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).

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

Missing Interval: $n=3$

... $s_{t-1}$ $s_t$ $s_{t+i}$ $s_{t+n+1}$ $s_{t+n+2}$ ...
Data Processing

Data Smoothing

- Weighted moving average function

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

**Why filtfilt?** Smoothing may introduce undesired horizontal shift into the time series, however in `filfilt`, 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**

**Bluetooth Data Processing**
- Bluetooth Field Data
  - Paired Travel Time Data
    - Filter
      - Paired Data without Outliers
        - Convert
          - Segment Travel Speed Data
            - Aggregate
              - One-min Speed Data
                - Interpolate
                  - Speed Data without Gap
                    - Smooth
                      - Smoothed Speed Data

**GPS-probe Data Processing**
- GPS-probe Data
  - Filter (rule out data with low confidence score)
    - Path data
      - Interpolate
        - GPS-probe Data without Gap
          - Smooth
            - Smoothed GPS-probe Data

**Calculating Latency**
Objective

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

\[
\begin{align*}
\min f_1 &= \sum_{t=1}^{n} |S_t^{BT} - S_{t-\text{latency}}^{\text{probe data}}| & \text{Absolute vertical distance between two curves} \\
\min f_2 &= \sum_{t=1}^{n} (S_t^{BT} - S_{t-\text{latency}}^{\text{probe data}})^2 & \text{Gives more weights to the points that have bigger difference} \\
\min f_3 &= \text{corr}(S_t^{BT}, S_{t-\text{latency}}^{\text{probe data}}) & \text{Statistical representation of the linear relationship between two curves}
\end{align*}
\]

\(lb \leq \text{latency} \leq ub\)
Minimize Absolute Vertical Distance \( f(I) \) – as example

\[
\min f \frac{1}{n} \sum_{i=1}^{n} |S_{i}^{BT} - S_{i}^{probe\ data}|
\]
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 Dec 15, 2015

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

Travel Speed Comparison (AB-12/4)

Travel Speed Comparison (5 mins Latency)

Latency
Result
Some Afternoon Peaks

**CASE STUDY**
### Test One: Average Latency at Peak Periods

<table>
<thead>
<tr>
<th>Period</th>
<th>Number of Observations</th>
<th>Average Latency (minute)</th>
<th>f1 (AVD)</th>
<th>f2 (SVD)</th>
<th>f3 (COR)</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning</td>
<td>32</td>
<td></td>
<td>3.96</td>
<td>4.42</td>
<td>4.41</td>
<td>4.26</td>
</tr>
<tr>
<td>Afternoon</td>
<td>45</td>
<td></td>
<td>3.64</td>
<td>4.01</td>
<td>4.19</td>
<td>3.94</td>
</tr>
</tbody>
</table>

- 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.
Result

Test Two: Average Latency at Different Segments

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

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

Test Three: Average Latency at Slowdown and Recovery

![Travel Speed Comparison](image)

- **Travel Speed Comparison (CD-12/3)**
  - **Bluetooth data**
  - **GPS Probe data**

  - **Slowdown Starts**
  - **Transition from speed slowdown to recovery**
  - **Speed recovery ends**

  - **Slowdown period**
  - **Recovery period**
### Result

#### Test Three: Average Latency at Slowdown and Recovery

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Scenario</th>
<th>Number of Observations</th>
<th>Average Latency (minute)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>f1 (AVD)</td>
</tr>
<tr>
<td>Morning</td>
<td>Slowdown</td>
<td>32</td>
<td>3.55</td>
</tr>
<tr>
<td></td>
<td>Recovery</td>
<td>32</td>
<td>4.76</td>
</tr>
<tr>
<td>Afternoon</td>
<td>Slowdown</td>
<td>45</td>
<td>3.43</td>
</tr>
<tr>
<td></td>
<td>Recovery</td>
<td>45</td>
<td>4.70</td>
</tr>
<tr>
<td>Overall</td>
<td>Slowdown</td>
<td>77</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>Recovery</td>
<td>77</td>
<td>4.72</td>
</tr>
</tbody>
</table>

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