Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2016) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

International Conference on Intelligent Transport Systems in Theory and Practice, mobil.TUM

2017, 4-5 July 2017, Munich, Germany

Estimating Detectors' Systematic Error in Signalized Intersection Using Flow Conservation

Zahra Amini[a,]*, Alexander Skabardonis[a], Pravin Varaiya[b]

[a] *Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720 USA*
[b] *Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA*

Abstract

Real-time availability of traffic data at signalized intersections provides opportunities for development and implementation of advanced in traffic management strategies. However, the quality of the data from surveillance systems is often problematic. In this paper we focus on improving the quality and reliability of an available dataset by reducing the error in the detected values. We estimated the systematic error of the detectors using flow conservation approach. The intuition of this approach is to minimize the error corresponding to each detector by putting the overall input and output flow of the intersection equal to each other. We implemented our algorithm in an urban network located in Montgomery County in Maryland, USA consists of several consecutive intersections equipped with magnetic filed detectors. These detectors are collecting data, which is known as High Resolution (HR) data. HR data provides information on vehicles' arrival and departure along with control system's status. We implemented our algorithm in an intersection over different time intervals. Comparing the imputed values with some ground truth values, collected by observers, shows that the error between the true value and imputed value is up to 25% smaller than the error between the true value and estimated value.

Click here and insert your abstract text.

Peer-review under responsibility of the scientific committee of the International Conference on Intelligent Transport Systems in

Theory and Practice.

* Corresponding author. Tel.: +1-925-408-8336
E-mail address: zahraamini@berkeley.edu

Peer-review under responsibility of the scientific committee of the International Conference on Intelligent Transport Systems in Theory and Practice.

Keywords: Type your keywords here, separated by semicolons ;

1. Introduction

Recent assessment of the state of practice in signal management on urban networks indicates that on average intersections are poorly managed, according to the National Transportation Operations Coalition (NTOC) 2012 signal systems report card. The main reason for the poor performance is the lack of systematic data collection to estimate performance measures. According to FHWA and Purdue University's report by Day et al. (2014) recent advances in technology are making High Resolution (HR) data collection at traffic signals economical. Sensys Networks Inc. uses magnetic field sensors in signalized intersections to collect HR traffic data. Haoui et al. (2008) explains how a detector measures changes in the earth's magnetic field induced by a vehicle, then processes the measurements to detect the vehicle, and transfers the processed data via radio to the Access Point (AP). The AP combines data from the detectors into information for the local controller or the Traffic Management Center (TMC). Muralidharan et al. (2014) describes a HR system called SAMS (Safety and Mobility System) that detects and records the lane, speed, signal phase and time when each vehicle enters and leaves the intersection based on the available information in the AP. HR data at signalized intersections refers to continuous acquisition of detector data that provide information on the vehicles approaching and leaving the intersection plus simultaneous information on the signal status (signal display). The processing of the HR data provides extended set of metrics that can be used for different purposes. Coogan et al. (2017) uses HR data to predict traffic volume in future days based on historical database. Another important use of HR data is for the development and assessment of signal settings at signalized intersections.

HR data could be extremely beneficial only if they provide reliable and accurate information about the traffic condition. Krogmeier et al. (1996) and Hughes Transportation Management Systems (1996) study errors in data caused by failure of detectors and communication links. These errors could be investigated by running a health inspection on the equipment and the system. There are errors due to detector malfunction, which are challenging to investigate because the source of the error is not clear and the amount of the error is usually small compared to the detected value. Coifman et al. (1999) uses some microscopic approaches, which focus on characteristic of the dual loop detectors to investigate the error. Some studies use macroscopic-level approaches such as Jacobson et al. (1990) that applys different threshold test on the data to improve the accuracy. PeMS (Caltrans Performance Measurement System) is using a data-filtering algorithm, which was developed by Chen et al. (2003) to detect bad loop detectors from their output and impute the missing data from neighboring good loops. Moreover, Peeta et al. (2002) uses Fourier transform-based techniques to detect the data errors dues to malfunctioning detectors. Anastassopoulos (2000) and Peeta et al. (1998) use neural network model and genetic algorithms to address detector's issues. Smiths (1995) studies forecasting techniques such as historical average, nearest-neighbor algorithm, and autoregressive integrated moving average models for data correction and substitution.

In this paper we used flow conservation approach to estimate the systematic error of the detectors. At any time interval, the overall number of the vehicles that enter the intersection is equal to the sum of the vehicles that exit the intersection. Our approach has several advantages over the methods that we mentioned. Most importantly, it could be used real time and it is possible to include the data-filtering algorithm in the data collection process. Moreover, with applying some simple modifications to the algorithm it can work for intersections with different detector systems and geometric. Finally, in addition to estimating the error value, it can substitute any missing value by using available data.

In the next section of this paper, we explain the approach first, and then we introduce the data-filtering algorithm. In section 2 we use a typical intersection to explain the problem and the process of error detection. Sensys Networks Inc. provided a full database to us for an urban network located in Montgomery County in Maryland, USA, which in section 3 we present the result from implementing the data-filtering algorithm in to that network. Eventually in the last section we provide a conclusion about the performance of the algorithm and introduce some ideas for future researches.

2. Data Filtering

In an isolated intersection with 4 legs we have 4 inputs, 4 outputs, and 12 movements flow values as it is shown in Figure 1. The same figure also shows the location of the detectors. According to Figure 1 advance detectors are located at the upstream of the intersection and they measure the input flow. The output flow is measured by the departure detectors, which are located in the departure lanes, downstream of the intersection. Stop-bar detectors, which are located at the stop line, count the number of vehicles in each movement (right, left, through). For any time interval, we can collect the traffic volume for each of the flow values.

Therefore, we will have 20 values in total and each value will have different amount of error. There are multiple sources of errors in traffic data; some errors could be due to vehicle's movement such as lane change and driving over the lane, while some are due to environmental condition such as rain. Accidents can also cause error in traffic data and interrupt the data collection process. Some detectors miscount the vehicles when there are very large or very small in dimension or when the vehicles are traveling too close to each other. In addition, there are systematic errors in detector's data collection process that are impossible to eliminate. Using different detectors to count the same vehicle several times would improve the counting process and provide accurate values. Figure 1 shows an example where each vehicle was been counted 3 times and each measured value is equal to the true value plus some error. Our goal is to minimize the summation of all the error terms.

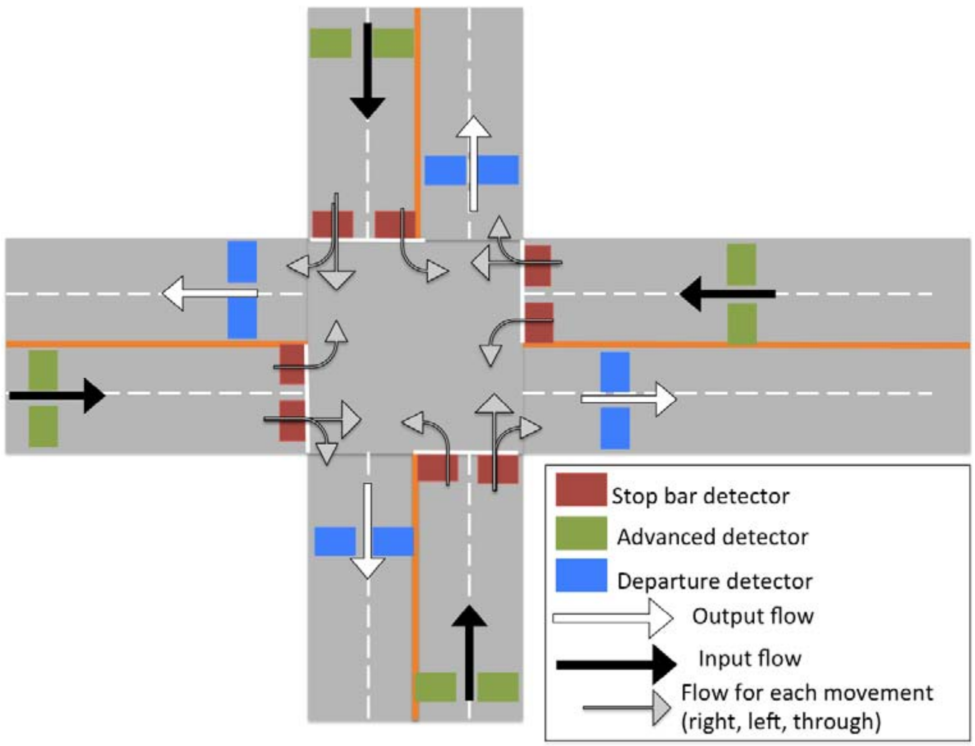

Figure 1. An Intersection with 4 legs

2.1. Flow conservation Approach

As it is shown in Equation 1 for our problem formulation we assume measured flow value, $\hat{f}_i$, equals to the estimated flow value, f_i, plus the error term, ε_i. Therefore, a negative error term means the detected value is less than the actual value and our detector is undercounting. While a positive error term implies that the detector is over counting.

$$\hat{f}_i = f_i + \varepsilon_i \quad for\ all\ i \tag{1}$$

Estimated value, f_i, is not necessarily equal to the actual value because ε_i only captures the amount of the systematic error and not all the error values that exist in the detected value. However, our claim is that the estimated value would be closer to the actual value compared to the detected value. In section 3 we will test this hypothesis by doing an experiment on a real database with ground truth result available.

In the example that we introduced in section 2, Figure 1, there are 20 variables. The cost function, C, in Equation 2, is corresponding to Figure 1 example and it includes 20 error terms, one for each variable. In the cost function C, we are summing the ε_i^2 values to include both positive and negative errors and make sure that they do not cancel each other out. Also the squared error is an appropriate fit for our cost function because it is differentiable. Moreover, each ε term has a corresponding ω_i value, which is the weight of each ε_i term and its value depends on level of accuracy of $\hat{f}_i$. In an equal situation ω_i would be same for all the variables, but it is possible that we have more confident in one or some detectors' performance than others. For example, it is reasonable to say that advanced detectors have a better accuracy in their counting process than stop-bar detectors, because of the lane change and stops that happen at the stop lines. Based on available information about the detectors and previous studies we can assign appropriate value for ω_i.

$$C = \sum_{all\ i} \omega_i * \varepsilon_i^2 \tag{2}$$

Also, we have a set of constraints based on the flow conservation rule. Equation 3 requires the input flow in each leg to be equal the sum of turn-movement flows in the same leg. Next constraint, Equation 4, requires the output flow in each leg to be equal the sum of turn-movement flows that are feeding the output direction. Finally, it is important to mention that all the detected and estimated flow values, $\hat{f}_i$ and f_i, should be equal or greater than zero.

$$f_{in} = \sum_{all\ movements} f_{movement} \quad for\ all\ i \tag{3}$$

$$f_{out} = \sum_{all\ feeding\ movements} f_{movement} \quad for\ all\ i \tag{4}$$

Ultimately, we have the detected flow values and we want to estimate the errors by minimizing C, which is a convex function, subject to linear constraints so we have minimum point where the cost should be at its lowest value.

2.2. Data Filtering Algorithm

After formulating the cost function, assigning appropriate ω values, and inputting the detected flow values we can run the optimization problem. This process could be done real time as well. After minimizing C we get the ε_i for each $\hat{f}_i$ value, then we can estimate f_i, which is our imputed flow value. The time interval that we are running the optimization needs to be long enough to minimize the impact of the incomplete trips, but at the same time it should be small enough so we can assume the traffic volume is uniform during each time step. Also, the cycle length can impact the length of the time interval as well.

3. Case Study

3.1. Site description

In this section we used flow conservation approach to estimate the detector's systematic error in intersection of Montrose Rd. and Tildenwood Dr., in Montgomery County. Figure 2 shows the schematic of the detector layout for the signalized intersection. Each approach has stop bar detectors denoted by grey circles, and an advance detector (not shown). In addition there is a detector in each departure lane, denoted by red circles, which permits measurement of turn movements. There is a network-monitoring card that provides the signal phase. The vehicle detectors are magnetometers that send the detection events wirelessly to the Access Point (AP). The AP also receives the signal phase events. The AP time-stamps all events with an accuracy of 10ms. The event stream is sent via a cellular modem to a server. The synchronous data collection permits matching different vehicle movements and the signal phase(s) serving them.

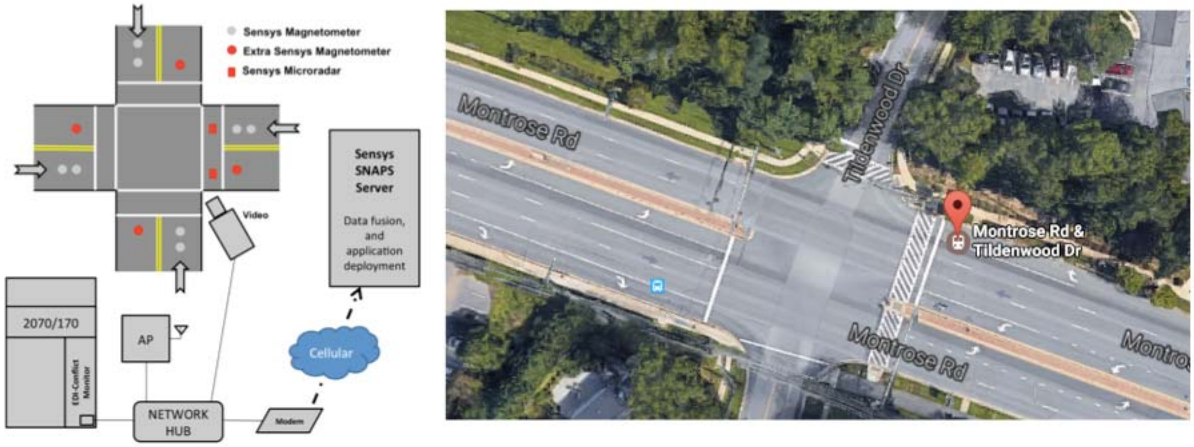

Figure 2. Schematic of detector layout (left) and aerial photo of intersection (right)

3.2. Problem formulation

The intersection has four approaches (legs) and there are three movements per approach. As it shown in Figure 3, in each leg there are 3 different types of detectors and each one count vehicles independently. Advanced detectors (located approximately 200-300 ft upstream from the stop-line) record the input values, which is shown by $f_{i,n}$ where n is the leg number. The output value, $f_{o,n}$ is recorded by departure detectors. On the same figure we can see the flows from the turn movements presented by $f_{(n,m)}$ where left, right, and through movement are respectively m=1,2, and 3. The intuition of our approach is to minimize the error corresponding to each variable. We have 4 inputs, 4 outputs, and 12 turn movement variables, which gives us 20 variables in total. The measured value, $\hat{f}$, equals to the estimated value, f, plus the error term, ε (Equations 5-7).

$$\hat{f}_{(n,m)} = f_{(n,m)} + \varepsilon_{f_{(n,m)}} \quad for\ all\ (n,m), n = 1,2,3,4\ and\ m = 1,2,3 \tag{5}$$

$$\hat{f}_{i,n} = f_{i,n} + \varepsilon_{f_{i,n}} \quad for\ all\ n = 1,2,3,4 \tag{6}$$

$$\hat{f}_{o,n} = f_{o,n} + \varepsilon_{f_{o,n}} \quad for\ all\ n = 1,2,3,4 \tag{7}$$

The cost function in Equation 8 includes the square of all the error terms multiply by the weights, α_n, β_n, and $\gamma_{(n,m)}$. As we mentioned in section 2, weights reflect confidence in the accuracy of the measurements. Flow values should be positive so we add lower bounds condition (Equations 9-11). Also, we have a set of constraints based on the flow conservation rule, which shows each input counts should be equal the turn movement counts of that approach and each output counts is equal the sum of all the movements that enter that approach (Equations 12-16). For example, in Equation 13 output flow in leg 1 equals to the right movement flow from leg 3 plus the left movement flow from leg 4 and the through movement from leg 1.

$$\min_{f_{i,n}, f_{o,n}, f_{(l,m)}} \sum_{n=1}^{4} \alpha_n \varepsilon_{f_{i,n}}^2 + \sum_{n=1}^{4} \beta_n \varepsilon_{f_{o,n}}^2 + \sum_{n=1}^{4} \sum_{m=1}^{3} \gamma_{(n,m)} \varepsilon_{f_{(n,m)}}^2 \tag{8}$$

$$s.t.\ 0 \le f_{n,m}\ for\ all\ (n,m) \tag{9}$$

$$0 \le f_{i,n}\ for\ all\ n \tag{10}$$

$$0 \le f_{o,n}\ for\ all\ n \tag{11}$$

$$f_{i,n} = \sum_{m=1}^{3} f_{(n,m)} \qquad for\ n = 1,2,3,4 \tag{12}$$

$$f_{o,1} = f_{(2,3)} + f_{(3,2)} + f_{(4,1)} \tag{13}$$

$$f_{o,2} = f_{(1,3)} + f_{(4,2)} + f_{(3,1)} \tag{14}$$

$$f_{o,3} = f_{(4,3)} + f_{(2,2)} + f_{(1,1)} \tag{15}$$

$$f_{o,4} = f_{(3,3)} + f_{(1,2)} + f_{(2,1)} \tag{16}$$

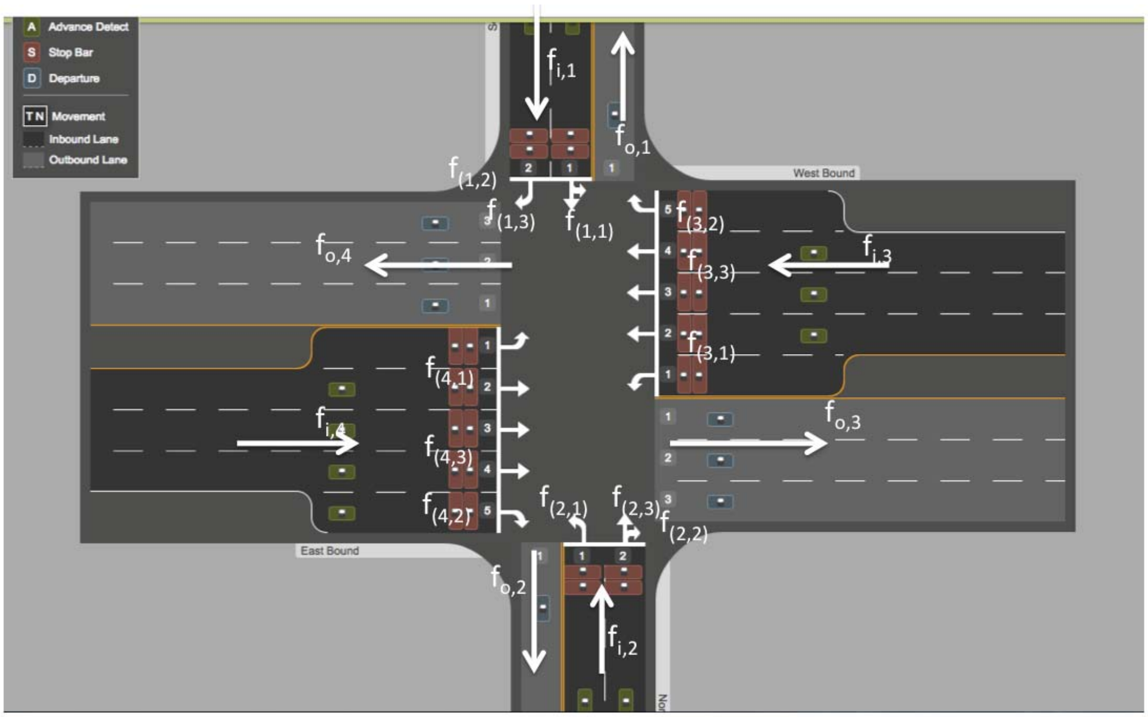

Figure 3. Intersection of Montrose Rd. and Tildenwood Dr. layout

3.3. Algorithm implementation result

We implemented the data-filtering algorithm on the detector data for every 15 minutes interval starting from 6AM until 7PM on June 14, 2016. We put all the weights equal to 1 and assumed all detectors are in equal condition and have same level of accuracy. For the same intersection and the same day, we have the ground truth turn movement flow values available. Observers collected ground truth data over a 13-hour interval (6AM to 7PM) for every 5 minutes, so we aggregated the data to get the vehicle counts for every 15 minutes, and then compared it with the turn movement vehicle count from detectors and imputed result that was estimated by data-filtering algorithm. The first column of Table 1 shows different movements in each leg and it separated the major movements from minor ones. Major movements have very high traffic demand, while the traffic volume in minor approaches is minimal. The second column shows the difference between the true and detected value divided by the true value in percentage and the last column is the difference between the true and estimated value in percentage. For Table 1 we estimated the error for each 15-minute time interval and then found the average error value over 13-hour.

Based on the result, in the major movements we can reduce the error up to 25%, (100*(4-3)/4=25%), when we implement the algorithm, which is about 100s of vehicles per day or few vehicles per cycle. For the minor

Movement		Error for true vs. detected (%)	Error for true vs. estimated (%)
Major Movements	Leg 3 through	4%	3%
	Leg 4 through	6%	5%
Minor Movements	Leg 3 Right	6%	9%
	Leg 3 Left	-43%	-35%
	Leg 4 Left	-41%	-11%
	Leg 4 Right	-10%	7%
	Leg 2 Left	6%	2%
	Leg 2 Right	5%	-20%
	Leg 1 Left	-8%	-37%
	Leg 1 Right	7%	-19%
	Leg 2 through	-75%	115%
	Leg 1 through	-50%	198%

Table1. Detected, imputed, and true values comparison

movements there is no warranty for improvement, however the traffic volume is very small so the error is about few vehicles per day. Now we want to focus on one movement by looking at Figure 4, which presents the cumulative vehicle counts over the 13-hour interval. In this figure each curve presents the cumulative vehicle count for a different method. Blue curve is the estimated value from the algorithm, the red one is the detected values before we apply the algorithm and the green curve is the true value from the observers. It is clear that the algorithm is improving the detected value and making it closer to the true value.

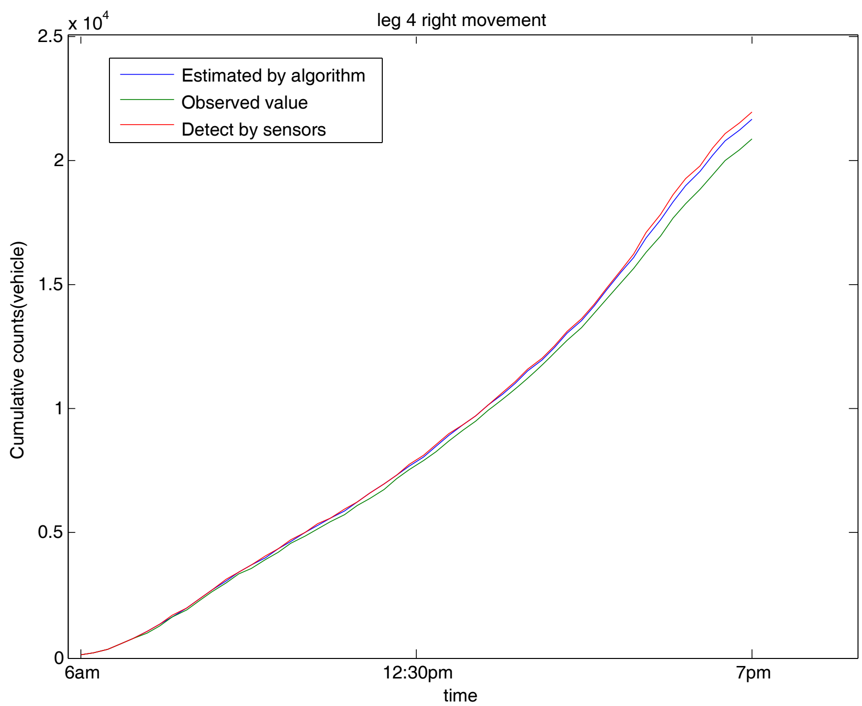

Figure 4. Leg 3 through movement cumulative count curves

Another conclusion that we can make from Table 1 is that detected values from minor movements are closer to the true values compared to the imputed values. Therefore, we ran an experiment and increased the weights of minor movements, which means that we assumed detected vehicle counts in minor movements are more accurate compared to the major movements. Figure 5 shows that by calibrating the algorithm and changing the weights we can achieve improvements for movements with small traffic volume as well. In Figure 5 the dark blue curve is the imputed value in a situation that all the weights are equal to 1 and the light blue curve is the calibrated case. With the new weights we see improvements in the major movement direction as well. However, for better calibration we need more ground truth result.

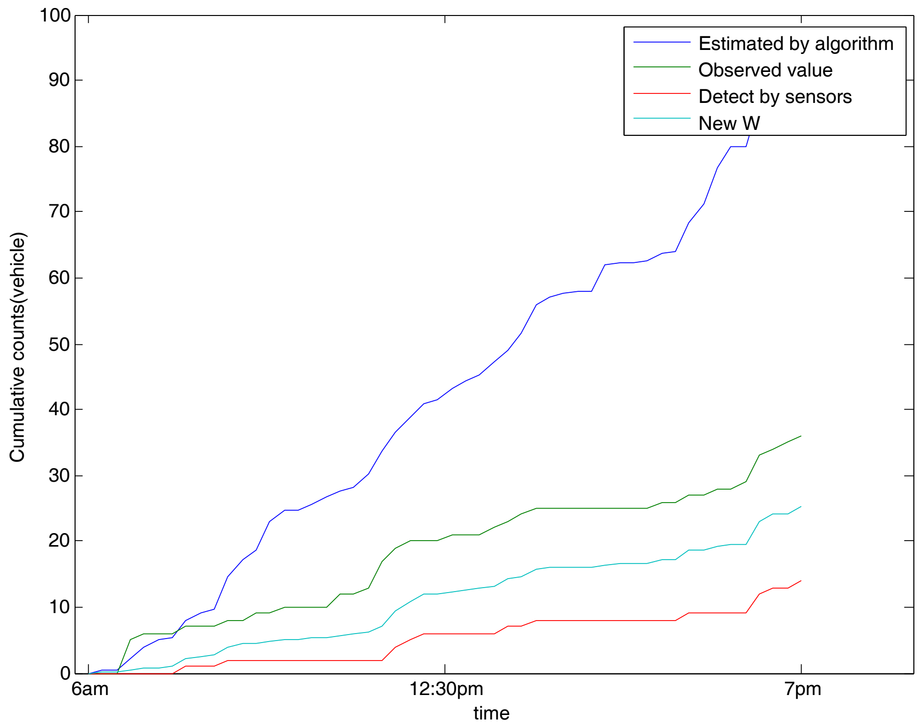

Figure 5. Leg 2 through movement cumulative count curves

4. Conclusion and Future work

HR data can provide useful information about the traffic condition. Having access to HR data enable us to improve the control system at urban network and increase the capacity at intersection. Therefore, the accuracy of these dataset is substantial. Data filtering algorithm is a fast and efficient way for estimating the systematic error of the detectors' data and amending the accuracy of the dataset. The advantage of this algorithm over similar methods is that it works real time and it can be a part of data collection process. In this paper we suggested that using the imputed value instead of the detected value improves the accuracy of the dataset. As we showed with a real world experiment, data filtering algorithm can reduce the error of the data point up to 25% in traffic approaches with high traffic volume. Data filtering algorithm could work for minor approaches as well, if we have enough ground truth data to calibrate it. Ground truth data let us to determine the reliability of each detector set and assign appropriate weight to each error term in our cost function.

There are different sources of error in traffic data. For future works, we would extend the cost function in a way that it includes multiple error terms for each detector set. Each error term can have specific distribution model and caused by different sources. The cost function that we provided in this paper is a general format and we can extend it or modify it for other type of detector layout and data collection process. Having more information about a site and its detection system helps us to model a more realistic error detection algorithm.

Acknowledgements

We would like to thank Sensys Networks Inc. for providing the database to us. Also we thank Dr. Christopher Flores and Dr. Rene Sanchez for assistance with data collection process.

References

National Traffic Operations Coalition. 2012 National Traffic Report Card, 2012. http://www. ite.org/reportcard/

Dowling , R., and S. Asiabor, 2012. Traffic Signal Analysis with Varying Demands and Capacities. Final Report NCHRP Project 03-97, Transportation Research Board, 2012.

C.M. Day, D.M. Bullock, et al. 2014. Performance measures for traffic signal systems: An outcome-oriented approach. Purdue University, West Lafayette, IN, doi: 10.5703/1288284315333.

FHWA. Measures of Effectiveness and Validation Guidance for Adaptive Signal Control Technologies. http://www.ops.fhwa.dot.gov/publications/fhwahop13031/index.htm

Haoui, A., Kavaler, R., and Varaiya, P., 2008. Wireless magnetic sensors for traffic surveillance. Transportation Research Part C, 16(3), 294 –306.

Muralidharan, A., C. Flores, and P. Varaiya, 2014. High-Resolution Sensing of Urban Traffic. 17th IEEE ITSC Conference, Qingdao, China.

Coogan, S., Flores, C., Varaiya, P., 2017. Traffic predictive control from low-rank structure. Transportation Research Part B: Methodological Volume 97, March 2017, Pages 1–22.

Krogmeier, J. V., K. C. Sinha, M. P. Fitz, S. Peeta, and S. Y. Nof., 1996. Borman Expressway ATMS Equipment Evaluation. Technical Research Report IVHS-9418 (301), FHW A/IN/JHRP-96/15. Joint Transportation Research Program, Purdue University, West Lafayette, Ind.

Hughes Transportation Management Systems. 1996. Borman Expressway Advanced Traffic Management System (ATMS) Phase I Final Report. Prepared for Indiana Department of Transportation, Indianapolis.

Coifman, B, 1999. Using Dual Loop Speed Traps to Identify Detector Errors. Transportation Research Record: Journal of the Transportation Research Board, No. 1683, TRB, National Research Council, Washington, D.C., pp. 47–58.

Jacobson, L. N., N. L. Nihan, and J. D. Bender, 1990. Detecting Erroneous Loop Detector Data in a Freeway Traffic Management System. Transportation Research Record 1287, TRB, National Research Council, Washington, D.C., pp. 151–165.

Chen, C., J. Kwon, J. Rice, A. Skabardonis, P. Varaiya, 2003. Detecting Errors and Imputing Missing Data from Single-Loop Surveilance Systems. Transportation Research Record 1855 Paper No. 03-3507, TRB, National Research Council, Washington, D.C. .

Peeta, S., I. Anastassopoulos, 2002. Automatic Real-Time Detection and Correction of Erroneous Detector Data with Fourier Transforms for Online Traffic Control Architectures. Transportation Research Record 1811 No. 02-2244, TRB, National Research Council, Washington, D.C. .

Anastassopoulos, I, 2000. Fault Tolerance and Incident Detection Using Fourier Transforms. M.S. thesis. School of Civil Engineering, Purdue University, West Lafayette, Ind. .

Peeta, S., and D. Das, 1998. Continuous Learning Framework for Freeway Incident Detection. Transportation Research Record 1644, TRB, National Research Council, Washington, D.C., pp. 124–131.

Smith, B. L, 1995. Forecasting Freeway Traffic Flow for Intelligent Transportation Systems Applications. Ph.D. dissertation. Department of Civil Engineering, University of Virginia, Charlottesville.