Identifying Lane Mapping Errors at Freeway Detector Stations

Lixin Lang
Graduate Research Assistant
Department of Statistics,
The Ohio State University,
Columbus, OH 43210,
e-mail: Lang.151@OSU.edu

Benjamin Coifman
corresponding author,
Associate Professor
Joint appointment with the Department of Civil and Environmental Engineering and Geodetic Science, and the Department of Electrical and Computer Engineering,
The Ohio State University,
Columbus, OH 43210,
phone: 614-292-4282;
fax: 614-292-3780;
e-mail: Coifman.1@OSU.edu

Abstract

With many loop detectors, numerous junctions, and site-specific configuration files, at a freeway detector station there are many opportunities for errors to occur in the mapping from a loop detector input to its physical location and lane in the freeway. If these configuration errors are not caught they propagate to the station's data and any control decisions based on those data. Great care is necessary to prevent such errors, yet they persist in the field even after the added expense to prevent them.

This paper develops and tests two algorithms to identify such errors, without using a priori knowledge about the detector station configuration. By correlating events between two loop detectors, the first algorithm tests all active detectors to both match pairs from dual-loop detectors and identify any single-loop detectors at the station. The second algorithm uses a hypothesis-test based K-means method to group lanes from common directions based on time series velocity. Both algorithms exhibited very good performance, as shown herein, they were tested over many days and detector stations. Although the analysis focuses on loop detectors, the methodology should also apply to emerging technologies that mimic loop detectors, e.g., wayside mounted microwave radar.

Keywords

Loop detectors, Freeway traffic, error detection, lane grouping

4965 words + 10 figures and tables = 7465 words

Introduction

Loop detectors are the most widely used sensors for collecting traffic surveillance data [1]. Each detector consists of one or more loops of wires embedded in the pavement, whose inductance is changed by the presence of vehicles. Using this phenomenon the electronics driving a resonant circuit through the loop detector can identify the presence or absence of vehicles over the loop and thus, report event data, i.e., individual vehicle arrival and departure times in a traffic lane. These event data are typically aggregated to measure traffic flow and occupancy in the lane. Vehicle speed can then be estimated from a single-loop using an assumed average vehicle length (e.g., see [2] for more information) or measured directly from the traversal time between two loops a few yards apart in the same lane, i.e., a dual-loop detector. Throughout this paper we use *loop detector* to refer to an individual loop that may be a single-loop detector or either of the two loops in a dual-loop detector.

While a single- or dual-loop detector is normally deployed in a single lane, a detector station uses detectors in each lane to monitor one or both directions of a freeway, potentially with additional detectors on ramps, connectors and other locations for additional information. Simple detector stations may have as few as two loop detectors while a complicated detector station could have 32 loop detectors or even more. Although operating agencies strive for standards to minimize the difference between detector stations, they have to be flexible enough to accommodate the wide array of geometries seen on freeways. To this end, the controller driving the detector station uses a configuration file to map each active electrical input to a location that includes direction, lane, and in the case of a dual-loop detector either to the upstream or downstream loop, e.g., controller input 17 might be mapped to the upstream loop in the eastbound lane 3 dual-loop. Any discrepancy between the configuration file and the physical wiring of the detector station will lead to measurement errors, e.g., continuing the example, contrary to the hypothetical configuration file perhaps controller input 17 is actually connected to the queue detector of an on-ramp. If these lane mapping errors are not caught they will propagate to the station's data and any control decisions based on those data. In some cases it might be easy to identify such errors, e.g., a queue detector should never exhibit speeds close to freeway free flow, though even this problem is potentially difficult to diagnose exactly which loops are mapped incorrectly. In other cases it might not even be easy to identify the presence of such errors, e.g., if an eastbound lane is swapped with a westbound lane.

Lane mapping errors may arise due to software, e.g., using the wrong configuration file, or hardware. In the latter case, there are multiple electrical junctions between the controller and the physical loops in the pavement. Each of these junctions is wired manually and each junction offers a possibility that wires may literally become crossed. Great care is necessary to prevent software and hardware errors, yet such errors persist in the field even after the added expense to prevent them. Such problems caused Caltrans to launch the detector fitness program, requiring considerable resources to manually verify the configuration of all detector stations.

This paper develops two algorithms to automatically identify many lane mapping errors with the goal of improving data quality and control actions taken to manage the freeway while decreasing the cost of identifying, diagnosing and preventing such configuration errors. The first algorithm reads the data from all reporting loops at the station over one day then checks to see if there are any dual-loop detectors at a given station and if so, pair the upstream and downstream loops in each lane. Any loops that go unmatched are then considered single-loop detectors. The second algorithm seeks to combine lanes (i.e., the output from the first algorithm) into different groups based on the direction of travel. Again, this algorithm uses an entire day to make the decision, but in this case the performance is enhanced by independently repeating the trials on several days. This paper examines 40 stations in detail and the methodology correctly found previously unknown configuration problems at two of them.

The algorithms require event data in the analysis. Although most agencies do not collect event data from their traffic monitoring systems, these data exist in almost all traffic controllers and it is envisioned that these tests would be incorporated into modern controller software.

The remainder of the paper is organized as follows, the first algorithm is developed to find dual-loop detectors and identify the upstream and downstream loops. Then the second algorithm is developed to group lanes by direction. Both algorithms are evaluated using field data from 45 detector stations. Finally the paper closes with conclusions.

Matching dual-loop detectors

The event data from a loop detector can be represented as a series of pulses, within each pulse the rising and falling edges denote a vehicle's arrival and departure times at the loop detector, respectively. The event data from a loop detector can be represented by a Kx2 matrix, X, whose elements $X(k,1)$ and $X(k,2)$ are the time of the rising and falling edges, respectively, of the k-th pulse. Let $X_1, X_2, \ldots, X_n$ denote the event data from all n loop detectors at a station. Depending on the configuration of the station, all, some or none of these n loops could come from dual-loop detectors.

The matching algorithm is developed so that each pair of operating detectors in a dual-loop at the station can be found and correctly identified as upstream and downstream without any knowledge of the configuration file. The algorithm progresses successively through the data from each of the n loop detectors and tests whether the given loop could be in the same lane upstream of one of the n-1 remaining loops.

Because the loops are closely spaced in a dual-loop detector, most vehicles will be recorded by the two loops, e.g., Figure 1. As discussed below, the traversal time for a vehicle should be proportional to the pulse duration at each of the detectors and the algorithm tallies the percentage of pulses between each pair of detectors that fall within a tolerance. If this percentage is high enough, the pair of detectors is considered to be a dual-loop detector and any detector that is not paired in this manner is considered a single-loop detector. This matching procedure consists of two steps, pulse pairing and pulse validating as described below.

Pulse pairing

The first step of the matching algorithm tests whether two pulses, one from a potential-upstream detector and the other from a potential-downstream detector, could have been generated by the same vehicle passing a dual-loop detector. The first pulse at the potential-downstream detector with a rising edge time after the rising edge of the pulse at the potential-upstream detector is found. If the detectors are working reliably and are indeed in the same lane, the delay between these two rising edges, Δt_a, as shown in Figure 1, is bounded by some minimum traversal time,

$$\delta = \frac{L}{V_{\max}} \quad (1)$$

where $V_{\max}$ is the maximum feasible vehicle speed and L is the dual-loop detector spacing (or the minimum feasible spacing if the exact spacing is not known). For this paper $V_{\max}$ was set to 85 mph and L to 20 feet. Similarly, the minimum feasible delay between the two falling edges, Δt_d will also be bounded by Equation 1.

Pulse validating

The next step of the matching algorithm further checks the feasibility of the matched pairs and discards the unlikely ones. The pulse duration (also known as the on-time), t_p, for a vehicle passing over a loop detector is a function of that vehicle's speed, v_i, and effective length, l_i, namely,

$$t_p = \frac{l_i}{v_i} \quad (2)$$

While it is impossible to measure l_i from a single-loop detector without knowing v_i, most vehicles on urban roadways are passenger vehicles and l_i should be close to 20 ft for these vehicles (the exact effective length depends on the detector sensitivity). Most of the remaining vehicles will be trucks and busses, with an effective length longer than 20 ft. Meanwhile, the traversal time between the two loops of a dual-loop detector is given by,

$$T_a = \frac{L}{v_i} \quad (3)$$

If the two loops from the pulse pairing happen to indeed be correctly oriented as a dual-loop detector, then $\Delta t_a = T_a$. Otherwise, since the traversal time is much smaller than the headway, in general Δt_a will be much larger than T_a. Thus for a dual-loop detector,

$$\Delta t_a \leq \lambda \cdot t_p \quad (4)$$

where λ is a scale factor to limit the disparity between Δt_a and t_p arising from varying l_i and fixed L. After calibrating λ for the passenger vehicles this equation uses conditions at one detector to establish an upper-bound on the traversal time between the two detectors. This paper uses $\lambda=3$ because $L \sim l_i$ for passenger vehicles and $L<l_i$ for longer vehicles, thus, no true matches should be rejected by Equation 4. On the other hand, if the two loops under consideration are not oriented as a dual-loop detector then Equation 4 will be violated frequently.

Formalizing this process, each pair of pulses from the pulse pairing that meet the criterion of Equation 4 is taken to support the supposition that the two loops are a dual-loop with the assumed orientation. Figure 2 provides an example of this filtering process using data from a dual-loop detector. In the first case, the two loops are oriented correctly and all of the pulses are retained. In the second case, the two loops are swapped from their correct orientation, for many of the presumed-upstream pulses note the long time lag until the next presumed-downstream pulse, now only 4 of 13 pulses are retained.

Final decision and Reliability index

For a given pair-wise test, let X_i denote 24 hrs of event data from the presumed-upstream detector and X_j the corresponding event data from the presumed-downstream detector. The matching ratio is defined as,

$$R_{ij} = M_{ij} / K_i \tag{5}$$

where K_i is the number of pulses observed at the presumed-upstream detector and M_{ij} is the number of those pulses that are retained by both Equations 1 and 4. Let

$$\begin{aligned} H(i) &= \arg\max_{j \neq i}\{R_{ij}\} \\ R_i &= R_{iH(i)} \end{aligned} \tag{6}$$

the pair (X_k, $X_{H(k)}$) is declared to be dual-loop detector only if R_k is larger than a threshold value, where X_k is the upstream loop and $X_{H(k)}$ is the downstream loop. For this paper the threshold value was set empirically to 0.80. Few R_{ij} were close to this threshold for the λ used in this study, correctly matched dual loop detectors generally had R_{ij} over 0.90, while incorrectly matched loops generally had R_{ij} below 0.50. One can then define the reliability index, R_k, for the pair (X_k, $X_{H(k)}$). The larger the index is, the more reliable the conclusion is.

Using data from I-71 in Columbus, Ohio, Figure 3A shows the schematic for station 1009, which is just downstream of a diverge between I-71 and I-670. The latter split also has an on-ramp entering from the Broad St. overpass, denoted with prefix "r". The loops on I-71 are denoted with prefix "n" for freeway northbound and the loops on I-670 are denoted with prefix "w" for freeway westbound. Similarly, the upstream loops have suffix "u" and downstream loops have suffix "d". Figure 3B shows the matching ratio matrix $\{R_{ij}\}$ from this station on an arbitrarily chosen date, August 10, 2002. All R_{ij}s larger than 0.80 are all identified with values shown above the corresponding bars in Figure 3B, in fact all upstream/downstream pairs are correctly identified in this example with R_i's greater than 0.90. For example, the first tall bar on the left, with 0.999 shows that presumed-upstream ru is indeed upstream of presumed-downstream rd in the same lane with very high confidence. One may note that the loops on I-71 tend to have larger R_{ij} for the non-matching loops than the loops on I-670. All of these non-matches on I-71 are below 0.40, the fact that they are larger than those on I-670 is due to the fact that the average daily traffic on the I-71 lanes is about ten times greater than that on the I-670 lanes, so any presumed-upstream pulse has a much greater chance of finding a matching pulse within the search window when one of the I-71 lanes is the presumed-downstream loop. Also worth noting on this figure is the fact that w1u has a high R_{ij} with w2d, almost 0.57. We believe this transient is due to splash-over, i.e., w2d is detecting some of the vehicles that are actually passing over w1d, although we have not validated this hypothesis yet. If it stands the test of further scrutiny, then it would indicate a valuable extension of this methodology to also detect operational errors in addition to configuration errors.

Figure 4 repeats this process for another station, 1003, on the same day. This station is somewhat unusual in the fact that the eastbound lanes (prefix "e") are all single-loop detectors while the westbound lanes (prefix "w") are all dual-loop detectors. The matching ratio matrix $\{R_{ij}\}$ in Figure 4 shows that the matching algorithm correctly identified the dual-loop detectors while correctly leaving the single-loop detectors unmatched.

Extending this process to all 45 detector stations on the I-71 corridor, and repeating the process on seven consecutive days, Table 1 tallies the results. Note that about one third of the detector stations are equipped with dual-loop detectors, with the remaining stations being single-loop detector stations. The number of loops changes slightly over the observation period due transient errors impacting the operational status of some of the loop detectors. Across over 420 loop detectors, all of the dual-loop detectors that it identified were indeed dual-loops. While on four of the seven days it failed to match one particular dual-loop detector pair out of 107 dual-loops reporting. This one error is due to the fact that the upstream detector reported much fewer pulses than its corresponding downstream detector in the same lane, i.e., a pre-existing detector fault.

Grouping detectors of like direction

In order to classify groups of lanes at a detector station by their direction, the algorithm uses time series speed in each lane at the station, where a lane is either a single-loop or dual-loop resulting from the loop matching algorithm. Generally speaking, vehicle speeds among different lanes in a given direction exhibit similar trends, decreasing and increasing at similar times. For example, Figure 5A-B shows the time series from the six lanes at detector station 1002 on Aug. 10, 2002, with each sub-plot showing the three lanes in the given direction. As can be seen, the curves in one direction are much more similar than they are to the other direction, the most evident feature being the fact that the eastbound speeds drop around 11:00 while the westbound do not. Note that this figure deliberately excludes all of the data between midnight and 6:00, when prolonged, extremely low flows can lead to large errors in speed measurement or estimation.

The similarity of speed drops and recoveries between lanes in a common direction provides a good measure for grouping them, albeit not a perfect one. More formally, large correlation values among a group of lane speed data is used as an indicator that these data come from a particular direction. Thus, grouping correlation values into different clusters is equivalent to grouping lanes based on the above assumption.

The K-means method is a classical approach used in cluster analysis [3]. It iteratively partitions a data set into k distinctive clusters to minimize the sum of squared distance between each datum to its closest cluster centroid (Euclidean distance). During the iteration, clusters are updated until there is no change to any of the k clusters. While both a simple and powerful tool, the method is complicated by the fact that the cluster quantity, k, needs to be pre-decided. In this paper a hypothesis-test method is used to determine the best value of k for a given data set. The details of this process are presented below.

Grouping method

Traffic speed data are taken as the basis for calculating correlation between lanes and the subsequent grouping of lanes. This study takes the median of the individual vehicle speeds over a sampling period of 1 minute. If the detector matching established that a given lane has a dual-loop detector, the measured speeds from these paired detectors are used, otherwise, the speed is estimated from the single-loop detector [2]. Each time a dual-loop detector is used it has the additional advantage of reducing by one the number of detectors (or lanes) remaining to be clustered. If no datum is available during a given period this sample is replaced with the mean value of its immediately adjacent time intervals to smooth over short communication outages or drops in demand of a few samples[1]. Though this process of finding the median speed has a smoothing effect, the resulting time series is still noisy, e.g., the data shown as points in Figure 5A-B. Thus a hamming-window filter of length 11 is applied to further smooth the data while conserving the data trend [4], as shown with a solid line for each lane in Figure 5C-D.

Lane grouping proceeds by forming groups of lanes with the similar velocity time series one lane at a time and then combines the obtained groups to make an overall clustering. Specifically, for the speed data time series on lane i, i = 1, 2… m, where m is the number of lanes (including both mainline and ramps), the grouping method first calculates the correlation value γ_{ij} between the two smoothed time series for all $j \neq i$. Then it uses the following steps, commonly known as hypothesis-test clustering [5], to form a set, S_i, which contains detector i and any detectors deemed to be of similar direction.

Step 1: Gather the correlation values to be clustered for lane i:
$$\mathrm{Z} = \{\gamma_{i1}, \cdots, \gamma_{i,(i-1)}, \gamma_{i,(i+1)}, \cdots, \gamma_{im}\}$$

Step 2: Remove any negative entries from Z, denote the resulting set
$$\mathrm{Z}^{+} = \{\alpha_1, \alpha_2, \cdots, \alpha_n\},\ n \le m-1$$

Step 3: Repeat the following substeps 3a-3c for k = 1, 2, 3,

3a) Cluster set Z^{+} into k groups using K-means [3], the resulting groups,
$$\mathrm{Z}_1^{+} = \{\beta_{11}, \beta_{12}, \cdots, \beta_{1N(1)}\},\ \ldots,\ \mathrm{Z}_\mathrm{k}^{+} = \{\beta_{k1}, \beta_{k2}, \cdots, \beta_{kN(k)}\},\ \text{where } \sum_{I=1}^{k} N(I) = n$$

3b) Calculate the Sum of Squared Error, SSE_k,
$$\text{let } \mu_\ell = \frac{1}{N(\ell)}\left(\beta_{\ell 1} + \beta_{\ell 2} + \cdots + \beta_{\ell N(\ell)}\right),\ \ell = 1, \cdots, k$$
$$\mathrm{SSE}_k = \sum_{I=1}^{k} \sum_{J=1}^{N(I)} \left(\beta_{IJ} - \mu_I\right)^2$$

3c) Without loss of generality, define,
$$\mathrm{Z}_\mathrm{k}^{*} = \left\{\mathrm{Z}_\ell^{+} : \mu_\ell = \max(\mu_1, \cdots, \mu_k)\right\}$$

Step 4: Hypothesis testing
$$\mathrm{H}_0:\ \mathrm{k} = 2$$
$$\mathrm{H}_1:\ \mathrm{k} = 1$$

[1] If large gaps are present in the data one should first fix any chronic detection problems before applying these tools. If large gaps are due to transient errors, one should choose time periods when the large gaps are not present.

Test statistic: $t = \frac{SSE_1 - SSE_2}{1} \div \frac{SSE_2}{n-2}$

If $F_{1,n-2}(t) \geq 0.70$, accept H_1 and proceed to Step 6, otherwise, continue to Step 5, where $F_{m,n}(t)$ is the cumulative distribution function (CDF) of the F distribution with degree of freedom (m, n).

Step 5: Hypothesis testing

$H_0: k = 3$

$H_1: k = 2$

Test statistic: $t = \frac{SSE_2 - SSE_3}{1} \div \frac{SSE_3}{n-3}$

If $F_{1,n-3}(t) \geq 0.80$, accept H_1, otherwise accept H_0.

Step 6: Set $S_i = \{j : \gamma_{ij} \in Z_k^*\} \cup \{i\}$

In Step 3, k is limited to a maximum of three because at a single-loop detector station m may be under 6 and it is uncommon that a given station will have more than two mainline directions and assorted ramps. After repeating steps 1 through 6 for each lane, the set S_i contains lane i and lanes of like direction chosen from the hypothesis-test clustering above. For each lane i, a superset, g_i is formed of all lanes j such that $S_i \in j$ and $S_j \in i$. If two or more supersets intersect, the union is taken to form a group of lanes and then these supersets are removed from further consideration. Each remaining, ungrouped superset containing more than one lane also becomes a group of lanes and is removed from further consideration. All of the supersets that have not been grouped by this stage must contain only one lane and these remaining supersets are classified as ungrouped. In this manner, the grouping algorithm chooses the number of groups for a given station without a priori knowledge of the configuration or geometry.

Grouping performance

To improve the fidelity of the grouping algorithm several successive days are processed independent of one another and then the aggregate results at a given station are tallied. Table 2 shows the results for two single-loop detector stations. Each row shows the number of times the lane is grouped with the other lanes at that station over a 14 day period, August 11-24, 2004. Except for the final column that lists the number of times that lane went ungrouped, the matrices must be symmetric. Station 1004 is on I-70/71 and one can see that all of the eastbound lanes, e-, were grouped together on at least 10 of 14 days while the westbound lanes, w-, were grouped together on at least 12 of 14 days. But this first matrix also illustrates the fact that the algorithm is not perfect, e.g., e3 is grouped with w2 on 2 of 14 days. By sampling several days independently the methodology becomes more robust to such transient errors.

But some systematic problems persist even after aggregating over many days. The first arises when speeds normally differ across neighboring lanes in the same direction, e.g., immediately upstream of a diverge or lane drop. At these locations the algorithm is likely to split one direction into two groups. But the two groups should usually be distinct from the traffic in the opposing direction, e.g., the algorithm produces three groups when there are only two directions. Or in the case of station 2 on I-71, the second matrix in Table 2, it splits both directions into two groups.

There is a lane drop approximately 500 ft downstream of the station in the northbound direction, n-, and as a result lane n4 usually exhibits different traffic patterns than the other three lanes. A southbound lane, s-, is added from an on-ramp approximately 50 ft upstream of the station and as a result, s4 also exhibits patterns different than the other lanes in the given direction. These facts are evident in the high number of times that n4 and s4 are ungrouped, shown in the final column. The remaining lanes in the given directions are grouped correctly for at least 10 of 14 days.

Extending this procedure to all of the operational stations on the I-71 corridor during the same 14 days, the first five columns of Table 3 show the results for the mainline lanes. Here a detector is considered grouped with other detectors if they were grouped together for at least 70 percent of the days that the station reported data. The choice of this threshold is simply for brevity of presentation across 40 stations, in practice one should study the detailed results such as presented in Table 2. Many stations also had ramp detectors that were included in the grouping algorithm making it more difficult to correctly group the mainline lanes (again for brevity, the grouping results for the ramp detectors are not shown on the table). Three stations did not report any data over this 14 day period (21, 34 and 1008) and two stations have chronic faults that prevent accurate speed estimation (8 and 12) and thus also preclude grouping, all five of these stations are excluded from the table.

The table shows that over the 40 stations and 255 main-line lanes monitored, the grouping algorithm successfully segmented all of the lanes in opposing directions. But as already noted, the grouping algorithm occasionally over-segments the lanes in a given direction. In fact it does so an average of 33 percent of the time, as noted at the very bottom of the table.

Which leads to the second systematic problem. Obviously if neither direction of a two way freeway experiences a delay in a given day, both directions will exhibit similar speed time series and will likely be correlated to one another in the grouping algorithm. Keeping track of whether any congestion was observed at the station and discounting the results if none occurred can reduce this problem. Table 3 is sorted by the sixth column, which tallies the number of days that had at least three 5 min samples with speeds below 40 mph, in either direction, between 6:00 and 24:00. The first sixteen rows met this criterion at least 5 of 14 days. Restricting the analysis to just these rows the performance improves, with only 16 percent of the lanes going ungrouped. Although included in the total, seven of the ungrouped lanes at these 16 stations are due to known geometric features as discussed above with respect to station 2.

Now consider the lower 24 stations in the table. On most days each of these stations saw free flow conditions throughout the entire day, making difficult the task of differentiating between directions. In fact all of the stations with 3 or fewer days of congestion (out of 14 days) all had at least one erroneously ungrouped lane. But the free flow traffic will inevitably be broken occasionally by non-recurring congestion, which can be a valuable clue. Consider station 15 on I-71, the top of Table 4 shows the results over all 14 days. One can see that the lanes were generally grouped correctly, though none of the northbound lanes met the 70 percent criterion to be listed as grouped in Table 3. Compared to Table 2, opposing lanes were grouped together more often, e.g., s3 and n3 were grouped together 5 of 14 days. On three of the days, August 19, 20 and 24, long queues stretched through several of the stations that otherwise saw free flow conditions. The final column of Table 3 shows the number out of these three days that had at least three 5 min samples with speeds below 40 mph, in either direction, between 6:00 and 24:00.

Indeed, the two congested days from station 15 were in this smaller group. Repeating the analysis on just these three days yields the matrix at the bottom of Table 4. The difference between the opposing directions becomes much more pronounced through this judicious choice of days.

Table 5 shows the 13 stations impacted by the queues. The first five columns repeat the results from Table 3, still sorted by the number of days out of 14 that were congested, while the last three columns repeat the analysis on just the three incident days with the criterion for grouping dropped slightly to 66 percent, or 2 of 3 days. Once more, no lanes are grouped with the opposing direction and the ungrouped lanes drop to only 4 percent.

Conclusions

There are many opportunities for errors to occur in the mapping from loop detector to physical location and lane in the freeway. If these configuration errors are not caught they will propagate to the station's data and any control decisions based on those data. This paper developed and tested two algorithms to identify such errors, without using a priori knowledge of the detector station configuration.

The matching algorithm looks for possible dual-loop detectors based on the similarities between event data at any two detectors at the given station. It then matches any dual-loops it finds, classifying the remaining detectors as single-loops. On a trial of over 420 loops, on seven successive days, the algorithm only missed one out of 107 dual-loop detectors and this omission was due to a fault in one of the loops. Meanwhile, the algorithm made no incorrect dual-loop matches. In fact at two dual-loop detector stations (1002 and 1005) the matching algorithm yielded results that differed from the configuration file, upon consultation with the operating agency these two stations proved to be wired incorrectly and the algorithm found the correct mapping. Figure 3 also suggests the potential for the matching algorithm to find splash-over problems, though developing such an application is left for future research.

Next, the grouping algorithm uses the correlation between the time series velocity in different lanes to identify lanes that are in the same direction. The grouping algorithm successfully segmented the lanes in opposing directions for all 40 detector stations examined. However, it also over-segmented approximately 33 percent of the lanes. These errors could be manually checked, e.g., comparing the grouping output as in Table 2 against the station configuration. The error was reduced to 16 percent when limited to stations that showed at lest a few samples below 40 mph for 5 of 14 days and almost half of these remaining errors were due to the freeway geometry causing neighboring lanes to exhibit different traffic patterns, e.g., just upstream of a lane drop. At other stations that were generally freely flowing, the error rate was further reduced to 4 percent by limiting the analysis to only the few days with non-recurring congestion.

The algorithms could easily be incorporated into modern controller software to continuously monitor conditions. Of course the lane mapping errors only occur when the configuration is modified, so the value of the algorithms come primarily after each field visit by operating agency personnel, but the cost of running the algorithms at other times is trivial. The value of continuously running the algorithms may be largest at stations with no recurring congestion, e.g., at the end of the month an operating agency could go back and review the grouping results for just those days that exhibited non-recurring congestion. Although it is impossible to address all contingencies, if deployed correctly the algorithm should catch many configuration errors. Fortunately, the configuration of a detector station should rarely change without human intervention, so if a false alarm does occur it can be silenced after verification and remain silenced until after the next time a technician visits the detector station. Nonetheless, the algorithms could also be coded on a field test unit such as a laptop PC that monitors the performance of the controller and the configuration of the detector station for only a short while and is then removed.

Finally, the algorithms presented herein are intended to catch configuration errors. Any unexpected results from the algorithms should be checked to verify that there are no detector faults that preclude a detector from providing accurate measurements, which may also impact the performance of these algorithms, e.g., [6].

Acknowledgements

This material is based upon work supported in part by the California PATH (Partners for Advanced Highways and Transit) Program of the University of California, in cooperation with the State of California Business, Transportation and Housing Agency, Department of Transportation. The Contents of this report 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. This report does not constitute a standard, specification or regulation.

References

[1] ITE (Institute of Transportation Engineers), (1998), *Traffic Detector Handbook. 2nd Edition*, Washington D.C.

[2] Coifman, B., Dhoorjaty, S., and Lee, Z. "Estimating Median Velocity Instead of Mean Velocity at Single Loop Detectors", *Transportation Research: Part C*, vol 11, no 3-4, 2003, pp 211-222.

[3] Hartigan J. A., *Clustering Algorithms*, Wiley, New York, 1975.

[4] Oppenheim A.V., Schafer R.W., and Buck J.R., *Discrete-time signal processing*, Prentice Hall, N.J., 1999.

[5] Kutner M. H., Nachtsheim C., Neter J., *Applied linear regression models, 4th edition,* McGraw-Hill/Irwin, Boston, New York, 2004.

[6] Coifman, B., Dhoorjaty, S., "Event Data Based Traffic Detector Validation Tests", *ASCE Journal of Transportation Engineering*, Vol 130, No 3, 2004, pp 313-321.

List of Figures

List of Tables

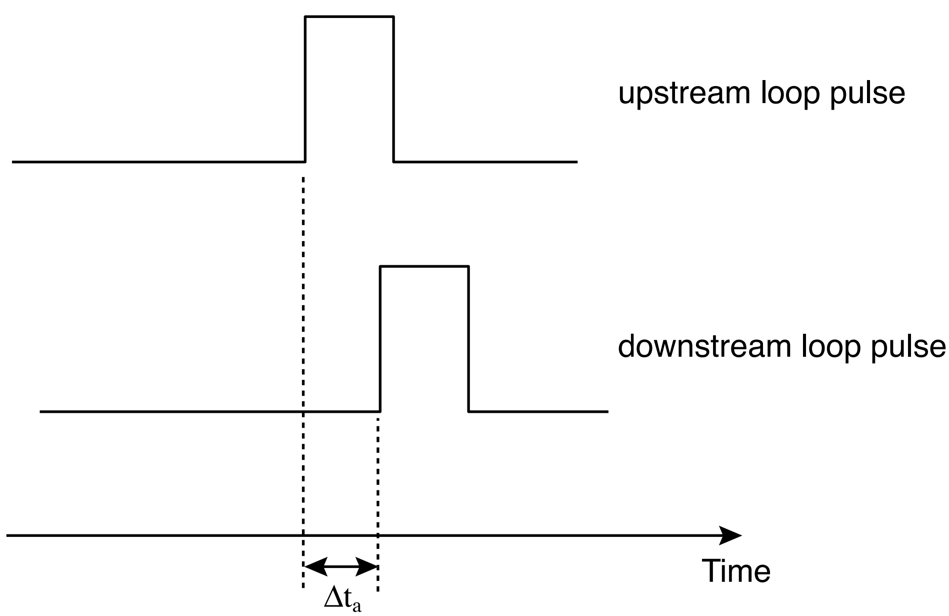


Figure 1, A hypothetical example of the detector response as a vehicle passes over a dual-loop detector.

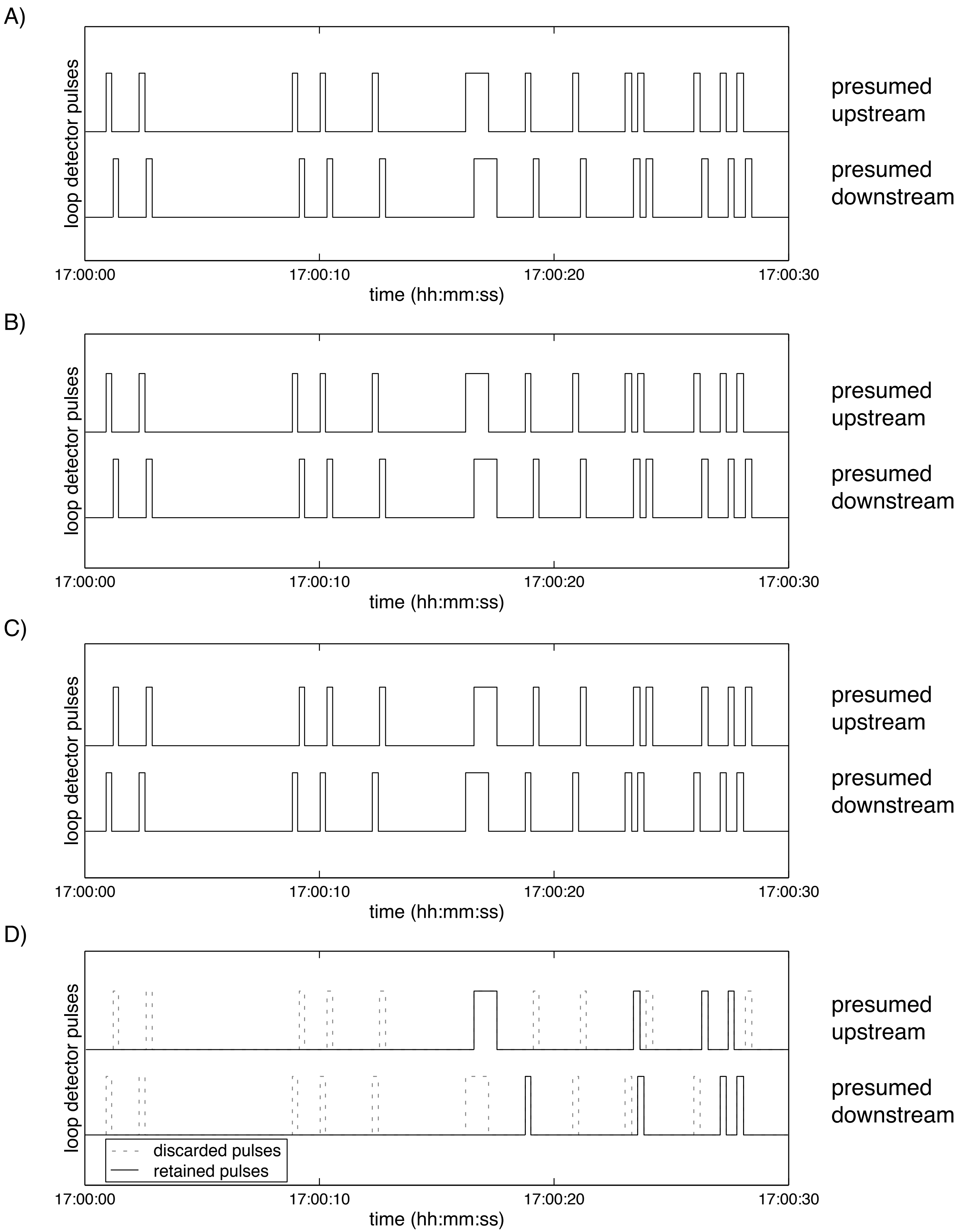


Figure 2, An example of matching pulses for a dual-loop detector with the correct orientation (A) before matching (B) retained pulses. Then repeating the process with the two loops swapped (C) before matching and (D) retained pulses.

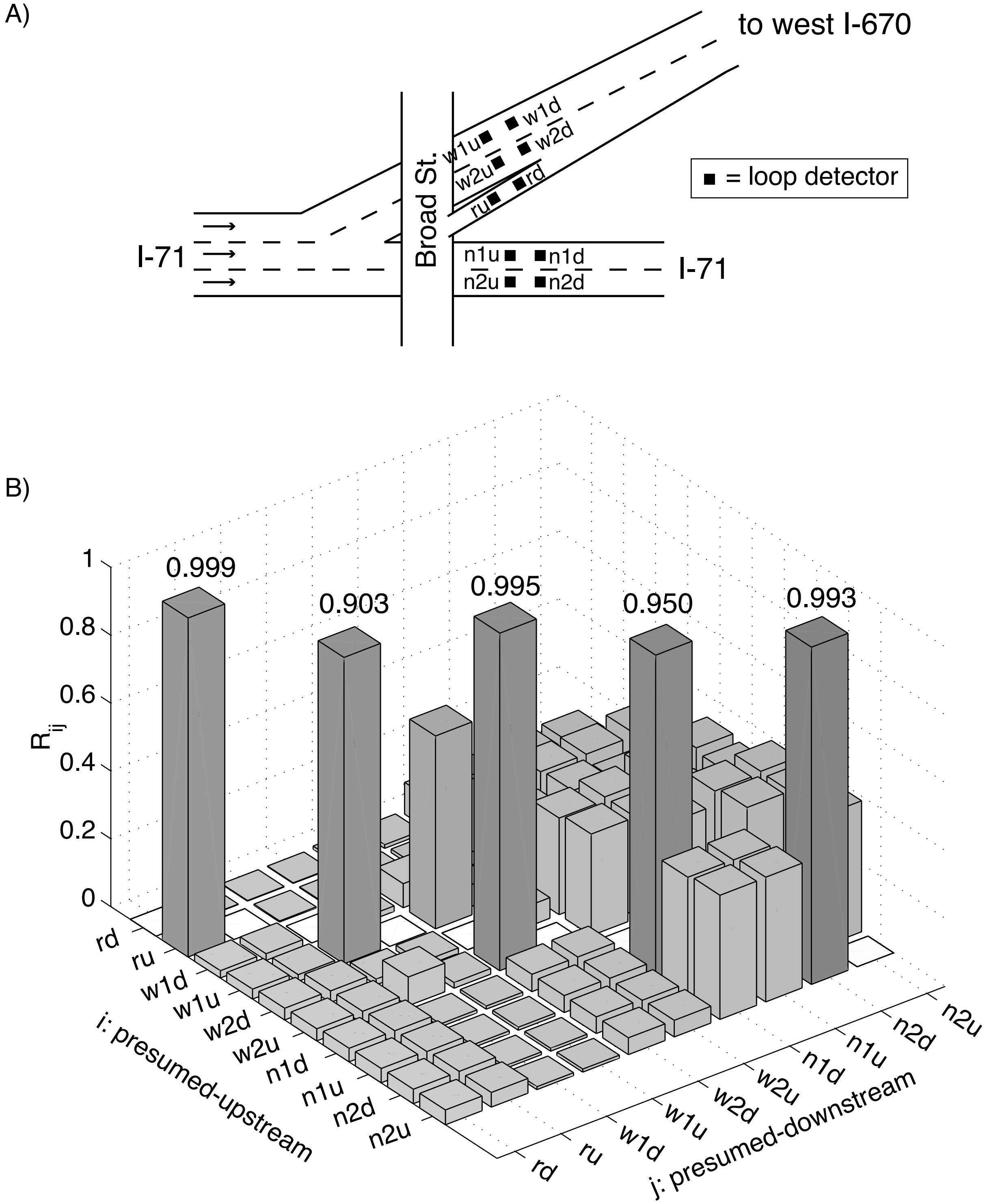
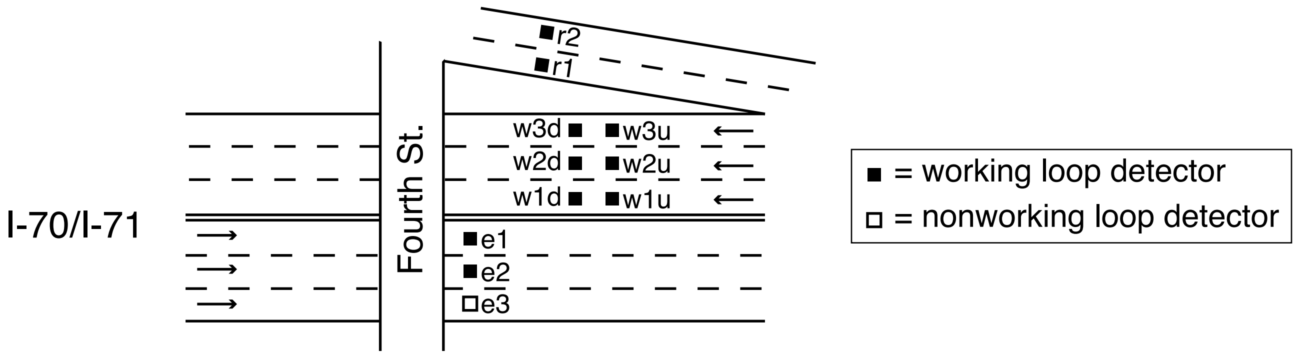


Figure 3, I-71 dual-loop detector station 1009 (A) Schematic, (B) matching ratio matrix from one day of data.

A)



B)

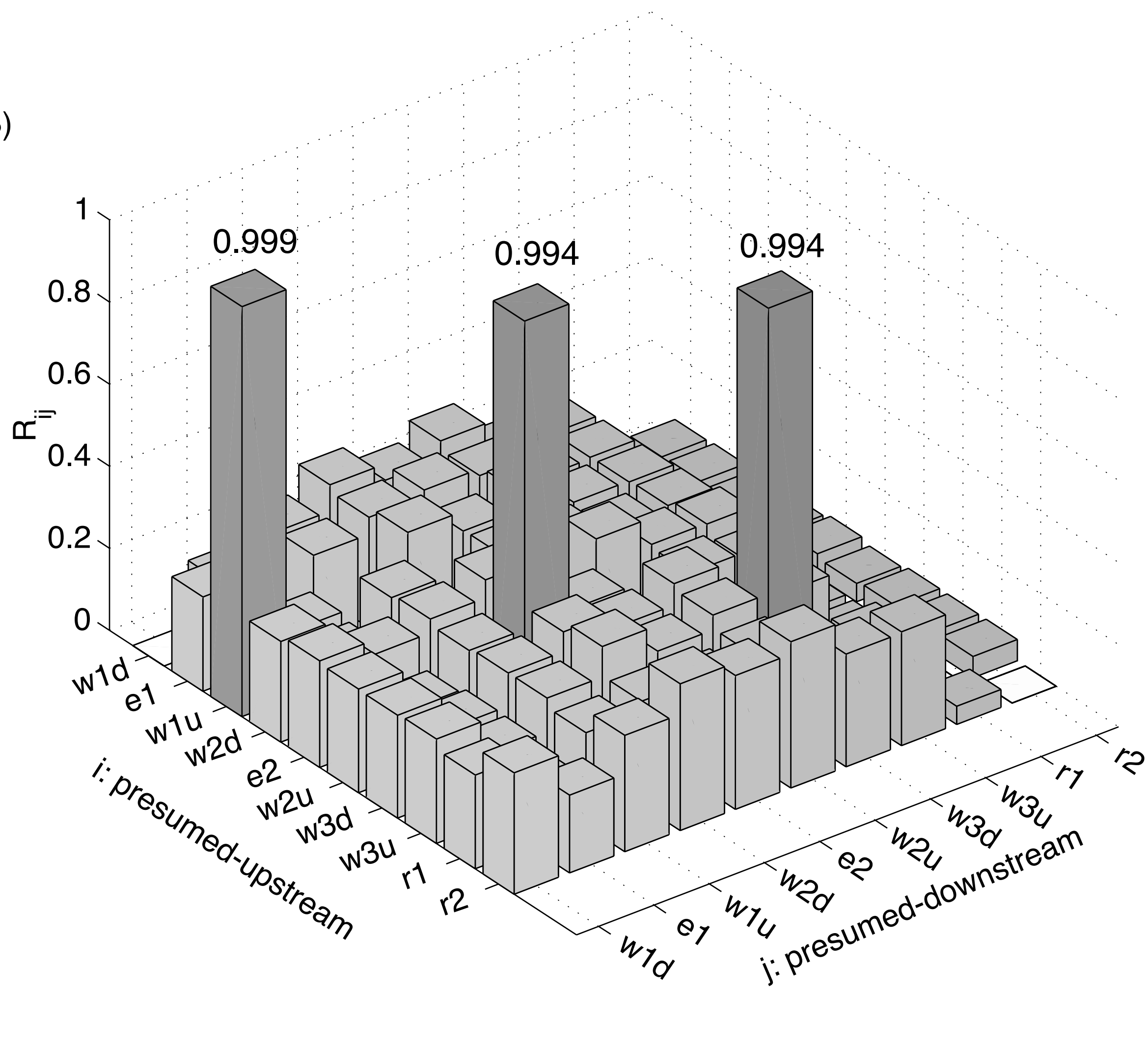


Figure 4, I-70/71 dual-loop detector station 1003 (A) Schematic, (B) matching ratio matrix from one day of data.

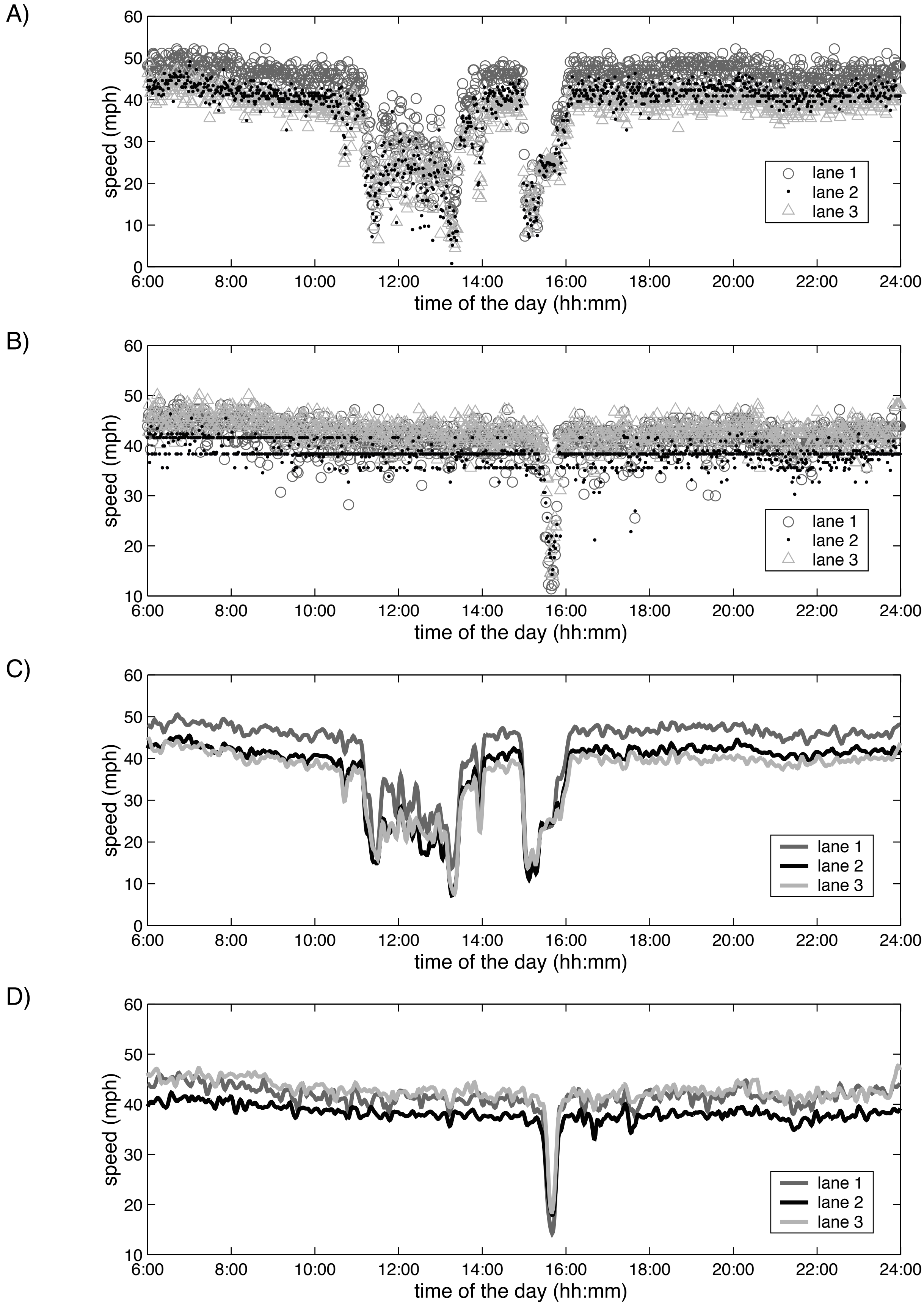


Figure 5, I-70/71 dual-loop detector station 1002 (A) raw speed eastbound, (B) raw speed westbound, (C) smoothed speed eastbound, (D) smoothed speed westbound.

Table 1, Summary of dual-loop detector matching results over 7 days across 45 detector stations.

Date	Number of loop detectors reporting	Pairs of dual-loop detectors reporting	Number of single-loop detectors reporting	Matching Error: Number of incorrectly matched loop detectors	Matching Error: Number of unmatched dual-loop detectors
August 10, 2002	422	107	208	0	0
August 11, 2002	423	107	209	0	1
August 12, 2002	422	107	208	0	1
August 13, 2002	426	107	212	0	1
August 14, 2002	426	107	212	0	1
August 15, 2002	426	107	212	0	0
August 16, 2002	426	107	212	0	0

Table 2, Sample grouping output at two stations over 14 days, August 11-24, 2004.

station 1004	e1	e2	e3	ramp (e off)	w1	w2	w3	ungrouped
e1	X	10	13	10	1	1	1	1
e2	10	X	10	10	1	1	1	4
e3	13	10	X	10	1	2	1	0
ramp (e off)	10	10	10	X	1	1	1	4
w1	1	1	1	1	X	12	12	2
w2	1	1	2	1	12	X	13	0
w3	1	1	1	1	12	13	X	1

station 2	n1	n2	n3	n4	s1	s2	s3	s4	ungrouped
n1	X	13	12	4	0	1	1	0	1
n2	13	X	12	4	0	1	1	0	1
n3	12	12	X	4	0	1	1	0	2
n4	4	4	4	X	0	0	0	0	10
s1	0	0	0	0	X	10	10	2	4
s2	1	1	1	0	10	X	14	2	0
s3	1	1	1	0	10	14	X	2	0
s4	0	0	0	0	2	2	2	X	12

Table 3, Grouping reslts from 14 days, August 11-24, 2004.

	mainline lanes				Days with at least three 5 min samples below 40 mph, either direction		
station	operational lanes	grouped correctly	ungrouped	grouped with opposing direction	all days	three incident days	notes
1002	6	3	3	0	14	3	
1005	4	4	0	0	14	3	
1004	6	6	0	0	13	3	
3	8	8	0	0	12	3	
2	8	6	2	0	10	3	(a)
4	8	7	1	0	10	3	(a)
1006	6	4	2	0	9	2	
1	8	7	1	0	8	3	(a)
1003	3	3	0	0	8	2	(b)
5	8	6	2	0	7	3	(a)
9	6	5	1	0	6	3	
1007	4	3	1	0	6	2	(a)
1009	4	2	2	0	6	2	
1010	4	4	0	0	6	2	
7	6	6	0	0	5	3	
1012	2	2	0	0	5	2	
6	6	5	1	0	4	3	
10	8	8	0	0	4	3	
11	6	6	0	0	4	3	
13	6	6	0	0	4	3	
32	10	5	5	0	4	0	
17	6	2	4	0	3	2	
18	6	5	1	0	3	2	
20	6	4	2	0	3	2	
14	6	5	1	0	2	2	
15	6	3	3	0	2	2	
16	6	5	1	0	2	2	(c)
19	6	5	1	0	2	2	
22	8	7	1	0	2	1	
23	7	3	4	0	2	1	
29	7	5	2	0	2	0	
31	9	3	6	0	2	0	
24	8	0	8	0	1	0	
28	7	0	7	0	1	0	
30	8	7	1	0	1	0	
25	8	6	2	0	0	0	
26	8	0	8	0	0	0	
27	6	2	4	0	0	0	
33	6	4	2	0	0	0	
1011	4	0	4	0	0	0	
first 16 stations	91	76 84%	15 16%	0 0%			
all 40 stations	255	172 67%	83 33%	0 0%			

(a) In these cases the ungropued lanes correspond to geometric features like lane drops or adds.
(b) Several eastbound lanes have gone down at 1003 since the data were collected for the earlier example in the text
(c) This ungrouped lane is due to the dual loop matching problem, mentioned in the text, occuring on most days.

Table 4, Sample grouping output at one station over 14 days, August 11-24, 2004 and again over just three of these days, including the two that exhibited congestion.

station 15 (all 14 days)	n1	n2	n3	ramp (n on)	s1	s2	s3	ramp (s off)	ungrouped
n1	X	8	8	0	1	2	2	1	5
n2	8	X	8	0	1	3	2	0	4
n3	8	8	X	2	4	4	5	2	0
ramp (n on)	0	0	2	X	0	0	0	0	12
s1	1	1	4	0	X	10	11	2	2
s2	2	3	4	0	10	X	10	2	2
s3	2	2	5	0	11	10	X	2	2
ramp (s off)	1	0	2	0	2	2	2	X	11

station 15 (3 incident days)	n1	n2	n3	ramp (n on)	s1	s2	s3	ramp (s off)	ungrouped
n1	X	3	3	0	1	1	1	0	0
n2	3	X	3	0	1	1	1	0	0
n3	3	3	X	0	1	1	1	0	0
ramp (n on)	0	0	0	X	0	0	0	0	3
s1	1	1	1	0	X	3	3	1	0
s2	1	1	1	0	3	X	3	1	0
s3	1	1	1	0	3	3	X	1	0
ramp (s off)	0	0	0	0	1	1	1	X	2

Table 5, Grouping results from three incident days, August 19, 20, and 24, 2004.

station	operational lanes	All days: grouped correctly	All days: ungrouped	All days: grouped with opposing direction	Three incident days: grouped correctly	Three incident days: ungrouped	Three incident days: grouped with opposing direction
9	6	5	1	0	6	0	0
7	6	6	0	0	6	0	0
6	6	5	1	0	6	0	0
10	8	8	0	0	8	0	0
11	6	6	0	0	6	0	0
13	6	6	0	0	6	0	0
17	6	2	4	0	4	2	0
18	6	5	1	0	5	1	0
20	6	4	2	0	6	0	0
14	6	5	1	0	6	0	0
15	6	3	3	0	6	0	0
16	6	5	1	0	6	0	0
19	6	5	1	0	6	0	0
all 13 stations	80	65 81%	15 19%	0 0%	77 96%	3 4%	0 0%