2013

Impact of Data Resolution on Peak Hour Factor Estimation for Transportation Decisions

Jan-Mou Li

Lee D. Han

Chung-Hao Chen
Old Dominion University, cxchen@odu.edu

Original Publication Citation

Li, J. M., Han, L. D., & Chen, C. H. (2013). Impact of data resolution on peak hour factor estimation for transportation decisions. *Central European Journal of Engineering*, 3(4), 732-739. doi:10.2478/s13531-013-0121-x

Cent. Eur. J. Eng. • 3(4) • 2013 • 732-739
DOI: 10.2478/s13531-013-0121-x

VERSITA

Central European Journal of **Engineering**

Impact of data resolution on peak hour factor estimation for transportation decisions

Research Article

Jan-Mou Li[1], Lee D. Han[2], Chung-Hao Chen[3]

1 *Oak Ridge National Laboratory, 2370 Cherahala Blvd. Knoxville, TN 37932, USA*

2 *Department of Civil & Environmental Engineering, The University of Tennessee, 112 Perkins Hall, Knoxville, TN 37996-2010, USA*

3 *Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, Virginia 23529, USA*

Received 24 March 2013; accepted 29 July 2013

Abstract: Inductance loop detection systems serve as a primary data source to contemporary traffic information systems. Measures like 20-second or 30-second average velocity, flow, and lane occupancy can be aggregated from individual loop detector actuation sampled at 60 Hz typically. Practically, these measures would sometimes be further aggregated into a much lower, e.g. 15-minute, resolution and then the raw data were lost. Valuable traffic information like flow variation may be distorted when the lower resolution aggregation is practiced. A biased conclusion could be drawn from a data integration system consisted of this kind of distortions. Three approaches estimating a peak hour factor based on traffic volume from loop detection systems are introduced in this paper to explore such a quality issue for data integration systems. Peak hour factor is commonly used in Highway Capacity Manual for determining and evaluating future system needs. By processing the raw data with the introduced approaches, different PHFs can be determined from a same traffic dataset. It is found that 2% to 5% (about one standard deviation from the mean) reduction in PHF may have 5 to 20 seconds increase in control delay estimation. The results suggest that distortion of control delay estimation at a signalized intersection exists due to an improper aggregation. That is, data quality might not be good enough for a right decision if the data were not processed appropriately.

Keywords: Process-Oriented • Data Quality • Data Integration • Data Resolution • Peak Hour Factor

1. Introduction

Highly organized and integrated data can support informed and comprehensive transportation decision-making. With the development of data integration system (DIS), data/information quality is becoming a first class property which is more and more required by end-users [1]. Bringing compatibility to the disparate data sets is challenging because each system usually is characterized by limited capability. Much information and guidance is needed to address the technical and other organizational challenges involved in data integration. Inductive loop detectors are widely used in the United States to provide traffic data for advanced traffic management systems (ATMS) and advanced traveler information systems (ATIS) as well as for actuate intersection signal controllers. Measures like 20-second or 30-second average velocity, flow, and lane occupancy can be aggregated from individual loop detector actuation sampled at 60 Hz typically. Much useful data regarding individual vehicles is possibly drawn with this level of resolution (e.g. [2]). However, data in this level of resolution would sometimes practically be further aggregated into a much lower, e.g. 15-minute, level of resolution and then the raw data were lost. Valuable traffic information like flow variation might be distorted once the data were aggregated in a lower resolution. A biased conclusion could be drawn from a data integration system consisted of this kind of distortions. In order to explore this kind of data quality issue for transportation data integration systems, three approaches estimating a peak-hour factor (PHF) based on traffic volume from loop detection systems are introduced in this paper. Peak-hour factor is commonly used in the Transportation Research Board's Highway Capacity Manual (HCM) [3] for the purpose of converting peak-hour traffic volume to the design hour flow rate, which is in turn used to assess various measures of effectiveness (MOEs) as well as the level of service (LOS) for transportation facilities. Because it is used in virtually all HCM methodologies, any change in the value of PHF could have a very strong impact on the analysis results as well as any subsequent engineering decisions and treatments [4]. Traffic demand and flow patterns fluctuate for a wide variety of reasons (e.g. [5]), ranging from human psychology to weather conditions. The variation dominates operational performance at signalized intersections [6]. A commonly accepted "catch-all" explanation for traffic-flow fluctuation is the stochastic nature of a system involving a multitude of autonomous agents. Tarko and Perez-Cartagena [7] investigated the variability of PHF over time and across locations, finding that day-to-day variability is as strong as site-to-site variability. They recommend that PHF be estimated on the basis of several days of vehicle counting to improve the precision of the average PHF estimate. There is still some debate as to how PHF is and should be calculated (e.g. [8, 9]). Decades ago, when the concept of PHF was first established, the data resolution of common automated traffic counters was limited to 15-minute counts due to technical constraints such as device memory and transmission bandwidth. The peak hour was established as the four consecutive 15-minute periods with the highest total volume for the day, and the one period with the highest volume among these four 15-minute periods was described as the peak 15 minutes. The design hour flow rate is, then, four times the peak 15-minute volume. Using the HCM definition, PHF can then be calculated in a straightforward fashion:

$$PHF = \frac{\text{Peak-Hour Volume}}{\text{Design Hour Flow Rate}} = \frac{V}{4 \times V_{15}} \quad (1)$$

where V is peak hour volume and V_{15} is volume during the peak 15 minutes of flow.

As technology has advanced and overcome the storage space and communication bandwidth issues, detector data resolution has greatly increased. Nowadays, it is not uncommon for transportation agencies to routinely collect and archive traffic data at 30-second intervals [10]. With such a wealth of data available, better calculation of the peak 15-minute volume and peak-hour factor should be within the grasp of today's traffic engineers. However, the common practice is still to aggregate the 30-second data into 15-minute volumes and in that process, more or less "throw away" about 97%, or 29 out of every 30 pieces, of data. Such aggregation of data is not only wasteful, but also leads to discrepancies in the identification and calculation of the true peak hour, the peak 15 minutes, and the PHF. While such aggregations may be appropriate for serving as inputs to control system algorithms and save disk space for archiving the volume data, much useful data regarding traffic variation which may further result cycle failures are lost. A distortion of information may be suggested due to the lost information. This paper presents the sources of the discrepancies in PHF calculation, proposes alternative ways to calculate PHF and peak 15-minute volume, explores the magnitude of the discrepancies, and demonstrates the impact on performance estimation when data were processed improperly.

2. Discrepancies in PHF

The discrepancies in PHF calculation when 30-second data are aggregated into 15-minute volumes result primarily from two sources: *aggregation clock offset* and *non-inclusive peak flow*. These concepts are presented below.

2.1. Aggregation clock offset

Depending on the beginning point of the aggregation process, that is, if the aggregation starts at exact midnight, or 30 seconds after that, or 5 minutes thereafter, the average count for the first and every subsequent 15-minute period would be different for each case. Figure 1 shows a one-day worth of real-world 30-second data from a randomly selected detector in Minnesota. The two curves tracking the middle of the data are 15-minute average volumes. The only difference, in terms of calculation, between them is that one starts at exactly midnight and the other is offset by 7 minutes and 30 seconds.

It is easy to discern that the two aggregation results, though they both track the average of the 30-second data,

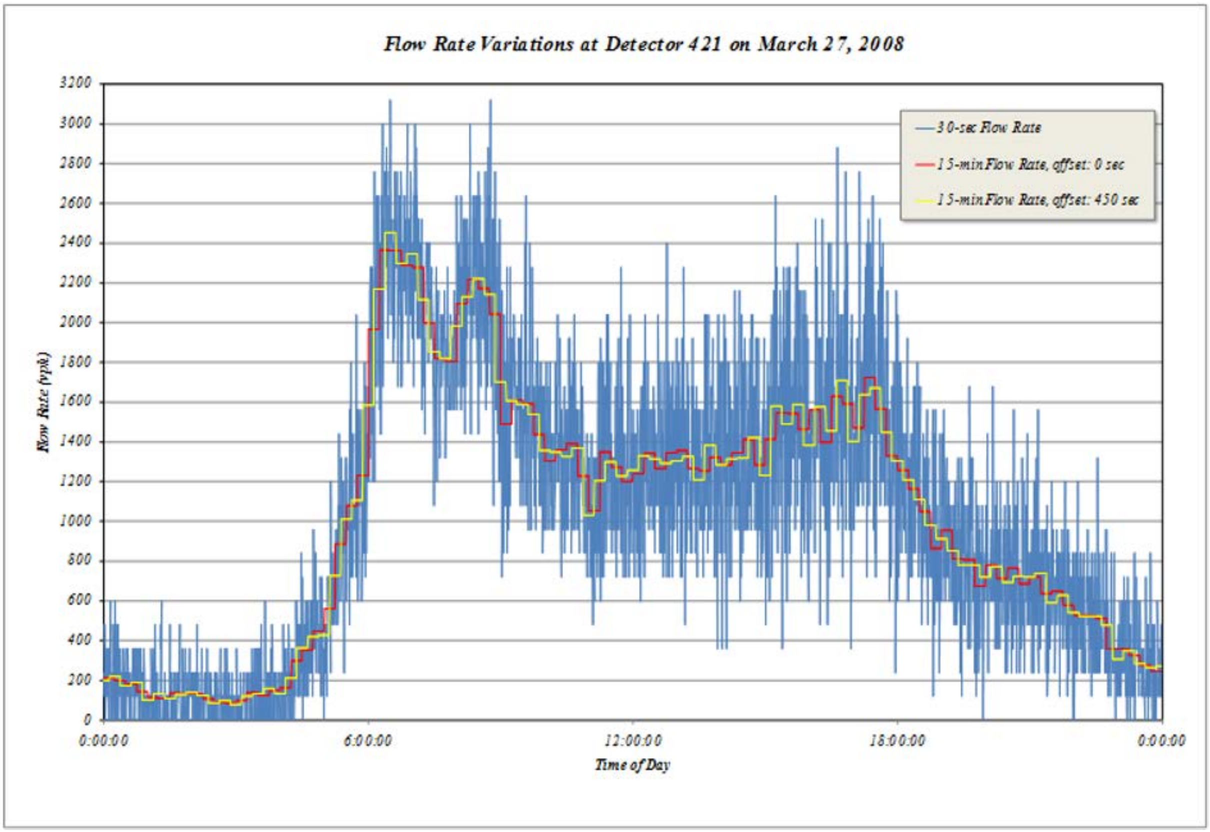

Figure 1. 30-second and 15-minute Flow Rates from Real Data

are different. The 15-minute data peaks at 6:15:00 AM with a two-lane flow rate of 2,364 vph for one case and at 6:22:30 AM with a rate of 2,452 vph for the other. In fact, there could be 30 different ways, or offset values, of aggregating the 30-seond data. Chances are good that the one offset that yields the highest peak 15-minute flow rate and the one that yields the highest peak hour volume are not the same; in fact, neither will be picked most of the time.

This realization means that the peak 15-minute flow rate used for HCM analysis is often underestimated. In other words, the situation is no better and often far worse than HCM calculation reports.

2.2. Non-inclusive peak flow

The conventional approach to PHF involves first identifying the peak hour of the day, and then the peak 15 minutes within the peak hour, which is simple and easy to follow. With the availability of 30-second data, one can still identify the peak hour by searching for the consecutive 120 30-second data points with the highest sum; similarly, the peak 15-minute period is represented by the consecutive 30 data points with the highest sum.

The problem is that the peak 15 minute period is not guaranteed to be entirely contained in the peak hour. A quick look at the one-day data from 1,669 detectors on Minnesota's Twin City Metro freeway reveals that about 18% of the detectors have their peak 15 minutes at least partially outside the peak hour. Of these, about 58% have their peak 15 minutes entirely outside the peak hour. By restraining the search duration within the peak hour and settle for the "local" peak 15 minutes, a certain proportion of all peak hour factors are artificially increased and as a result, the actual traffic is underestimated.

3. PHF Calculations

To address the discrepancies in PHF values resulting from the aggregation of high-resolution detector data, this paper proposes two alternative approaches to the current HCM calculation shown earlier in Eq.1 and presented herein with more details.

3.1. Aggregation approach (A)

The conventional method of PHF calculation uses 15-minute traffic counts. Since many existing computer programs and procedures use this protocol, simply aggregating the 30-second data into 15-minute volume and the old procedures and software can be done as before.
To aggregate the 30-second data, v_j, into 15-minute volume, V_i, assuming data were collected for the entire 24 hours by simply using the following equation:

$$V_i = \sum_{j=30i-29+k}^{30i+k} +v_j \quad (2)$$

where
V_i is the volume of the ith 15-min period of the day, i = 1,2, Ĕ, 96;
v_j is the volume of the jth 30-sec period of the day, j = 1,2, Ĕ, 2880
k is aggregation clock offset by a multiple of 30 sec., k = 0,1, Ĕ, 29.
Once the data are aggregated the peak hour, defined as the highest consecutive 15-minute counts, can be identified:

$$V = max\left[\sum_{i=n}^{n+3} V_i\right], n = 1, 2, ..., 93 \quad (3)$$

where V is the peak-hour volume.
Within the peak hour, the peak 15-minute volume, V_{15}, can also be identified easily.

$$V_15 = max[V_{i*}V_{i*+1}, V_{i*+2}, Vi * +3 \quad (4)$$

where $i*$ is the first 15-minute data point of the identified peak hour. Simply applying Eq. 1 would yield PHF for this approach.

3.2. Constrained approach (B)

The constrained approach also seeks to identify the peak hour first and then locate the peak 15-min period within that hour. The main difference between this approach and the aggregation approach is the searching steps, which depend on the resolution of data. Since 30-second data are used, the 120 consecutive data points (out of the total of 2,880) with the highest volume can be identified.

$$V = max\left[\sum_{j=n}^{n+119} v_j\right], n = 1, 2, ..., 2761 \quad (5)$$

After that, the peak 15-min period within the hour can be identified as:

$$V_{15} = max\left[\sum_{j=j*+n}^{j*+n+29} v_j\right], n = 0, 1, ..., 90 \quad (6)$$

Where $j*$ is the first 30-second data point of the identified peak hour.

3.3. Unconstrained approach (C)

The distinction of the unconstrained approach is that it searches for both the peak hour and the peak 15 minutes independently from each other. That is, the peak 15-minute period does not have to lie within the peak hour. This approach guarantees the true peak hour, not aggregated to the nearest 15 minutes, and the true peak 15 minutes, not within any confine of time. The calculation of the peak hour volume, V, is the same as Eq. 5. Thus, the peak 15 minutes can be identified as follows:

$$V = max\left[\sum_{j=n}^{n+29} v_j\right], n = 1, 2, ..., 2851 \quad (7)$$

4. Real-Word Data Descriptions

Traffic-count data collected by the Minnesota Department of Transportation at 30-second intervals were used for the analysis. Data from 60 detectors widely distributed on the Twin City Metro freeways network (see Fig. 2) were chosen to avoid close proximity and, potentially, high correlations.
For the temporal horizon, sixty days in twelve consecutive weeks in 2008 were selected. Since the peak periods are the focus of this study, only weekday data were considered. In total, there are 10,368,000 data points in 3600 detector-day combinations for the analysis.

5. Results and discussion

All peak periods (hourly and 15 minutes) were identified for each day of count data using approaches A, B, and C, respectively. Statistics of PHFs from the three approaches are summarized in Table 1. For the data used in this study, about 17% (611 out of 3,600 cases) have the true peak 15 minutes partially or entirely outside of the peak hour.
If the resultant PHF are viewed from the three above approaches alone, the average underestimation in PHF by the conventional aggregation approach (A) of actual traffic is not great, only about 2% or 3% less than that from approaches B and C, respectively (see Figs. 3 and 4). This is because even though the peak 15-minute volume is on average underestimated by 7 to 8% (see Figs. 5 and

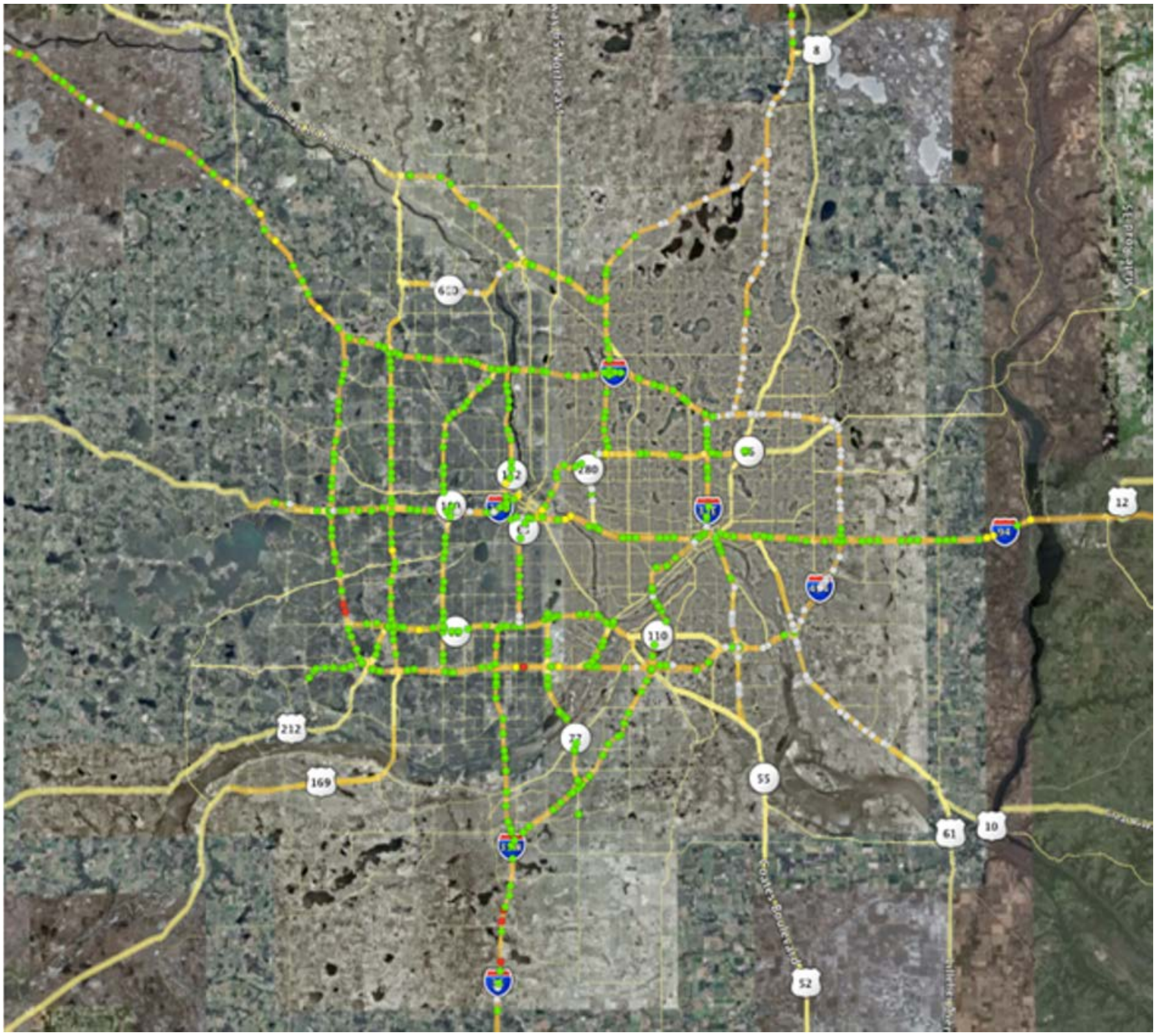

Figure 2. Twin City Metro Network System and Detector Locations

Table 1. Statistics of the PHFs by Calculation Approaches

PHF	Aggregation	Constrained	Unconstrained
Mean	0.8898	0.8723	0.8653
Std. Div.	0.0751	0.0741	0.0759
Max	0.9969	0.9824	0.9783
Min	0.2772	0.3590	0.3590

6), the peak hour volume is also underestimated. As the PHF is a function of the ratio of peak hour volume to the peak 15-minute volume, when both are underestimated by a similar magnitude, the underestimation in the resultant PHF appears to be less significant.

The cumulative frequency in Figure 7 suggests that more than 50% of the counters saw a 5% underestimation in their peak 15-minute volumes and more than 20% saw a 10% underestimation when approach A was used instead of the unconstrained approach C. Figure 8 also tells a similar story when approach A was used instead of the constrained approach B. These are very significant discrepancies.

In the absence of field measurements, a lower PHF leads to a higher design hour flow rate for a HCM analysis [6]. As shown previously, the PHF estimated by an alternative approach, say approach B, is on average about 2% lower than the conventional approach, A. Figure 9 illustrates the impact of 2% reduction in PHF on the estimation of control delay at a signalized intersection. Under higher

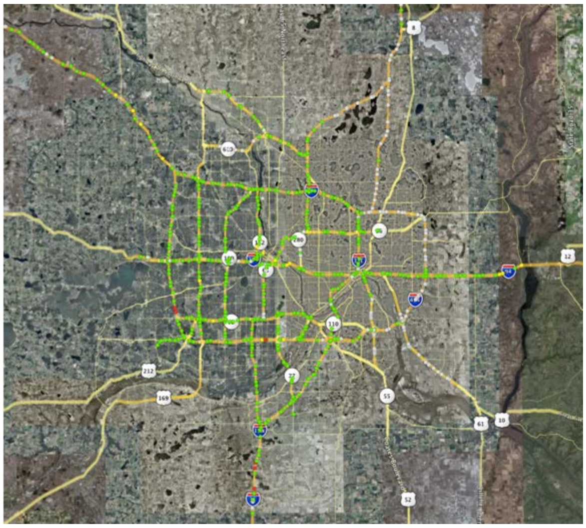

Figure 3. Comparison of PHF Computed by Approaches A and B

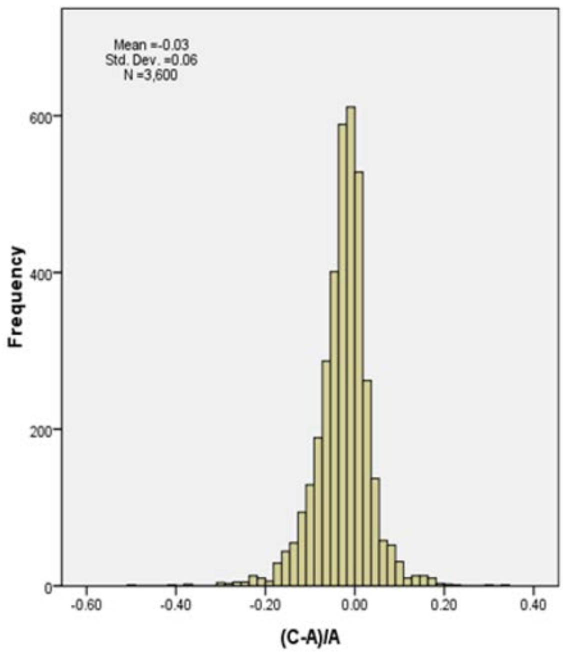

Figure 4. Comparison of PHF Computed by Approaches A and C

volume and lower PHF conditions, a 2% reduction in PHF can make an appreciable increase in the delay estimation. For example, control delay is estimated to increase over 5 seconds when the volume is greater than 800 vph and PHF less than 0.92. A drop in level of service, LOS, is possible for such a case.

Now, what if there is a reduction of 5% (only about one standard deviation from the mean) in PHF? Figure 10

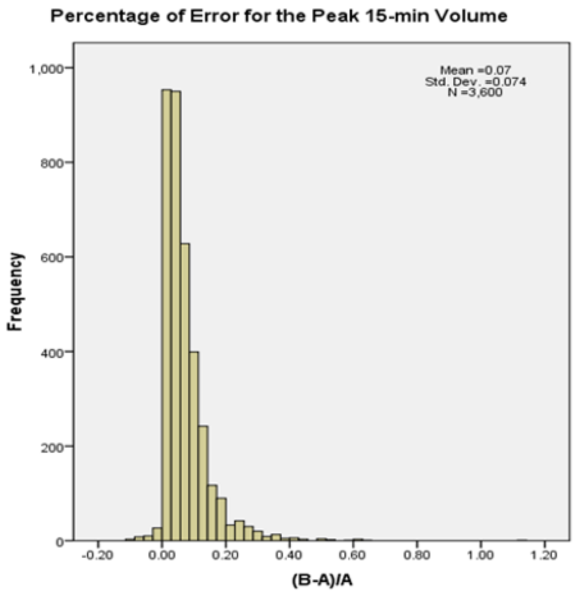

Figure 5. Comparison of Percentage Error between B and A

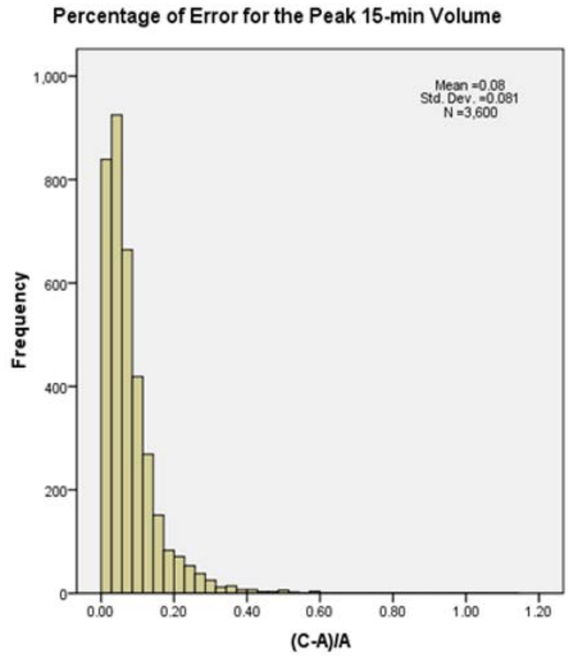

Figure 6. Comparison of Percentage Error between C and A

shows that much larger control delay can result. For the same case with 800 vph and PHF around 0.92, an increase of about 20 seconds could result. The drop in LOS would be quite significant.

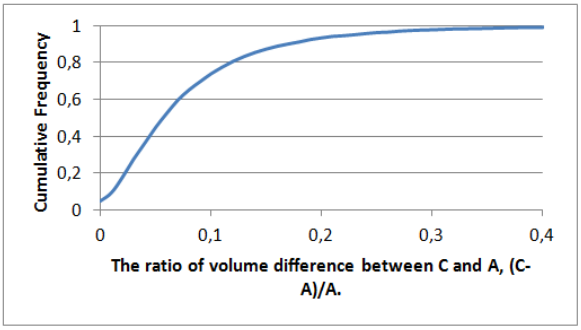

Figure 7. Cumulative Frequency of the Ratio of Volume Difference between C and A

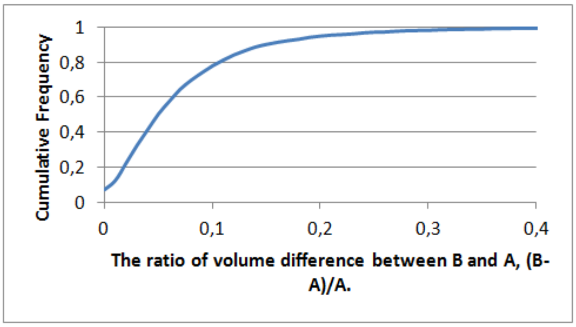

Figure 8. Cumulative Frequency of the Ratio of Volume Difference between C and A

6. Summary and recommendations

While it is expected that PHF may vary over time and by locations, the impact of increased data resolution has not been previously recognized or explored. While it may be tempting to just aggregate the count data of finer resolution, e.g. 30-second data, to a more familiar level, i.e. 15-minute data, the insights lost and the errors introduced by this simple act may be significant, as demonstrated in this paper. Both non-aggregated approaches, constrained (B) or unconstrained (C) identify the true peak hour and more realistic peak 15-minute volume for a set of count data. It is evident that the unconstrained approach provides the true peak 15-minute flow rate, but the authors also acknowledge that the constrained approach with the peak 15-minute with the true peak hour is, arguably, more in line with HCM's original spirit. With these considerations in mind, the authors make the following recommendations.

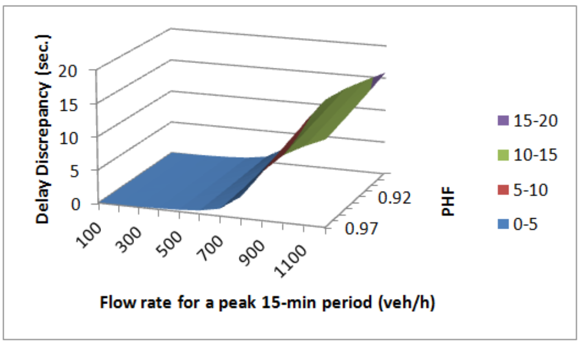

Figure 9. Impact of 2% Reduction in PHF on Control Delay

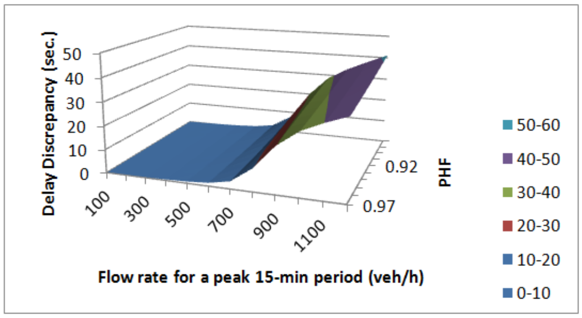

Figure 10. Impact of 5% Reduction in PHF on Control Delay

1. If the count is natively 15-minute volume data, the conventional approach for PHF calculation is still the best alternative.

2. When applying the new PHF to obtain design hour flow rate, one should use it with the "true" peak hour calculated from 30-second data. If such is not available, a 2 to 5% reduction on PHF may be exacted to compensate for the use of conventional peak hour volume.

3. TRB's Highway Capacity and Quality of Service committee may want to look into various MOE methodologies to take into consideration the impacts of increased data resolution.

Highly organized and integrated data can support informed and comprehensive transportation decision-making if appropriate details can be presented. This paper shows the data quality would not be good enough for PHF estimation if they were not processed properly. An inappropriate PHF estimation may conclude a poor decision. Therefore, data quality should be evaluated carefully with what they had been processed for transportation data integration systems.

References

[1] Marotta A., Cancela H., Peralta V., Ruggia R., Reliability Models for Data Integration Systems, Simulation Methods for Reliability and Availability of Complex Systems, Springer Series in Reliability Engineering, Part II, 123-144, 2010

[2] Zhang, X., Wang, Y., Nihan, N. L., and Hallenbeck, M. E., Development of a System for Collecting Loop-Detector Event Data for Individual Vehicles, Transportation Research Record, No.1855, 2003, 168-175

[3] Transportation Research Board (TRB). Highway Capacity Manual, Special Report 209, National Research Council, Washington, D.C., 2000

[4] Dowling R. G. Use of Default Parameters for Estimating Signalized Intersection Level of Service. Transportation Research Record, No.1457, 1994, 82-95

[5] Yi P., Wolfenbarger J., Raavi V.S., Yu L., Investigation of Peak Flow Variations: Preliminary Field Studies, Transportation Research Record, No. 2130, 2009, 75-82

[6] Han D. L., Li J.-M., Short or Long - Which Is Better? Probabilistic Approach to Cycle Length Optimization. Transportation Research Record, No. 2035, 2007, 150-157

[7] Tarko A.P., Perez-Cartagena R. I. ,Variability of Peak Hour Factor at Intersections. Transportation Research Record, No.1920, 2005, 125-130

[8] Ji X., Prevedouros P.D., Probabilistic Analysis of Highway Capacity Manual Delay for Signalized Intersections, Transportation Research Record, No.1988, 2006, 67-75

[9] Akcelik R., Rouphail N., Estimation of Delays at Traffic Signals for Variable Demand Conditions. Transportation Research, Part B, Vol. 27B, 1993, 109 - 131

[10] Choe T., Skabardonis A., Varaiya P., Freeway Performance Measurement System: Operational Analysis Tool, Transportation Research Record, No.1811, 2002, 67-75