Estimating the vehicle accumulation: Data-fusion of loop-detector flow and floating car speed data

Paul B.C. van Erp (corresponding author)
Victor L. Knoop
Serge P. Hoogendoorn
Department of Transport & Planning
Faculty of Civil Engineering and Geosciences
Delft University of Technology
Stevinweg 1, 2628 CN Delft, The Netherlands
Tel: +31-15-278-4912
Email: {p.b.c.vanerp,v.l.knoop,s.p.hoogendoorn}@tudelft.nl

1706 words (excl. title page and references) + 2 figures + 1 table

Acknowledgments
This research was conducted as part of a collaboration between the Amsterdam Institute for Advanced Metropolitan Solutions (AMS) and Google. We thank these institutions and the other data provider, i.e., the Dutch road Authority Rijkswaterstaat, for helping us to successfully conduct this research. This research is co-funded by AMS and the Netherlands Organisation for Scientific Research, i.e., NWO in Dutch.

October 27, 2017

INTRODUCTION

Queue lengths, either in distance or number of vehicles, are important in traffic management applications. Intersection controllers can minimize delays based on real-time queue length estimates (1), (2). Furthermore, a capacity drop can be avoided by limiting inflow into a freeway, and thus for instance storing vehicle on an on-ramp, e.g., (3), (4). This control approach requires real-time queue length estimators are needed to determine available buffer-space while preventing spill-back from the buffers.

We describe the queue length on a link in terms of the number of vehicles, i.e., vehicle accumulation N. The vehicle accumulation can be estimated based on the link arrivals a and departures d within a period p. These can be observed using road-side detectors placed at the upstream (arrivals) and downstream (departures) end of the link:

$$N(p) = N(p-1) + a(p) - d(p) \quad (1)$$

Where $N(p)$ denotes N at the end of p. Given a known initial number of vehicles $N(0)$ and correct arrival/departure observations, the vehicle accumulation can be estimated without error. However, under observation errors a problem arises. Due to the recursive nature of this estimation methodology, the estimation error is the sum of all past observation errors. This is known as the cumulative error problem. Prior researchers have addressed this problem by using an additional data-source, e.g., (5), (6), (7).

In this research we propose a new methodology to use aggregated floating car speed data to address the cumulative error problem. The use of aggregated floating car data instead of individual observations is a result of privacy restrictions. The real data used in this research ensures privacy. We use these data for error recovery and online learning of the error characteristics.

DATA CHARACTERISTICS

This research uses three types of real data. These are: (1) loop-detector, (2) floating car and (3) radar data. The characteristics of these data-types are shown in Table 1. We assume that there is no data latency, i.e., the data is available directly at the end of the related time-period. However, the methodologies presented in the next section can be adjusted to cope with a potential data latency in the floating car data.

TABLE 1 Characteristics of the three data-types used in this research

Name	Floating car (Google)	Loop-detector	Radar
Traffic variable	Speed	Passings (flow)	Speed
Spatial char.	Road stretch	Location	Road stretch
	120 and 94 m	Up- and downstream boundaries	7 m
Temporal char.	Period	Period	Period
	300 s	10 s	10 s
Spatial/temporal	Area	Line	Area
Used for	Estimation	Estimation	Evaluation

The data are obtained from two sources. Google has provided the floating car speed data as part of its Better Cities program (8). The purpose of this program is to investigate the added value of Google data in tackling mobility problems. The source of these data are mobile devices using Google services, e.g., Android or Google Maps, for which the users explicitly agreed to share their location data with Google. To ensure privacy, Google aggregates the data and applies a differential privacy filter. The effect of this filter is limited in high traffic demand, e.g., peak, periods as studied in this research. The loop-detector and radar data are collected as part of the Praktijkproef (field operational test) Amsterdam (9).

METHODOLOGY

Each floating car data period, i.e., 300 s, we use the data to estimate the mean vehicle accumulation, see equation (4). This estimate serves as a reference estimate to the loop-detector data-based recursive estimates, i.e., equation (2), which is obtained each loop-detector data period, i.e., 10 s. The reference estimator allows us to address the cumulative error problem.

In our methodology, the reference estimator is used to (1) reduce the size of the cumulative error build-up in the past and (2) reduce the size of the cumulative error build-up in the future. We denote these two processes as (1) error recovery and (2) online learning of the bias term. In error recovery, we recover the cumulative error build up in past periods. This influences the initial vehicle accumulation used to estimate the vehicle accumulation for the next periods. In online learning of the bias term, we opt to learn the size of the structural error. We expect that the future loop-detector data also contains this structural error and take it into account in the estimator, i.e., by defining b in equation (2). Below we provide the estimation sequence of the proposed estimation methodology. This estimation sequence describes the methodology in a concise manner. For additional explanations the reader is referred to the full article (to be published soon).

To describe the time period, three variables are used, i.e., ι, p_s and p_l. Starting from $t = 0$, p_s and p_l respectively denote the short and long period with durations $\Delta t = 10$ s and $\Delta\tau = 300$ s. Within a long period p_l, ι is used to describe the short periods starting from $\iota = 1$. This means that $p_s = p_l n + \iota$, where the number of short periods within a long period is given by $n = \Delta t/\Delta\tau$. Two notations are used to define the time-period of a variable. For N these are (1) $N(p_s)$ and (2) $N_{p_l}(\iota)$. Furthermore, in the error recovery process, the prior (initial) vehicle accumulation N and related error variance σ^2 are corrected. To denote whether we are dealing with the estimate before or after correction, we respectively add a minus $(-)$ or plus $(+)$ to these variables. Other important variables which are not explained in the estimation sequence are σ^2_{ind}, $\mathbf{R}$ and $\hat{\mathbf{d}}$. These are respectively the detector error variance per departure, the error variance of $\overline{N}_{FL}$ and the fractional contributions of individual errors to the cumulative error.

Estimation sequence

1. Set $p_l = 1$ and define initial vehicle accumulation $N_1^-(0)$ and error variance $\sigma^2_{1,-}(0)$.
2. Inner loop:
   (a) $\iota = 1$.
   (b) Estimation vehicle accumulation $N_{p_l}^-(\iota)$ based on loop-detector arrivals $\hat{a}$ and departures $\hat{d}$:

$$N_{p_l}^-(\iota) = \max\left[N_{p_l}^-(\iota-1) + \hat{a}(\iota) - \hat{d}(\iota) + \hat{d}(\iota)b(p_l-1), 0\right] \quad (2)$$

$$\sigma^2_{p_l,-}(\iota) = \sigma^2_{p_l,-}(\iota-1) + \hat{d}(\iota)\sigma^2_{ind} \quad (3)$$

   (c) Real-time estimate: $N(p_s) = N^-_{p_l}(\iota)$, where $p_s = p_l n + \iota$.
   (d) If $\iota = n$ go to step 3, else $\iota = \iota + 1$ and go back to step. 2b
3. Estimation of long-period time-mean vehicle accumulation $\overline{N}_{FL}(p_l)$ based on floating car mean speed u and loop-detector mean flow $\bar{q}$:

$$\overline{N}_{FL}(p_l) = \frac{L}{u(p_l)}\bar{q}(p_l) \tag{4}$$

4. Error recovery, this yields the corrected estimates, i.e., $N^+_{p_l}(n)$ and $\sigma^2_{p_l,+}(n)$:
   (a) Prior estimate $\mathbf{x}^-$:

$$\mathbf{x}^-(p_l) = \left[N^-_{p_l}(0) \cdots N^-_{p_l}(n)\right] \tag{5}$$

   (b) Prior long-period estimate $\overline{N}^-$:

$$\overline{N}^-(p_l) = \frac{1}{2}\frac{\Delta t}{\Delta \tau}\sum_{\iota \in 1}^{n}\left(N^-_{p_l}(\iota - 1) + N^-_{p_l}(\iota)\right) \tag{6}$$

   (c) Prior error covariance matrix $\mathbf{P}^-$:

$$\mathbf{P}^-(p_l) = \begin{bmatrix} \sigma^2_{p_l,-}(0) & \cdots & \sigma^2_{p_l,-}(0) \\ \vdots & \vdots & \vdots \\ \sigma^2_{p_l,-}(0) & \cdots & \sigma^2_{p_l,-}(n) \end{bmatrix} \tag{7}$$

   (d) Correction vector $\mathbf{K}$:

$$\mathbf{K}(p_l) = \frac{\mathbf{P}^-(p_l)\mathbf{C}^T}{\mathbf{C}\mathbf{P}^-(p_l)\mathbf{C}^T + \mathbf{R}} \tag{8}$$

   where

$$\mathbf{C} = \frac{\partial \overline{N}^-}{\partial \mathbf{x}} \tag{9}$$

   (e) Posterior estimate $\mathbf{x}^+$:

$$\mathbf{x}^+(p_l) = \mathbf{x}^-(p_l) + \mathbf{K}(p_l)\left[\overline{N}_{FL}(p_l) - \overline{N}^-(p_l)\right] \tag{10}$$

   (f) Posterior error covariance matrix $\mathbf{P}^+$:

$$\mathbf{P}^+(p_l) = \left[I - \mathbf{K}(p_l)\mathbf{C}\right]\mathbf{P}^-(p_l) \tag{11}$$

5. Online learning of the bias term, this yields the updated bias term estimate, i.e., $b(f)$:
   (a) Individual estimate of the bias term $\tilde{b}(p_l)$:

$$\tilde{b}(p_l) = \frac{\Delta\overline{N}_{CV}(p_l) - \Delta\overline{N}_{FL}(p_l)}{\hat{\mathbf{d}}^T(p_l)\mathbf{o}} \tag{12}$$

   (b) Update combined estimate of the bias term $b(p_l)$:

$$b(p_l) = b(p_l - 1) + \alpha\left[\tilde{b}(p_l) - b(p_l - 1)\right] \tag{13}$$

6. $p_l = p_l + 1$, $N^-_{p_l}(0) = N^+_{p_l - 1}(n)$, $\sigma^2_{p_l,-}(0) = \sigma^2_{p_l-1,+}(n)$ and go back to step 2.

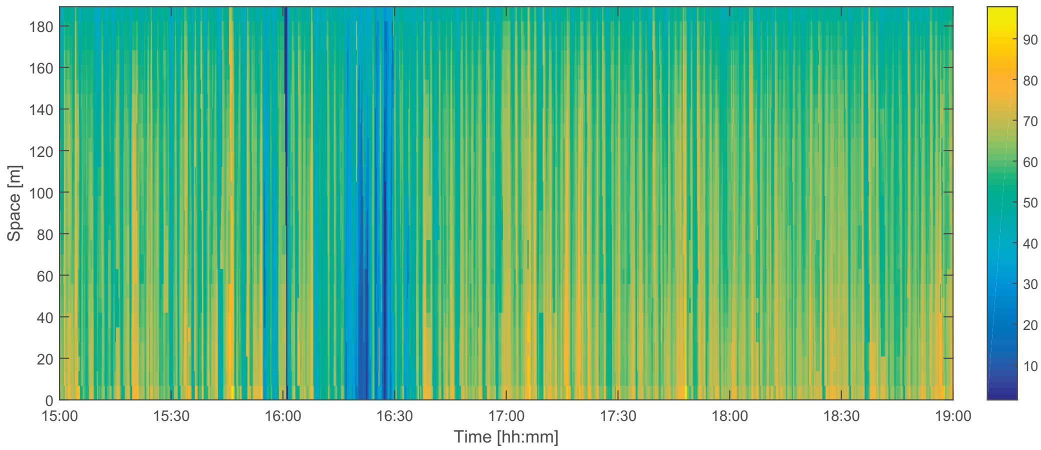


FIGURE 1 Mean speed over the two on-ramp lanes retrieved from the radar speed data

FINDINGS

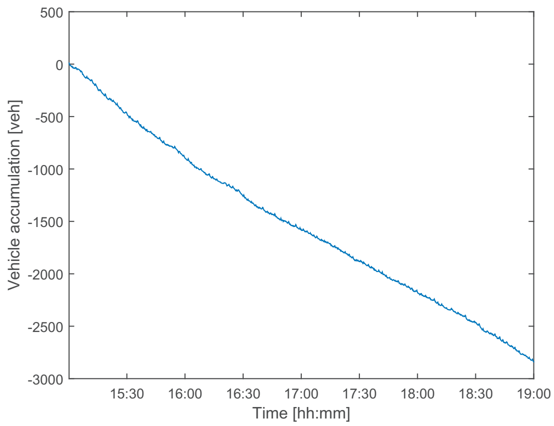
The three data-types shown in Table 1 are all available for an on-ramp in the Metropolitian Region Amsterdam in the evening peak, i.e., 15:00-19:00h, on July 6th, 2016. The detector observing the upstream end of the on-ramp was out-of-order in this period. Instead we use the relevant outflows of the upstream intersections, which cover all legal inflows and are separated from other flows.

The radar speed plot (Figure 1) allows us to evaluate the estimation performance. From this plot we derive the periods in which a queue was present on the on-ramp. Two queues stand out. Firstly, the largest queue is observed in the period 16:15-16:30. Secondly, we observe lower speeds around 15:55.

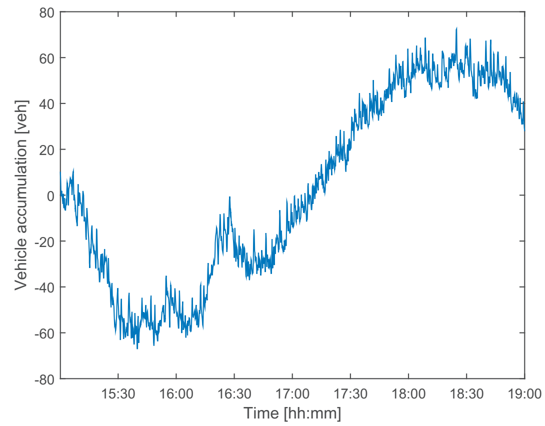
The methodology discussed in the previous section is applied to estimate the vehicle accumulation. For this purpose, three parameters have to be defined, i.e., σ^2_{ind}, $\mathbf{R}$ and α. The following values are used in this research: $\sigma^2_{ind} = 0.125$ veh^2, $\mathbf{R} = 5.0$ veh^2 and $\alpha = 0.25$. Furthermore, we define $N_1^-(0) = 10$ veh and $\sigma^2_{1,-}(0) = 5.0$ veh^2.

Figure 2 shows six different estimates of the vehicle accumulation over time. The difference between the six estimators lies in the use floating car (Google) data and the (initial) bias term. The (initial) bias is set to 0.0 (no bias) and 0.36, where the latter is based on the measured cumulative in- and outflows over the four-hour peak-period. Therefore, the estimators having an (initial) bias term of 0.36 actually incorporate additional information of the data. In reality, this information was not available for real-time vehicle accumulation estimates.

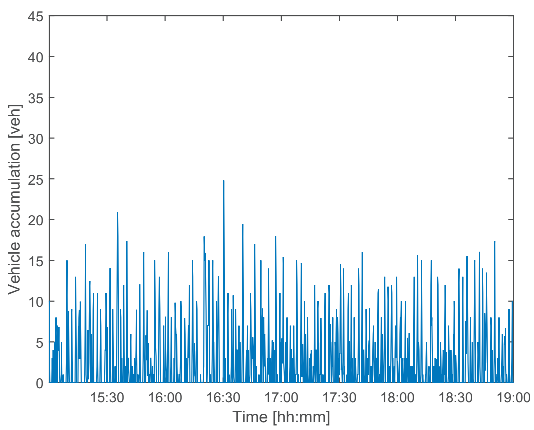
The estimators which do not use floating car data, i.e., Figure 2a and 2b, show that the loop-detector data is biased and that the bias changes over time. If we use the floating car data for error recovery, the estimates improve. However, it is still important whether we also use the floating car data for online learning of the bias term. The four figures, i.e., Figures 2c till 2f, show that online learning of b has two advantages. Firstly, our (initial) guess of b is less critical. Therefore, without having any prior information related to the b the estimator which applies online learning still yields good estimates, see Figures 2c and 2e. Secondly, online learning is able to capture changes in b. In our case study, the bias seems to be lower in the period 15:30-18:00 than before and after this period, see Figure 2b. As a result, the estimators applying online learning seem to be more accurate in estimating the build-up of N than the one having a fixed b of 0.36. In Figure 2d the second-highest peak is observed around 17:35, while in Figures 2d and 2d the second-highest peak is observed around 15:55, which coincides with the radar speed data.



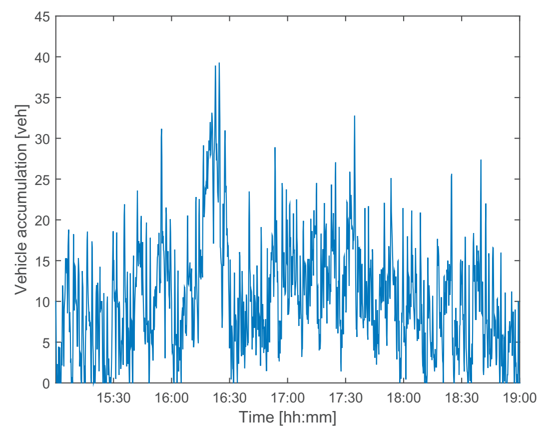
(a) No floating car data and negative N allowed, $b = 0.0$



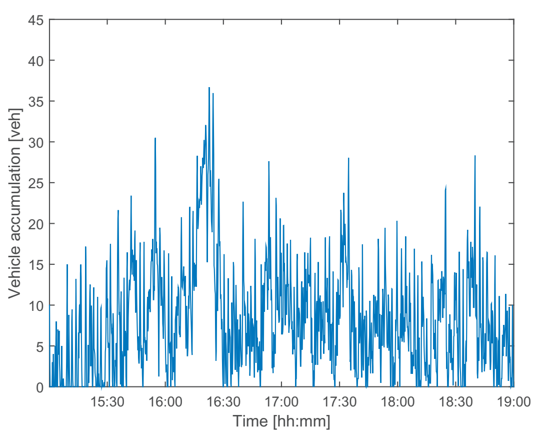
(b) No floating car data and negative N allowed, $b = 0.36$



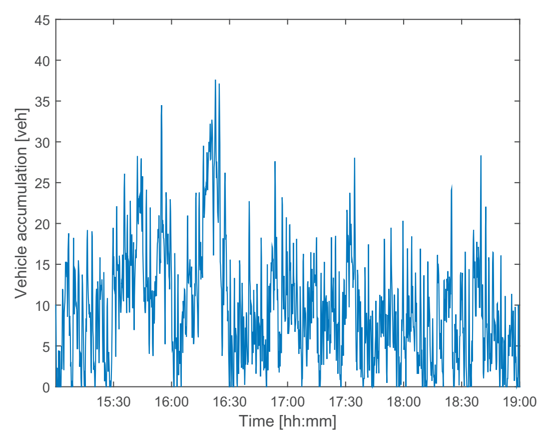
(c) Error recovery + fixed bias term, $b = 0.0$



(d) Error recovery + fixed bias term, $b = 0.36$



(e) Error recovery + online learning of b, $b(1) = 0.0$



(f) Error recovery + online learning of b, $b(1) = 0.36$

FIGURE 2 Vehicle accumulation estimates

CONCLUSIONS

In our case study we observe that adding information in the form of mean speed to detector measurements successfully addresses the cumulative error problem. The estimates no longer drift away from feasible values, and capture the build-up of a queue on the on-ramp. Both applications of the mean speed information, i.e., error recovery and online learning of the bias, contribute to improving the quality of the vehicle accumulation estimates.

Improvements in the real-time vehicle accumulation estimation is beneficial for the effectiveness of dynamic traffic management systems. The estimates are, for example, important in traffic management systems which temporarily store traffic on buffers to temporarily reduce the inflow to a downstream bottleneck. Accurate estimates can prevent spill-back or underusing the buffer capacity.

The mean speed information used in this research is provided by Google as part of their Better Cities program. The results are thus dependent on both the quality of the Google data and the proposed methodologies. However, if other providers are able to provide reliable mean speed information, our methodologies can be applied with their data.

REFERENCES

[1] Varaiya, P., Max pressure control of a network of signalized intersections. *Transportation Research Part C: Emerging Technologies*, Vol. 36, 2013, pp. 177–195.

[2] Le, T., P. Kovács, N. Walton, H. L. Vu, L. L. H. Andrew, and S. S. P. Hoogendoorn, Decentralized signal control for urban road networks. *Transportation Research Part C: Emerging Technologies*, Vol. 58, 2015, pp. 431–450.

[3] Papageorgiou, M., H. Hadj-Salem, and J. Blosseville, ALINEA: A Local Feedback Control Law for On-ramp Metering; A Real-life Study. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1320, 1991, pp. 58–64.

[4] Hoogendoorn, S. P., R. Landman, J. Van Kooten, and M. Schreuder, Integrated Network Management Amsterdam: Control approach and test results. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 2013, pp. 474–479.

[5] Vigos, G., M. Papageorgiou, and Y. Wang, Real-time estimation of vehicle-count within signalized links. *Transportation Research Part C: Emerging Technologies*, Vol. 16, 2008, pp. 18–35.

[6] Van Lint, H. and S. Hoogendoorn, A generic methodology to estimate vehicle accumulation on urban arterials by fusing vehicle counts and travel times. In *94th AnnualMeeting of the Transportation Research Board*, 2015, January.

[7] Amini, Z., R. Pedarsani, A. Skabardonis, and P. Varaiya, Queue-Length Estimation Using Real-Time Traffic Data. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016, pp. 1476–1481.

[8] Eland, A., Tackling Urban Mobility with Technology. *https://europe.googleblog.com/2015/11/tackling-urban-mobility-with-technology.html (accessed on 2017-07-26)*, 2015.

[9] Praktijkproef Amsterdam (2013 - 2015). *www.praktijkproefamsterdam.nl (accessed on 2017-07-26)*, 2015.