

PARTNERS FOR ADVANCED TRANSPORTATION TECHNOLOGY
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

Connected Corridors: I-210 Pilot Integrated Corridor Management System

Processing Received Loop Data for the ICM System

May 22, 2017



Partners for Advanced Transportation Technology works with researchers, practitioners, and industry to implement transportation research and innovation, including products and services that improve the efficiency, safety, and security of the transportation system.

This page left blank
intentionally

PROCESSING RECEIVED LOOP DATA FOR THE ICM SYSTEM

The ICM system receives important data about traffic conditions from inductive loop detectors embedded in the road:

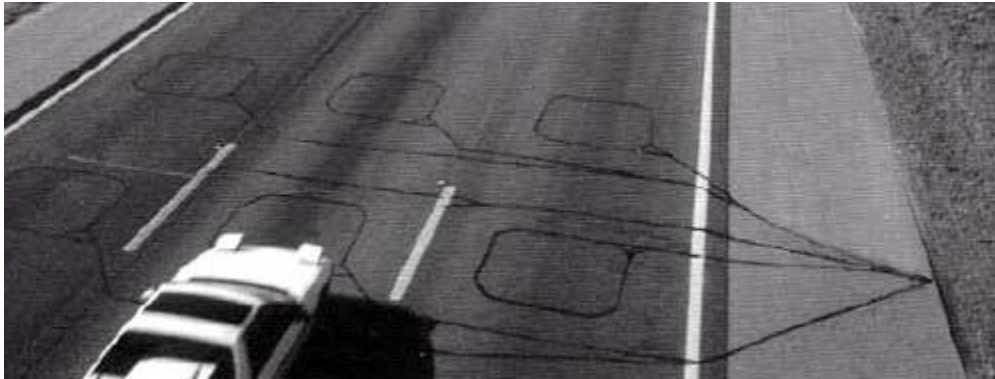


Figure 1: Loop detectors embedded in the road

For loop data to be useful for modeling, estimation, prediction, response plan evaluation, and more, it must be sufficiently accurate, timely, reliable, and complete. Key aspects to consider include:

Basic detector health	Data accuracy
<ul style="list-style-type: none">• Do we know and agree where the sensors are?• Are the sensors labeled and configured correctly (no sensors on the wrong side of the freeway, no HOV sensors mistaken for mainline sensors, etc.)?• Do the sensors capture a full cross section of traffic flow?• Are the sensors turned on and communicating data regularly?	<ul style="list-style-type: none">• Is the data provided by a detector trustworthy?• When we compare traffic flowing into and out of each section of freeway, do the numbers make sense (flow balance)?• Are data consistent with traffic engineering expectations?

The data must therefore be processed and analyzed to verify its quality. The steps in that process are:

1. Identify the hardware and software components of the system.
2. Get metadata.
3. Get physical world data.
4. Compare metadata to physical world data.

5. Validate the data path.
6. Get initial sensor data.
7. Validate the metadata and data path.
8. Check sensors' trustworthiness.
9. Filter the data.
10. Fill in blanks where possible.
11. Convert the data into a usable form.
12. Revalidate data quality over time.

1. IDENTIFY THE COMPONENTS OF THE SYSTEM

Understand what elements are involved in the data communication process. This can include:

- Roads (geometry, number of lanes, postmiles, freeways)
- Loop detectors
- Communication connections to agency
- Agency data management system
- System software

2. GET METADATA

Metadata is information about the loops, received from the agency that owns them:

- Initial list from the agency of where loops are located according to their system
- VDS (vehicle detection station) IDs, including name, how they are grouped, latitude/longitude of the control box, the nearest bridge/exit/cross-street, etc.

3. GET PHYSICAL WORLD DATA

Using Google's satellite and street view images, locate the loops in the physical world:

1. Find the loops in Google Earth or Google Street View.
2. Get the latitude/longitude of the loop itself and the VDS it is part of.
3. Create a shareable Google map identifying the precise location of each loop.



Figure 2: Satellite view of loops



Figure 3: Street view of loops

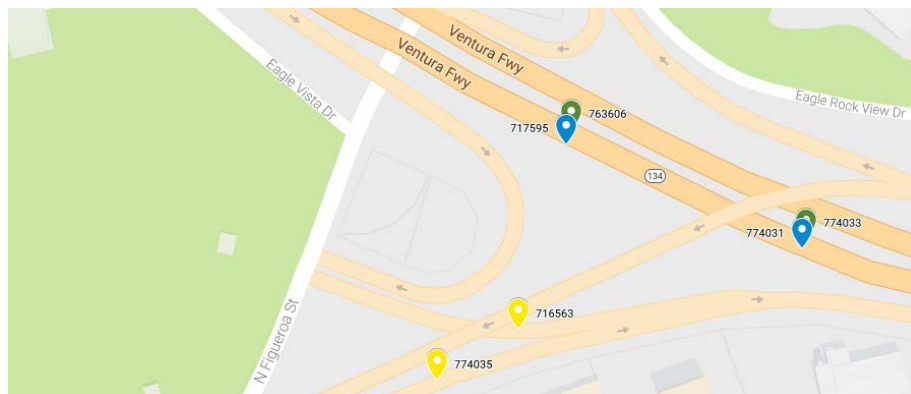


Figure 4: Loops on Google map

4. COMPARE METADATA TO PHYSICAL WORLD DATA

Compare the metadata from the agency with the physical world data to see if they contain the same information. A common issue, for example, is a mismatch in the number of lanes:



Figure 5: Satellite view showing five lanes

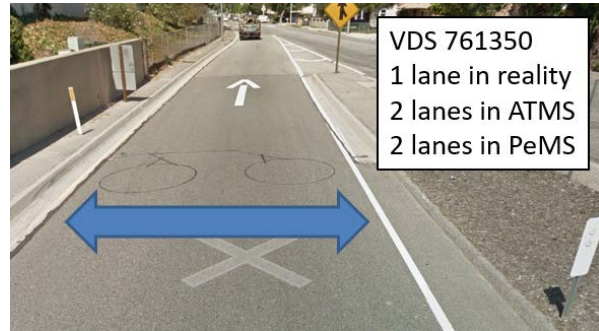


Figure 6: Street view showing one lane

1. Create a list of items that don't match.
2. Request that the agency review the list item by item to verify the correct locations and attributes of the sensors.
3. Revise the shareable Google map you have created to reflect actual conditions.
4. When the agency confirms the shareable Google map is accurate, request that they update the metadata in their system, if needed.

Note: Not every apparent mismatch is an error. For example, a picture may show four loops, while the metadata indicates only three because one is an old, unused loop. The metadata, in that case, is correct.

5. VALIDATE THE DATA PATH

Identify the exact path the data takes from point of origin to final destination, to verify it is being communicated properly all along the way. Delineating this communication path—how the components are connected—is important for troubleshooting data issues.

Figure 7 shows a simplified illustration of a data path, connected in this sequence:

1. VDS with three loop detectors in the roadway
2. Lead-in cable
3. Pull box
4. Cabinet and controller
5. Communication connection (cable, fiber, or wireless)
6. Agency system
7. ICM Core System

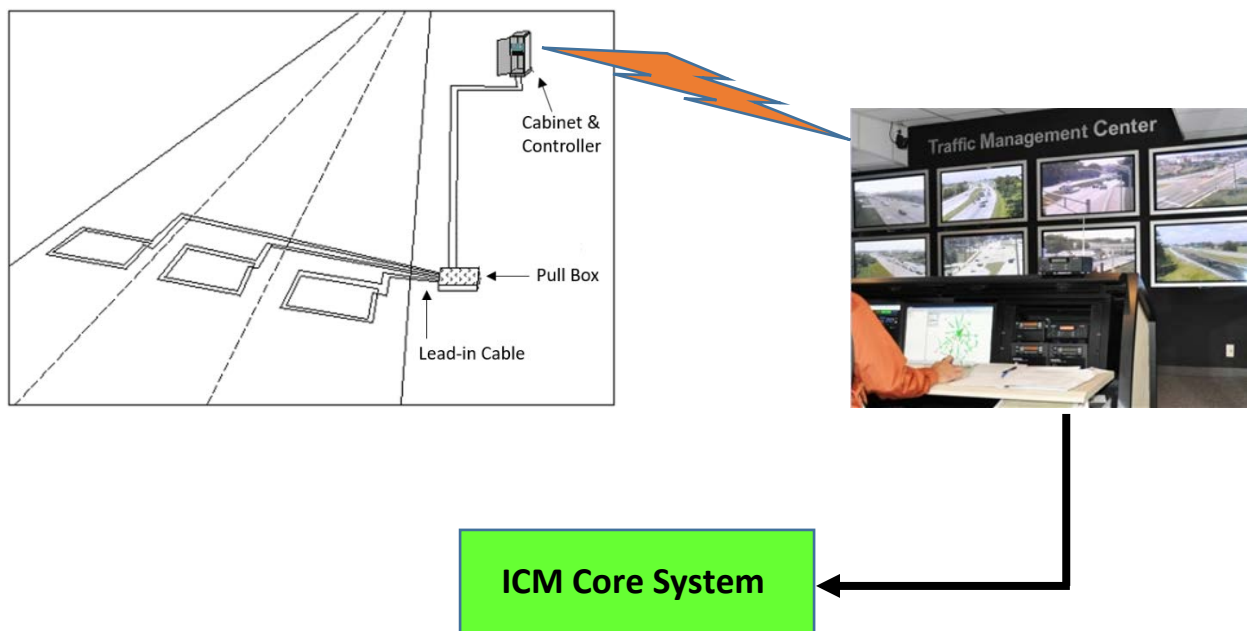


Figure 7: Simplified data path

6. GET INITIAL SENSOR DATA

The preceding steps are designed to ensure, to the extent possible, that data received from an agency matches the physical world and is communicated reliably. This step focuses on actually receiving initial data from loop detectors. The process is:

First:	Then:
<ul style="list-style-type: none"> • Check data availability • Do we have any? • Are sensors communicating? 	<p>If less than 100% of sensors are reporting, provide a list of non-reporting sensors to the owning agency and ask for help to:</p> <ul style="list-style-type: none"> • Turn on the controller • Provide power • Turn on communication • Fix communication failures

7. VALIDATE THE METADATA AND DATA PATH

After receiving sensor data, check it for internal consistency to see if the data patterns make sense (i.e., do they match known traffic patterns such as prevailing commute directions, etc.):

- Metadata categorizes sensors into HOV lanes, mainline, on/off-ramps.
- Does the data show the sensors in the right categories and locations?

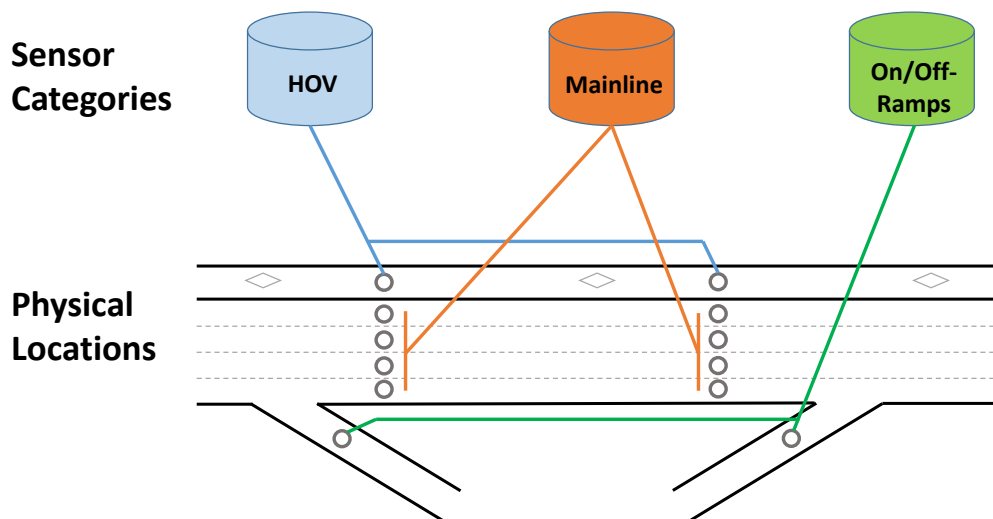


Figure 8: Data pattern match

1. Apply the k -nn (nearest neighbors) algorithm to automate the task of placing the sensor in the most likely category. (k -nn is a machine learning algorithm used for classification.)
2. Review the result and look for instances where the category assigned by the algorithm doesn't agree with the metadata.
3. If needed, request that agency personnel fix metadata where sensors are incorrectly assigned to the:
 - Wrong category
 - Wrong side of the road

8. CHECK SENSORS' TRUSTWORTHINESS

Can the sensor data be relied on? To check the trustworthiness of the sensors, examine the flow balance of the measured area over 24 hours.

- Is the traffic entering the area generally balanced by the traffic leaving the area?
- Do the numbers add up?
- Is there a noticeable bias toward input or output?

For example, if sensor data indicates that 500 vehicles enter a freeway segment but 300 exit the segment, that suggests a potential problem with the sensor measurements. In Figure 9, if the flow were completely balanced, then $A + D = B$, and $A + D = C + E$:

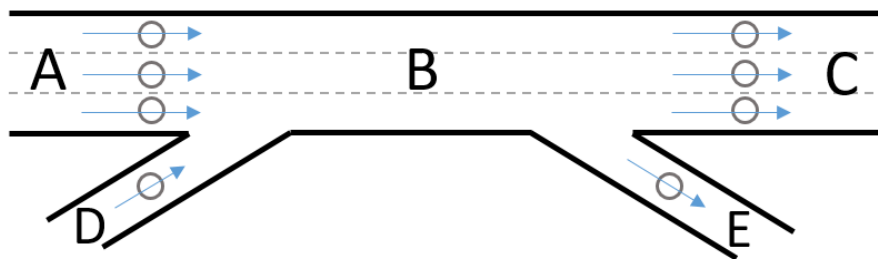


Figure 9: Sample freeway segment with traffic input and output

9. FILTER THE DATA

What data do we keep? What data should be permitted to flow through to be used by the ICM system?

- Provide a score for each sensor, based on:
 - Pattern match
 - Flow balance within 5%
 - Threshold

10. FILL IN BLANKS WHERE POSSIBLE

It is not unusual for data to be missing due to dead loops. For example, suppose only 4 of 5 loops in a VDS are working:

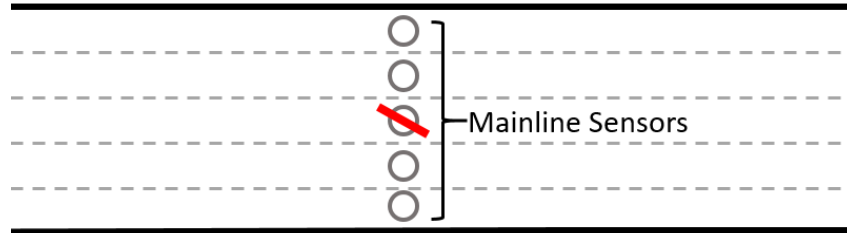


Figure 10: Freeway VDS with 4 of 5 loops working

To fill in the missing data:

1. Perform flow imputation (algorithm that estimates the missing data).
2. Scale the data.
3. See if it is still above the threshold.
4. Check the flow balance again (repeat step 8) to see if it looks reasonable.

11. CONVERT THE DATA INTO A USABLE FORM

The raw data received from the loops must be converted to data usable by the models.

Loops provide:	Which must be converted to:
<ul style="list-style-type: none"> • Occupancy • Counts 	<ul style="list-style-type: none"> • Density • Flows • Speeds

12. REVALIDATE DATA QUALITY OVER TIME

The corridor is a living system. Roadways change as lanes are added or removed. New loops are installed and old ones replaced or abandoned. A process is therefore needed for revalidating data quality over time.