

# FAULTY LOOP DATA ANALYSIS/CORRECTION AND LOOP FAULT DETECTION

Xiao-Yun Lu, California PATH, U. C. Berkeley; Tel: 510-665 3644; email: xylu@path.berkeley.edu  
Pravin Varaiya, EECS, U. C. Berkeley; Tel: 510-642-5270; email: varaiya@eecs.Berkeley.edu Prof. Roberto  
Horowitz, ME, U. C. Berkeley; Tel: 510-642-4675; email: horowitz@me.berkeley.edu  
Joe Palen, Caltrans DRI, Sacramento, Tel: 916-654-8420; email: Joe\_Palen@dot.ca.gov

## ABSTRACT

Inductive Loops are widely used in California for traffic detection and monitoring. Extensive studies have been conducted to improve loop detection system performance. This paper reviews previous work on faulty loop data analysis for correction and faulty loop diagnosis. It is necessary to distinguish faulty loop data analysis and loop fault detection according to the data level. According to the level of data used, it divides the work in three levels: (1) macroscopic level as in TMC (Transportation Management Center) or PeMS (Performance Measurement System) in California, which uses highly aggregated data to look at loop problems related wide range; (2) mesoscopic level, which involves synchronized data for a section of freeways involving several control cabinet such as Berkeley Highway Lab (BHL); and (3) microscopic level: or at a control cabinet level including all the loop stations involved. Corresponding data correction methods are also briefly reviewed.

Keywords: *inductive loop, faulty loop data, loop fault detection, data correction and imputing*

## INTRODUCTION

Traffic management, traveler information, and transportation planning heavily rely on traffic data. Inductive Loops are widely used in California for traffic monitoring. The statewide *sensor system* consists of 25,000 sensors located on the mainline and ramps, and grouped into 8,000 vehicle detector stations (VDS). Each sensor records the presence of a vehicle above it. The measurements are electronically processed in the VDS to produce 30-second averages of vehicle counts or volume and vehicle occupancy. Over 90 percent of the sensors use inductive loops, most of the remaining use radar detectors. A systematic approach for loop fault detection, maintenance and retrieving reliable data to support traffic operation, traveler information and Corridor Management is crucial. However, loop data are not reliable. The error in the loop data obtained at TMC may be caused by fault at any point or several points from the loop to the TMC database. This presents great challenge to loop fault detection. To eventually solve this problem, it is necessary to take a systems approach. This approach is composed of three mutually complementary tasks: (a) loop fault detection; (b) faulty loop data correction/imputation; (c) loop detection system maintenance. This paper will focus on (a) and (b).

(a) Loop fault detection: How to efficiently and accurately detect and isolate the fault(s) in the loop detection system through data analysis and/or portable diagnostic tool working at the control cabinet level.

(b) Faulty loop data correction/imputation for reliable traffic data: How to temporarily correct/cleanse faulty loop data at different levels (traffic control cabinet, TMC) to maximally achieve reliable and accurate traffic data at TMC level with minimum time delay. If some data

are missed due to any reasons, how to impute some data based some neighbor station data and/or historical data.

Many methods have been adopted for fault detection and data correction/imputation for loop fault detection, and data correction/imputation. Different methods worked on different level of data in different ways. For example,

- Time aggregated data versus sub-second data
- At TMC level versus at control cabinet level
- Single loop stations versus dual-loop stations
- Synchronized adjacent lane data versus downstream/upstream data
- Historical data versus real-time data
- Raw loop data versus filtered/aggregated data
- Statistical methods versus deterministic filtering

This paper focus on the following points: (a) Systematic review of previous loop fault detection and data correction methods; and (b) Systematic classification of possible faults and causes in different levels. Although this review did not exhaust all the publications in this area, the reader could trace other publications further from the papers reviewed. The objective this review is to find merits and weakness of those methods which will be used as the basis for the development of a Loop Fault Detection Tool.

## **TRAFFIC MONITORING SYSTEM BASED ON INDUCTIVE LOOPS**

To systematically consider loop fault detection and data correction/imputation, it is necessary to possible faults at different level of the traffic monitoring system based on loop station. The overall picture for the data flow from individual loops to the TMC and PeMS in California can be described as: loop → pull-box → control cabinet with 170 + modem (30s data packets) → Tele-co line or wireless (up to 20 cabinet share one line) → Front-End-Process (FEPT) of ATMS of District TMC → PeMS. The system can be divided into three level: (i) Macroscopic Level: such as TMC/PeMS; (ii) Mesoscopic Level: A stretch of freeway such as BHL; and (iii) Microscopic Level at a Control cabinet. Different data are available at different system levels. For example, in California PeMS (Figure xx1), 2Hz data is available at the TMC and PeMS level, which are aggregates into 5 minutes data.

Systematic fault detection of the traffic monitoring system composed of loops needs the combination of diagnostics at different levels. As is shown in Figure xx1, the traffic monitoring system has a hierarchical structure for data collection, processing and passing. Inductive loops and other sensors are in the lower level. Data analysis at any higher level through data analysis can only diagnose the loop fault indirectly. This is because the data fault at higher level may be caused by communication system or any other operation on the data. This also suggest that it is necessary to

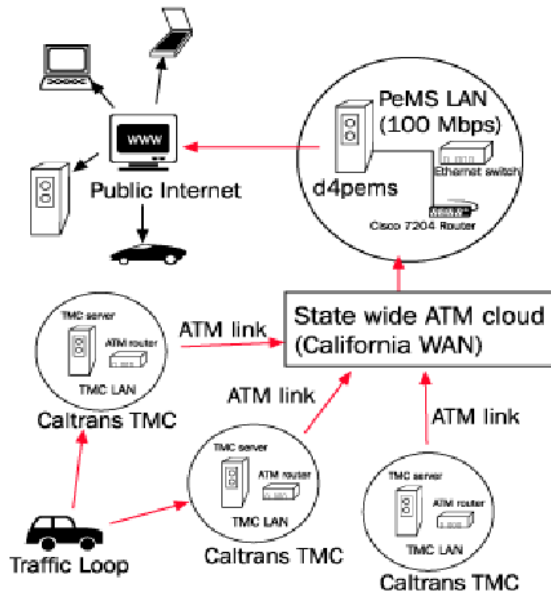
(a) to distinguish data analysis and data correction at higher level from loop fault detection since they are indirect;

(b) using higher level data analysis to identify suspicious control cabinet which may have potential faulty loop stations;

(c) to diagnose higher level problems such as communication system, power down, or data acquisition system or software;

(d) to combine higher level data analysis with lower level (onsite) loop fault detection using Portable Tool.

**Figure xx1.** PeMS in California



The error in the loop data obtained at TMC may be caused by fault at any point or several points from the loop to the TMC database: physical loop, connection of the loop and the control cabinet, loop card including sensitivity, traffic data (occupancy, count, speed) estimation method from the pulse signal, communication media between control cabinet and TMC. It can also be envisaged that data analysis at any point other than the control cabinet can only indirectly diagnose the loop fault in the sense that faults at higher level would interfere with fault loop data analysis. Thus those methods claimed for loop fault detection based faulty loop data analysis are essentially fault diagnosis of the monitoring system, which can be called *indirect methods*. Only those methods or tools

used at the control cabinet level can be called *direct methods*. Direct detection must be performed by a *portable tool* at the control cabinet level, which needs to have the following functions:

- (i) to generate ground truth based on some independent sensor;
- (ii) to synchronize the detection of the loops connected to the control cabinet with the ground truth detection;
- (iii) to compare the loop data with ground truth for diagnosis.

## **FAULTY DATA ANALYSIS/CORRECTION AND LOOP FAULT DETECTION**

This section classifies and reviews previous study on faulty loop data analysis and loop fault detection at different levels of the system, which corresponds to using different level of aggregated data. Since data correction/cleansing and imputation are usually closely related to data analysis and detection, they will be briefly reviewed and classified in parallel. Previous work on loop fault detection and data correction/imputation can be divided according to the data levels: macroscopic, mesoscopic and microscopic.

### **Macroscopic Level**

Typical example is the PeMS level or Caltrans District TMC level, which provide 30 second and 5 minute aggregated data. Each Caltrans District is composed of multiple highway corridors. The main characteristics of those data are that (a) they are the data practically used for traffic management such as ATMS and control; (b) heavy data aggregation are usually involved; (c) those data usually

need to pass long distance communication media to reach PeMS/TMC; (d) the data will be subject to small time delay due to data processing and passing through the communication.

### Faulty Loop Data Analysis Based on PeMS Aggregated Data

PeMS data DSA (Daily Statistics Algorithm) checking for data errors [1]:

- The number of samples in a day that have zero occupancy must be less than a certain threshold;
- The number of samples in a day that have occupancy greater than zero and flow equal to zero must be less than a certain threshold;
- The number of samples in a day that have occupancy greater than a certain value (PeMS uses 35%) must be less than a certain threshold;
- The entropy of occupancy samples must be greater than a certain threshold.

The definition of entropy is:

$$E = \sum_{x:p(x)>0} p(x) \cdot \log(p(x))$$

The idea is that constant value of flow will lead to low entropy. Thus entropy could be used to detect if the detector has constant value consistently.

[2] used adjacent loop point flow for comparison to detect possible erroneous data. It used the ratio of flows of upstream and downstream stations as the measure for test. The reason is that: for the same time  $t$ , the upstream and downstream have completely different clusters of vehicles. For free-flow traffic and 10-minute aggregated data, this makes sense.

Work in [1] is a systematic work in data based fault detection focusing on how to correct the data for the following two cases: data missing and bad data. It also proposed method for data correction. [3] used ARMA model for prediction of loop data, which was over time and could be used for fill in faulty data. But [1] commented that that its response was too fast. It suggested using good neighbor (same location but different lanes, or adjacent locations) data for patching the wholes. Averaging or interpolation over space methods were used for filling the whole. The mathematical foundation for this method was that occupancy and flow of neighbor loops were highly correlated. However, if several loop stations were down in a section of freeway, this method would become questionable. The algorithm developed in [1] is called Daily Statistics Algorithm (DSA) since it produces only one result using a whole day 30s  $q$  (volume) and  $k$  (occupancy) data: good or bad on that day. The Detection criterion is based on the value of 4 statistic parameters and the selected threshold. Each statistic parameter targets for one error type.

The main used in methods used [1] for data correction was to look at neighboring loops in adjacent lanes and/or up/downstream as well as historical data:

- Linear interpolation overtime of the loop itself
- Linear interpolation over space of neighboring loops
- Averaging overtime of the loop itself
- Averaging over space of neighboring loops
- Combinations of them all – in fact, averaging is a special case of interpolation

This method could not distinguish the case of temporal loop failures since the statistic over a whole

day will not tell temporal behaviors. The proposed method used threshold to identify 4 types of loop data errors:

- Occupancy and flow are mostly zero
- Positive occupancy and zero flow
- Very high occupancy
- Constant occupancy and flow

This has been achieved by classify a fixed loop daily data into 4 categories and then aggregate over time. Then threshold is defined for such error identification, which is based on some common knowledge. Those methods could not be used for the following faulty loops:

- Permanent isolated fault loop
- Temporal faulty data: such as those cases which are affected by weather and heavy traffic
- Individual loop fault such as sensitivity, crosstalk, etc.

This algorithm has been used in PeMS for several years. It proved to be reliable and better than other methods for higher level aggregated data for some larger range and or longer time loop problems.

Data correction methods were also proposed in [4], which was basically using historical data as well as adjacent station data for interpolation over distance and time. A Kalman filter is also designed for estimation of lane volume to filter out measurement noise. The filter performance showed that it was unbiased with discrepancy of 300vhr.

In the work of [2], Poisson distribution was used to describe the probability for the number of vehicles counted (flow) at a loop station every 30s interval.

$$p(y) = e^{-\mu} \frac{\mu^y}{y!}$$

$y$  – point flow: vehicle count at a given loop station. The probability for  $n$  continuous reading of a flow  $y$  was:

$$p^n(y) = e^{-n\mu} \frac{\mu^{ny}}{(y!)^n}$$

Then a threshold was set for data error checking:  $p^n(y) \leq P_{\min} = 0.0005$ . An accumulated Poisson distribution should be used to represent the point flows at a loop station.

$$P(0 < y \leq x) = \sum_{y=1}^x e^{-\mu} \frac{\mu^y}{y!}$$

Due to the stochastic property, the point flow  $y$  could be quite different for different traffic situations: AM peak, PM peak, off-peak, congested and non-congested cases. This idea is quite different from the entropy test of PeMS where constant flow will lead to very low entropy. This means that low entropy corresponds to invariance of traffic flow, which can happen only if the loop has faulty reading.

Time-of-day flow and occupancy ratio were used reflect vehicle types such as trucks and passenger

cars [2, 5]:

- This ratio could assume any value;
- Trucks corresponds to low flow and high occupancy
- Passenger cars the other way around in the same time period
- Low flow and high occupancy may indicate congested traffic in another time period (cause by AM peak, PM peak and incident/accident)

[6] used loop data to calculate average vehicle length: 2.7m~18.0m. This threshold is used for data error checking. It is obvious that such check can only tell if the data is reasonable or not. It could not tell what was wrong exactly with the system.

The Detector Fitness Program (DFP) [7] looked at the loop station in 3 Caltrans District: D4, D7 and D11. It defined some measurement parameters. The study proposes and calculates three metrics of system performance: *productivity* is the fraction of days that sensors provide reliable measurements; *stability* is the frequency with which sensors switch from being reliable to becoming unreliable; and *lifetime and fixing time* — the number of consecutive days that sensors are continuously working or failed, respectively. Productivity measures the performance of the sensor system; stability measures the reliability of the communication network; lifetime and fixing time provide more detailed views of both components of the sensor network. The evaluation method first uses PeMS 30s data. The second data set comprises records from the Detector Fitness Program (DFP) for Districts 4 and 7. These records were created by crews following a field visit to a loop. *Fault States* looked at included: *line down*, *controller down*, *no data*, *insufficient data*, *card off*, *high value*, *intermittent*, *constant value*, and *feed unstable*. Detection methods involved was mainly *Data Threshold Checking*. This work also looked at the possible higher level fault caused by communication systems involved in data passing for TMC/PeMS, which include: Caltrans owned fiber optics, wireless GPRS modem (UDP, TCP), telephone line and wireless cell-phone lines. The main idea is to tell if the communication system is healthy from the status of all the loop data related to the same communication system such as those belonging to the same control cabinet.

**Summary:** The problems to be looked at for macroscopic data analysis are:

- Communication Down: No samples were received for the loop between 5:00 am to 10:00 pm;
- Mis-assignment: Mismatch between the real location and the location assigned in the map in control cabinet;
- Insufficient Data: PeMS receives too few samples to determine the loop health;
- Card Off: Too many samples have zero occupancy;
- High Occupancy: Too many samples with occupancy above 70%;
- Intermittent: Too many samples with zero flow and non-zero occupancy;
- Constant: The loop is stuck on a particular value;
- Feed Unstable: The detector failed in the past, and its current status can not be determined due to problems in the data feed;
- Systematic failures: Systematic differences in failure rates by freeway and by lane which could be affected by vehicle types;
- Electrical failures such as splicing problems or detector card faults;
- Synchronized failures: District-wide synchronized failures; e.g., unusually many loops in a District fail on the same day;

Methods used at this for direct loop fault detection include: (a) statistic, (b) Entropy, (c) threshold checking based on some known physical limits and empirical values, and (d) Comparing with neighboring (adjacent lanes, upstream/downstream) stations. Method used at this level for data correction/imputation include: linear interpolation or moving window averaging over time, space (adjacent lanes, upstream and downstream)

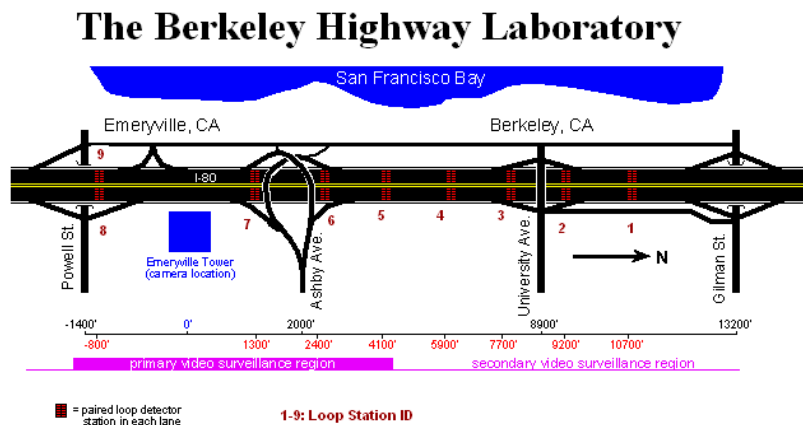
### Mesoscopic Level

System in this level involves a section of freeways which has more than one control cabinet with multiple loops. The characteristics in this level are:

- Sub-second data of each are available;
- Loops connected with the same control cabinet are time synchronized;
- Loops connected with different control cabinets are time synchronized;
- Minor communication system is involved in data synchronization and data passing.

Thus the communication system fault can be easily determined by some simple ad hoc method such as check sum. In this way, the communication system fault could be isolated from the loop fault.

Berkeley Highway Lab (BHL) is a typical example of such system. BHL has 9 loop stations with 164 loop detectors for both side of Interstate I-80 between Gilman St. and Power St. Figure xxx2.



**Figure xx2.** Berkeley Highway Lab

Work in [8, 9] considered loop fault detection systematically based on BHL system. A two-level, nine diagnostic scheme has been developed including dynamic diagnostics based on speed and vehicle composition. The developed algorithms were implemented software and currently running in BHL system. This work separated *detector deficiencies* and detector fault. The fault detection system used 1/60s data from loop, which were basically some threshold tests:

- activity test: test criterion: continuous 15 minute constant signal;
- Minimum on-time test for at least 100 vehicles (fail criterion: 5% vehicles occupancy < 8/60s);
- Maximum on-time test for at least 100 vehicles (fail criterion: 5% vehicles occupancy > 600/60s);
- Dynamic Minimum/ Maximum on-time test: similar to minimum/maximum adjust those time

interval test threshold based on speed and vehicle length;

- Minimum Off-Time – If 5% or more of the off-times in a sample of 100 vehicles are less than 25/60 seconds, the test fails;
- Dynamic Maximum Off-Time – This is one of the new diagnostics. If 5% or more of the off-times in a sample of 100 vehicles are greater than a threshold value which is a variable depending on the calculated average time headway, the test fails;
- Mode on-time test: test for 1000 vehicles; Test criterion: calculate mode of the distribution is outside of the interval [10/60s, 16/60s];
- Dual loop on-time difference test: test for 1000 vehicles; Test criterion: if the difference between the Upstream and downstream loop is outside the time interval [-3.5s, +3.5s]; it only valid in free-flow condition; not well-designed yet;
- Refining those tests in two aspects:
  - Predicting that the detector passes the tests when in fact the detector data is not good
  - Predicting that the detector fails the tests when in fact the detector data is good.

The test fail will need to account for the situations when there is little traffic such as in the early morning. This study also identified that some data problems are due to Verizon CDPD modem network connection instead of loop stations faults, which means that the communication fault could not fully separated. It indicated the necessity of direct loop fault detection at control cabinet level.

This work recognized the importance of using low level sub-second data instead of aggregated data. Conventional traffic monitoring aggregates the event data to fixed period samples of flow, velocity and occupancy before transmitting the data to the Transportation Management Center (TMC). The sampling period is typically on the order of 30 sec or 5 min. This relatively coarse aggregation can obscure features of interest and is vulnerable to noise. Both of these factors delay the identification of resolvable events, the former due to the need to wait until the end of a given sample period and the latter due to the need to wait for multiple sample periods to exclude transient errors. The Nyquist sampling criteria from basic signal processing dictates that one can only resolve features that last two sampling periods and the need to tolerate noise in the measurements further reduces the response time. As such, it is necessary to have a trade off between cost and data passing frequency. This study also suggest to pass all the event (low level sub-second) data to MTC as well as all the data processing. It mentioned that link travel time for BHL is based on vehicle re-identification.

A methodology for substituting for missing data (imputation) was also developed in [9]. The missed data is imputed based on the data of adjacent lanes using interpolation.

[10] also looked at 20s and 5minute data. The work used reasonable interval for flow density and speed to test if the data were reasonable: if they fell into the interval, then they are considered good data. Otherwise, they were considered bad data. The thresholds of those intervals were specified based on experiences on historical data. Similar idea was used for  $k$ - $q$  plane for specifying a criterion region by [JACO\_1]. The boundary of the region is determined by some parameters which need to be calibrated according to the site situation. This idea is slightly better because the relationship between  $k$  and  $q$  is taken into consideration. However, they did not take the advantage of using historical data as well temporal data relationships in detection and correction. [1] indicated that those methods were difficult to use in practice since the thresholds were difficult to calibrate. Due to those factors, several situations were incorrectly detected: false positive and false negative happened.



[4] uses FSP data which is composed of three parts: loop detector data, probe vehicle data and incident data of approximately two months. The loop detector data includes 30s data and 5 min aggregated data for data error checking. The loop locations are divided into mainline, HOV lane and on-ramp, which have different traffic characteristics. 14 error checking criteria based on the two types of data sets are proposed. Parameters taken into consideration are volume, occupancy and average speed. The data needs pass 10 consecutive tests. Those checks include bounding checking – traffic parameters must be within certain physical bound; contradictory check – two traffic parameters such as occupancy and value, occupancy and speed from the same loop station must be consistent. The seriousness of erroneous data have been analyzed according percent of time in malfunction, percent of station and percent of time in malfunction, etc. It has been found that data missing is the most significant error which appeared for blocks are sensors/stations. This may suggest that such error is caused by data transmission or communication system. It was found that for I-880 FSP data, malfunction stations are about 21% on average even the stations are well-maintained for proper function.

**Summary:** Highway Section/Corridor: Typical example is the data from Berkeley Highway Lab. In this level, 60Hz data is available every second. The characteristics of those data are that (a) sub-second data could be obtained at this level; (b) time synchronized sub-second data are available for loop stations on a stretch of highways; (c) only a short distance communication system is involved; (d) the detection could be near real-time in the sense that the time delay for data passing were in the level of few seconds. Loop problems looked at this level include: *mis-assignment, temporary data missing, crosstalk, no data or constant for a period of time, broken cable, chattering, card broken, card sensitivity too high or too low, pulse broken, mismatch of ON/OFF time instant between upstream and downstream loops for dual loop stations.*

Methods used at this for direct loop fault detection include: analyzing sub-second data, threshold checking, and vehicle re-identification. Method used at this level for data correction/imputation include: linear interpolation or moving window averaging over time, space (adjacent lanes, upstream and downstream). It is noted that even if in this system level, some detailed loop fault still cannot be detected. He advantage for such system is that one could compare the synchronized upstream station and downstream station data for diagnosis and data correction which could not be achieved at control cabinet level.

### **Microscopic Level**

Many operating agencies use specialized loop testers to assess the quality of the wiring [12, 13], but these tools bypass the controller and loop sensors; thus, they do not analyze the entire detector circuit, nor do they analyze the circuit in operation. To this end, most operating agencies employ simple heuristics methods such as if the loop sensor indicator lights is on as a vehicle passes. Such tests are typically employed when the loops are installed close to the control cabinet. Many practitioners and some researchers [11, 14, 3] have worked to formalize the latter heuristic by looking at if the time series 30 second average flow and occupancy within statistical tolerance.

Low level loop data correction could be trace back to the Freeway Service Patrol study in 1990s [15, 18]. It looked at the transition times in sub-second of dual loop stations with 20ft distance

between upstream and downstream loops. It noticed some problems in low level data including:

- data missing
- un-matching of those data which results in unreasonable occupancy and speed;
- on-time and off-time are not always related;
- no-flow and no-speed but with positive occupancies;
- existing pulses in both up and down streams

The author mentioned that some of the phenomenon could be explained as caused vehicle changing lane. However, there is no systematic diagnosis in [15] for loop fault, nor systematic methods for lower level data correction.

Work in [16] considered detect for a single loop. It uses the number pulses as the vehicle counts to verify loop data. If pulse breaks, it will cause the data problem. It developed automated loop fault detection system which uses aggregated data. They must accept a large sample variance and potentially miss problems altogether. For example, the systems have to tolerate a variable percentage of long vehicles in the sample population. Their methodology examines the distribution of detector *on-time*, i.e., the time the detector is occupied by a vehicle. Unlike conventional aggregate measures, their approach is sensitive to errors such as "pulse breakups", where a single vehicle registers multiple detections because the sensor output flickers off and back on. This is the main disadvantage to use vehicle count for a single loop for fault detection: one cannot isolate other loop fault from the pulse flickering problem.

### Using Multiple Loop

Studies in [17, xxxc2] use dual loop information for comparison to detect loop fault. This paper focuses on evaluation of loop sensors and detection of cross-talk. It was developed for off-line data analysis but could possibly be used for on-line in the future. It can be summarized in three steps:

- (i) Record a large number of vehicle actuations during free flow traffic;
- (ii) For each vehicle, match actuations between the upstream and downstream loops in the given lane;
- (iii) Take the difference between matched upstream and downstream on-times and examine the distribution on a lane-by-lane basis. Assuming the loops are functioning properly, only a small percentage of the differences should be over 1/30 seconds. Otherwise, "Cross-talk" fault is announced.

Using dual loop speed traps to identify detector errors is another approach conducted in [17, xxxc2]. At free flow, on-time difference and off-time difference should be the same if no hardware problem. So if they are not the same, there may be hardware and/or software problem. But this is not true if it is not free-flow speed.

**Summary:** This is only level one could conduct direct loop fault detection and isolate the loop faults with possibly other system faults. Data in this level can be either data processed by loop detector card, which will be loop ON/OFF time instant or occupancy; or the raw loop pulse signal before the

loop card. The main characteristics of those data are that (a) they do not pass any communication media and thus there is no possibility of communication fault which usually pollute or loose the data stream; (b) all the raw information is available if proper interface with the control cabinet is available; (c) real-time data are available; and (d) most importantly, ground truth could be obtained at this level thus loop fault detection could be conducted through comparison between the loop detector reading and the ground truth.

Loop faults to be look at the microscopic level include any loop card faults: *mis-assignment, temporary data missing, crosstalk, no data or constant for a period of time, broken cable, chattering, card broken, card sensitivity too high or too low, pulse broken, mismatch of ON/OFF time instant between upstream and downstream loops for dual loop stations.*

Methods used at this for direct loop fault detection include: using 60 Hz data, using pulse signals which is by pass loop card. There is no data correction/imputation in this level well-documented.

### **CONCLUDING REMARKS**

Traffic monitoring system based loop detection stations may be divided into three levels which lead to different viewpoints for loop fault detection and data correction: macroscopic such as: (a) TMC/PeMS level; (b) mesoscopic – a stretch of freeway; and (c) microscopic – control cabinet level. Those three levels of methods look at the problem from different aspects using different level of data, but they are complementary to each other. In fact, higher level data analysis could be effective for systematic errors which are caused by communication system and power outage. Data analysis in the highway section/corridor level could be used to locate suspicious loop stations if the communication system can be made reliable. Those two methods can only be called indirect since they are affected by communication system for data passing. Only the detection at control cabinet level can be called direct loop fault detection. Although not many effective results in this level have been documented, a Portable Tool which compares the loop data with ground truth generated from video camera vehicle tracking for systematic loop fault detection and data correction/imputation are under development by the authors, which will be addressed in future work.

### **ACKNOWLEDGEMENT**

This work was performed as part of the California PATH 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 the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This paper does not constitute a standard, specification, or regulation.

## REFERENCES

- [1] Chen, C., Kwon, J., Rice, J., Skabardonis, A., and Varaiya, P., Detecting errors and imputing missing data for single loop surveillance systems, 82<sup>nd</sup> TRB Annual Meeting, January 2003, Washington, D.C.
- [2] Robinson, S. and Polak, J. W., ILD data cleaning treatments and their effect on the performance of Urban Link Travel Time models, 85<sup>th</sup> TRB Annual Meeting, Jan. 2006
- [3] Nihan, N., Aid to Determining Freeway Metering Rates and Detecting Loop Errors, *Journal of Transportation Engineering*, Vol 123, No 6, ASCE, November/December 1997, pp 454-458.
- [4] Payne, H.J. and Thompson S., Malfunction Detection and Data Repair for Induction-Loop Sensors Using I-880 Data Base, TRB, *Transportation Research Record*, 1570, p.191-201
- [5] Coifman, B., Improved velocity estimation using single loop detectors, *Transportation Research A*, 35, 863-880, 2001
- [6] Turochy, R.E. and B.L. Smith. A New Procedure for Detector Data Screening in Traffic Management Systems. *Transportation Research Record* 1727, Transportation Research Board, Washington, D.C., 2000, pp 127-131
- [7] Rajagopal, R. and Varaiya, P., Health of California's Loop Detector System: Final Report for PATH TO 6300, 2007
- [8] A. May, Coifman, B., Cayford, R., Merritt, G., Automated diagnostics of loop detectors and the data collection system in the Berkeley Highway Laboratory, California PATH Research Report, **UCB-ITS-PRR-2004-13**
- [9] A. May, R. Cayford, L. Leiman, G. Merritt, BHL Traffic Detector Analysis, Consolidation of BHL Detector System at CCIT, and Development of Portable Detector Diagnostic Tool, California PATH Research Report, UCB-ITS-PRR-2005-24
- [10] Payne, H.J., E.D. Helfenbein and H.C. Knobel. Development and testing of incident detection algorithms, FHWA-RD-76-20, Federal Highway Administration, Washington DC 1976.
- [11] Jacobson, L, Nihan, N., and Bender, J., Detecting Erroneous Loop Detector Data in a Freeway Traffic Management System, *Transportation Research Record* 1287, TRB, Washington, DC, 1990, pp 151-166.
- [12] Kell, J., Fullerton, I., and Mills, M., *Traffic Detector Handbook, Second Edition*, Federal Highway Administration, Washington, DC, 1990.
- [13] Ingram, J., *The Inductive Loop Vehicle Detector: Installation Acceptance Criteria and Maintenance Techniques*, California Department of Transportation, Sacramento, CA, 1976.
- [14] Cleghorn, D., Hall, F., and Garbuio, D., Improved Data Screening Techniques for Freeway Traffic Management Systems, *Transportation Research Record* 1320, TRB, Washington, DC, 1991, pp 17-31.
- [15] Karl Petty, Freeway Service Patrol (FSP) 1.1: The Analysis Software for the FSP Project, California PATH Research Report, UCB-ITS-PRR-95-20
- [16] Chen, L., and May, A., Traffic Detector Errors and Diagnostics, *Transportation Research Record* 1132, TRB, Washington, DC, 1987, pp 82-93.
- [17] Coifman, B., Using Dual Loop Speed Traps to Identify Detector Errors, 78<sup>th</sup> TRB Annual Meeting, Washington, DC., Jan. 1999
- [18] Skabardonis, A., Petty, K., Noeimi, H., Rydzewski, D. and Varaiya, P. (1996). I-880 Field Experiment: Data-Base Development and Incident Delay Estimation Procedures, *Transportation Research Record* 1554 , TRB, pp 204-212.