Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 10 (2015) 113 – 123

Transportation Research Procedia

www.elsevier.com/locate/procedia

18th Euro Working Group on Transportation, EWGT 2015, 14-16 July 2015, Delft, The Netherlands

Estimation of missing flow at junctions using control plan and floating car data

Xiao Xiao[ab] , Yusen Chen[ab]*, Yufei Yuan[a]

[a]*Delft University of Technology, Stevinweg 1, 2628CN Delft, Netherlands*
[b]*TNO, van Mourik Broekmanweg 6, 2628XE Delft, Netherlands*

Abstract

This paper proposes a new and consistent approach for estimating missing flow by analyzing data from SCATS (containing both flow and timing plan at junctions) and FCD (Floating Car Data). SCATS system provides flow data and timing plan at a 5-minute interval, while FCD contains information of taxi trajectories with speed and position for each vehicle at a 30-second interval.
Two objectives are defined in this paper: 1) to summarize major methods of flow estimation and create a generalized framework in flow estimation, 2) to research for the possibility of improving utilization of traffic flow data, by comparing methods from multiple aspects, to provide accurate and reliable source for traffic research and application.
The paper devises three consistent methods to estimate missing flow at junctions. Firstly, historical flow data of a specific lane is used to make an initial estimation. Flow estimation values are estimated from each single lane and normalized to further complement Secondly, flow values from adjacent lanes with similarities are processed to compare with the estimated lane, for which timing plan is applied to identify relevant control group (same turning lanes) with their relative phase time proportions. Proportions and flow rate from observed lanes on the same control group are normalized to make an estimation at missing lanes. Thirdly, the information of FCD (such as speed) is used to estimate corresponding flow value. Typically detected flow and FCD speed relationship is established from junction streams. This relation is then applied back to the stream to calculate traffic flow. Each of the methods is expected to perform under varying situations. If all the results are proved to be reliable to a certain degree, they can be iterated for mutual verification and consistency.
This methodology has been applied to Changsha municipality in China. Initial results indicate that suggested methods give promising indication that almost all missing lane flow could be recovered using these three methods. Further research is ongoing to investigate specific data fusion mechanism or interchangeable data source for traffic state estimation and its quality. Further research will also consider adjacent junctions within a given area to understand how data and flow relationship works at the network level.

Peer-review under responsibility of Delft University of Technology

* Corresponding author. Tel.: +31 651490662.
E-mail address: Yusen.chen@tudelft.nl

Peer-review under responsibility of Delft University of Technology
doi:10.1016/j.trpro.2015.09.061

Keywords: Control timing plan; junction loop flow, FCD; missing flow estimation; data fusion.

1. Introduction

1.1. Background

The signal control scheme at junctions is influenced by traffic flow and traffic status significantly. Especially for adaptive traffic control system such as SCATS (Sydney Coordinated Adaptive Traffic System), which adapt signal plan based on the traffic volume and saturation degree, while both factors are derived from the flow data detected at loop near junction. Thus, Traffic flow detection is crucial for accurate traffic state estimation and efficient traffic management in urban areas and on freeways. However, due to system or device failure, traffic flow obtained from the urban traffic control system shows frequently irregularity. For example, 5-10% on average of the available data from dual loop systems in Dutch freeway networks are missing or unreliable (Van Lint & Hoogendoorn, 2009). Similar situations exist in urban traffic monitoring systems. This fact can be further demonstrated in section 1.2 by using the data SCATS as a case. Reliable flow data are crucial no matter in ex-post or real-time situation; if the availability of flow data is not sufficient, the missing data estimation at junctions or on road section is necessary. Traffic flow acts as a significant role in accurate traffic state estimation and efficient traffic management both in urban areas and on freeways.

Previous studies have sought to improve the quality and comprehensiveness of raw observation data from monitoring systems. They consider the problem from varying aspects. On the one hand, some researchers deal with the problem by focusing on the coherence of traffic flow on road section; for instance, Chen et. al. (2003) apply an imputation method to filter out bad data samples and to form a complete clean data set from single-loop systems. This method considers the relations of detected flow corresponding to their relative locations, and it can be applied in both urban and freeway networks. On the other hand, some tackle the problem by emphasizing the similarities of flow pattern over a time cycle, which is more commonly used; for example, Wall et. al. (2003) present a time-series algorithm for correcting errors in freeway traffic management system archived loop data. Except for getting reference traffic values directly from time or space, some use more complex way to estimate flow and achieve much more insight at the same time. For example, Treiber and Helbing (2002) develop an adaptive smoothing method based on the notions from the first-order traffic flow theory, to reconstruct and clean flow observations from dual-loop systems. This approach has been further generalized by Van Lint and Hoogendoorn (2009) to fuse multiple data sources. Yuan et. al. (2012) apply a regression analysis from inductive statistics to estimate multi-class and multi-lane flow counts from generic freeway surveillance systems. It is based on the correlation between lane and class disaggregated counts. These researchers and studies involve the estimation or prediction of traffic flow to some extent. However the topic is seldom raised separately and comprehensively to be generalized into a framework. In this paper, the focus will be given on the estimation of missing traffic flow at urban junction. It follows a concept to estimate the traffic state by checking the consistency and generic pattern of multiple data sources, which is similar to the concept derived by Ou et. al. (2013) who developed a series of data fusion methods using data-data consistency.

1.2. Research question

The paper aims at developing a generalized and cross-compared methodology for missing flow estimation at junctions, using available data sources. The main research question then is: what are the good ways to estimate missing flow at junction when loop detectors fail? Several sub-questions are formulated as: what is the data quality at junctions in a traffic control system such as SCATS system? what are available methods for missing flow estimation? how do the methods work in missing flow estimation?

The question of data quality will be answered in the following part of section 1; methodology the paper is introduced in section 2, while experiments setup of case study is given in section 3. In section 4, the results of experiment are provided followed by analysis and comparison. Finally the conclusion is given in section 5.

1.3. Data analysis

The data used in the paper is obtained from the Changsha municipality in China. Two data sources are available; loop data from SCATS systems and GPS data from floating car (FCD, taxi data). The former one can provide traffic flow volume every 5 min and the latter one can give FCD records of speeds and positions at a 30-second-interval. The quality of traffic flow data from the record of SCATS system is analysed first. Detectors for SCATS system are implemented to junctions in Changsha city in two consecutive terms (time periods). 102 junctions are equipped with detectors in the 1st term (2011) and 104 junctions are equipped with detectors in the 2nd term (2013).

The ratio of data availability at one junction is defined as number of observed data availability over a whole day. Fig. 1 shows sample results of the analysis of current data quality in the urban area for one day (23-April-2013) for all junctions. Each bar in Fig. 1 represent one dataset from a junction, the number on X-axis (1 to 102) refers to the junction number. The value of availability for every bar ranges from 0 to 1, in which 0 shows no data available at a junction, and 1 shows all detectors works well over the whole day.

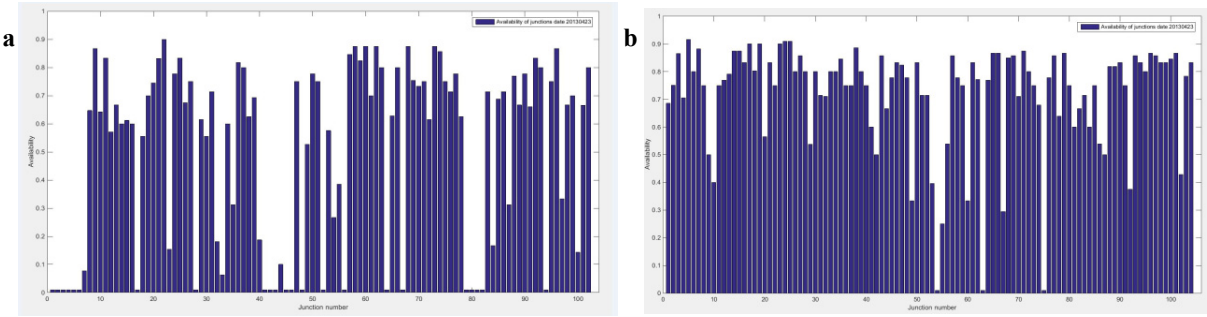

Fig. 1. Date 20130423 (a) Data quality for all 1st term junctions in SCATS system, (b) data quality for all 2nd term junctions in SCATS system

It is depicted that data missing occurs at different levels, the worst case is that data from a whole day is not recorded. After calculation, the average rate of data availability is 51% over all the 1st term junctions and 75 % over all the 2nd term junctions in 2013. Since the junctions from the 2nd term have generally better data availability than the ones in the 1st term, test cases are picked up from the 2nd term junctions for the convenience of comparison of estimated data and ground-truth. And the developed methods are applied to both the 1st term and the 2nd term junctions for validation purposes.

As for FCD data, in previous relative research work of the authors, the records are formed into consecutive trajectories, these trajectories can cover the roads concerned, and the speed or location information provided by FCD can be used in the paper. .

2. Methodology

This section presents a generic concept of traffic flow estimation followed by three specific methods (or aspects of the general framework, in all the sentence below, they are referred to as 'methods'), while for each method, there are multiple ways of implementation. Without lacking of generality, only some part for each method will be introduced in this paper.

Traffic flow (q) is defined as the number of vehicles passing through a given location within a given period of time. When the flow on a detector is missing, a direct way is to estimate it referring to other flow observations from a same system. These observations can be chosen from records over other space or time, only if they can be proved to have a close relation to the detection with the missing data. Another way is to consider the relation of traffic flow to other traffic state values, such as density and speed. A definition has been commonly used in traffic research that flow equals to the product of density and speed and it can be expressed by: $q = k * u$. Thus, only if density k and speed u can be obtained, flow q can be computed. If flow is estimated from this aspect, usually the data of speed need to be obtained or computed from other resources directly or indirectly.

To combine these aspects, a general concept in this paper is formulated. It describes how the flow on specific lane may come from: directly from other observation of the same traffic states (flow) or indirectly calculated from other traffic states. Flow, speed and density can be observed at different location (expressed as l), date (expressed as d) and time of a day (expressed as t). While a suitable combination of them is expected to make reliable estimation for the traffic states (flow) at left hand.

$$q(l,t) = f(q(l,d,t), u(l,d,t), k(l,d,t)) \tag{1}$$

q: Traffic flow;
l : Location or position of a traffic state where it is detected;
d: Date or DOW (day of the week) when it is detected;
t : Time (time of day) of a traffic state when it is detected;
f: Relationship between flow and dependent factors;
u : Average vehicle speed;
k : Density, number of vehicles per unit length of the roadway;

The general concept, with its redundancy and alternative, is a main means to reconstruct or to find a missing value. It can make combinations fitting for multiple situations according to available sources, as long as systems or devices can somehow provide at least part of the data needed in this concept: Loop detectors can provide flow measurement at specific junction approaches or links during a given time period while GPS records can provide vehicle speed and trajectory at the same location but discretized time instances. Other sensors such as Camera (CCTV) would be able to provide trajectory travel time and section density. These sources are all related to the same targets 'flow' while reaching it differently.

In this paper, as stated in section 1 data analysis part, the current dataset available to be conducted to support the general concept can be described as follows: SCATS registers flow per time interval, which offers the possibility to address a missing value from a historical perspective. Its timing plan links different lanes to the same signal group. FCD provides the speed of a vehicle. Based on the data, three methods are chosen and are applied independently and conjointly to estimate a missing flow. These are (1) historical pattern: the historical flow value on a same lane over days of week provides typical pattern for the lane and its relation to the adjacent lanes; (2) timing plan: flow from each turning direction within the same signal group provides reference flow rate and inter-relationship among junction approaches; and (3) FCD: its speed and trajectory reflect traffic state independently. To examine the property of the estimation framework, each method will be tested separately, and then tested conjointly.

For Method 1 – Historical Pattern

Only the time aspect is considered. The assumption is that the historical data at the same lane could preserve similar flow pattern as the missing flow data. The flow observations from historical days could then be normalized or directly used. The equation expression of this method can be derived from general framework by only applying the observations of q in right side, while the q is from different days at the same location, thus factor l remain the same for q in both sides, so we omit it. This method can be represented as follows:

$$q_1(t) = f(q(d,t)) \tag{2}$$

For ex-post analysis, there are two ways of implementing method 1 concerning the inputting flow from two dimensions of the historical pattern:

(1) Use the historical info f_1: Average value of other full observed flows from other days in a week. The weekdays and weekends are separately considered empirically due to their different patterns.

$$f_1: q_1(d_i,t) = \begin{cases} \sum_{j\neq i, j\in N}^{n} q_j(d_j,t)/n & \forall\, i \in N \\ \sum_{j\neq i, j\in M}^{m} q_j(d_j,t)/m & \forall\, i \in M \end{cases} \tag{3}$$

Where: i: the date with missing flow; j: the date with available data;
N: week days in one or several weeks containing date i; n: number of j ∈ N;
M: weekend days in one or several weeks containing date i; m: number of j ∈ M.

(2) Use the historical info f_2: Values from same DOW (day of week).The example expression can be seen in the following equations:

$$f_2: q_1(d_i,t) = \sum_{j\neq i, j\in S}^{n} q_j(d_j,t)/n \quad \forall\, i \in S \tag{4}$$

Where: i: the date with missing flow; j: the date with available data;
S: days within a period of time with a same week of same DOW as date i; n: number of j ∈ S.

Considering the application of the methods, in reality, for an on-line control approach, usually there is no access to the information for the upcoming days of, so only the data from previous week days are used. These two ways of method 1 are then revised to: (1) Use the historical info f'_1: Average value of other full observed flows from days from the last week (2) Use the historical info f_2': Values from same DOW (day of week) from previous weeks. The example expression can be seen in the following equations:

$$f_1'\colon q_1(d_i,t) = \begin{cases} \sum_{j\neq i, j\in N}^{n} q_j(d_j,t)/n \ \forall\, i \in N \\ \sum_{j\neq i, j\in M}^{m} q_j(d_j,t)/m \ \forall\, i \in M \end{cases} \tag{5}$$

Where: i: the date with missing flow; j: the date with available data;
N: previous week days in one or several weeks; n: number of j ∈ N;
M: preivous weekend days in one or several weeks ; m: number of j ∈ M .

$$f_2'\colon\ q_1(d_i,t) = \sum_{j\neq i, j\in S}^{n} q_j(d_j,t)/n \ \forall\, i \in S \tag{6}$$

Where: i: the date with missing flow; j: the date with available data;
S: previous days within a period of time with a same week of same DOW as date i; n: number of j ∈ S.

For Method 2 Timing Plan

The relations of lane locations are used. In this paper, it is assumed that similar flow rate patterns exist flow and these relations between lanes are resulted from timing plan. Thus turning direction/lane information as well as control plan from traffic control (e.g. SCATS) system is used to define similar pattern groups. The equation expression of this method can also be derived from general framework by only applying the observations of q in right side, while q is from different locations at the same day, thus the factor d remain the same for q in both sides, so we omit it .

$$q_2(l,t) = f(q(l,t)) \tag{7}$$

Two ways of implementation are defined:

(1) Use the location info f_3 : The first function f_3 gives the estimation using the average value of all other lanes within the same control group;

$$f_{3:}\ q_1(l_i,t) = \sum_{j\neq i, j\in G}^{n} q_j(l_j,t)/n \ \forall\, i \in G \tag{8}$$

Where: i: the lane with missing flow; j: the lane with available data;
G: lane in the same signal control group as lane i; n: numberof j ∈ G.

(2) Use the location info f_4: The second function f_4 gives the estimation using average value of the other lanes from the same turning direction in a same control group.

$$f_{4:}\ q_1(l_i,t) = \sum_{j\neq i, j\in D}^{m} q_j(l_j,t)/m \ \forall\, i \in D \tag{9}$$

Where: i: the lane with missing flow; j: the lane with available data;
D: lane in the same turning direction as lane i; m: numbers of j ∈ D.

For Method 3 -- FCD

A data fusion concept is applied from external data sources aspect. The relation between aggregated loop flow and FCD speed is investigated. Then the relation is applied to estimate missing flow by FCD speed value, while the locations and date are chosen to be the same. From this viewpoint, the general formulae has become:

$$q_3(t) = f(u(t)) \tag{10}$$

For the FCD speed, it can to some extend represent the average speed of a road section, thus a fundamental relation is expected. Here the speed uses the same date and same position as only the flow needs to be estimated.

The generic concept can be fully or partly used in different kinds of situations, in which there are suitable ways of implementation methods. When applying the methodology into practice, their results can be cross-compared and

combined. In the case that all methods are to be used for flow estimation, an iterative process with generic formulae could be applied to achieve the convergence under reliable estimation, to ensure the consistent estimation and sound results. The three methods will be compared, while the suitable situation of each method and the iteration of three methods can only be discussed and implemented in the future related work. The way of iteration is firstly briefly introduced in this paper as follows:

Step 0: set $\Delta q(0) = 0$;

Step 1: estimate q_1, and relevant weight factor for its reliability w_1; (reliability w_1 is based on the closeness the estimated value using method 1 to the real value);

Step 2: estimate q_2, and relevant weight factor for its reliability w_2; (reliability $w_2$is based on the closeness the estimated value using method 1 to the real value);

Step 3: estimate q_3, and relevant weight factor for its reliability w_3 and scaling factor s_3; (here the scalling factor is applied to make the flows on approach level and lane level comparable);

Step 4: estimate q, based on averaging of each components; (equation 11)

Step 5: calculate the total difference Δq *(1),* based on averaging of each components; (equation 12)

Step 6: calculate the convergence: (equation 13)

$$q = w_1 * q_1 + w_2 * q_2 + w_3 * s_3 * q_3 \tag{11}$$

$$\Delta q = w_1 * (q - q_1) + w_2 * (q - q_2) + w_3 * s_3 * (q - q_3) \tag{12}$$

$$c = (\Delta q(1) - \Delta q(0))/(\Delta q(1) + 0.001) \tag{13}$$

If a convergence is reached, stop, otherwise set $\Delta q(0)$ equals to $\Delta q(1)$and repeat from Step 1. Weighting factors and scaling factor would need still to obtain from data processing and experiment.

3. Experiment setup

In this paper, only one junction among all the 102 1st term and 104 2nd term junctions is picked up for the validation of the proposed methodology. The junction is located at Road Wanjiali - Road Laodong, Changsha, China, which is marked as ID 31616 in Changsha 2nd term SCATS system. It has four approaching directions, for each approach there are six lanes, from left to right, namely two left turning lanes, three straight lanes and one right turning lane respectively. It has a full observation on each detector of each lane, while FCD also cover this area, so sufficient ground truth values can be obtained as a reference for each method. The layout of the junction is shown in Fig. 2, it should be noted that, light blue numbers shows adjacent junction ID.

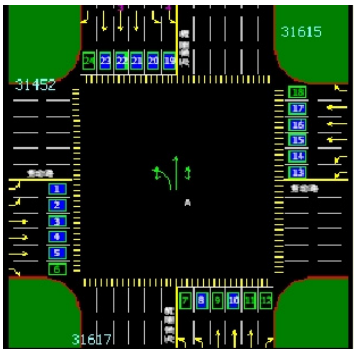

Fig. 2. Layout of a case junction in SCATS system

Firstly, method 1 and method 2 are implemented, by assuming the flow is missing from the detector on an individual lane; and then their indicators are calculated by comparing the total flow on the selected approaches. The computation updating interval is 5 minutes. Secondly, method 3 is implemented by assuming the flow is missing on approach level, and indicators are calculated by comparing the total flow on approach level, too. Computation updating interval is 30 minutes due to relative lower records of FCD. Thirdly, to compare with method 3, method 1 and method 2 are implemented by assuming that flows from detectors on all lanes are missing, thus every lane are

estimated separately and then added up to be compared with ground-truth value of total flow on the approach. The resolution of 30 min is used for all methods to ensure fair comparison. It should be noted that, as presented in Fig. 2 junction layout in experiment setup, there are 6 lanes on an approach, two left turning lanes, three straight lanes, and one right turning lane. Method 1 can be implemented to all of them while method 2 cannot be fully applied to right turning lane since it does not have any group members in turning direction.

As for the range of experiment time, data of fourteen days are selected, they are from 2013.4.15 to 2013.4.28. Among these days, file for raw flow data on one day (2013.4.22 Monday) is totally broken, to make the range of the experiment two complete weeks, data from the next Monday 2013.4.29 is used. For method 3, due to the limited availability data sources, only one day (2014.4.23) is chosen to do the experiment.

The evaluation indicators are chosen as MAE (Mean absolute error), MAPE (Mean absolute percentage error) and RMSE (Root-mean-square deviation) respectively. They are expressed as:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y - y_t| \quad \text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}|\hat{y}_t - y_t/y_t| \quad RMSE = \sqrt{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2 / n} \tag{14}$$

In the equation, estimated values are $\hat{y}_t$ while actual value is y_t.

4. Results and discussion

The results presented in this section contain three methods and the compassion results. In first two methods, only results of lane 1 over are demonstrated in figures, and others are averaged to a table.

4.1. Results of estimation using method 1

Firstly an example result on a lane is presented, the three figures show the performance of two ways from method 1, they are MAE, MAPE and RMSE over all fourteen days respectively. Notice that here MAE & RMSE refer to average traffic count error for 5 minute period. Secondly, to show more generalized result, average results indicators for all lanes on different approaches are demonstrated in Table 1

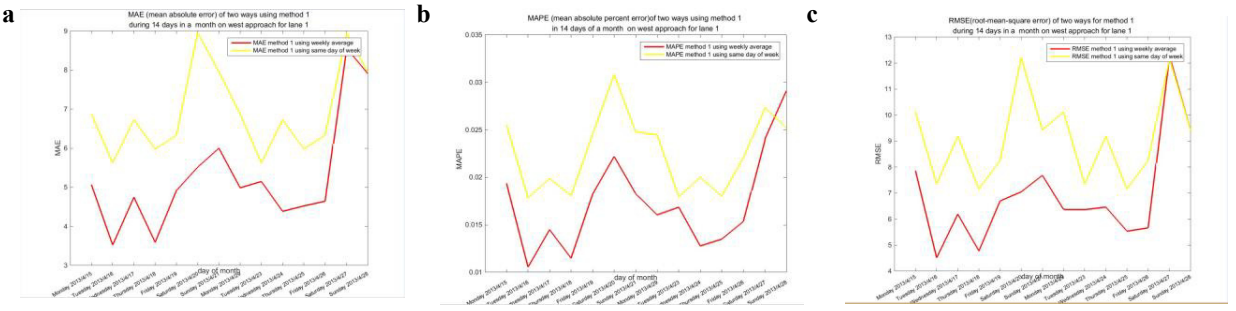

Fig. 3. Indicators of two ways of using method 1 on lane 1 over 14 days in a month (a) MAE; (b) MAPE (c) RMSE. Notice: Red line shows method 1 using weekly average, Yellow line shows method 1 using same day of week

From the MAE (Mean absolute error) results, qualitatively, no matter on which lane the flow is missing, both ways of implementing method 1 have given relative regular results over days of month. Method 1.1: using weekly average, Method 1.2: using same day of week.

It can be seen that the trends of MAPE are similar for different approaches. RMSE (Root-mean-square deviation) result have given another evidence of estimation results from absolute deviation perspective. On the west and east approach, average RMSE on a lane is around 4 on south and north approach it is around 8 which is similar as the phenomena in MAE. It can be supposed that, on the north and south approach, holding a larger amount of traffic volume, it may provide more fluctuating flow over day of week or day of month, which gives a difficulty in estimation.

Table 1 indicators for flow estimation on a single lane using method 1, average value on each approach

	MAE		MAPE		RMSE	
The approach of lanes	method 1.1	method 1.2	method 1.1	method 1.2	method 1.1	method 1.2
The west approach	2.55	3.24	0.35	0.45	3.41	4.43
The south approach	4.95	6.19	0.29	0.36	6.70	8.38
The east approach	1.78	2.23	0.38	0.47	2.41	3.07
The north approach	4.95	6.34	0.42	0.52	6.93	8.99
overall	3.56	4.50	0.36	0.45	4.86	6.22

,.

It can be seen that the trends of MAPE are similar for different approaches. RMSE (Root-mean-square deviation) result have given another evidence of estimation results from absolute deviation perspective. On the west and east approach, average RMSE on a lane is around 4 on south and north approach it is around 8 which is similar as the phenomena in MAE. It can be supposed that, on the north and south approach, holding a larger amount of traffic volume, it may provide more fluctuating flow over day of week or day of month, which gives a difficulty in estimation.

Apart from this, it can be found that, the first way in Method 1 using weekly average generally shows a lower mean absolute error than the second way using the same DOW(day of week). This probably resulted from its larger amount of data input of the estimation, since for the second ways in method 1, only one data from previous week same day of week can be used due to the limitation of dataset. The result is expected to be better for the second way if they have a same amount of input as the first way. For example, if there is available dataset during a whole month, data of three Tuesday can be normalized to make an estimation.

4.2. Results of estimation using method 2

The same as method 1, an example result on a lane over all fourteen days is presented followed by a table showing the average value of indicators of estimated flow on lanes on all approaches. Notice that here MAE & RMSE refer to the average traffic count error for 5-minute period.

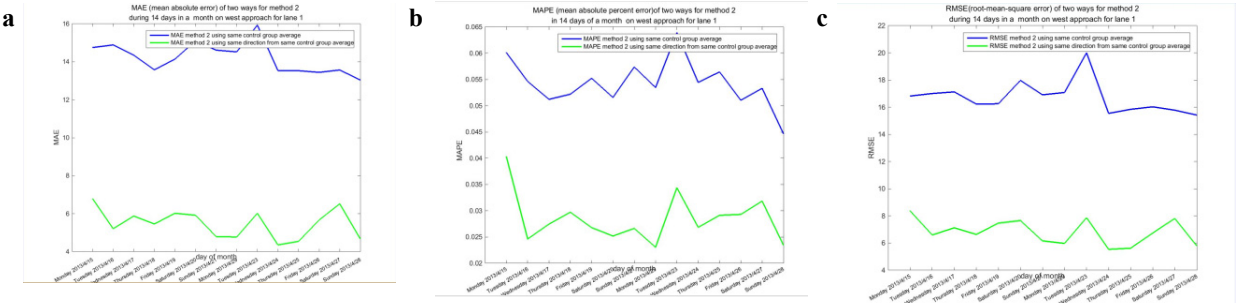

Fig. 4. Indicators of two ways of using method 2 on lane 1 over 14 days in a month (a) MAE; (b) MAPE (c) RMSE. Notice: Blue line shows method 2 using same control group average, Green line shows method 2 using same direction from control group average.

It can be seen from results that, MAE (Mean absolute error) shows regular performance over day of month. Besides, although it is not shown in the figure, two different ways of implementing method 2 have shown converse performance on different turning direction lane groups: using the values from lane in the whole control group seems perform better on left turning direction (lane 1) and only using value from lanes in same direction in a control group perform better on straight (lane 3).These phenomena can tell that, flow on left turning lane have less similarity than that on a straight lane; the flow on a left turning lane may be very different from the flow on another left turning lane. Method 2.1: using same control group average, Method 2.2: using same direction from control group average.

Table 2 indicators for flow estimation on a single lane using method 2, average value on each approach

	MAE		MAPE		RMSE	
	method 2.1	method 2.2	method 2.1	method 2.2	method 2.1	method 2.1
The west approach	2.12	1.90	0.31	0.31	2.78	2.49
The south approach	4.38	3.28	0.27	0.22	5.47	4.33
The east approach	1.68	1.57	0.36	0.33	2.19	2.09
The north approach	7.22	6.72	0.75	0.47	9.02	8.67
overall	3.85	3.37	0.42	0.33	4.87	4.39

The MAPE (Mean absolute percentage error) have shown good percentage under 10% for each approach. This indicates that the method 2 works well on this junction. From the figures of RMSE (Root-mean-square deviation), it can be read that, the method 2 works well and there are not too many differences for two ways.

To make a conclusion, the general output of method 2 is reliable at this junction, with less fluctuation over day of week, which is another way round for method 1. It could be concluded that, this method may be more suitable than method 1 under situations that there are big fluctuation of flow over day of week .Results of estimation using method 3.

Firstly, the relation of SCATS flow and FCD speed from the west approach and south approach are put together to get curve fittings. These are implemented by plotting the results of aggregated flow on 30 min interval and corresponding FCD speed on one Map, the average speed is calculated by counting FCD records of instant speed for every 30 minutes. A Polynomial relation with degree of 2 is assumed here to give the curve fitting . The results of curve fittings are then given by following figures:

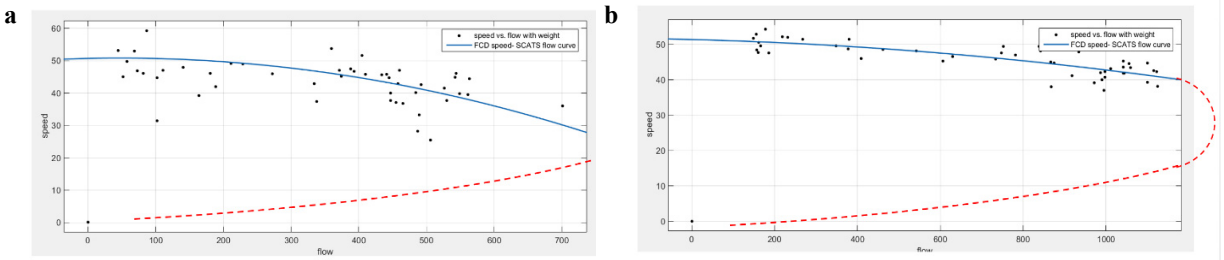

Fig. 5.two fitting curves of speed-flow relation on (a) the west approach and (b) the south approach

On the Fig. 5, blue lines are the fitting curves of fused SCATS flow and FCD speed, which represent the free flow branch of fundamental diagrams, while red dashed lines are the congestion branches of fundamental diagrams that are manually added due to the absence of congested data. These relations are applied to estimate the missing flow on approach level. The error indicators of estimated flows are as follows. Notice that now MAE & RMSE refer to the average traffic count error for 30-minute period.

Table 3 performance in method 3

Relation come from	MAE	MAPE	RMSE
West approach	160.58	1.26	212.51
South approach	168.33	0.52	201.70

Then the performance of the method 3 is shown in Table 3.With relative poorer indicators, method 3 does not look as well as the previous two methods. However, the fitting curves inherit the data fusion concept to link two data sources and provide reasonable relation – in free flow situation, when flow is higher, speed becomes lower (It can be

assumed that flow on an outbound of an approach is near to free flow, although some vehicles may still in process of accelerating). These relations make sense in carrying out estimating flow using other traffic states. However, the speed used in the experiment is the average speed of FCD, which may be a reason of large deviation. Since the relations are the main estimation tool in this method, more precise relations can lead to more reliable results, and good relations calls for more precise expressions of the input speed. And these are also in further researches of the authors. A conclusion can be drawn that, providing the situation that little flow information can get from the system itself, extra data sources are useful in estimating missing flow using this method.

4.3. Comparison of three methods

Three methods are cross-compared in this part. All the lanes are assumed missing when applying method 1 and method 2, which is comparable to the situation that flow missing on approach level in method 3. It should be noted that, since all the flow on an approach will be assumed to be missing, the second way of method 2 using average value of all lanes in a control group cannot give results, so it will not be counted in results comparison. For validation, approach level instead of single lane level is used here. Note that now MAE & RMSE refer to average traffic count error for 30-minute period.

Table 4 Performance of three methods on the west approach and the south approach for one day 2013.4.23

			The west approach			The south approach
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Method 1.1	30.48	0.15	35.85	111.39	0.20	125.83
Method 1.2	36.22	0.17	44.54	117.79	0.21	133.73
Method 2.2	68.60	0.23	91.83	93.93	0.13	114.91
Method 3	160.58	1.26	212.51	168.33	0.52	201.70

It can be seen that method 1 and method 2 show good performance in estimating flow on the west and east approach, while they do not perform well on the south approach. This is probably due to larger uncertainty and fluctuation of flow pattern on the south approach. When looking at method 3, though it does not show satisfactory performance on both approaches, this method is theoretically sound since it establishes a relationship between two data sources- FCD speed and SCATS flow. The results of method 3 could still be an important reference for flow estimation and the method requires further investigation and tuning. Moreover, it can be assumed that, after iteration of three methods, the performance will turn out to be better.

5. Conclusion

This paper has proposed three consistent methods (aspects) to estimate missing flow at junctions. These methods are based on (1) historical pattern; (2) timing plan and loop location; and (3) FCD and flow relationship- data fusion concept. They all fall into the generic flow estimation framework presented in the methodology part. Each method has its own scope, accuracy, and physical meaning. From the experimental studies, some methods show better results than others under varying situations. It is suggested that the framework be used according to data sources available. It can also be expected that these methods are promising to show reliable estimation results after iteration process. Further research is under way to get the results and performance of the generic estimation formula by applying weighting factors and scaling factor derived from estimation results from the three methods. By applying reliable estimation methods in this paper, traffic flow data quality can be improved to some extend for further applications.

Acknowledgement

This research is supported by TNO, The Netherlands. Data used in experiment comes from SCATS system from Changsha Traffic Police and FCD from Changsha Taxi Companies, China.

References

Treiber, M. Helbing, D., 2002. Reconstructing the spatio-temporal traffic dynamics from stationary detector data. Cooper@tive Tr@nsport@tion Dyn@mics 1, 3.1-3.24.

Van Lint, J.W.C. Hoogendoorn, S.P., 2009. A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways. Computer-Aided Civil and Infrastructure Engineering 25 (8), 596-612.

Chen, C., Kwon, J., Rice, J., Skabardonis, A., & Varaiya, P. (2003). Detecting errors and imputing missing data for single-loop surveillance systems.Transportation Research Record: Journal of the Transportation Research Board, 1855(1), 160-167.

Wall, Z. R., & Dailey, D. J. (2003). Algorithm for detecting and correcting errors in archived traffic data. Transportation Research Record: Journal of the Transportation Research Board, 1855(1), 183-190.

Cathey, F. W., & Dailey, D. J. (2003). Estimating corridor travel time by using transit vehicles as probes. Transportation Research Record: Journal of the Transportation Research Board, 1855(1), 60-65.

Wang, Y., van Schuppen, J. H., & Vrancken, J. (2014). Prediction of traffic flow at the boundary of a motorway network. Intelligent Transportation Systems, IEEE Transactions on, 15(1), 214-227.

Yuan, Y., Wilson, R. E., Van Lint, H., & Hoogendoorn, S. (2012). Estimation of Multiclass and Multilane Counts from Aggregate Loop Detector Data.Transportation Research Record: Journal of the Transportation Research Board, 2308(1), 120-127.

Ou, Q., van Lint, J. W. C., & Hoogendoorn, S. P. (2010). Data-Data Consistency: A New Approach to Traffic Data Fusion.

Jie, L., Van Zuylen, H., Chunhua, L., & Shoufeng, L. (2011). Monitoring travel times in an urban network using video, GPS and Bluetooth. Procedia-Social and Behavioral Sciences, 20, 630-637.

Deng, W., Lei, H., & Zhou, X. (2013). Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach. Transportation Research Part B: Methodological, 57, 132-157.