Hybrid Data Implementation
Thursday, April 1, 2021

Dr. Anthony Patire, Research Engineer at California Partners for Advanced Transportation Technology (PATH), UC Berkeley

Disclaimer: The contents of this presentation are disseminated in the interest of information exchange. The contents of this presentation reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration nor does it constitute a standard, specification, or regulation.
What is the need for this research?

- Caltrans had 40,000 individual vehicle detection zones
- Gigabytes of data is collected every day
- Caltrans programmed over $150 million of SHOPP funds for failed or failing detection stations

Pavement failure near installation

Copper wire theft

Rural area with no detection
Benefits of Third-Party Data

- Reduce the use of traditional vehicle detection
- Reduce maintenance cost
- Limit exposure of construction, maintenance, and operations personnel to live traffic
- Provide broader coverage of the state routes to include those areas not currently monitored (Arterials, Rural)
Presentation Highlights

- New methodologies for using third party data
- Benefits to Caltrans for using third party data in established performance measurement, including reduced costs and increased coverage
- Impacts that data sources have on performance measurement
Acknowledgements

Sakib Mahmud Khan, Ph.D., Post-Doctoral Scholar
Nicholas Fournier, Ph.D., Post-Doctoral Scholar
Michael Mauch, Ph.D., Research Engineer
Alex Skabardonis, Ph.D., Professor In-Residence
6 Outline
Outline

- Introduction
- Methodology for delay calculation
- Challenges
- Evaluation of methods
- Goals and next steps
Hybrid Data Question

- Is it possible to use third-party traffic data to augment or replace existing infrastructure for collecting point-based traffic data?

Here & Now
Point-based sensors (loops)

Transition

Future
Hybrid data
Leverages private sector
Yes. Third-party travel time data are useful and complementary to data from point-detectors.

Point-detector data should focus on quality over quantity:
- Lane specific
- Obtains complete cross-section of flow

This evaluation pertains strictly to the measurement of delay:
- Third-party data can compensate for loss of point-based sensors
- Third-party data can be used to roughly estimate delay with limited instrumentation

Must overcome challenges related to legacy PeMS meta information (configuration information)
## Comparison of Data Vendors

<table>
<thead>
<tr>
<th></th>
<th>FHWA NPMRDS</th>
<th>HERE</th>
<th>TOMTOM</th>
<th>INRIX</th>
<th>STREETLIGHT</th>
<th>CITILABS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Sources</td>
<td>HERE Data</td>
<td>CELL, GPS, CV - MANY</td>
<td>GPS</td>
<td>GPS, some CV - MANY</td>
<td>GPS, CELL</td>
<td>Multiple (GPS, CELL, Traffic Counts)</td>
</tr>
<tr>
<td>Data Collection Method(s)</td>
<td>HERE Method</td>
<td>Purchased from App providers, OEM vehicles</td>
<td>INTERNAL GPS DEVICES</td>
<td>CV, INRIX APP</td>
<td>INRIX METHOD</td>
<td>Proprietary process combining multiple data types and sources</td>
</tr>
<tr>
<td>Main Product</td>
<td>Auto and Truck Speeds and TT</td>
<td>SPEED</td>
<td>SPEED</td>
<td>SPEED VOLUME</td>
<td>O-D VOLUME</td>
<td>SPEED VOLUME O-D</td>
</tr>
<tr>
<td>Real-time Delivery Capability</td>
<td>NO</td>
<td>Yes, Real-time and predictive</td>
<td>Yes, Real-time and predictive</td>
<td>Yes, Real-time and predictive</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Historical Delivery Capability</td>
<td>Historic, delivered monthly</td>
<td>Historic, delivered daily</td>
<td>Historic</td>
<td>Historic, delivered daily</td>
<td>Historic, delivered daily</td>
<td>Historic</td>
</tr>
<tr>
<td>Data validation reports?</td>
<td>YES</td>
<td>YES</td>
<td>?</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Mapping Capability</td>
<td>HERE mapping</td>
<td>have a map product</td>
<td>have a map product</td>
<td>Previously using OSM and TomTom, migrating to HERE</td>
<td>NO</td>
<td>?</td>
</tr>
</tbody>
</table>

In general, they have different maps
Grid of Data Types

Real World
- Third-party Data
  - Raw GPS Points
  - Link travel times / speeds
- Caltrans Field Data
  - Vehicle Detector Station (VDS)
    - Flow, Occupancy, Speed
    - Annual Average Daily Traffic (AADT)

Simulated
- Synthesized Data
  - Link travel times / speeds
- Synthesized Data
  - Flow, Occupancy, Speed

3rd PeMS
- Real
- Simulated
Methodology for DVHD

Calculation of Daily Vehicle Hours of Delay (DVHD)
The existing MPR uses data from fixed point-sensors called Vehicle Detection Stations (VDS).

The pavement covered by each VDS extends from upstream midpoint to downstream midpoint.
Hybrid Calculation: Overview

- Use flow and density measurements from VDS
- Use travel time measurements from 3rd party vendors
- Different vendors may have different maps
Hybrid Calculation: Domain of Analysis

- Create evenly spaced grid
Hybrid Calculation: Data Projection

- **Fill in the blanks using VDS**
  - Populate grid with flow and density data
  - Confined Generalized Adaptive Smoothing Method (C-GASM)*

Hybrid Calculation: Use 3rd Party Data

- Conflate third party travel time information onto the grid
- Calculate desired metrics
Challenges

What challenges impede a hybrid data approach?
Existing PeMS meta-information

- Provides enough sensor location information for maintenance
- Does not provide adequate sensor location information for an algorithm to automatically conflate third-party data with PeMS data
- The location information in PeMS corresponds to the location of the controller instead of the location of the pavement being monitored
**Interpretation of Meta Information**

**Name**: Closest cross-street or feature

**Abs PM**: Absolute postmile of sensor’s controller

**Fwy**: Physical freeway associated with the sensor’s controller

**Type**: Relationship of sensor to physical freeway

<table>
<thead>
<tr>
<th>Abs PM</th>
<th>ID</th>
<th>Name</th>
<th>Fwy</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>36.29</td>
<td>769723</td>
<td>NB 605 TO WB 210 CON</td>
<td>I210-W</td>
<td>HOV</td>
</tr>
<tr>
<td>36.29</td>
<td>769722</td>
<td>NB 605 TO WB 210 CON</td>
<td>I210-W</td>
<td>Mainline</td>
</tr>
<tr>
<td>36.89</td>
<td>773204</td>
<td>NB 605 TO MT. OLIVE</td>
<td>I210-W</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>36.89</td>
<td>773205</td>
<td>EB 210 TO MT. OLIVE</td>
<td>I210-W</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>36.89</td>
<td>773206</td>
<td>SB 605 FROM WB 210</td>
<td>I210-W</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>36.89</td>
<td>773207</td>
<td>NB 605 TO EB 210</td>
<td>I210-W</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>36.89</td>
<td>775795</td>
<td>NB 605 TO WB 210</td>
<td>I210-W</td>
<td>Mainline</td>
</tr>
<tr>
<td>36.90</td>
<td>775796</td>
<td>EB 210 TO SB 605</td>
<td>I210-W</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>27.95</td>
<td>774264</td>
<td>NB 605 TO MT OLIVE</td>
<td>I605-S</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>27.95</td>
<td>774261</td>
<td>EB 210 TO MT OLIVE</td>
<td>I605-S</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>27.95</td>
<td>774262</td>
<td>WB 210 TO MT OLIVE</td>
<td>I605-S</td>
<td>Off Ramp</td>
</tr>
<tr>
<td>27.95</td>
<td>774263</td>
<td>MT OLIVE TO WB 210</td>
<td>I605-S</td>
<td>On Ramp</td>
</tr>
<tr>
<td>27.95</td>
<td>774258</td>
<td>MT OLIVE TO EB 210</td>
<td>I605-S</td>
<td>On Ramp</td>
</tr>
<tr>
<td>27.95</td>
<td>774260</td>
<td>MT OLIVE TO SB 605</td>
<td>I605-S</td>
<td>On Ramp</td>
</tr>
</tbody>
</table>

- VDS inherit their Abs PM and freeway (Fwy) association from the controller they are connected to.
- This works well for maintenance purposes.
- But has confusing consequences at freeway interchanges.
Meta Information at Interchanges

- One single control box may handle multiple freeways at an interchange, but it can only be associated with one freeway.
- The description of freeway location, connectivity and type get condensed into the name.

<table>
<thead>
<tr>
<th>Abs PM</th>
<th>ID</th>
<th>Name</th>
<th>Fwy</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>36.29</td>
<td>769723</td>
<td>NB 605 TO WB 210 CON</td>
<td>I210-W</td>
<td>HOV</td>
</tr>
<tr>
<td>36.29</td>
<td>769722</td>
<td>NB 605 TO WB 210 CON</td>
<td>I210-W</td>
<td>Mainline</td>
</tr>
<tr>
<td>36.29</td>
<td>769724</td>
<td>NB 605 TO WB 210 CON</td>
<td>I210-W</td>
<td>On Ramp</td>
</tr>
<tr>
<td>36.89</td>
<td>773204</td>
<td>NB 605 TO MT. OLIVE</td>
<td>I210-W</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>36.89</td>
<td>773205</td>
<td>EB 210 TO MT. OLIVE</td>
<td>I210-W</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>36.89</td>
<td>773206</td>
<td>SB 605 FROM WB 210</td>
<td>I210-W</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>36.89</td>
<td>773207</td>
<td>NB 605 TO EB 210</td>
<td>I210-W</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>36.89</td>
<td>775795</td>
<td>NB 605 TO WB 210</td>
<td>I210-W</td>
<td>Mainline</td>
</tr>
<tr>
<td>36.89</td>
<td>775796</td>
<td>EB 210 TO SB 605</td>
<td>I210-W</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>27.95</td>
<td>774264</td>
<td>NB 605 TO MT OLIVE</td>
<td>I605-S</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>27.95</td>
<td>774261</td>
<td>EB 210 TO MT OLIVE</td>
<td>I605-S</td>
<td>Fwy-Fwy</td>
</tr>
<tr>
<td>27.95</td>
<td>774262</td>
<td>WB 210 TO MT OLIVE</td>
<td>I605-S</td>
<td>Off Ramp</td>
</tr>
<tr>
<td>27.95</td>
<td>774263</td>
<td>MT OLIVE TO WB 210</td>
<td>I605-S</td>
<td>On Ramp</td>
</tr>
<tr>
<td>27.95</td>
<td>774258</td>
<td>MT OLIVE TO EB 210</td>
<td>I605-S</td>
<td>On Ramp</td>
</tr>
<tr>
<td>27.95</td>
<td>774260</td>
<td>MT OLIVE TO SB 605</td>
<td>I605-S</td>
<td>On Ramp</td>
</tr>
</tbody>
</table>
Proposed Solution

- **Question:** How could configuration meta-data be improved at major junctions?

- **Best answer:** Associate sensors on the pavement with a network map to show exact locations

- **Minimal answer:** Add additional information
  - Each sensor (pavement location) should get its own latitude and longitude coordinates
  - Add one additional table to PeMS to organize VDS around fully accounted traffic volumes (FATVs)
Fully Accounted Traffic Volume (FATV)

- FATV with VDS 773205 as an input flow sensor
Fully Accounted Traffic Volume (FATV)

- FATV with VDS 773205 as an output flow sensor
Advantages of FATV approach

- FATVs would help clarify locations of VDS
  - Everywhere along a freeway
  - Especially useful at major junctions

- Over the course of one day, input flow should roughly equal output flow
  - Enables automated checking of configuration
  - Enables automated checking of data integrity

- FATVs would improve ability to fill in missing data

- Partially accounted traffic volumes (PATVs) are also useful to know what kind of data to expect
Evaluation of Methods
Microsimulation

Study Corridor: 16 miles of I-210 WB

Data taken from outside the geographic boundaries of the model
Simulated Scenarios

- Selection of four time periods
  - Before AM Peak
  - AM Peak
  - Noon
  - PM Peak

- Approximate flow: no instrumentation within model geography

- Reduced instrumentation, removing VDS pairs
Approximate flow

- **Available information**
  - Travel times from third party data
  - Annual Ave Daily Traffic (AADT)

- **Convert AADT to hourly flow:**
  - Case 1: Generic flow profile
  - Case 2: Measured flow profile from nearby sensors

![Diagram showing AADT and hourly flow profile with vendor data and delay]

- Vendor data
- Hourly flow profile
- Delay
Delay Estimation Error Distribution

Error distribution of segmented by time of day

Error from Approximate Flow

Case 1
(Generic profile)

Case 2
(Measured profile)

Abs % Error

Before morning
Morning peak
Noon
Afternoon peak
Reduced Instrumentation (Sensor Removal)

- Systematically remove sensors along corridor
- Repeat for all pairs of VDS

VDS pairs to be removed
Delay Estimation Error Distribution

- Error distribution segmented by time of day

Error From Reduced Instrumentation

- Abs % Error

Traditional vs Hybrid

- Before morning
- Morning peak
- Noon
- Afternoon peak
Error Distribution

- **Error distribution of all evaluation scenarios**

![Graph showing error distribution with different scenarios and labels for Traditional and Hybrid data with different flow cases.](image-url)
Summary Results

- The ability to leverage third-party data to calculate delay depends on the quality of the point-detector data, not the quantity.

- Point-detectors are needed where lane specific information is required, such as HOV lanes.

- Must overcome challenges related to legacy PeMS meta information (configuration information).

- This evaluation pertains strictly to the measurement of delay.
  - Third-party data can compensate for loss of point-based sensors.
  - Third-party data can be used to **roughly** estimate delay with limited instrumentation.
Recommendations for Delay Calculations

<table>
<thead>
<tr>
<th>Calculation Methods</th>
<th>Mainline</th>
<th>HOV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional data and calculation</td>
<td></td>
<td>3rd party data not widely available</td>
</tr>
<tr>
<td>Hybrid calculation</td>
<td>Obtained best performance</td>
<td>Potential for the future</td>
</tr>
</tbody>
</table>

- **Adjustments for limited instrumentation**
  - Applicable where data is limited
  - Appropriate where high fidelity is not required
Goals and Next Steps
Implementation Roadmap

- **Step 1: Limited pilot**
  - Select well-studied freeways with excellent data
  - Use the pilot period to
    - Determine accuracy with real-world data, not simulation
    - Compare data quality of alternative 3rd parties

- **Step 2: Full-scale pilot in selected district**
  - Assess cost and difficulty of data integration over a limited geographical region
  - Assess value of hybrid, integrated traffic information
    - Delay and other performance measures
    - Situational awareness for TMC operators
    - Traffic management applications
Next steps

- The future of point-detector data should focus on quality over quantity

- Key research related tasks that could inform pilot
  - Create an initial set of freeways with high quality and reliable data.
  - Pre-select sites for an initial pilot
  - Perform an initial FATV assessment
  - Obtain precise location information at freeway-freeway connectors
  - Redundancy analysis to prioritize existing sensors
Questions?

Contact DRISI to discuss any research needs

DRISI.Communications@dot.ca.gov