Calibration and validation of cellular automaton traffic flow model with empirical and experimental data

Jin, Cheng Jie; Knoop, Victor L.; Jiang, Rui; Wang, Wei; Wang, Hao

DOI
10.1049/iet-its.2016.0275

Publication date
2018

Document Version
Final published version

Published in
IET Intelligent Transport Systems

IET Intelligent Transport Systems

Research Article

IET Journals
The Institution of Engineering and Technology

Calibration and validation of cellular automaton traffic flow model with empirical and experimental data

ISSN 1751-956X
Received on 24th October 2016
Revised 1st December 2017
Accepted on 6th January 2018
E-First on 14th February 2018
doi: 10.1049/iet-its.2016.0275
www.ietdl.org

Cheng-Jie Jin[1,2,3] ✉, Victor L. Knoop[3], Rui Jiang[4], Wei Wang[1,2], Hao Wang[1,2]

[1]*Jiangsu Key Laboratory of Urban ITS, Southeast University of China, Nanjing, Jiangsu 210096, People's Republic of China*
[2]*Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Nanjing, Jiangsu 210096, People's Republic of China*
[3]*Delft University of Technology, Transport & Planning, Stevinweg 1, 2628 CN Delft, The Netherlands*
[4]*MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, People's Republic of China*
✉ *E-mail: yitaikongtiao@gmail.com*

Abstract: For traffic flow models, calibration and validation are essential. Cellular automaton (CA) models are a special class of models, describing the movement of vehicles in discretised space and time. However, the previous work on calibration and validation does not discuss CA models systematically. This study calibrates and validates a stochastic CA model. The authors use a simple CA model, which only has two important parameters to be calibrated. The methodology for optimisation is to minimise the relative root mean square error between two properties: the averaged velocity and the variation of velocities in a platoon at a given density. Three different sites are used as cases to show the methodology, for which different types of data (video trajectories or GPS data) are available. The authors find that the best model parameters vary for the different locations. This may result from various driving strategies and potential tendencies. Thus, it is concluded that for CA models, various traffic flow phenomena need to be simulated by various parameters.

1 Introduction

The study of traffic flow has a long history [1–3]. The traffic flow models can help us in the field of traffic control, management and operations, which are very important for both the scientists and engineers. Usually, traffic flow models can be divided into two categories: macroscopic models which describe traffic in aggregated quantities, and microscopic models which describe the status of each vehicle, including velocities and time-headways.

A special category of microscopic models is cellular automaton (CA) model. It is discrete in both space and time, which makes it is easy to use and can run fast in the computer simulations [4]. Since the proposal of Nagel–Schreckenberg (NS) model [5], in this field there are many CA models with various features. Especially, with the development of Kerner's three-phase traffic theory [6–9], many three-phase CA models which can reproduce synchronised flow phenomena are proposed and studied in the recent years, including Kerner–Klenov–Wolf model [10–12] and Kerner–Klenov–Schreckenberg model [13].

In order to make traffic flow models useful, the model calibration and validation are necessary. In the past years there are many studies concerning this topic, especially for the microscopic car-following models [14–19]. Many traffic flow data can be used: some of them are empirical ones, e.g. the Next Generation SIMulation (NGSIM) video data. Some others are experimental ones, e.g. the precise data from GPS equipped [20–22].

However, until now there are nearly no specific studies about the systematic calibration and validation of CA traffic flow models. In many previous studies [23–31], the choice of parameters in CA models is usually done by some qualitative comparison of different simulation results, e.g. the fundamental diagrams or spatiotemporal diagrams. This process is not systematic, and a complete framework is lacked. Usually the differences and the errors are not easy to measure. This also becomes an important disadvantage of various CA models, which has been criticised much before. (Some related work has been done in one CA pedestrian flow model [32]. However, its modelling mechanism is different from that of CA traffic flow model.). Therefore, how to make CA traffic flow models more realistic and how to choose the best parameters systematically become an important topic, which will be our main task.

Therefore, in this paper we put the CA model through a complete calibration and validation process. To this end, we use a simple and flexible CA model: it only has two important parameters to be calibrated, and the change of them can lead to significant differences for the macroscopic characteristics of the model. For quantifying the error between the model and reality, we use the averaged value and standard deviation of single-vehicle velocities. The parameters which can minimise the error will be the best results. This can become a general process for the calibration work of CA model. Besides, different types of traffic flow data will be used for the calibration, and the validation process, including holdout validation and cross-validation will also be presented.

This paper is organised as follows. In Section 2, we briefly introduce the rules and parameters of our CA model, and present the basic model properties. In Section 3, we introduce the calibration methodology for the CA model. In Section 4, we present the testing methodology for the calibration, including the introduction of three different types of traffic flow data. In Section 5, we show and discuss the calibration results. In Section 6, we do the validation for the calibrated parameters, and the conclusions are presented in Section 7.

2 CA model

2.1 Revisit of the model rules and parameters

In this paper, we use our model [28, 29] as an example for the calibration and validation of a CA traffic flow model. In the simulation, each cell is set to 1 m and each time step corresponds to 1 s. Each vehicle occupies eight cells. A special property of our rules is the parameter named *anticipated deceleration* (AD), which means drivers prefer to decelerate with this value when necessary. Note that AD can be chosen from a continuous distribution. The anticipated velocity V_{anti} is defined as the highest v in which a vehicle can stop in the gap in front satisfying the inequality

$$B(v, \text{AD}) \leq \text{gap} \tag{1}$$

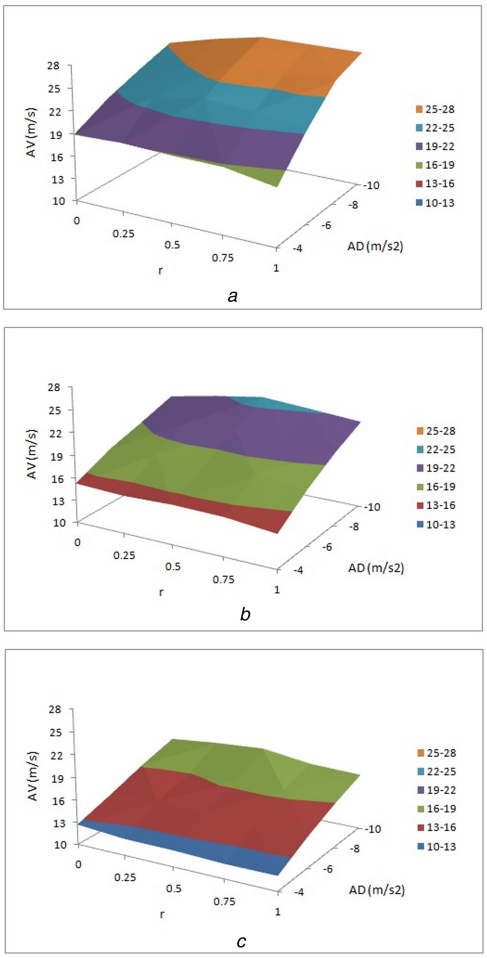


Fig. 1 *AV values under different periodic boundary conditions*
(a) $\rho = 25$ veh/km, ***(b)*** $\rho = 35$ veh/km, ***(c)*** $\rho = 45$ veh/km

Here $B(v, \mathrm{AD})$ is the distance that a vehicle travels before stopping if it keeps decelerating with AD from an initial velocity. We suppose $m = \mathrm{int}(v/|\mathrm{AD}|)$, then $B(v, \mathrm{AD})$ can be calculated as

$$\begin{aligned} B(v, \mathrm{AD}) &= v + (v + \mathrm{AD}) + (v + 2\mathrm{AD}) + \ldots + (v + m\mathrm{AD}) \\ &= (2v + m\mathrm{AD})(m + 1)/2 \end{aligned} \tag{2}$$

Note that since m is proportional to $v/|\mathrm{AD}|$, B scales with the square of v and the inverse of $|\mathrm{AD}|$. However, since CA models are discrete and the decelerations occur before the movement, the stopping distance does not equal $v^2/(2\mathrm{AD})$.

The model rules are executed on all vehicles in parallel as follows:

1. Estimating the virtual velocity of the front vehicle (vehicle $n-1$)

$$v'_{n-1} \rightarrow \min\{v_{\max} - A_n,\ \max\{0, V_{\mathrm{anti}}(\mathrm{AD}, \mathrm{gap}_{n-1}) - A_n\}, v_{n-1}\} \tag{3}$$

2. Deterministic acceleration or deceleration (see (4))
3. Randomisation

$$v_n \rightarrow \max\{v_n - A_n, 0\} \text{ with probability } p \tag{5}$$

4. Updating positions

$$x_n(t+1) \rightarrow x_n(t) + v_n(t) \tag{6}$$

In the above formulations, $x_n(t)$ and $v_n(t)$ are the position and velocity of vehicle n at time t; vehicle $n-1$ precedes vehicle n; gap_n is the distance between vehicle n and vehicle $n-1$; A_n is the normal acceleration value, which is set as 1 m/s^2; $v_{\max}$ is the velocity limitation of the vehicles; p is the randomisation probability which is chosen between 0 and 1.

It should be noted that in step 2, we consider two possible choices: the one is to accelerate, when the gap is large enough to ensure the safety. The calculation method is similar to that used in many previous CA models (e.g. NS model). The other one is to decelerate, when the gap is not large enough. As mentioned before, the calculation of V_{anti} is one new idea, which is different from previous CA models. Besides, the r and $1-r$ represent the proportion of conservative and radical strategy adopted by drivers, since $r = 1$ is a very conservative condition for acceleration, and $r = 0$ is an aggressive one. This parameter can be roughly interpreted as the aggressiveness in some other car-following models [33].

Actually, one difficult aspect of calibrating CA models is that they include a randomisation term which stops or slows down vehicles at random times. This is not realistic at the vehicle level, but it represents the collective phenomena observed in traffic. When including random effects, the calibration of the parameter indicating the probability of random decelerations (p in our formulation) becomes difficult: unlike some other factors (e.g. the reaction time in car-following models), the effect of p cannot be found at the level of an individual driver. This makes some usual techniques for the calibration of car-following models unavailable for the calibration of CA models, including the use of vehicle trajectories.

However, in our model, this is not so much a problem, since within a certain range the effect of p is not so important as in some other CA models. For example, when the other parameters are all fixed, for the situations when $p = 0.1$ and when $p = 0.01$, the fundamental diagrams will be similar [30, 31]. This is due to the mechanism that AD and r have a much more significant influence on our model. Thus in this paper, we will always set p as a constant value: $p = 0.1$, and the focus is on AD and r. Only dealing with two important parameters will make the work easy to control.

2.2 Discussion of the model properties

In this paper, we calibrate the model on two properties of single-vehicle velocities. They are the averaged velocity (AV for short) of all the vehicles in one platoon, and the standard deviation of these velocities (SDV for short). They are simple yet important, since the velocities are the most effective indicators for the traffic flow status. The remainder of this section shows the effect of AD and r on AV as well as SDV. These effects are computed using a simulation under the periodic boundary conditions. The length of the circular road is 80 km, which can contain 10,000 vehicles, and the free flow velocity is set as 32 m/s.

An overview of the resulting AV and SDV is shown in Figs. 1 and 2, in which three global densities are considered, 25, 35 and 45 veh/km. We can find that the basic trends are quite similar: from Fig. 1 it is clear that r has nearly no relation with AV, and AV mainly depends on AD. Larger $|\mathrm{AD}|$ leads to larger AV. The trend in Fig. 2 is different: AD has nearly no relation with SDV, and SDV mainly depends on r. Smaller r leads to larger SDV. Here all the trends are monotonic, which is important in light of the sequel of calibration. Moreover, the fact that the two outcome measures are decoupled (AD is linked to AV and r is linked to SDV) will ease finding the right value, even if one needs to search manually. In this paper, we will explore all combinations of r and AD.

We also check the distribution of the velocities of the vehicles, since they can help to understand how AV and SDV change. Fig. 3 depicts this distribution with a 25 veh/km density as example. Here r can determine the shape of the distribution: larger r can make it much narrower, which implies the traffic flow is more stable and

$$v_n \rightarrow \begin{cases} \min\{v_n + A_n, v_{\max}\} & \text{when} \quad (1-r)v_n + rB(v_n, \mathrm{AD}) < \mathrm{gap}_n + v'_{n-1} \\ V_{\mathrm{anti}}(\mathrm{AD}, \mathrm{gap}_n + v'_{n-1}) & \text{when} \quad (1-r)v_n + rB(v_n, \mathrm{AD}) \geq \mathrm{gap}_n + v'_{n-1} \end{cases} \tag{4}$$

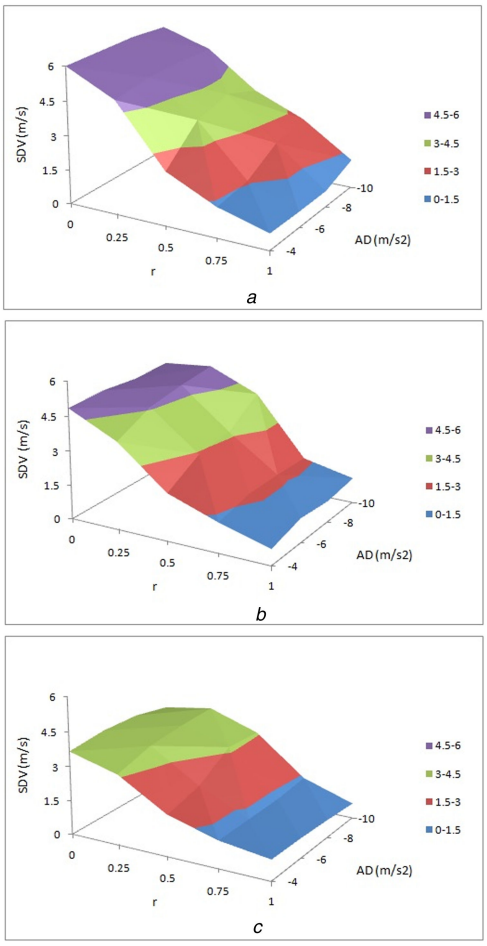


Fig. 2 *SDV values under different periodic boundary conditions*
(a) $\rho = 25$ veh/km, ***(b)*** $\rho = 35$ veh/km, ***(c)*** $\rho = 45$ veh/km

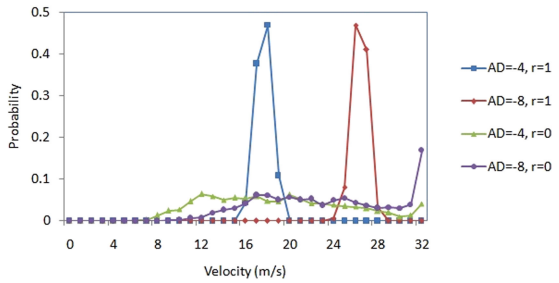


Fig. 3 *Some typical velocity distributions under different simulation conditions, ρ = 25 veh/km. Four results of (AD, r) are shown*

the interactions between vehicles are strong. For the situation when $r = 1$, there is only one obvious peak and the traffic flow can be considered as one platoon. The criterion of defining one platoon is that inside the platoon, all time-headways should be 6 s or less [34], which will be suitable for the calibration methodology introduced in the next section. AD can decide the velocity values, and larger |AD| can make them also larger. For the situation when $r = 1$, with the growth of |AD|, the single peak just moves to righter side. However for $r = 0$, larger |AD| can produce the second peak at high velocities, which even becomes dominant. It implies the existence of free flow, and the traffic flow is not stable in time or space. For this situation, it cannot be considered as one single platoon again, and cannot be used for calibration.

Besides, it is clear that too large |AD| values can lead to some too large velocities, which should not be considered during the model calibration. This also shows the problem of some early CA models. For example, NS model can be considered as our model with $\mathrm{AD} = -32\ \mathrm{m/s^2}$, since vehicles always can stop in one time step, even at the maximum velocity. This configuration will produce some unrealistic simulation results.

3 Methodology for the calibration

The traffic flow we observed on the real highways can be considered as the situation under open boundary conditions. However, it is difficult to find the precise traffic demand in a real situation, which might be fluctuating. Therefore, we cannot directly compare the empirical data with the simulation results under open boundary conditions, since the demand cannot be set according to the real-life situation. Instead, we simulate traffic under periodic boundary conditions, and the dynamics for the certain density can be found. For the calibration work, we use the following steps:

(1) Find stable platoons in empirical traffic data. Here 'stable' means the velocities and time-headways of all the vehicles are similar, and no headways are longer than 6 s (cut-off value for platoons).
(2) Combine small platoons with the same densities into one large platoon for calibration, if necessary. The new flow and velocity may be different, while the density will remain the same. This process will be explained by cases B and C in Sections 4 and 5.
(3) Determine the basic properties of this platoon: flow (Q), velocity properties (AV and SDV) and density (ρ). Due to the perspective and distortion, the spatial density in the empirical video data is more accurate closer to the cameras than further away. Therefore, in this paper we choose to use the temporal density, which is obtained by the locally observed flow and velocity: $\rho = Q/\mathrm{AV}$.
(4) Initialise the simulation on the circular road ($L = 80$ km). The vehicles are homogeneously distributed at the global density of ρ.
(5) Run the simulation to $T = 10{,}000$ s to reach a stable state. After that, the single-vehicle velocities are recorded at one fixed location, and the velocity distributions can be obtained.
(6) Find the minimum of the objective function (measure of performance [35]). To define the function there are many possible options [36], and here we only present one simple form like relative root mean square error (RMSE)

$$E = \sqrt{\left(\frac{\mathrm{AV_s} - \mathrm{AV_e}}{\mathrm{AV_e}}\right)^2 + k \times \left(\frac{\mathrm{SDV_s} - \mathrm{SDV_e}}{\mathrm{SDV_e}}\right)^2} \tag{7}$$

Here $\mathrm{AV_s}$ and $\mathrm{SDV_s}$ represent the simulation results of AV and SDV, while $\mathrm{AV_e}$ and $\mathrm{SDV_e}$ represent the empirical results. $\mathrm{AV_s}$ and $\mathrm{SDV_s}$ are the functions of ρ, AD and r. For sake of simplicity we set $k = 1$ here. In order to find the minimum E, we draw some three-dimensional figures to show how E changes with AD and r, which will be presented in Section 5.

It should be noted that in the empirical data, sometimes the traffic flow is stable, while sometimes it is not. To discuss the reason why is out of the scope of this paper. The methodology discussed here is intended to be used for stable platoons. Actually, inside one unstable platoon, there may be two or more different states, which can be calibrated separately. However, it is difficult to automatically divide one large unstable platoon into several small stable platoons.

4 Calibration datasets

For the calibration, we use data from three different sites, which are described here. The calibration results for each of these sites are presented in Section 5. Since they have different densities, we have to calibrate them separately.

4.1 Case study A: empirical data from two-lane traffic

Traffic flow data is the base of the calibration work. First we use the empirical video data from Nanjing Airport Highway [37]. This highway is the main road connecting the urban district with Nanjing Lukou Airport. There are 40 cameras along the 28 km-long highway, and in each direction there are two lanes. The speed limit is 120 km/h.

The data we used in Sections 4 and 5 are all single-vehicle data, in which the velocities and time-headways of all the vehicles can be obtained. The single-vehicle data are extracted by the equipment

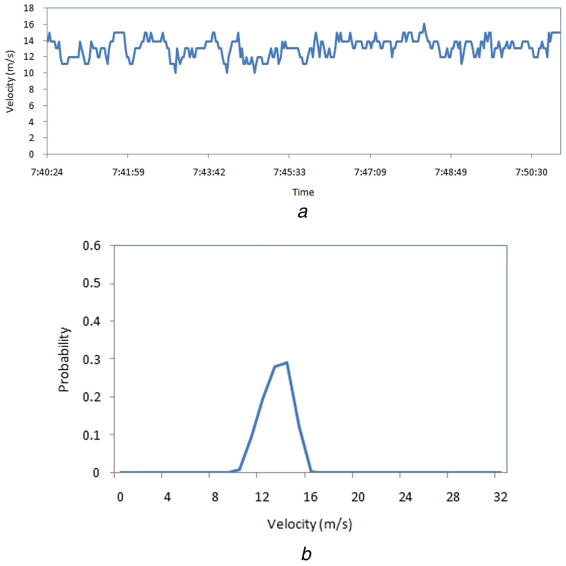


Fig. 4 *Platoon A in our empirical video data. It was collected at the left lane, on the morning peak of 27 August 2010*
(a) Single-vehicle velocities, ***(b)*** Velocity distribution

Table 1 Statistical information of four small platoons, and their combination results (Platoon B) in our empirical video data. They were all collected at one location in the single-lane section, on the afternoon of 13 August 2010

Platoon	Length, veh	Passing time, s	Flow, veh/h	AV, km/h	Averaged density, veh/km
1	27	48	2025	59.9	33.8
2	21	37	2043	60.6	33.7
3	18	33	1964	58.0	33.9
4	22	37	2141	64.5	33.2
combination	88	155	2044	61.2	33.4

named *Autoscope Rackvision Terra*. We performed error correction on the data through checking the videos by the software named *Tracker* [38]. The error of each single-vehicle velocity is no more than 1 m/s (3.6 km/h).

The single-vehicle velocities of one typical platoon (named Platoon A) in the left lane are shown in Fig. 4*a*, and the corresponding distribution is in Fig. 4*b*. The data was collected during the morning peak, while the traffic demand was high and no stop-and-go waves can be found. It is clear that in Fig. 4*a*, the fluctuation of the velocities is quite small, and most of them are all between 11 and 14 m/s (40–50 km/h). Thus in Fig. 4*b*, only one obvious peak can be found. For Platoon A, there are 318 vehicles passing one location during 643 s, which means the averaged flow is 1780 veh/h. Here AV = 13.1 m/s and SDV = 1.18 m/s, so the averaged density is about $\rho = 37.7$ veh/km.

4.2 Case study B: combined empirical data from single-lane traffic

The second dataset is obtained from another empirical study. From July to September 2010, a large-scale reconstruction project was carried out in Nanjing Airport Highway: everyday a 5–10 km-long road section was under reconstruction, and one of the two lanes was temporarily closed. Thus, we obtain some empirical single-lane traffic flow data. For more details, we refer to earlier work [39, 40].

The traffic demand in this section was usually low. Therefore, when there is no other interference (e.g. a slow truck which became an obvious moving bottleneck), we cannot easily find long platoons on this single-lane highway, and only short platoons can be seen. Nevertheless, as mentioned in Section 3, short platoons with the same averaged density can be combined for calibration. For example, the statistical results of four small platoons are shown in Table 1. The averaged density of these platoons is all between 33 and 34 veh/km. Therefore, we combine all the single-vehicle velocities and obtain a new 88-vehicles platoon named Platoon B. The flow, AV and averaged density are 2044 veh/h, 61.2 km/h and $\rho = 33.4$ veh/km, respectively. So there are AV = 17.0 m/s and SDV = 1.56 m/s for Platoon B.

4.3 Case study C: single-lane traffic experiments

In the third dataset, the GPS data is obtained from the traffic experiment conducted in the suburban area of Hefei [41, 42]. During the experiment with 25 vehicles, the driver of the leading vehicle was required to control the velocity at certain pre-determined constant values, and other drivers were asked to drive their cars as they normally do, but overtaking is not allowed. All the drivers were requested to continue driving several rounds.

The road in the experiment is 3.2 km long, and we only use the single-vehicle velocity data collected in the section where $x = 500$–2500 m. This is because the influences of start-up at the beginning (0–500 m), and that of turning and decelerating at the end (2500–3200 m) are obvious. In many experimental rounds we find four of them have the similar averaged densities: they are all between 50.8 and 51.8 veh/km. Two platoons which belong to one round (Platoon 1 in Table 2) are collected at $x = 500$ m and $x = 1500$ m, while the other two platoons which belong to another round (Platoon 2 in Table 2) are collected at $x = 2000$ m and $x = 2500$ m. The four platoons are obtained from four locations. As we do in Section 4.2, we also combine them as a new 100-vehicles platoon named Platoon C. Here the flow, AV and averaged density are 1894 veh/h, 36.9 km/h and $\rho = 51.3$ veh/km, respectively. AV = 10.2 m/s and SDV = 0.96 m/s for Platoon C.

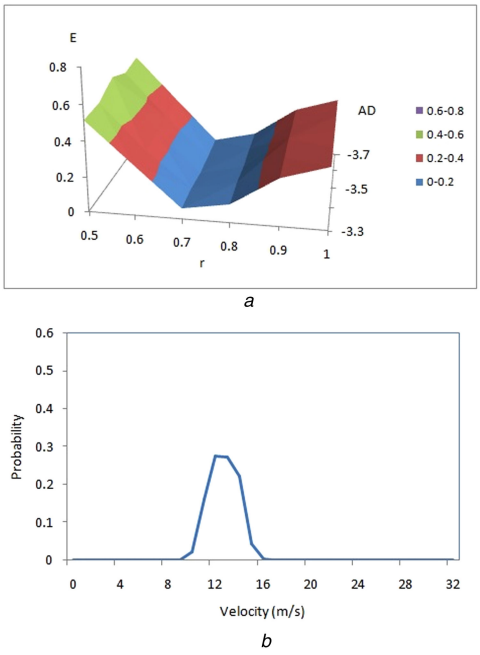
5 Calibration results and discussions

5.1 Results for case study A

At the density of $\rho = 37.7$ veh/km, we perform the simulations, as mentioned in Section 3. Fig. 5 shows the errors between the measurements and the simulations. In normal conditions, accelerations up to 3–4 m/s^2 are observed. Accelerations stronger than −4 m/s^2 will occur only in emergency conditions, rather than everyday driving. The explorative tests show that the model will

Table 2 Statistical information of the two platoons at four different locations, and their combination results (Platoon C) in our experimental GPS data. They were all collected on 19 January 2013

Platoon	Leading velocity, km/h	Location, m	Passing time, s	Flow, veh/h	AV, km/h	Averaged density, veh/km
1	41	500	47.6	1891	36.5	51.8
1	41	1500	45.4	1982	38.3	51.8
2	37	2000	48.5	1856	36.5	50.8
2	37	2500	48.6	1852	36.1	51.3
combination	/	/	190.1	1894	36.9	51.3



a

b

Fig. 5 *Calibration results for Platoon A*
(a) Errors under different simulation conditions, $\rho = 37.7$ veh/km, ***(b)*** Simulated velocity distribution when $(AD, r) = (-3.5, 0.7)$

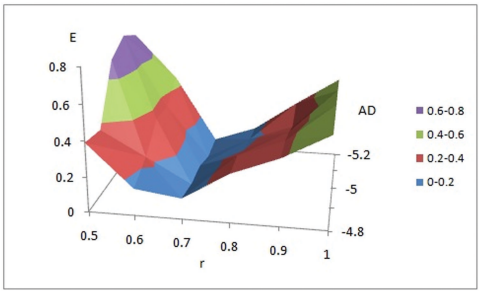


Fig. 6 *Calibration results for Platoon B: the errors under different simulation conditions, ρ = 33.4 veh/km*

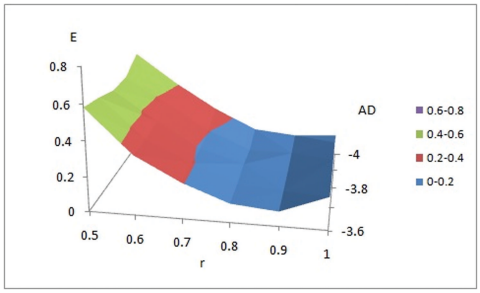


Fig. 7 *Calibration results for Platoon C: the errors under different simulation conditions, ρ = 51.3 veh/km*

yield non-realistic results when AD is not in the range of −3.7 to −3.3 m/s^2. Thus, calibration is limited to this range. It is found that in this range, the errors mainly depend on the value of r, rather than AD. Here we find the minimum value is $E_{A,\min} = 0.039$, at the point of $(AD, r) = (-3.5, 0.7)$. Therefore, $(-3.5, 0.7)$ should be the best parameters for the modelling of Platoon A. Besides, the velocity distribution is shown in Fig. 5*b*, which is similar to Fig. 4*b*. In all the three cases, the velocity distributions only have one peak.

5.2 Results for case study B

The results at $\rho = 33.4$ veh/km are shown in Fig. 6. By the comparison with Fig. 5, we can find that the shape is more complex, and the effect of r on E is larger. Here the minimum value is $E_{B,\min} = 0.036$, at the point of $(AD, r) = (-5.1, 0.7)$. It is interesting that for the two different types of platoons, the best values of r are just the same, while for AD they are different.

5.3 Results for case study C

We repeat the calibration process as in Sections 5.1 and 5.2, and get the results shown in Fig. 7. Here the minimum value is $E_{C,\min}$ 0.079, and the best parameters should be $(-3.9, 0.9)$. Now the differences between the empirical data and the experimental one seem large.

5.4 Discussions

By comparison we find there are three different (AD, r) results for the three cases: $(-3.5, 0.7)$, $(-5.1, 0.7)$ and $(-3.9, 0.9)$. These differences may be due to two main aspects:

i. *The driving strategies*: In our model, r represents the proportion of conservative strategy chosen by drivers. It is possible that in the conducted experiments (case C), all the drivers know they have no chance to overtake, so there is no need to hurry at all. For this situation, the model drive less aggressively, and should have larger r value ($r = 0.9$). On the contrary, in real life (cases A and B) drivers will not be so relaxed, and sometimes they want to save as much time as possible. This can lead to smaller r value ($r = 0.7$). Actually, the results are not due to the differences between real life and conducted experiments, but the impact of environments.
ii. *The potential tendencies*: For example, for the case B of single-lane traffic, |AD| is larger than that in cases A and C. However, it is not reasonable to simply make the conclusion that under single-lane conditions, drivers will prefer a larger AD. This difference may result from some other factors, e.g. in the single-lane section of Nanjing Airport Highway the traffic demand is always low. Even when one lane is temporarily closed, the flow value per lane is still lower than that in the multi-lane sections. Thus, drivers know in this section they can run much faster, and have the potential tendency to choose a much larger |AD| value.

6 Validation of the CA model

6.1 Methodology

Next we need to validate the parameters calibrated in the previous sections. First we check the validity of the parameters inside the platoon. We can divide one large platoon (e.g. Platoon A) into two small parts: Platoon A1 and Platoon A2. Platoon A1 can be selected for calibration and Platoon A2 can be used for validation. Here we set the function as (see (8)) This methodology shows whether the same calibration results will be found for two different datasets with the same conditions (same density, road stretch and

$$E_{t1}(\mathrm{AD}, r) = E_A(\rho_A, \mathrm{AD}, r) + E_{A1}(\rho_{A1}, \mathrm{AD}, r) + E_{A2}(\rho_{A2}, \mathrm{AD}, r) \quad (8)$$

Table 3 Holdout validation results of Platoon A

Platoon	(AD, r)	E_A (ρ_A, AD, r)	E_{A1} (ρ_{A1}, AD, r)	E_{A2} (ρ_{A2}, AD, r)	E_{t1}
A	(−3.5, 0.7)	0.039	0.041	0.262	0.342
A1	(−3.6, 0.7)	0.046	0.026	0.213	0.285
A2	(−3.5, 0.8)	0.097	0.167	0.057	0.321

Table 4 Cross-validation results of three platoons

Platoon	(AD, r)	E_A (ρ_A, AD, r)	E_B (ρ_B, AD, r)	E_C (ρ_C, AD, r)	E_{t2}
A	(−3.5, 0.7)	0.039	0.253	0.228	0.520
B	(−5.1, 0.7)	0.492	0.036	0.280	0.808
C	(−3.9, 0.9)	0.279	0.528	0.079	0.886

even platoon) This will result in three different optimal parameter sets: one for E_A, one for E_{A1} and one for E_{A2}. We can compute the combined error E_{t1} for each of these three parameter sets for each of the platoons. Obviously, the error is minimum for the optimal parameter set for that dataset. The interesting part is to find out how far off the errors are if the other parameter sets are being used. This validation technique belongs to 'holdout validation', as introduced by Treiber and Kesting [43].

Another option is to compare the parameter sets for the different datasets, thereby checking the consistency of the parameter sets for different conditions. We do something similar: we check the error of the model with the parameters optimised for one parameter set applied to another parameter set. Moreover, we compute a total error for a parameter set, consisting of the sum of the errors in the three datasets

$$E_{t2}(\mathrm{AD}, r) = E_A(\rho_A, \mathrm{AD}, r) + E_B(\rho_B, \mathrm{AD}, r) + E_C(\rho_C, \mathrm{AD}, r) \quad (9)$$

This validation technique belongs to 'cross-validation', which can make maximal use of the available data.

6.2 Results

The results of holdout validation are shown in Table 3. Here Platoon A1 corresponds to the earlier 159 vehicles of Platoon A, while Platoon A2 corresponds to the later 159 ones. Their flows, velocities and densities can be calculated, respectively: here for A1, AV = 12.8 m/s, SDV = 1.24 m/s and ρ = 38.8 veh/km. The best parameters are (−3.6, 0.7). For A2, AV = 13.5 m/s, SDV = 1.00 m/s and ρ = 37.0 veh/km. The best parameters are (−3.5, 0.8). As expected, the fitted values fit best on the parts they are optimised for, confirming the calibration procedure. Overall, all the three parameters of (AD, r) perform well. Therefore, we conclude that a parameter set of (−3.5, 0.7) for Platoon A is correct for the whole platoon as well as for each of its parts.

Then the results of cross-validation are shown in Table 4. Here again, the parameters fitted the situations they are calibrated for (the diagonal of the table), are the best. The numbers off the diagonal are much higher, which is contrary to the holdout validation. Here we can say the parameters of (−3.5, 0.7) for Platoon A are a little better than the others. Nevertheless, none of them can fit *all* three datasets very well. Comparing with Table 3, if we set 0.5 as one critical value for E, we can find in Table 3 they are all smaller than 0.5, while in Table 4 they are all larger than it. These results indicate that it is impossible to simulate all the traffic flow phenomena by one model with fixed parameters. This situation could also be found in some other field, including pedestrian flow [44].

7 Conclusion

In this paper, we show how to do calibration and validation for CA models. Due to the existence of randomisation probability, some frequently used methodologies for the calibration and validation of car-following models are not possible for CA models. Nevertheless, we find a new macroscopic way. We define the objective function E, which has the form like relative RMSE of the AV and SDV. Two important parameters in our CA model, the AD and the fraction of aggressive drivers are calibrated, and for the minimum error we can find the best combination of (AD, r). This process is easy to understand and follow.

Then, three different types of traffic flow data are used for calibration: the single-vehicle data of empirical traffic in Nanjing Airport Highway, including the situations of single-lane traffic and multi-lane traffic, and the GPS data obtained from traffic experiments conducted in Hefei. For the small platoons with the same densities, we can combine them together. It is not strange that we find different best parameters for different datasets.

Here the differences mainly result from the driving strategies and the potential tendencies in different environments. Finally, after the validation we find none can fit all of them well at the same time, which implies it may be impossible to simulate all the traffic flow phenomena with one model with fixed parameters.

Although a complete process of calibration and validation has been presented in this paper, this is just a beginning. For more complex CA models, more parameters need to be calibrated, including some common factors in all the models (vehicle length, time step etc.) and some specific parameters in our model (e.g. An). On the other hand, some complex scenarios need to be considered, including the lane-changing behaviours, and the influence of heterogeneity of traffic flow, especially the proportion of large trucks. These works will be done in the future.

8 Acknowledgments

This work was funded by the National Natural Science Foundation of China (grant nos. 11302022, 11422221, 71371175, 71621001), the Natural Science Foundation of Jiangsu Province (grant no. BK20150619) and the Netherlands Organization of Scientific Research (NWO).

9 References

[1] Greenshields, B.D.: 'A study of traffic capacity', *Proc. Highway Res. Board*, 1934, **14**, pp. 448–477
[2] May, A.D.: '*Traffic flow fundamentals*' (Prentice Hall, New York, 1989)
[3] Daganzo, C.F.: '*Fundamentals of transportation and traffic operations*' (Pergamon Press, Oxford, 1997)
[4] Wolfram, S.: 'Statistical mechanics of cellular automata', *Rev. Mod. Phys.*, 1983, **55**, pp. 601–644
[5] Nagel, K., Schreckenberg, M.: 'A cellular automaton model for freeway traffic', *J. Phys. I*, 1992, **2**, pp. 2221–2229
[6] Kerner, B.S.: '*The physics of traffic*' (Springer, Heidelberg, 2004)
[7] Kerner, B.S.: 'Three-phase traffic theory and highway capacity', *Physica A*, 2004, **333**, pp. 379–440
[8] Kerner, B.S.: '*Introduction to modern traffic flow theory and control: the long road to three-phase traffic theory*' (Springer, Berlin, 2009)
[9] Kerner, B.S.: 'Criticism of generally accepted fundamentals and methodologies of traffic and transportation theory: a brief review', *Physica A*, 2013, **392**, pp. 5261–5282
[10] Kerner, B.S., Klenov, S.L.: 'A microscopic model for phase transitions in traffic flow', *J. Phys. A*, 2002, **35**, pp. L31–L43

[11] Kerner, B.S., Klenov, S.L., Wolf, D.E.: 'Cellular automata approach to three-phase traffic theory', *J. Phys. A*, 2002, **35**, pp. 9971–10013
[12] Kerner, B.S., Klenov, S.L., Wolf, D.E.: 'Microscopic theory of spatial–temporal congested traffic patterns at highway bottlenecks', *Phys. Rev. E*, 2003, **68**, p. 036130
[13] Kerner, B.S., Klenov, S.L., Schreckenberg, M.: 'Simple cellular automaton model for traffic breakdown, highway capacity, and synchronized flow', *Phys. Rev. E*, 2011, **84**, p. 046110
[14] Ossen, S., Hoogendoorn, S.P.: 'Validity of trajectory-based calibration approach of car-following models in presence of measurement errors', *Transp. Res. Rec.*, 2008, **2088**, pp. 117–125
[15] Ossen, S., Hoogendoorn, S.P.: 'Heterogeneity in car-following behavior: theory and empirics', *Transp. Res. C*, 2011, **19**, pp. 182–195
[16] Kesting, A., Treiber, M.: 'Calibrating car-following models by using trajectory data: methodological study', *Transp. Res. Rec.*, 2008, **2088**, pp. 148–156
[17] Treiber, M., Kesting, A.: 'Validation of traffic flow models with respect to the spatiotemporal evolution of congested traffic patterns', *Transp. Res. C*, 2012, **21**, pp. 31–41
[18] Ciuffo, B., Punzo, V., Montanino, M.: 'Global sensitivity analysis techniques to simplify the calibration of traffic simulation models. Methodology and application to the IDM car-following model', *IET Intel. Trans. Sys.*, 2014, **8**, pp. 479–489
[19] Punzo, V., Montanino, M., Ciuffo, B.: 'Do we really need to calibrate all the parameters? Variance-based sensitivity analysis to simplify microscopic traffic flow models', *IEEE Trans. Intel. Transp. Sys.*, 2015, **16**, pp. 184–193
[20] Sugiyama, Y., Fukui, M., Kikuchi, M., *et al.*: 'Traffic jams without bottlenecks – experimental evidence for the physical mechanism of the formation of a jam', *New J. Phys.*, 2008, **10**, p. 033001
[21] Nakayama, A., Fukui, M., Kikuchi, M., *et al.*: 'Metastability in the formation of an experimental traffic jam', *New J. Phys.*, 2009, **11**, p. 083025
[22] Tadaki, S., Kikuchi, M., Fukui, M., *et al.*: 'Phase transition in traffic jam experiment on a circuit', *New J. Phys.*, 2013, **15**, p. 103034
[23] Jiang, R., Wu, Q.S.: 'Cellular automata models for synchronized traffic flow', *J. Phys. A*, 2003, **36**, pp. 381–390
[24] Jiang, R., Wu, Q.S.: 'Spatial–temporal patterns at an isolated on-ramp in a new cellular automata model based on three-phase traffic theory', *J. Phys. A*, 2004, **37**, pp. 8197–8213
[25] Jiang, R., Wu, Q.S.: 'First order phase transition from free flow to synchronized flow in a cellular automata model', *Eur. Phys. J. B*, 2005, **46**, pp. 581–584
[26] Gao, K., Jiang, R., Hu, S.X., *et al.*: 'Cellular-automaton model with velocity adaptation in the framework of Kerner's three-phase traffic theory', *Phys. Rev. E*, 2007, **76**, p. 026105
[27] Gao, K., Jiang, R., Wang, B.H., *et al.*: 'Discontinuous transition from free flow to synchronized flow induced by short-range interaction between vehicles in a three-phase traffic flow model', *Phys. A*, 2009, **388**, pp. 3233–3243
[28] Jin, C.J., Wang, W., Jiang, R., *et al.*: 'On the first-order phase transition in a cellular automaton traffic flow model without a slow-to-start effect', *J. Stat. Mech.*, 2010, **2010**, p. P03018
[29] Jin, C.J., Wang, W., Gao, K., *et al.*: 'Effect of acceleration threshold on the phase transition in a cellular automaton traffic flow model', *Chin. Phys. B*, 2011, **20**, p. 064501
[30] Jin, C.J., Wang, W.: 'The influence of nonmonotonic synchronized flow branch in a cellular automaton traffic flow model', *Phys. A*, 2011, **390**, pp. 4184–4191
[31] Jin, C.J., Wang, W., Jiang, R.: 'On the modeling of synchronized flow in cellular automaton models', *Chin. Phys. B*, 2014, **23**, p. 024501
[32] Feliciani, C., Nishinari, K.: 'Cellular automata model to simulate the behavior of high density crowd and validation by experimental data', *Phys. A*, 2016, **451**, pp. 135–148
[33] Laval, J.A.: 'Hysteresis in traffic flow revisited: an improved measurement method', *Transp. Res. B*, 2011, **45**, pp. 385–391
[34] Hoogendoorn, S.P.: 'Unified approach to estimating free speed distributions', *Transp. Res. B*, 2005, **39**, pp. 709–727
[35] Daamen, W., Christine Buisson, C., Hoogendoorn, S.P.: '*Traffic simulation and data: validation methods and applications*' (CRC Press, Boca Raton, 2014)
[36] Van Hinsbergen, C.P.I.J., Schakel, W.J., Knoop, V.L., *et al.*: 'A general framework for calibrating and comparing car-following models', *Transportmetrica A*, 2015, **11**, pp. 420–440
[37] Jin, C.J., Wang, W., Jiang, R., *et al.*: 'Understanding the structure of hyper-congested traffic from empirical and experimental evidences', *Transp. Res. C*, 2015, **60**, pp. 324–338
[38] Tracker, Version 4.90. Available at http://physlets.org/tracker/
[39] Jin, C.J., Wang, W., Jiang, R., *et al.*: 'Spontaneous phase transition from free flow to synchronized flow in traffic on a single-lane highway', *Phys. Rev. E*, 2013, **87**, p. 012815
[40] Jin, C.J., Wang, W., Jiang, R., *et al.*: 'An empirical study of phase transitions from synchronized flow to jams on a single-lane highway', *J. Phys. A*, 2014, **47**, p. 125104
[41] Jiang, R., Hu, M.B., Zhang, H.M., *et al.*: 'Traffic experiment reveals the nature of car-following', *PLoS ONE*, 2014, **9, p**. e94351
[42] Jiang, R., Hu, M.B., Zhang, H.M., *et al.*: 'On some experimental features of car-following behavior and how to model them', *Transp. Res. B*, 2015, **80**, pp. 338–354
[43] Treiber, M., Kesting, A.: '*Traffic flow dynamics: data, models and simulation*' (Springer, Berlin, 2013)
[44] Duives, D.C., Sparnaaij, M., Knoop, V.L., *et al.*: 'Multi-objective calibration framework for pedestrian simulation models: study on the effect of movement base cases, metrics and density levels'. Transportation Research Board, 2018