Automatic Speed-Bias Correction with Flow-Density Relationships

Yufei Yuan, J. W. C. van Lint, Th. Schreiter, S.P. Hoogendoorn, J. L. M. Vrancken

Abstract—Accurate and reliable speeds measured with local traffic sensors are critically important for many applications, such as travel time estimation, state estimation and prediction and dynamic traffic control. Many local sensors (e.g. the dual induction loops used in The Netherlands, UK and in Germany) calculate average speed by arithmetic time averaging, which leads to strongly biased estimates. This paper proposes a speed-bias correction algorithm based on notions from first order traffic flow theory and empirical flow-density relationships, using only a limited number of parameters. On the basis of a microscopic simulation study, it is demonstrated that this algorithm is able to correct the speed bias due to time averaging considerably. A more extensive version of this paper can be found at http://dux.ict.tbm.tudelft.nl/Research/SpeedCorr.pdf

I. INTRODUCTION

Traffic data collected by traffic sensors are critically important for many applications, both for online purposes (traffic information, management and control) and for offline use (travel time estimation, model calibration and validation, incident/bottleneck detection and ex-post policy evaluation). Traffic speed is one traffic variable which can be measured by many different sensors, both locally (with e.g. induction loops and radar technology) as well as over short or longer distances (e.g. automated vehicle identification systems or floating car data). In particular, knowledge of the speed profile and/or the flow profile is useful to determine bottleneck locations, physical extent of queues, travel times, turn fraction, fundamental relations and many other parameters used for traffic control systems.

In many countries, the traffic network monitoring is largely based on the stationary loop detector. Single loops (mainly implemented in the US and in the urban environment) provide flow and occupancy information, which in turn can be used to estimate speed (using average vehicle length). However, these estimates are quite noisy due to unobserved variables (e.g. vehicle length) and measurement errors. Many researchers have sought better estimates of speed from single loop detectors (1-4). Additional to flow and occupancy, dual loops also collect (average) speed. The Dutch freeway network in the Netherlands, for example, is monitored by dual loops which are located at about every 500m on the western part of the Dutch freeway network. Of the available data from this dual loop system (named MoniCa) on average 5-10% is missing or otherwise deemed unreliable (5). Similar traffic monitoring systems, which are equipped with inductive loops, can be found in England (6) and in Germany (7).

Locally measured and averaged speeds are not necessarily representative for spatial properties of traffic streams, due to two reasons. First of all, a local average speed at some cross section $x_i$ over a time period $p$ is equal to the space mean speed $u_{M}$ on a road section $m$ (including $x_i$) over some time period $p$ only, in case traffic conditions are homogeneous and stationary over $m$ during $p$ (which does not mean all vehicles drive with the same speed). Secondly, the former holds only in case speeds of passing vehicles are averaged harmonically, and not arithmetically with $\bar{u} = \Sigma v_i / N$, where $N$ depicts the number of observations in time period $p$. The latter average (often referred to as time mean speed) is biased due to an overrepresentation of faster vehicles in a time sample. In contrast, the harmonic mean speed

$$u_H = \frac{N}{\sum_{i=1}^{N} \frac{1}{v_i}}$$

essentially represents the reciprocal value of average “slowness” $1/v_i$, i.e. the average time per unit space each vehicle spends while passing the detector. Save for the error due to the assumption of homogeneous and stationary traffic, this average is an unbiased estimator of the space mean speed (see many textbooks, e.g. (8)). The relationship between space mean speed ($u_{M}$) and local arithmetic time mean speed ($u_{L}$) can be analytically expressed by the following equation (9):

$$u_L = u_M + \frac{\sigma^2}{u_M}$$

where $\sigma^2$ denotes instantaneous speed variance.

The implication of eqn [2] is firstly that time mean speed is always equal or larger than space mean speed. This is a structural difference proportional to instantaneous speed variance. Secondly, the effect of arithmetic time averaging cannot be “undone” afterwards, that is, there is no straightforward (analytical) formula to derive space mean speeds from time mean speeds directly, since the bias depends on a second unknown quantity (instantaneous speed variance). Many empirical studies have revealed that the bias ($\sigma^2/u_M$) is significant and may result in speed overestimation of up to 25% in congested conditions, and even larger errors in for example densities or travel times derived from these speeds (10-13). A typical result from using arithmetic time averaged speeds in estimating densities (via $k = q/u_k$) is that the estimated traffic states $(q,k)$ (flow and density) do not lie on a nice semi-linear line in the
congested branch of the flow-density diagram but scatter widely below this line, resulting in a “P-shape” distortion of the estimated fundamental \((q-k)\) diagram (10), where the congested phase line looks like a convex curve as the upper part of the letter “P”. This effect has also been analytically proven in (11). Additionally, the biased speeds may cause underestimates of the route travel time (12, 13).

Clearly, in case monitoring systems (such as the Dutch MoniCa system, the British and the German highway monitoring systems) collect time averaged speeds, there is a need for algorithms and tools, which are able to correct for the inherent bias. In this paper we will present such an algorithm, based on the flow-density diagram and notions from traffic flow theory. The paper is organized as follows. In the next section we will briefly overview methods to eliminate this bias and derive the new method proposed in this paper. Next we will present the setup of a simulated experiment on the basis of which the algorithm and two alternative approaches will be evaluated. The paper will close with conclusions and recommendations for further research.

II. SPEED-BIAS CORRECTION ALGORITHMS

Based on eqn [2], there are two ways of correcting for the bias. One could consider the bias term \(B = (\sigma_{M}^{2}/u_{L})\) as a whole, that is

\[
u_{M} = u_{L} - B
\]

where \(B\) might be a constant or some function of quantities which are available. One can also solve \(u_{M}\) from eqn [2] which gives us

\[
u_{M} = \frac{1}{2}(u_{L} + \sqrt{u_{L}^{2} - 4\sigma_{M}^{2}}), \quad \sigma_{M} \leq \frac{1}{2}u_{L}
\]

Once the instantaneous speed variance \(\sigma_{M}^{2}\) is known, the space mean speed can be deduced. In the next two sections we will recall and propose methods in each of these two categories.

A. Estimating speed variance

The simplest method to estimate speed variance is to consider the speed variance (standard deviation) equal to certain percentage of the local mean speed, that is \(\sigma_{M} = \beta u_{L}\), with \(\beta\) in the order of 0.05-0.2 (5-30%, according to (12)). In (12) Lindveld and Thijs also illustrate that \(\beta\) is not a constant, but rather a function of traffic flow and speed. This relationship is derived using empirical data, in which the ratio standard deviation over mean speed increases with increasing with time mean speeds. This can be easily understood. In congestion vehicles are constraint in their choice of speed by one another. The predominant cause of speed variation over space will be (collective) acceleration and deceleration e.g. due to passing stop and go waves. Under free-flow conditions, speed variance will logically increase due to (a) traffic heterogeneity (trucks versus person cars) and (b) the fact that samples (number of vehicles per space and time unit) will become increasingly smaller with decreasing density (and higher speeds). The first method proposed by Van Lint (14) is simply a bi-linear function fitted through these empirical data, illustrated by the solid line in figure 1(a). In (12) the following bi-linear relationship is derived

\[
\hat{\sigma}_{M}^{2} = \begin{cases} 
0.5\cdot u_{L} - 34 & \text{if } u_{L} \geq 74 \\
0.02\cdot u_{L} + 5 & \text{otherwise}
\end{cases}
\]

Figure 1 (b) shows a second scatter plot, now of instantaneous speed variance and “local density” \(k_{L}=q/u_{L}\) and “time series” approximation, both adapted from (14).

In completely free-flowing conditions, i.e. when \(q/k_{L}\) that is, flow divided by time mean speed \((q/u_{L})\). The second approach Van Lint (referred to as SVE-TS further below) introduces in (12) uses an approximation of this relationship, illustrated by the dashed line in figure 1(b). Without going into detail, this method assumes that in congested conditions the instantaneous speed variance \(\sigma_{M}^{2}\) in some period \(p\) is of the same order as the variance \(\Theta_{i_{p}}^{2}\) of a time series of \(P\) consecutive time mean speeds \(u_{i}\) around \(p\). Due to autocorrelation in this time series, one can apply first order differencing to remove the time dependence, which leads to the following lower bound for within period variance

\[
\sigma_{M,\text{low}}^{2} \geq \Theta_{i_{p}}^{2} = \text{var}[u_{i_{p}} - u_{i_{p-1}}], i \in \{p-\frac{P}{2}+1, p-\frac{P}{2}+2, \ldots, p+\frac{P}{2}\}
\]

In completely free-flowing conditions, i.e. when consecutive time mean speeds will be only very weakly correlated to one another, the central limit theorem provides the following upper boundary for the within period speed variance

\[
\sigma_{M,\text{f low}}^{2} \leq \frac{P}{2}\Theta_{i_{p}}^{2}.
\]

The relationship illustrated in figure 1(b) is then used to predict speed variance between these two extreme cases as follows.
\[ \bar{\sigma}_{u,p} = a_p \bar{\Theta}_{u,p} + (1-a_p) \frac{P}{2} \bar{\Theta}_{u,p}, \quad \text{with} \quad a_p = \min\left( \frac{k_{\text{crit}}}{k_L}, 1 \right). \]  

In eqn [8], \( k_{\text{crit}} \) denotes the critical local density, typically around 0.02 veh/m/lane, which essentially distinguishes between free and congested conditions. Finally, in (12) this estimate is smoothed with an exponential filter

\[ \bar{\sigma}_{u,p} = (1-\gamma) \cdot \bar{\sigma}_{u,p-1} + \gamma \cdot \bar{\sigma}_{u,p}. \]  

Note that this procedure contains three parameters (\( P, k_{\text{crit}} \) and \( \gamma \)), which need to be tuned with respect to the site-specific characteristics.

The methods outlined above essentially correct speed bias based on local relationships found in the available detector data. Alternatively, one could estimate speed-bias also on spatiotemporal relationships found in the data, using for example the fundamental diagram, which relates average flow (a local quantity) to average density (a spatial quantity). In (3), for example speed estimates from single loops are validated using flow-occupancy relationship. Also Coifman (15) exploits basic traffic flow theory and spatiotemporal propagation characteristics of traffic patterns to estimate link travel time using local dual loop data. Although these methods do not apply to the problem of estimating the bias in time mean speed directly, they provide inspiration for an alternative method to solve the problem.

**B. A new correction algorithm based on flow-density**

According to the kinematic wave theory (16), perturbations (low speeds, high densities) propagate through a traffic stream at a speed equal to \( \frac{dj}{dk} \), with \( q = f(k) \) depicting an equilibrium relationship (the fundamental diagram) between average flow and density. This characteristic speed is typically positive in free flow and negative in congested conditions. A simple, but according to some authors (17-21) still reasonable approximation is to assume only two main characteristic speeds, one for congested traffic and one for free-flowing traffic, which results in a triangular flow-density relationship (17, 18, which reads

\[ q = f(k) = \begin{cases} 
\text{c
_\text{free}} \cdot k & k \leq k_{\text{crit}} \\
q_{\text{cap}} + \text{c
_\text{cong}} \cdot (k-k_{\text{crit}}) & \text{otherwise}
\end{cases} \]  

where \( q_{\text{cap}} \) is the capacity and \( \text{c
_\text{free}} \) and \( \text{c
_\text{cong}} \) depict the characteristic propagation speed in free-flow and congested conditions respectively. Note that \( \text{c
_\text{cong}} \) is often parameterized with \( \text{c
_\text{cong}} = \frac{q_{\text{cap}}}{(k_{\text{jam}}-k_{\text{crit}})}, \) where \( k_{\text{jam}} \) and \( k_{\text{crit}} \) depict the jam density and critical density respectively.

\[ \text{c
_\text{cong}} \]

**Figure 2.** Speed contour plot measured at the Dutch A13 freeway southbound between Delft North (km6.5) and Rotterdam Airport (km14.5) on Monday 11/6/2009 from 3:00 PM to 7:00 PM. Colors correspond to one-minute arithmetically averaged speeds.

Figure 2 presents a typical speed contour taken from a Dutch freeway, in which those approximate constant characteristic propagation speeds can be identified. In congestion (low speeds), for example, perturbations in a traffic stream move upstream with remarkably constant speeds, illustrated by the thick dark stop and go waves (low speed areas) in figure 2, which propagate upstream over the entire freeway stretch of 8 km. Note that inside the speed waves themselves, the individual vehicle speed is not uniform.

Using the same detector data, one would expect these phenomena to translate into traffic states on a straight line in the \( q-k \) diagram. Nonetheless, as already mentioned above and in (10, 11), the straight or semi-straight line for congested traffic is not observed when we use “local density” \( k_{\text{crit}} = \frac{q}{u_L} \) as a proxy for true (but unobserved) density \( k \). Instead, in that case, we see a “P-shape” distortion, attributed to the non-zero bias term (\( \sigma^2_{\text{m}}/u_\text{m} \) in eqn [2]). We can, however, with the assumption of a (approximately) straight congested branch of the fundamental diagram, straightforwardly estimate and correct this error in density, and as a result correct the bias in speed.

To do so, a few more conditions need to be met. First of all, if we assume that in congestion the true traffic state \( \{q, k\} \) lies on a straight line with slope equal to the propagation speed \( c_{\text{cong}} \), we need to be able to estimate this parameter \( c_{\text{cong}} \) – e.g. on the basis of spatiotemporal plots (figure 2) or otherwise. Secondly, we need to assume that the measured flows are unbiased, in which case correcting \( k_{\text{crit}} \) (= \( q/u_L \)) boils down to estimating the error in \( u_L \), which of course equals to \( \sigma_{\text{m}}^2/u_\text{m} \). Figure 3 illustrates the basic correction principle.

\[ \sigma_{\text{m}}^2 \]

The essence of this method is considering the bias term as a whole (B, refer to eqn [3]) to correct local mean speed, based on notions from traffic flow theory and traffic propagation characteristics. The detailed working principle is described in below, followed by a schematic procedure shown in figure 4.
Data input (The first step presented in figure 4.). The data source could be the raw observed data directly from the monitoring system or the pre-processed traffic data in the pattern of spatiotemporal (x-t) matrix. The raw data can be pre-processed by means of the method proposed in (5, 22) to filter out high frequency noise and measurement errors. Meanwhile, the propagation characteristic speed \( c_{\text{cong}} \) needs to be determined as the slope of the congested flow line used in this algorithm. This parameter can be estimated based on the historical data or set as default.

Data selection (The second stage in figure 4). Based on speed \( (u_t) \) and flow \( (q_{\text{input}}) \) data, the density data are generated (via \( k_{\text{input}} = q_{\text{input}} / u_t \)). As our focus is on the congested traffic, the free-flow state and the congested state need to be distinguished. In macroscopic traffic flow models, critical density on a section discriminates between free-flow and congested flow conditions. Here, the critical density value \( (k_{\text{cri}}) \) and/or the capacity value \( (q_{\text{cap}}) \), which relate to the peak of a \( q-k \) diagram, serve the same purpose. \( k_{\text{cri}} \) and \( q_{\text{cap}} \) values are easily estimated from the input data, The datasets in terms of flow and density, which meet the condition \( k_{\text{input}} > k_{\text{cri}} \) are categorized as congested state data.

Additionally, we need to ensure that the data selected (from congested conditions) represent stable traffic states, that is, traffic states which are likely to lie on the (linear) congested branch of the fundamental diagram. This implies we cannot select detector data located downstream (“in”) a bottleneck. On those locations where congestion is resolving one typically finds flows far below capacity and large spacing (low density) due to accelerating vehicles.

Density-wise correction (The third step in figure 4). This step is the core of the correction algorithm. The mapping of this operation is expressed by:

\[
k_{\text{input}} = K_{\text{fit}}(q_{\text{input}}) + \varepsilon \rightarrow k_{\text{corrected}} = K_{\text{c_cong}}(q_{\text{input}}) + \varepsilon
\]  

Where, \( \varepsilon \) denotes the residual deviation of each density \( (k_{\text{input}}) \) scatter point away from the fitted curve \( (K_{\text{fit}}) \) of the congested data, \( k_{\text{corrected}} \) denotes the corrected density values, \( K_{\text{c_cong}} \) denotes the targeted congested branch.

Given the peak point \( (\text{fit-origin}: k_{\text{cri}} \text{ and } q_{\text{cap}}) \) and congested traffic scatters \( (k_{\text{input}} \text{ and } q_{\text{input}}) \), fit function \( (K_{\text{fit}}) \) applies. In \( q-k \) space, fit function would fit the congested data, regarding the peak point \( (q_{\text{cap}} \text{ and } k_{\text{cri}}) \) as the origin point. Density values \( (k_{\text{input}}) \) are then expressed as a function \( K_{\text{fit}} \) of the input flow values \( (q_{\text{input}}) \) with residual \( (\varepsilon) \). We then shift these scatters in congested phase part of the \( q-k \) diagram from the fitted curve \( (K_{\text{fit}}) \) to the targeted congested characteristic line \( (K_{\text{c_cong}}) \).

The fitted curve can be either linear or polynomial. This leads to two variations of the correction algorithm. The linear fit function is defined as

\[
K_{\text{fit}}(q_{\text{input}}) = a \cdot q_{\text{input}} + b
\]

with two fit coefficients \( (a \text{ and } b) \).

Intuitively, the polynomial fitting could describe the feature of the “P-shape” distortion presented in (10, 11) better than the linear one. Here, a quadratic polynomial form of fit function is considered, written as

\[
K_{\text{fit}}(q_{\text{input}}) = c \cdot q_{\text{input}}^2 + d \cdot q_{\text{input}} + e, \quad c > 0
\]

with three fit coefficients \( (c, d \text{ and } e) \).

The two variations are referred to as Linear-fit and Polynomial-fit respectively further below.

For the targeted congested characteristic line, if only one degree of freedom for the congested state is considered, the \( K_{\text{c_cong}} \) is linear featuring by congested propagation speed \( (c_{\text{cong}}) \), expressed by

\[
K_{\text{c_cong}}(q_{\text{input}}) = \frac{q_{\text{input}} - q_{\text{cap}}}{c_{\text{cong}}} + k_{\text{cri}}
\]

The more degrees of freedom of parameters for the targeted characteristic line are taken into account, the more accurate correction will be.

Based on corrected density and the invariant flow values, the corrected speeds are generated directly (via \( u_{\text{corrected}} = \frac{q_{\text{corrected}}}{k_{\text{corrected}}} \)). In face, the correction presented here is another form of eqn \([3]\):

\[
u_{\text{corrected}} = u_t \cdot \frac{q_{\text{corrected}} - q_{\text{input}}}{k_{\text{corrected}} - k_{\text{input}}}
\]

In which, the bias term \( (B) \) is expressed as a function of the corrected density \( (k_{\text{corrected}}) \).

Post-processing (The last step in figure 4). In some cases, samples with high-speed (e.g. larger than 120 km/h) are identified as erroneous. Moreover, in the speed space, the generated speed values \( (u_{\text{corrected}}) \) of congested state contain discontinuity with respect to those in free-flow state, so a linear smoothing is used here. The smoothing is a neighborhood operation which has already proposed in (5, 22). The operation for each point is conducted with a small neighborhood which has already proposed in (5, 22). The operation for each point is conducted with a small neighborhood operation which has already proposed in (5, 22).
III. SIMULATION STUDY

To test the properties of the proposed algorithm, we will perform a simulation study, since this provides us with both (synthetic) detector data as well as ground-truth data (space mean speed). The microscopic simulation model ‘FOSIM’ (23) is used for this purpose. This model is developed at the Delft University of Technology, specially designed for the detailed analysis of discontinuities in the freeway network. It has been calibrated and validated for the Dutch freeway in terms of driving behaviors. From the FOSIM model, traffic data from any type of detectors can be simulated. Afterwards, the data processing and analyses are further conducted in MATLAB 7.5.0.

A. Model description

As mentioned above, the current monitoring system on the Dutch motorways consists of dual inductive loops located about every 500 meters, collecting one minute average time mean speeds and 1 minute aggregate flows. In order to assess the proposed algorithm, a part of the Dutch freeway A13 (figure 5(a)) has been modeled in FOSIM, as shown in figure 5(b). It is an 11.5 km long road stretch from Delft North to Rotterdam Airport. The simulation period is afternoon peak hours from 14:00 to 20:00 (six hours). The traffic volume is modeled to match a typical weekday afternoon, with 10% truck fraction. From the FOSIM model, MoniCa data (time mean speeds, flows) can be obtained, in the spatiotemporal pattern of 500(m)*60(s). In our case, the ground truth data (space mean speeds, that is, harmonic mean speed $u_H$ in eqn [1]) are derived from FOSIM over equidistant spatiotemporal regions of size 100(m)*30(s). The data are regarded as reference. The raw input for the correction algorithm has been pre-processed in order to match the same spatiotemporal grid (100(m)*30(s)) as the ground truth data.

B. Scenario definition

Based on speed contours derived from simulation, a constant value of -21km/h is used as congested propagation speed throughout the rest of the paper. That means in the proposed speed-bias correction method, the processed data are corrected to approach a straight (linear) congested phase line featuring this speed value. The two variations of the new proposed method are evaluated, i.e. Linear-fit and Polynomial-fit. For the purpose of comparison, we will compare our method to one of the two speed-variance estimation (SVE) methods, the “time series” method, proposed in (14) and discussed above. We will use the default parameter settings proposed in (14), namely $P = 30$, $k_L \text{cri} = 0.02 \text{veh/m/lane}$ and $\gamma = 0.24$.

In our study, the robustness of the proposed algorithm has also been considered. Two scenarios are considered. First, the input data are added with some white Gaussian noise to emulate the noise from traffic sensors. Meanwhile, the parameter ($c_{cong}$) used in correction is changed within a small range (from -18km/h to -24km/h).

C. Performance measures

In order to assess different scenarios, the performance measures need to be determined. The raw/processed traffic data are compared with the ground truth (reference) data in terms of root mean of squared error (RMSE) and three relative error indicators, namely mean percentage error (MPE) which reflects structural bias, mean absolute percentage error (MAPE) which gives a combined indication of the relative error and standard deviation of the percentage error (SPE) which is an index for the variability around the MPE. Since the bias/error mostly occurs at low speeds
(congestion), these relative error indicators are more informative than RMSE error which is calculated by the absolute speed values which are relatively low in congestion. Nonetheless, RMSE error can still provide an overview of performance on the whole dataset. The results are discussed in the following section.

IV. RESULTS AND DISCUSSION

This algorithm is tested by FOSIM synthetic data (figure 6). The results are categorized for three scenarios, the normal correction, the noisy-input case and the deviation of parameter test. The most important indicator is considered as MPE error, which provides an indicator for structural bias. So our analyses mainly focus on the MPE error, aiming at lowering structural errors (speed bias).

In the case of normal correction, the results are presented for four methods, namely speed variance estimation based on time-series of local mean speed (SVE-TS), Linear-fit correction and Polynomial-fit correction. Obviously, the polynomial-fitting outperforms the other two methods, with an absolute improvement of 2% on MPE and 1% on MAPE, compared to the input speed data. The relative improvement by polynomial-fit correction in terms of MPE is 47% compared to the uncorrected data, which is quite substantial. In MPE, the negative errors and positive errors can cancel each other out. So the higher improvement on MPE than on MAPE demonstrates that, the corrected data by polynomial fitting is able to lie closer around the ground truth data (or the congested phase line in q-k space). The linear fitting is slightly inferior to the polynomial one. Moreover, the SVE-TS method shows a negligible improvement on MPE, however, this method increase the RMSE error by 8% as a negative effect compared to the other methods, which is not preferable. The similar observations are also found in the noisy-input scenario.

In order to further analyze the correction effect on the congested data, the speed samples from normal correction case are subdivided into five speed ranges in terms of MPE and MAPE errors. First, it is numerically demonstrated that the biases/errors are mainly derived from the traffic in congestion as the large error values just related to low speed regions. And the large values for SPE can be accounted for by the noticeable change in the MPE (and/or MAPE) with respect to the different speed ranges. In terms of speed-bias correction, the decrease in the error indicators (MPE/MAPE) is predominantly contributed by the operation on the congested speed data (speeds lower than 60km/h). For the SVE-TS method, the correction effects distribute over different speed ranges, where the speed variances are able to estimate.

For the new proposed method, the improvement is proportional to the direction of decreasing on speed (or increasing on density as shown in figure 3). Moreover, compared to the linear fit method, the major achievement on decreasing error values by the polynomial correction is reflected at the lower speed ranges (significantly at ranges 0-20km/h and 20-40km/h). The explanation is that traffic data at low speed region, which relates to the lower part of congested data scatters in q-k diagram (figure 3), can be fit much closer by polynomial curve, thus the correction is better.

In the speed range of 0-20km/h, the original MPE/MAPE error of the raw data is quite high (more than 90%), nonetheless, the relative improvement on MPE/MAPE by polynomial-fit is still more than 50%. In the range of 20-40km/h, the relative improvement by polynomial-fit is even higher than 79%, because most of congested traffic samples in congestion are related to this speed range. However, it is noticed that the improvement by the new proposed method for the speed range of 60-80km/h is marginal. This can be explained as follows. At high speed range (related to the upper part of congested traffic in figure 3), the discrimination between the fit curve of the measured data and the targeted corrected line is marginal, since in our case, only one degree of freedom of parameter is considered. If more degrees of freedom of parameters for the congested characteristic line are taken into account, to capture the feature of the dynamics of congestion, except for a linear relation, the correction would be improved.

So far we can safely conclude that the proposed correction algorithm is effective on the input traffic data source in terms of reducing bias impact and pragmatism. And the polynomial fitting is better than linear fitting and time-series approach.

In terms of the robustness study, two test cases are considered here. First, if the input data from traffic sensors become noisy, the error values of raw data increase compared to the ground truth. Nonetheless, the correction is still effective on this noisy data, resulting in relatively low MPE and MAPE values. Of 46% the improvement on MPE by the polynomial fitting is comparable to the normal correction case.

In the new proposed algorithm, only one degree of freedom of parameter, the congested propagation speed, is embedded. The simulation results indicate that the correction effects are still positive when this speed parameter is changed around the default value within a small range. The
results just vary in a reasonably accepted range in terms of MPE/MAPE. In reality, the propagation speed just varies slightly depending on the accepted safe time clearance, the average vehicle length, traffic composition and weather conditions (17, 19, 20 and 22), so the affect of sensitivity of this parameter would be marginal. In sum, this algorithm is quite robust.

V. CONCLUSIONS AND RECOMMENDATIONS

In this paper a new effective and robust method is put forward for correcting the speed-bias caused by the common practice of arithmetic time averaging of speeds. Based on the simulation, this algorithm is able to correct this speed-bias to generate relatively low error values compared to the original input and performs better than the preceding proposed method. The relative improvement reaches almost 50%. The robustness of this algorithm with regards to noisy inputs from traffic sensors and a small deviation in the embedded parameter (congested propagation speed) has been verified during the study.

The idea of the speed correction is straightforward and easy to implement. The processing speed of the correction algorithm is rather fast (in MATLAB). These features are beneficial for pragmatic real-time operation. In sum, this algorithm may provide practitioners with a simple but effective and efficient tool to overcome the speed-bias problem in the empirical data.

The algorithm developed for correcting speed-bias derived from dual loop detectors, but it can also be extended and applied to densities measured from single loop detectors that are unreliable. As known, the density profiles are inferred from the measured occupancy information by making some assumption on the average vehicle length. This assumption is sensitive to low flow status (at both free-flow end and heavy congestion end refer to q-k diagram), since the variance in average vehicle length increases as flow decreases (2). That implies the density estimation at both two ends of q-k diagram could be unreliable. For correcting noisy density estimates, similar rationale is applied. One could correct density values to approach the congested and/or free flow phase line(s) while keeping flow values invariant in the flow-density space.

Let us finally propose some recommendations for further research. In the correction, the targeted characteristic curve for the congested traffic can be better described by adding more degrees of freedom of parameters besides the congestion propagation speed. The further study is needed on the improvement on the travel time estimation based on this new proposed method.

VI. REFERENCES


