



Contents lists available at ScienceDirect

## Transportation Research Part B

journal homepage: [www.elsevier.com/locate/trb](http://www.elsevier.com/locate/trb)

# From heterogeneous drivers to macroscopic patterns in congestion

Nicolas Chiabaut\*, Ludovic Leclercq, Christine Buisson

Université de Lyon, INRETS / ENTPE – Laboratoire Ingénierie Circulation Transports, Rue Maurice Audin, F-69518 Vaulx En Velin CEDEX, France

### ARTICLE INFO

#### Article history:

Received 8 April 2008

Received in revised form 28 July 2009

Accepted 28 July 2009

Available online xxxxx

#### Keywords:

Macroscopic patterns

Heterogeneous drivers

Congestion

Waves

### ABSTRACT

This paper first presents a method to estimate Newell's car-following model parameters in congestion at a microscopic scale. I-80 NGSim data analysis shows that at this scale driving behaviors vary from one vehicle to the another. Then, relations between stochastic Newell's model with heterogeneous drivers and its associated macroscopic pattern are established. This proves that the mean jam spacing is the arithmetic mean of individual jam spacings whereas the mean wave speed is the harmonic mean of individual wave speeds. These new insights are valuable for (i) calibrating a macroscopic model from individual observations and (ii) establishing the equivalence between the stochastic and the well-calibrated deterministic versions of Newell's model for a sufficiently large number of vehicles.

© 2009 Elsevier Ltd. All rights reserved.

## 1. Introduction

Macroscopic analysis of microscopic car-following (CF) models has received a lot of attention in traffic flow literature (Chandler et al., 1958; Kometani and Sasaki, 1961; Gipps, 1981). Such studies establish the connection between CF rules and related continuum models (Zhang, 1998; Aw et al., 2002; Daganzo, 2006; Leclercq et al., 2007) or associated fundamental diagrams (May and Keller, 1967; Rakha and Crowther, 2002). However, in all these studies individual dispersion in driver behaviors is disregarded since CF parameters are averaged. Indeed, distributing these parameters prevents from analytically deriving the related continuum model. Thus, for typical microscopic models, the only way to depict the corresponding macroscopic pattern is to gather the results obtained from several simulation runs.

This paper proposes an alternative method to connect a simple CF rule that accounts for individual driver behaviors with its related macroscopic pattern. This method studies the propagation of waves generated by the CF model in congested situations. Thus, the main macroscopic properties of the microscopic model can be achieved even if the related continuum model is unknown. The stochastic version of Newell's model (Newell, 2002) will support this study. This model was selected because (i) in congestion it only requires two parameters per vehicle (ii) it was shown to suitably fit on-field data at a macroscopic scale (Ahn et al., 2004) and (iii) the related continuum formulation for a homogeneous fleet of vehicles is the well-known Lighthill–Whitham–Richards (LWR) model (Daganzo, 2006; Lighthill and Whitham, 1955; Richards, 1956).

First of all, a paired analysis of vehicle trajectories extracted from NGSIM database will determine the experimental distributions of the Newell's CF (NCF) model parameters (Section 2). This study will also provide new convincing results about NCF model ability to accurately reproduce individual trajectories in congestion. Section 3 will focus on the macroscopic pattern associated to NCF model with respect to wave propagation and queue length. A brief discussion is included in Section 4.

\* Corresponding author.

E-mail address: [nicolas.chiabaut@entpe.fr](mailto:nicolas.chiabaut@entpe.fr) (N. Chiabaut).

## 2. An experimental analysis of NCF model parameters distributions

### 2.1. Newell's model

Newell's CF model (Newell, 2002) assumes that two consecutive vehicles are related in congestion by a shift  $-\bar{w}(s_x/w, s_x)$  (see Fig. 1), where  $s_x$  is the minimal or jam spacing between two vehicles that is observed at a null speed, and  $w$  is the wave speed in congestion. Note that  $w$  is the slope of  $\bar{w}$ . The first coordinate of  $\bar{w}$  corresponds to time and the second to space. The shift  $\bar{w}$  is uniform along the trajectory, i.e. the NCF parameters of a vehicle are constant in time. However, these parameters may vary from one vehicle to another (Mauch and Cassidy, 2002; Ossen et al., 2006; Cassidy and Windover, 1998). NCF with distributed  $\bar{w}$  among vehicles will be referred in the sequel as the stochastic NCF (SNCF).

### 2.2. Experimental data

Optimal  $\bar{w}$  will now be determined for consecutive experimental trajectories extracted from the I-80 NGSIM dataset (NGSIM, 2006). Video recording was performed during afternoon rush period on a six lane highway near San Francisco, California. Notice that data were collected during congested period. Thus, the leader speed is always below 50 km/h. We will focus on stable situations, i.e. vehicles pairs driving in the same lane without lane changing. A previous study has shown that lane-changes are less frequent in lanes 2 and 3 (Chiabaut et al., 2009). Thus, we only consider these two lanes. Finally, 971 pairs of trajectories have been found.

### 2.3. Calibrating at a microscopic scale

The literature on car-following models calibration is extensive. A method that is commonly adopted is to solve a lead-vehicle problem (LVP) given a set of feasible parameter values. Ossen and Hoogendoorn (2008) highlight two main drawbacks of such a method. (i) An optimal process is undertaken to determine the best fit between the simulated and the measured data of the follower. Measurement errors can trigger substantial bias in the estimation results. Consequently, optimal parameters do not best capture follow-the-leader dynamics. (ii) Optimization criterion are usually the root mean square error or the Theil's coefficients (Kesting and Treiber, 2008; Ossen et al., 2006; Duret et al., 2009; Brockfeld et al., 2005). However, such objective functions are really flat and may corrupt the optimization results because multiple optimal or quasi-optimal parameters can be found.

Theoretical properties of SNCF model provide a reliable method to determine shifts  $\bar{w}$  overcoming the above-mentioned drawbacks without simulation: (i) since the shift  $\bar{w}$  is uniform along the vehicles trajectories, its spatial and time coordinates (respectively),  $\bar{w}_{|x} = s_x$  and  $\bar{w}_{|t} = s_x/w$  (respectively) can be determined by following the shock path of speed  $-w$  from one trajectory to the next. These coordinates do not depend on the lead-vehicle speed,  $u$ . (ii) One can easily determine which value of  $w$  provides uniform coordinates of  $\bar{w}$  along the trajectories or at least minimizes discrepancy of the measurements (assessed by the standard deviation). This criterion appears not to be flat and prevent from identifying several optimal parameters. Goodness of the fit of the recorded trajectory is embedded in the variation of the measured  $\bar{w}$  coordinates.

Let  $\bar{w}$  denote now the optimal shift between a vehicle and its leader. For any point  $A$  of the leader trajectory, one can draw a line of slope  $v$ . This line intercepts the follower trajectory at point  $B$  (see Fig. 2). One can then calculate the coordinates of  $\bar{v} = \overline{AB}$ . These coordinates,  $\bar{v}_{|t}$  and  $\bar{v}_{|x}$ , are constant regardless the leader speed only if  $v$  is equal to  $w$ . To illustrate this property, Fig. 2 depicts a pair of trajectories where the leader changes its speed from  $u_1$  to  $u_2$  and finally reached speed  $u_3$ . Shift  $\bar{w}$

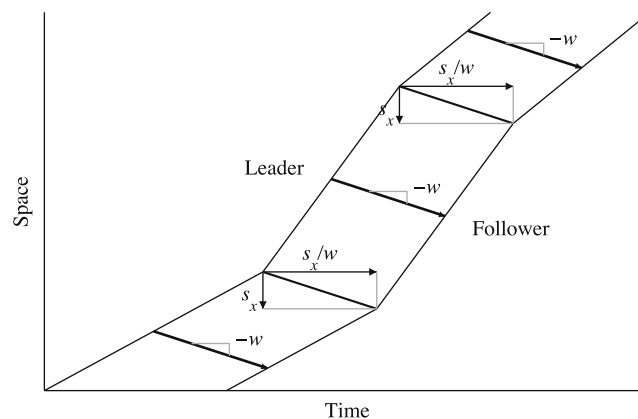


Fig. 1. Vehicles trajectories according to Newell's CF model.

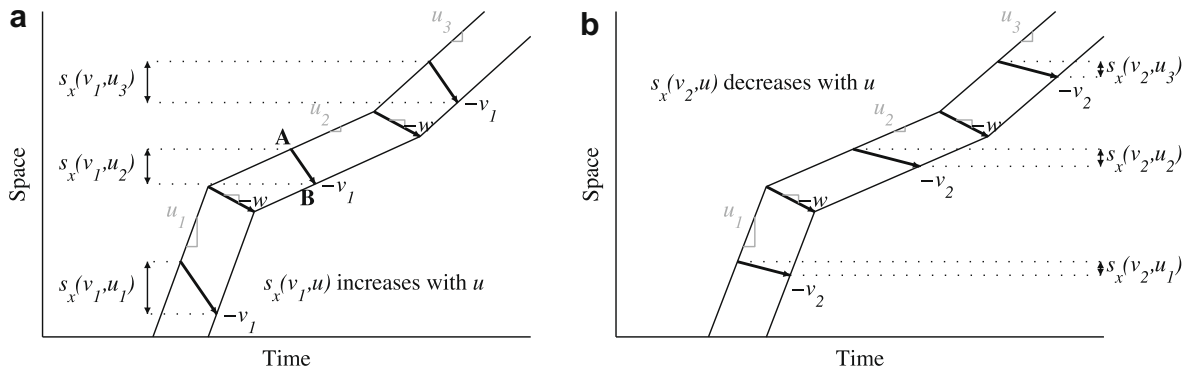


Fig. 2. Measurement process for (a)  $v \geq w$  and (b)  $v \leq w$ .

is uniform whatever the leader speed. Besides, shifts  $\bar{v} \neq \bar{w}$  are displayed. The spatial coordinate,  $\bar{v}_x$  is no longer the same in region  $u_1, u_2$  and  $u_3$ . In this case,  $s_x = \bar{v}_1|x$  increases with the leader speed whereas  $s_x = \bar{v}_2|x$  decreases.

This property is used to search the value of  $v$  that minimizes the variations in measurements of  $\bar{v}_x$  along the trajectories sample. Measurement of  $\bar{v}_x$  are performed along the vehicle trajectories for various values of  $v$ . Then, the standard deviation of these values is calculated to access to the invariant property of  $\bar{w}$ . Finally, for each pair of vehicles,  $\bar{w}$  is the shift  $\bar{v}$  whereof the associated value  $v$  minimizes the standard deviation of measurements of  $\bar{v}_x$ .

With classical criterion such as mean square error between recorded and simulated data, objective functions heavily weight prevailing situations. Impact of shockwaves on estimated parameters is smoothed by the function. Here, measurement of  $\bar{v}_x$  are gathered for range of lead-vehicle speeds and averaged. This emphasizes variations of  $\bar{v}_x$  for various values of  $u$  when  $v \neq w$ , but also permits to reduce measurements errors (Chiabaut et al., 2009). Indeed, inside a range of leader speed, the mean value is more representative than individual measurement of  $\bar{v}_x$ .

Thus,  $\bar{v}_x$  measurements are performed every  $\delta t$  along the leader trajectory ( $\delta t = 1/10$  second, the frequency of the videos of NGSIM dataset). Then,  $\bar{v}_x$  are averaged by range of speed centered in  $\bar{u}_i (u_i \in [0 \text{ km/h; } 50 \text{ km/h}])$  for each of the tested shift  $\bar{v}$ . Notice that an interval of 5 km/h ensures a sufficient number of range of speed for the selected trajectories of the NGSIM dataset and enough measurements in each range of speed. The mean value of a range of speed  $[u_i - 2.5 \text{ km/h; } u_i + 2.5 \text{ km/h}]$  is denoted  $\bar{v}_x(\bar{u}_i, v)$ . Then, the relative standard deviation  $s(v)$  of  $\bar{v}_x(\bar{u}_i, v)$  is calculated for each tested value of  $v$ :

$$s(v) = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n \left( \bar{v}_x(\bar{u}_i, v) - \overline{\bar{v}_x(\bar{u}_i, v)} \right)^2}}{\overline{\bar{v}_x(\bar{u}_i, v)}} \tag{1}$$

where  $\overline{\bar{v}_x(\bar{u}_i, v)}$  is the mean value of the  $\bar{v}_x(\bar{u}_i, v)$ . Note that the method can be carried out either on the time coordinate,  $\bar{v}_t$ , or on the spatial coordinate,  $\bar{v}_x$ .

The objective function  $s$  weights regions of different leader speed uniformly. Fig. 3 depicts evolution of  $s$  versus the value of  $v$  for a pair of vehicles of the NGSIM dataset. It confirms that the shape of the objective function is very sharp. Therefore, the parameters corresponding to the absolute extrema are determined without bias. Notice that this method is reliable if and

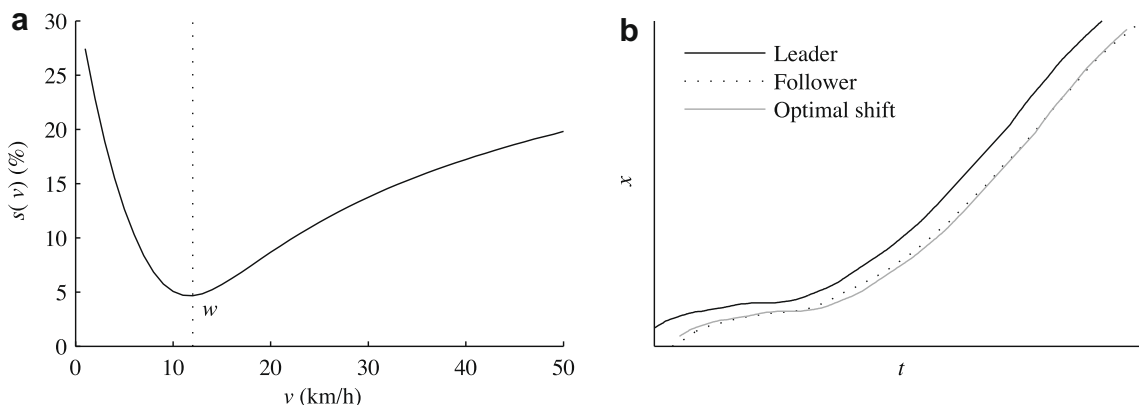


Fig. 3. (a) Evolution of  $s$  versus  $v$  for a pair of vehicles of the NGSIM dataset. (b) Trajectories of the vehicles.

only if Newell's model holds along the vehicles trajectories, i.e. parameters are constant in time. This key assumption will be verified in Section 2.5 on experimental data.

#### 2.4. Experimental results

For each vehicle pair of the experimental dataset, the value of  $v$  in [1 km/h; 100 km/h] that minimizes  $s$  (Eq. (1)) is calculated. It is worth noticing that some trajectories reveal measurement errors: sudden stop of a vehicle, low speed but not a lead-vehicle situation, sudden discontinuities in the leader trajectory. Instead of setting aside these vehicles by a visual inspection, it turns out that such case tends to  $w$  values bigger than 50 km/h or smaller than 3 km/h. In the same time, the associated value of  $s(w)$  is always bigger than 10%. Those vehicle pairs have then been disregarded from the analysis. The final set is constituted of 612 pairs (63% of the initial dataset). The criterion  $s$  is equal to 5.8%. Thus, Newell's model seems to accurately reproduces observed vehicle trajectories at a microscopic scale. It is then worth noticing that two parameters are enough to represent the behavior of an individual vehicle in congestion. Nevertheless, one should verify that these two parameters are constant in time for the trajectories of the final set (see Section 2.5).

Fig. 4 shows the histograms of  $w$  and  $s_x$  values for the 612 pairs of the final set. The parameters of the distribution of  $w$  are  $\bar{w} = 18.1$  km/h and  $\sigma = 10.3$  km/h. The values of  $w$  match what can be found in literature (between 10 km/h and 25 km/h) (Coifman et al., 2005; Windover and Cassidy, 2001; Chiabaut et al., 2009; Leclercq, 2005). Notice that the index of dispersion ( $\sigma/\bar{w}$ ) is equal to 0.6, which indicates that  $w$ -values are not far from the mean but that heterogeneity in drivers behaviors exists at a microscopic scale. The mean value of the distribution of  $s_x$  is  $\bar{s}_x = 6.9$  meters and the standard deviation is  $\sigma = 2.4$  m. Minimal jam spacings stand inside of the empirical range in the literature (Cassidy, 2003; Laval, 2005; May, 1988). The index of dispersion is much smaller (equal to 0.35). The distribution is sharply centered on the mean value  $\bar{s}_x$ . This is not surprising as most vehicles in the dataset are passenger cars.

Finally, the causes of the heterogeneity are explored. To this end, the vehicle size distribution is investigated. Fig 4c shows the histogram of the length  $l$  of the vehicles of the final dataset. The distribution is sharply centered on the mean value  $\bar{l} = 4.8$  m. The index of dispersion is very low (equal to 0.40). Thus, only one class of vehicles are involved in the NGSIM dataset. Heterogeneity in Newell's model parameters cannot be due to vehicle features. Hence, one can conclude that variations are generated by individual driving behaviors.

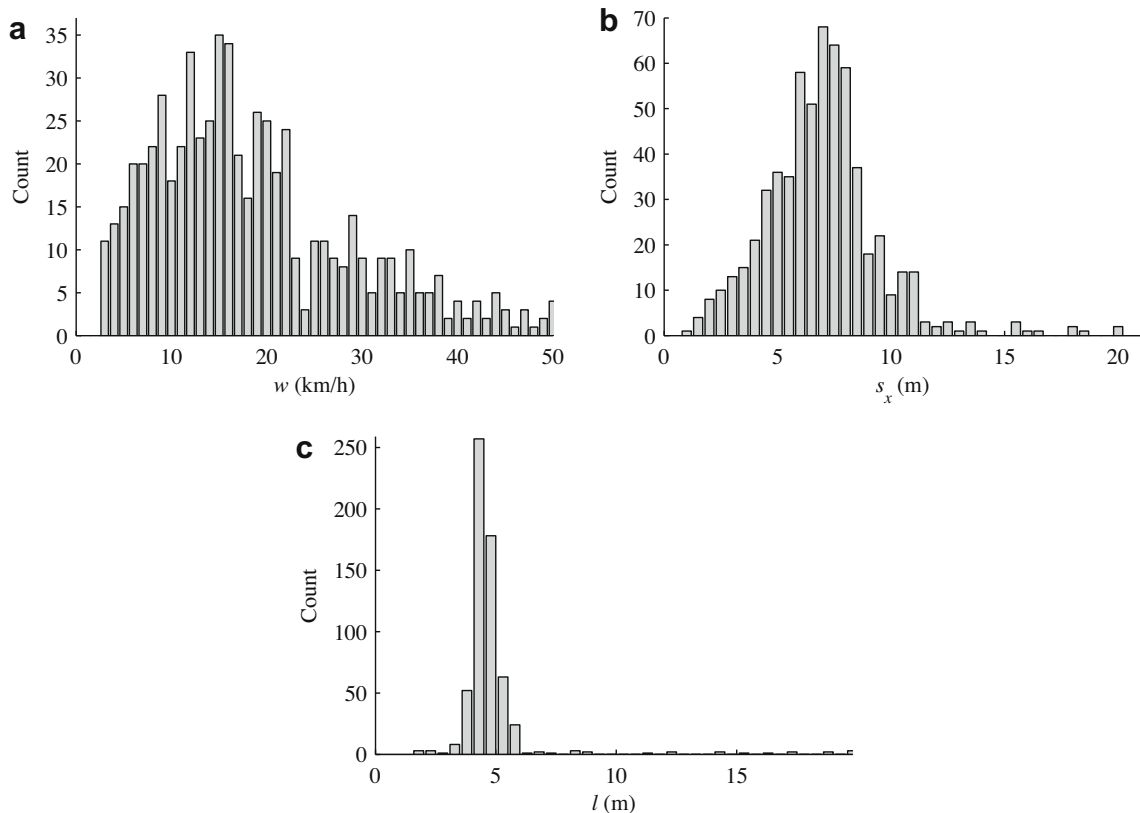


Fig. 4. Statistical analysis of results: (a) histogram for  $w$ ; (b) histogram for  $s_x$ ; (c) histogram for  $l$ .

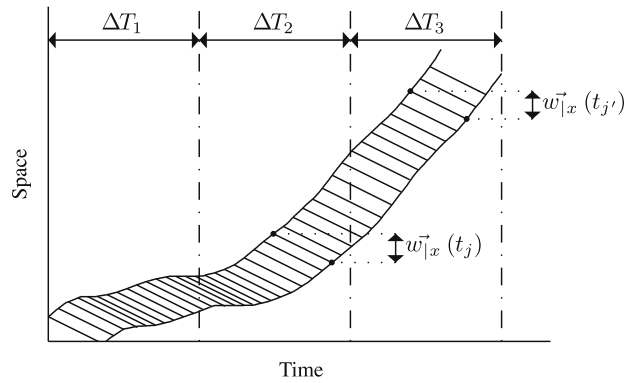


Fig. 5. Validation of the SNCF model.

2.5. Validation of SNCF parameters stationarity

The results obtained in Section 2.4 are reliable if and only if SNCF model holds on selected trajectories. This means that parameters are constant in time. Thus, even if driving behavior may vary from one vehicle to the next, estimated parameters  $w$  and  $s_x$  do not change along the trajectories for a pair of vehicles (Newell, 2002). This will be verified on experimental data.

Thus, the study is now focused on the evolution of the space-coordinate of the optimal shift  $\vec{w}$  with respect to time,  $\vec{w}_x(t)$ , for each pair of vehicles of the final set. To this end, trajectories have been divided in equal periods of time  $\Delta T_i$  (Fig. 5). Then, measurements performed in Section 2.4 are averaged by time period  $\Delta T_i$ . Mean value of  $\vec{w}_x$  are now calculated by range of time rather by range of speed:

$$\vec{w}_x(\Delta T_i) = \frac{1}{n} \sum_{j=1}^n \vec{w}_x(t_j) \tag{2}$$

where  $t_j$  are the measurement dates in time period  $\Delta T_i$  and  $n$  is the number of measurements during  $\Delta T_i$ .

If Newell’s model is valid for a pair of vehicles, then  $\vec{w}_x(\Delta T_i)$  have to be constant in time, i.e. variations are small from one period to the next. To assess the constancy in time of  $\vec{w}_x$ , the relative standard  $\sigma$  of  $(\vec{w}_x(\Delta T_i))$  is determined. This study is carried out for each pairs of vehicles of the final set.

Notice that only three time period are required to ensure that  $\Delta T_i$  is always longer than 15 s. Such a value permits (i) a sufficient number of measurement in each  $\Delta T_i$  always bigger than 150 and that (ii) vehicles change speed at least one time during the period. Otherwise, coordinates of any shift  $\vec{v}$  remains constant in time.

Fig. 6a shows the histograms of  $\sigma(\vec{w}_x(\Delta T_i))$  for the 612 pairs of vehicles of the final set. The mean value of the distribution of  $\sigma(\vec{w}_x(\Delta T_i))$  is equal to 0.07%. The cumulative distribution of  $\sigma(\vec{w}_x(\Delta T_i))$  (see Fig. 6b) reveals that 78% of the pairs of vehicles present a variation of  $\vec{w}_x(\Delta T_i)$  less than 10%. These values are consistent to conclude that parameters are constant in time for the trajectories of the final set. This confirmed that SNCF model accurately reproduces observed vehicle trajectories at a microscopic scale. Thus, the results obtained are suitable and furnish an estimation of the optimal shifts of SNCF model for the selected trajectories.

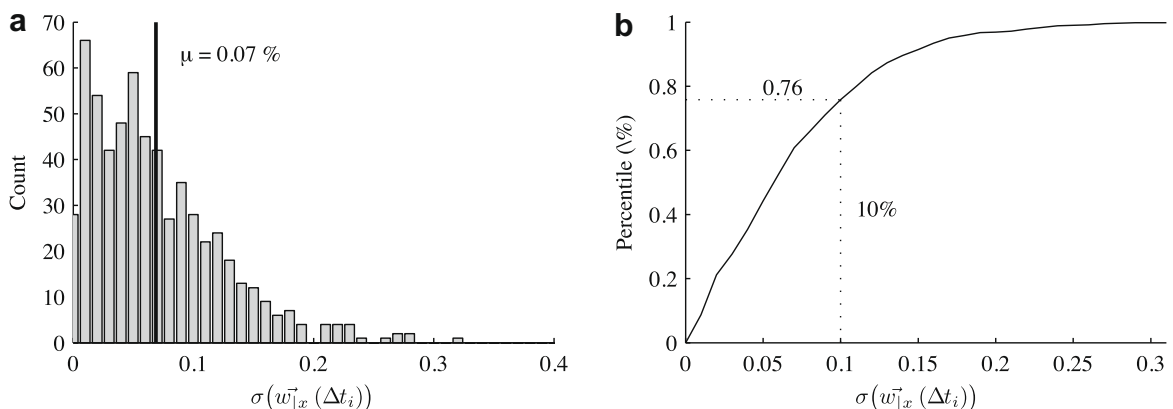


Fig. 6. Statistical analysis of results: (a) histogram for  $\sigma(\vec{w}_x(\Delta T_i))$ ; (b) cumulative distribution of  $\sigma(\vec{w}_x(\Delta T_i))$ .

### 3. Macroscopic patterns associated with SNCF model

Let  $\vec{w}_i$  denote the shift associated with vehicle  $i$ , where subscript  $i$  denotes the rank of the vehicle in the platoon. We are now going to consider platoons gathering  $n + 1$  vehicles and study wave propagation at a macroscopic scale, i.e. determining the mean shift  $\vec{W}_n$  between vehicle 1 and  $n + 1$ . Fig. 7 shows how wave propagates at a microscopic scale: shifts  $\vec{w}_i$  follows a random walk in the  $t-x$  plane (Newell, 2002). It follows that  $\vec{W}_n$  corresponds to a linear combination of  $\vec{w}_i (s_{x_i} / w_i, s_{x_i})$ .

$$\vec{W}_n = \sum_{i=1}^n \vec{w}_i \tag{3}$$

Its coordinates are:

$$\vec{W}_{n|t} = \sum_{i=1}^n \frac{s_{x_i}}{w_i} \tag{4a}$$

$$\vec{W}_{n|x} = \sum_{i=1}^n s_{x_i} \tag{4b}$$

The slope  $W_n$  of  $\vec{W}_n$  is given by:

$$W_n = \frac{\sum_{i=1}^n s_{x_i}}{\sum_{i=1}^n \frac{s_{x_i}}{w_i}} \tag{5}$$

This motivates the expression of the  $\vec{W}_n$  coordinates as the average of the individual parameters of SNCF model. We look for a model with equivalent wave propagation at a macroscopic scale. Its solution should match with  $\vec{W}_n$  for a platoon of  $(n + 1)$  vehicles. Such a model would be Newell’s model with appropriate uniform parameters (Leclercq, 2007). This last model is referred in the sequel as the deterministic NCF model (DNCF). To determine such parameters, one should look for the continuum limit of  $\vec{W}_n$  when  $n$  goes to infinity.

The case of  $\vec{W}_{n|x}$  is easily solved. The Central Limit Theorem states that the distribution converges toward the standard normal distribution  $\mathcal{N}$  in distribution. Thus, considering that the platoon is long enough, the parameters  $s_{x_i}$  can be estimated by the arithmetic mean of the distribution independently of its shape. Then  $\vec{W}_{n|x}$  is equal to  $n \cdot \bar{s}_x$  where  $\bar{s}_x$  is the arithmetic mean.

Since the distribution of  $s_{x_i}$  converges to its arithmetic mean, we will establish that the average wave speed,  $w = \lim_{n \rightarrow \infty} (W_n)$ , is the harmonic mean of individual wave speed  $w_i$ . To prove this formula, we will proceed in two steps. First, we will show that if the jam spacing  $s_{x_i}$  are uniform among drivers,  $w$  is actually the harmonic mean of  $w_i$ . In a second step, it will be proved that the latter remark is verified regardless the distributions of  $s_{x_i}$  and  $w_i$ .

#### 3.1. Uniform jam spacing

In this section  $s_{x_i}$  is assumed to be constant and equal to  $s_x$ . This corresponds to vehicles with the same length. Individual wave speeds are still considered. No restrictive assumption is taken on  $w_i$  distribution. Shift  $\vec{W}_n$  can now be expressed as:

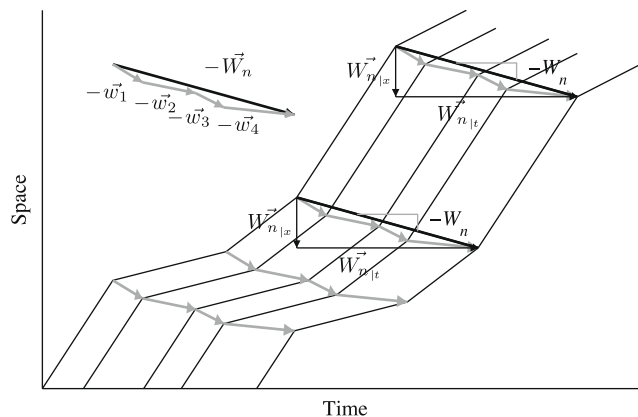


Fig. 7. Individualized wave propagation.

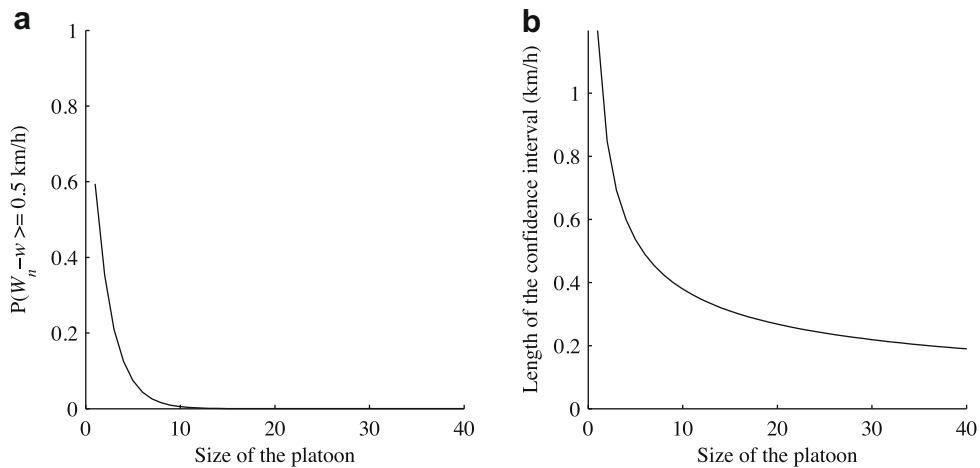


Fig. 8. (a) Evolution of the confidence level and (b) of the length of the confidence interval with respect to the number of vehicles.

$$\vec{W}_n|x = n \cdot s_x \tag{6a}$$

$$\vec{W}_n|t = \sum_i^n \frac{s_x}{w_i} \tag{6b}$$

It follows from the above-mentioned equations that the slope of  $\vec{W}_n$  is:

$$W_n = \frac{n}{\sum_i^n \frac{1}{w_i}} \tag{7}$$

This induces that the average wave speed of a fleet of  $n + 1$  vehicles is the harmonic mean of the  $n$  individual followers wave speeds. This is an important result because classically macroscopic behavior is estimated by arithmetic average. The last trajectory of a platoon simulated with SNCF model is identical to the trajectory computed by NCF model with uniform parameters corresponding to the mean shift  $\vec{\omega}_n = \frac{1}{n} \vec{W}_n: \vec{\omega}_n(s_x/W_n, s_x)$ . For an infinite platoon, wave speed can be determined by taking the limit of Eq. (7), which is the inverse of the expected value of parameters  $1/w_i$ .

$$w = \lim_{n \rightarrow \infty} (\omega_n) = \frac{1}{E[1/w_i]} \tag{8}$$

Thus, one should use  $w$  and  $s_x$  as parameters of an equivalent macroscopic model to produce the same macroscopic pattern.

However, the case of an infinite platoon remains hypothetical. The only way to estimate  $w$  is to approximate it in the case of a long platoon. Thus, it is appealing to study how  $W_n$  differs from  $w$ . To this end, the Hoeffding's inequality (9) provides an upper bound to likely estimate the wave speed, with respect to the platoon size, according to a desired accuracy  $x_w$  (in km/h) (Hoeffding, 1963).

$$P \left[ \left( \sum_{i=1}^n \frac{1}{w_i} - E \left[ \frac{1}{w_i} \right] \right) >= x_w \right] \leq e^{-\frac{2x_w^2}{n(b-a)}} \tag{9}$$

where  $[a, b]$  is the set of feasible values of  $1/w_i$ .

Fig. 8a displays the probability of estimating the harmonic mean with  $x_w = 0.5$  km/h, versus the platoon size  $n$  ( $1/a = 1$  km/h and  $1/b = 50$  km/h). Obviously the bigger the number of vehicles is, the better the estimate. Moreover, careful observation reveals that 10 vehicles in a row are sufficient to accurately assess  $w$  regardless of the distribution shape. In the same way, Fig. 8b presents evolution of the length of the confidence interval with respect to the number of vehicles. The point here is that small platoons suffice to estimate the mean wave speed. This is an important result to estimate  $w$  in practice because we can only consider quite small platoons. Indeed, in the I-80 NGSim dataset, platoons mostly gather around 7 vehicles.

### 3.2. Distributed jam spacing

The latter assumption is now relaxed to a continuous distribution of  $s_{x_i}$ . In the previous section, only one distributed parameter was involved in the calculus of the limit. We have now two distributed parameters and the convergence of their ratio,  $s_{x_i}/w_i$ , must be proved before taking the limit. Then, we are going to show that the associated macroscopic pattern

remains the mean shift  $\vec{W}_n(n.s_x/w, n.s_x)$  where  $s_x$  is the arithmetic mean of the distribution of jam spacing, and  $w$  is the harmonic mean of the distribution of wave speed.

The coordinates of  $\vec{W}_n$  are correspond to the general formula (4a) and (4b).

The point is to demonstrate that  $s_{x_i}/w_i$  converges. We have already mentioned that the distribution of  $s_{x_i}$  converges toward the standard normal distribution  $\mathcal{N}$  in distribution. One can likely estimate the parameters  $s_{x_i}$  by their arithmetic mean,  $s_x$ , when the size  $n$  of the platoon is big enough. The spatial coordinate of  $\vec{W}_n$  is thus equal to  $n \cdot s_x$ .

Next, according to the weak law of large numbers, the distribution of  $w_i$  converges in probability toward its expected mean value  $E[w_i]$ .

In accordance with the theorem of SLUTSKY (Billingsley, 1999, p. 21, Theorem 2.7), the pair of random variables  $(s_{x_i}, w_i)$  converges toward the pair  $(\mathcal{N}, E[w_i])$  in distribution. Noticing that if  $f$  is a continuous function and if the random variable  $X_i$  converges toward  $X$  in distribution, then  $f(X_i)$  converges toward  $f(X)$  in distribution. Thereby, the random variable  $(s_{x_i}/w_i)$  converges.

Thus, the average wave speed  $w$  can be written as the continuum limit of (4a). This equation can be formulated as the harmonic mean of parameters  $w_i$ , regardless the distributions of  $s_{x_i}$  and  $w_i$ .

$$w = \frac{1}{E[1/w_i]} \quad (10)$$

This proves that  $\vec{\omega}_n = \frac{1}{n} \vec{W}_n$  converges towards a single  $\vec{w}(\frac{s_x}{w}, s_x)$ . It is now appealing to determine how quickly  $\vec{\omega}_n$  tends to its limit. Identically to the case of uniform jam spacing, the Hoeffding's inequality (9) provide an accurate estimation of both coordinates of  $\vec{W}_n$ . Evolution of the confidence level and of the length of the confidence interval with respect to  $n$  follow the same trend as in Fig. 8.

From (9) one can calculate the minimal number of vehicle required to provide an appropriate estimation of the average parameter of SNCF model. Given a desired accuracy  $x_{s_x}$  for jam spacing, a desired accuracy  $x_w$  for wave speed, and the same level of certainty  $\alpha$  for both variables, the minimal number of required vehicles,  $n^*$  is:

$$n^* = \max \left[ -\frac{2x_{s_x}^2}{\log(\alpha)(b_{s_x} - a_{s_x})^2}, \frac{2x_w^2}{\log(\alpha)(b_{\frac{1}{w}} - a_{\frac{1}{w}})^2} \right] \quad (11)$$

where  $s_{x_i}, [a_{s_x}, b_{s_x}]$  is the set of feasible values of  $s_{x_i}$ ,  $[a_{\frac{1}{w}}, b_{\frac{1}{w}}]$  is the set of feasible values of  $1/w_i$  and where the probabilities in the Hoeffding's inequality (9) correspond to:

$$P \left[ \left( \sum_{i=1}^n s_{x_i} - E \left[ \sum_{i=1}^n s_{x_i} \right] \right) \geq x_{s_x} \right] = 1 - \alpha \quad (12a)$$

$$P \left[ \left( \sum_{i=1}^n \frac{1}{w_i} - E \left[ \sum_{i=1}^n \frac{1}{w_i} \right] \right) \geq x_w \right] = 1 - \alpha \quad (12b)$$

With the upper bound and the lower bound of the parameters estimated in Section 2,  $n^*$  is equal to 11 vehicles for a 95% confidence level and an accuracy of 95%.

### 3.3. Application of the theory

To implement the analytical formula demonstrated in the previous section, platoons of the dataset of Section 2 will be used. For each platoon, the shifts  $\vec{W}_n$  are estimated between vehicles 1 and  $n + 1$ . The method of Section 2 is performed on  $\vec{W}_{n|t}$  measurements. Distribution of  $W_n$  among platoons is obtained.

Then, average wave speeds,  $w$ , are calculated for each platoon as the harmonic mean of individual parameters of vehicles involved in the platoon. Section 2 provides values of  $w_i$ . Thus,  $w$  can be expressed as:

$$w = \frac{n}{\sum_{i=1}^n 1/w_i} \quad (13)$$

Results of the two approaches can thus be compared. The cumulative distribution of  $|W_n - w|$  is presented in Fig. 9a.

Fig. 9a shows that the model predictions for the mean wave speed are in an interval of 5 km/h for 71% of the tested platoons. However the dataset holds only small platoons (less than 10 vehicles). For each observed platoon size, the total number of platoons is shown in Fig. 9b. Notice that we excluded the single platoon formed by 3 vehicles. Nevertheless, the results are in good agreement with the analytical findings. Mean values of  $|W_n - w|$  by range of platoon size decreases with the length of the platoon (Fig. 9b). It enforces the analytical findings that  $W_n$  converges towards the harmonic mean of  $w_i$ . Thus, one requires longer congested stable platoons to estimate accurately the macroscopic wave speed.



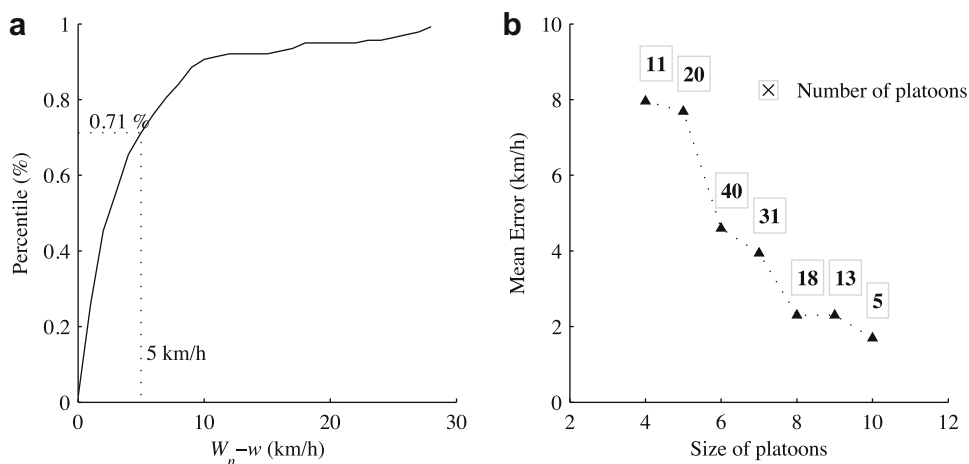


Fig. 9. Verification results: (a) cumulative distribution of approximation error of  $w$ ; (b) total number of platoons and mean values of  $|W_n - w|$  by range of platoons size.

#### 4. Discussion

The first insight of this paper is that a parsimonious method is proposed to estimate NCF parameters at a microscopic scale. I-80 NGSim data analysis shows that at this scale driving behaviors vary from one vehicle to the another. This is in agreement with previous observations (Ahn et al., 2004; Duret et al., 2009). Furthermore, driving behavior variability is appraised here through the  $s_x$  and  $w$  distributions.

The second insight is to formally establish relations between a stochastic microscopic model and its associated macroscopic pattern. Analytical formula shows that mean jam spacing corresponds to the arithmetic mean of individual jam spacing whereas the mean wave speed corresponds to the harmonic mean. These findings make it possible to: (i) calibrate a macroscopic model (e.g. DDCF model) from individual observations; (ii) establish that for large platoons (more than ten vehicles) SNCF model and DDCF model calibrated with mean shift provide exactly the same results. This explains why the LWR model (whose solutions are equivalent to DDCF model (Daganzo, 2006)) is known to provide accurate results in congestion at a macroscopic scale even if it does not account for individual driving behaviors.

An important corollary of the findings of the paper is that the order of vehicle creation has no impact in a microscopic stochastic simulation. Indeed, every permutation of a set of  $\tilde{w}_i$  leads to the same macroscopic pattern. This is a convenient property for a microscopic simulation tool.

This paper focuses on congested traffic where drivers heterogeneity appears to have no impact at a macroscopic scale. Further research should be conducted to study free-flow situations and nearly congested situation. Indeed, individual dispersion or vehicle type may affect travel time values but also trigger phenomenon such as lane changes, lane speed difference, or various merging behaviors (Laval and Daganzo, 2006; Laval and Leclercq, 2008). This is being investigated by the authors.

#### Acknowledgements

The statistical findings of the paper are the results of many discussions with MATTHIEU LERASLE of University Paul Sabatier (LSP, Toulouse, France). The authors are grateful to him for his valuable help. The authors thank ESTELLE CHEVALLIER for her careful reading.

#### References

- Ahn, S., Cassidy, M.J., Laval, J., 2004. Verification of a simplified car-following theory. *Transportation Research Part B* 38 (5), 431–440.
- Aw, A., Klar, A., Rasche, M., Materne, T., 2002. Derivation of continuum traffic flow models from microscopic follow-the-leader models. *SIAM Journal on Applied Mathematics* 63 (1), 259–278.
- Billingsley, P., 1999. *Convergence of Probability Measures*, second ed. Wiley-Interscience.
- Brockfeld, E., Kuhne, R., Wagner, P., 2005. Calibration and validation of microscopic models of traffic flow. *Transportation Research Record: Journal of the Transportation Research Board* 1934, 179–187.
- Cassidy, M.J., 2003. Freeway on-ramp metering, delay savings and the diverge bottleneck. *Transportation Research Record: Journal of the Transportation Research Board*, 1856.
- Cassidy, M.J., Windover, J.R., 1998. Driver memory: motorist selection and retention of individualized headways in highway traffic. *Transportation Research Part A* 32 (2), 129–137.
- Chandler, R.E., Herman, R., Montroll, E.W., 1958. Traffic dynamics: studies in car following. *Operation Research* 2 (6), 165–184.
- Chiabaut, N., Buisson, C., Leclercq, L., 2009. Fundamental diagram estimation through passing rate measurements in congestion. *IEEE Transactions on Intelligent Transportation Systems* 10 (2), 355–359.

- Coifman, B.A., Wang, Y., 2005. Average velocity of waves propagating through congested freeway traffic. In: Mahmassani, H.S. (Ed.), 16th International Symposium on Transportation and Traffic Theory. Elsevier, MD, USA, pp. 165–179.
- Daganzo, C.F., 2006. In traffic flow, cellular automata = kinematic waves. *Transportation Research Part B* 40 (5), 396–403.
- Duret, A., Chiabaut, N., Buisson, C., 2009. Estimating individual speed-spacing relationship and assessing the Newell's car following model ability to reproduce trajectories. *Transportation Research Record: Journal of the Transportation Research Board* 2088, 188–197.
- Gipps, P.G., 1981. A behavioural car-following model for computer simulation. *Transportation Research Part B* 15 (2), 105–111.
- Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* 58 (301), 13–30.
- Kesting, A., Treiber, M., 2008. Calibrating car-following models using trajectory data: a methodological study. *Transportation Research Record: Journal of the Transportation Research Board* 2088, 148–156.
- Kometani, E., Sasaki, T., 1961. Dynamic behavior of traffic with nonlinear spacing-speed relationship. In: Herman, R. (Ed.), *Theory of Traffic Flow*, General Motors Research Laboratory. Elsevier, Warren, MI, pp. 105–119.
- Laval, J.A., 2005. Linking synchronized flow and kinematic waves. In: Kuhne, R., Poeschl, T., Schadschneider, A., Schreckenberg, M., Wolf, D. (Eds.), *Traffic and Granular Flow'05*. Springer.
- Laval, J.A., Daganzo, C.F., 2006. Lane-changing in traffic streams. *Transportation Research Part B* 40 (3), 251–264.
- Laval, J.A., Leclercq, L., 2008. Microscopic modeling of the relaxation phenomenon using a macroscopic lane-changing model. *Transportation Research Part B* 42 (6), 511–522.
- Leclercq, L., 2005. Calibration of flow-density relationships on urban streets. *Transportation Research Record: Journal of the Transportation Research Board* 1934, 226–234.
- Leclercq, L., 2007. Hybrid approaches to the solutions of the “Lighthill–Whitham–Richards” model. *Transportation Research Part B* 41 (7), 701–709.
- Leclercq, L., Laval, J.A., Chevallier, E., 2007. The lagrangian coordinates and what it means for first order traffic flow models. In: Allsop, R.E., Bell, M.G., Heydecker, B.G. (Eds.), *Transportation and Traffic Theory 2007 (ISTTT17)*. Elsevier, London, UK, pp. 735–754.
- Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves ii. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society A* 229, 317–345.
- Mauch, M., Cassidy, M.J., 2002. Freeway traffic oscillations: observations and predictions. In: Taylor, M.A.P. (Ed.), 15th International Symposium on Transportation and Traffic Theory.
- May, A.D., 1988. *Traffic Flow Fundamentals*. Prentice Hall, Englewood Cliffs.
- May, A.D., Keller, H.E.M., 1967. Non integer car-following models. *Highway Research Record* 199 (1), 19–32.
- Newell, G.F., 2002. A simplified car-following theory: a lower order model. *Transportation Research Part B* 36 (3), 195–205.
- NGSIM (2006). Next generation simulation.
- Ossen, S., Hoogendoorn, S.P., 2008. Validity of trajectory-based calibration approach of car-following models in the presence of measurement errors. *Transportation Research Record: Journal of the Transportation Research Board* 2088, 117–125.
- Ossen, S., Hoogendoorn, S.P., Gorte, B.G.H., 2006. Interdriver differences in car-following: a vehicle trajectory-based study. *Transportation Research Record: Journal of the Transportation Research Board* 1965, 121–129.
- Rakha, H., Crowther, B., 2002. Comparison of greenshields, pipes, and Van Aerde car-following and traffic stream models. *Transportation Research Record: Journal of the Transportation Research Board* 1802, 248–262.
- Richards, P.I., 1956. Shockwaves on the highway. *Operations Research* 4, 42–51.
- Windover, J.R., Cassidy, M.J., 2001. Some observed details of freeway traffic evolution. *Transportation Research Part A* 35 (10), 881–894.
- Zhang, H.M., 1998. A theory of nonequilibrium traffic flow. *Transportation Research Part B* 32 (7), 485–498.