Calibrating Fundamental Diagram based on Two Types of Speed-Density Data

Jin Zhang, Xiaobo Qu

Griffith School of Engineering, Griffith University, Gold Coast, QLD, 4222, Australia.

Abstract

Speed – density relationship, which is usually referred to as the fundamental diagram, has been considered as the foundation of the traffic flow theory and transportation engineering. It has been long believed that single regime models could not well represent all traffic states ranging from free flow conditions to jam conditions. Qu et al. (2015; Transportation Research Part B, 73, 91-102) point out that the inaccuracy is not caused solely by their functional forms, but also by sample selection bias, and then apply a novel calibration approach to address the sample selection bias problem. In this technical note, we theoretically investigate how the functional forms of different models cope with the sample selection bias. A novel sampling approach is used to select uniform observational data (which would have the same characteristics with experimental data). With the selected uniform data, we can analyse the impact of functional forms themselves by eliminating the sample selection bias. As most speed – density models are derived from their corresponding car following models, we can use the aggregated macroscopic speed density data to evaluate the performance of microscopic car following models. In this way, we can establish the connection between macroscopic and microscopic traffic flow models.

1. Introduction

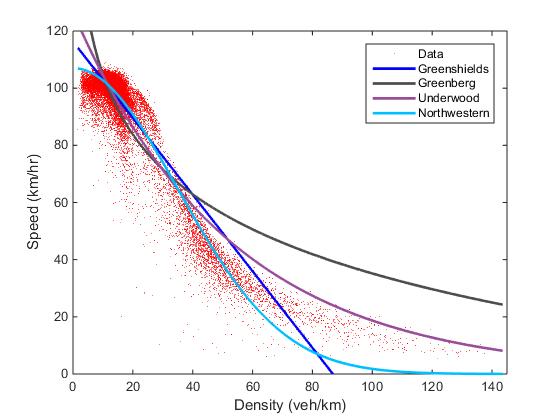
It has been 80 years since Greenshields (1935) proposed a parabolic function by suggesting a linear speed-density curve for the speed-density relationship. Numerous studies has been done to improve such an over-simplified relationship (e.g. Greenberg, 1959; Underwood, 1961; Newell, 1961; Northwestern, 1967; NoKerner and Konhäuser, 1994, Del Castillo, 1995a&b; MacNicolas, 2008; Wang et al., 2011). These models have been derived from different car following theories which are based on the follow-the-leader concept, in which rules how drivers follow their leading vehicles (Zhang and Kim, 2005). Many researchers believe that single-regime models cannot represent the empirical data consistently well both in the free-flow regime and the congested regime (Edie, 1961; Drake et al, 1967). As such, multiple regime models were proposed by using various different curves to represent distinct traffic states. Although the performance of multiple-regime model performs well in terms of accuracy, it has been criticized due to the lack of mathematical elegance. Table 1 lists four well-known single regime models with two parameters which will be discussed in the following sections.

**Table 1:** Four single-regime speed-density models

|  |  |  |
| --- | --- | --- |
| Single-regime models | Function | Parameters |
| Greenshields (1935) |  | *vf, kj* |
| Greenberg (1959) |  | *vf, kj* |
| Underwood (1961) |  | *vf, kj* |
| Northwestern (1967) |  | *vf, ko* |

We use speed density data collected on the Georgia State Route 400 (referred to as GA400 dataset hereafter). The calibration results by the traditional least square methods are presented in Figure 1. As all the functional forms are derived from car following models, researchers usually use the performances of calibrated equations to assess their corresponding car following models (Qu et al., 2015). In this regard, Greenberg was outperformed by all other models, especially at mid to high traffic conditions.

**Figure 1:** Performance of four single-regime speed-density models in Table 1



According to Qu et al. (2015), the inaccuracy of single-regime models is not caused solely by their functional forms, but also by the sample selection bias. Two types of data are defined: observational data and experimental data. Observational data has been measured or surveyed as it is. In contrast, experimental data is obtainable from measurements or survey which is designed by researchers. As a result, the observational data might be highly skewed and experimental data can be uniform. The data we use includes 47,815 observations of speeds and densities, 86.3% (41,259) of which refer to free flow conditions (< 20 veh/km). Evidently, most, if not all, freeway traffic data is a traditional observational data. Therefore, the calibrated models are likely to be dominated by free flow conditions, which result in poor performances for congested traffic states. In other words, the performances of calibrated models are determined not only by the function forms but also the data. Theoretically, the performances of microscopic car following models can be assessed by the performance of functional forms of the speed density relationship in a macroscopic way. It is possible that the functional form of Greenberg model is good. However, due to the feature of speed density data, the calibrated model, which is dominated by free flow conditions, does not perform as what it should. In this regard, only speed density equations calibrated from experimental data can be used to macroscopically assess the performance of their corresponding car following models.

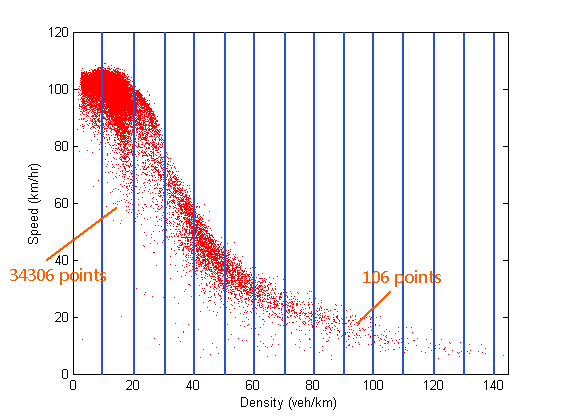
In this research, a uniform sampling procedure is proposed to generate experimental data from the collected data. Based on the generated experimental data, the performances of four two-parameter models are compared. It is found that Greenberg model performs very well and outperforms the Greenshields model and Northwestern model. A theoretical investigation is also presented to discuss why sample selection bias has a higher impact to Greenberg model compared the other three models.

2. METHDOLOGY

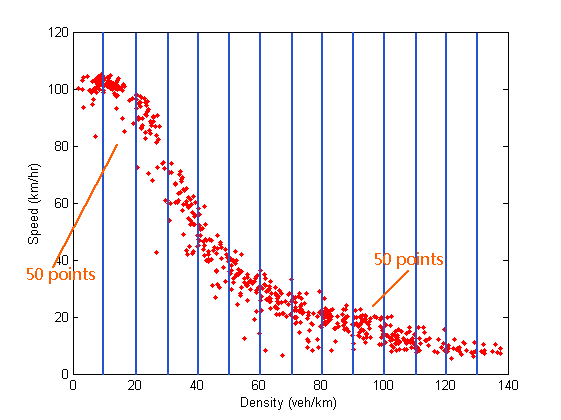
2.1 Uniform data sampling

In order to assess the speed-density functional forms, we intend to generate experimental data from the observational database. We first divide the observational data into several ranges based on their density, and then generate uniform number of data points for each group based on the observations. As can be seen in Figure 2, 50 data points for each range are generated from the uniform sampling process. The total number of the original observations in each range has been listed in Table 2 below,

**Figure 2(a):** Original data distribution



**Figure 2(b):** Experimental data distribution



**Table 2:** Number of observations in fourteen ranges

|  |  |  |  |
| --- | --- | --- | --- |
| Density Range  (veh/km) | Number of  observations (n) | Density Range  (veh/km) | Number of  observations (n) |
| 0-10 | 6989 | 70-80 | 378 |
| 10-20 | 34306 | 80-90 | 218 |
| 20-30 | 1824 | 90-100 | 150 |
| 30-40 | 1145 | 100-110 | 53 |
| 40-50 | 1557 | 110-120 | 23 |
| 50-60 | 800 | 120-130 | 22 |
| 60-70 | 378 | 130-140 | 9 |

2.2 Procedure

In this study, we adopt the following procedure to generate the experimental data which follow uniform distribution.

Suppose we have the *m* speed - density samples: .

Step 1: Rank the samples in terms of their densities. We thus have



where and  is the corresponding speed under traffic density of .

Step 2: Determine the total number of intervals through the whole database as considered each interval with a span of 10. That is



where  returns the value of a number rounded upwards to the nearest integer.

Step 3: Find the ranges of each intervals as



and count the number of points in each interval.

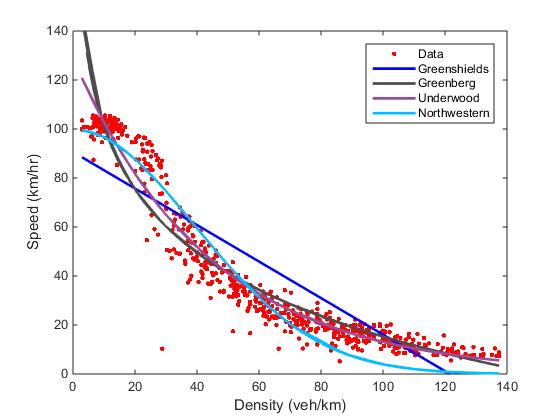
Step 4: Using Monte Carlo theory, generate a uniform random variable percentile between 0 and 1, and then find its accurate value of density according to its percentile in each interval.

Step 5: Repeat step 4 until you have 50 points in each interval. Then, find their corresponding speeds under each traffic density to create the new uniform data sampling.

3. Results

The four two-parameter speed-density models are listed in Table 1: Greenshields model (Greenshields et al., 1935), Greenberg model (Greenberg, 1959), Underwood model (Underwood, 1961) and Northwestern model (Drake et al., 1967) have been calibrated by least square method and the results are shown in Figure 3. The red dots are random uniform experimental data; the solid line represents the models calibrated by LSM. By contrast, quantitative comparisons will be reported in the next section.

**Figure 3:** Performances of two parameter models under experimental database



From the graph, all the models perform better than those in Figure 1, especially for the Greenberg Model which performs not well in Figure 1 but almost as well as others in Figure 3.

3.1 Discussion

As the four two parameter speed-density models have been calibrated under the uniform data sample, model validation is provided to directly show the validations regarding all the models. In order to have a clear view of the results, the results have been divided into fourteen ranges by the value of densities.

3.1.1. Mean square error (MSE)

The mean squared error (MSE) of a sample measures the average of the squares of the difference between the predicted value and the empirical data (Lehmann and George, 1998). The formula is shown below,



where is the value predicted by the model,  is the actual data, and  *n* is the number of observations in a range. The results are present in Table 3.

**Table 3:** Mean square error of six models by LSM

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| k (veh/km) | Greenshields  (1935) | Greenberg  (1959) | Underwood  (1961) | Northwestern  (1961) |
| 0-10 | 265.474 | 307.088 | 99.220 | 22.989 |
| 10-20 | 369.920 | 132.153 | 57.231 | 47.472 |
| 20-30 | 280.740 | 415.197 | 226.754 | 85.437 |
| 30-40 | 126.713 | 82.275 | 72.785 | 155.886 |
| 40-50 | 256.638 | 43.389 | 52.888 | 144.930 |
| 50-60 | 292.754 | 34.617 | 28.688 | 41.153 |
| 60-70 | 335.375 | 77.412 | 55.651 | 39.470 |
| 70-80 | 191.027 | 34.098 | 16.992 | 39.676 |
| 80-90 | 94.518 | 21.700 | 12.506 | 85.120 |
| 90-100 | 41.505 | 17.757 | 13.738 | 111.472 |
| 100-110 | 8.506 | 7.958 | 6.988 | 99.106 |
| 110-120 | 41.915 | 9.912 | 9.491 | 89.859 |
| 120-130 | 130.078 | 4.954 | 4.179 | 72.365 |
| 130-140 | 338.263 | 20.167 | 8.611 | 71.676 |

As can be seen in the above Table 3 and Figure 4, for the density from 20-50 veh/km is the hardest part for all the models to performs, as they all have larger errors in this ranges. Among all the four two-parameter models, the Underwood Model performs the best. The largest MSE is 226 with the density from 20-30 veh/km, which doubles the second largest error. If the density is more than 70 veh/km, the errors are pretty small as below 20. As for Greenberg Model, it has large errors in low traffic flow states (k<40 veh/km), other performances are acceptable. Compare to these two models, the errors of Northwestern model are bigger, but it still more accurate than Greenshield Model.

**Figure 4:** Mean square errors of six models by LSM

3.1.2. Ranking

Table 4 calculate the overall three types of errors of the six speed-density models. From the table, we can see that errors of Underwood Model are smaller than other models, which mean this calibrated model performs the experimental data best, followed by Greenberg Model, Northwestern Model, and Greenshields Model. These models are all calibrated under the generated experimental data. It should be pointed out that the sample selection bias significantly affects the performance of Greenberg model, evidenced by the difference between results from observational data and experimental data.

**Table 4:** Model validation of six speed-density models

|  |  |  |  |
| --- | --- | --- | --- |
| Single-regime models | Relative error  (per point) | Mean square error | Root-mean-square error |
| Greenshields (1935) | 58.52% | 198.102 | 14.075 |
| Greenberg (1959) | 22.88% | 86.334 | 9.292 |
| Underwood (1961) | 17.43% | 47.511 | 6.896 |
| Northwestern (1967) | 42.41% | 79.044 | 8.891 |

5. Theoretical investigation

The coefficient of determination (R2) is a measurement number that could indicate how well data fit a statistical model, which formula is (Glantz and Slinker, 1990):



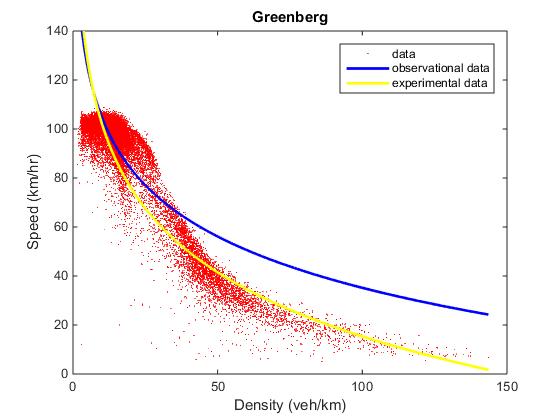
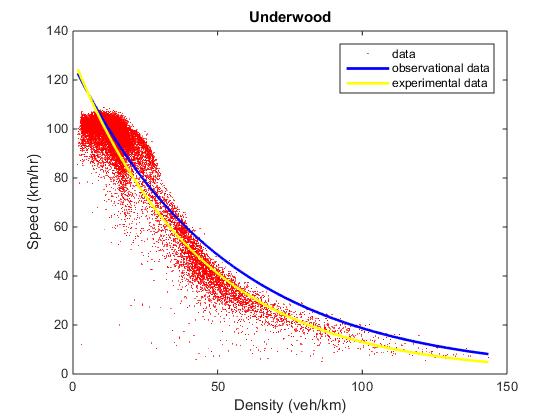
4.1 Linear two-parameter speed-density models

The three nonlinear two-parameter speed-density models can be linearized as shown in Table 5 (See appendix A). The parameter *k* and *v* has been transformed as, and during the linearization. The transformed density and speed are denoted as  and .

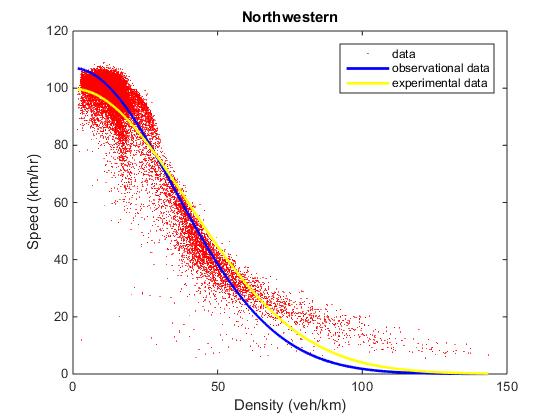
**Table 5:** Linear Function of two-parameter speed-density models

|  |  |  |
| --- | --- | --- |
| Single-regime models | Function | Linear Function |
| Greenberg (1959) |  |  |
| Underwood (1961) |  |  |
| Northwestern (1967) |  |  |

Figure 5 depicts the comparisons between the results calibrated by observational data and experimental data. The red dots are original observational data; the blue line represents the model calibrated by LSM based on the observational data; and the yellow line represents the model calibrated based on the generated experimental data.

**Figure 5 (a):** Performance of Greenberg Model **(b):** Performance of Underwood Model 

**Figure 5(c):** Performance of Northwestern Model



4.2 Results of *R2* of two-parameter speed-density models

As proven by Qu et al. (2015), the sample selection bias does not affect linear relationship. In other words, the performance of these models calibrated based on the observational data is determined by the coefficient of determination of the transformed parameters. In this regard, we calculate the the coefficient of determination (*R2*), and the results are shown in Table 6.

**Table 6:** *R2* of two-parameter speed-density models

|  |  |  |
| --- | --- | --- |
| Single-regime models | R2 | Average of the five ranges of largest MSE |
| Greenberg (1959) | 0.6864 | 176.7 |
| Underwood (1961) | 0.9020 | 102.3 |
| Northwestern (1967) | 0.8176 | 120.3 |

Evidently, the coefficient of determination of the sample and the performance of the calibrated function has a negative relationship. In other words, if the relationship between the transformed density and speed is linear or generally linear, the performances of these calibrated models are not significantly affected by sample selection bias. This is the reason why Greenberg model perform well based on the generated experimental data and poorly based on the observational data.

5. Conclusion

This research is an extension of Qu et al. (2015), in which paper the authors point out that the inaccuracy of single regime speed-density models is not caused solely by their functional forms, but also by sample selection bias. In this research, we theoretically investigate how the functional forms of different models cope with the sample selection bias. A novel sampling approach is used to select uniform observational data (which would have the same characteristics with experimental data). With the selected uniform data, we can analyse the impact of functional forms themselves by eliminating the sample selection bias. As most speed – density models are derived from their corresponding car following models, we can use the aggregated macroscopic speed density data to evaluate the performance of microscopic car following models. In this way, we can establish the connection between macroscopic and microscopic traffic flow models.

**References**

Del Castillo, J M and Benítez, F G (1995a) On the functional form of the speed-density relationship – I: General theory Transportation Research Part B 29 (5), 373–389

Del Castillo, J M and Benítez, F G (1995b) On the functional form of the speed-density relationship – II: Empirical investigation Transportation Research Part B 29 (5), 391–406

Drake, J S, Schofer, J L and May, A D (1967) A statistical analysis of speed–density hypotheses Highway Research Record 154, 112–117

Edie, L C (1961) Car-Following and Steady State Theory for Non-Congested Traffic Operations Research 9 (1), 66-76

Greenshields, B D, Bibbins, J R, Channing, W S and Miller, H H (1935) A STUDY OF TRAFFIC CAPACITY Highway Research Board Proceedings 14, 448–477

Greenberg, H (1959) An Analysis of Traffic Flow Operations Research 7, 255-275

Lehmann, E L and George, C (1998) Theory of Point Estimation (2nd ed.) New York: Springer

MacNicholas, M J (2008) A simple and pragmatic representation of traffic flow In: Symposium on The Fundamental Diagram: 75 years, Transportation Introduction Research Board, Woods Hole, MA

Newell, G F (1961) Nonlinear Effects in the Dynamics of Car Following Operations Research 9 (2), 209-229

Underwood, R T (1961) Speed, volume, and density relationship: quality and theory of traffic flow Yale Bureau of Highway Traffic, 141–188

Qu, X, Wang, S and Zhang, J (2015) On the fundamental diagram for freeway traffic: A novel calibration approach for single-regime models. Transportation Research Part B 73 (2015), 91-102

Wang, H, Li, H, Chen, Q and Ni, D (2011) Logistic modeling of the equilibrium speed–density relationship Transportation Research Part A 45 (2011), 554-566

Zhang, H, M and Kim, T (2005) A car-following theory for multiphase vehicular traffic flow. Transportation Research Part B 39 (2005), 385-399

Appendix A

Linear functions of two-parameter models

For Greenberg Model, as its equation is



Linear it by ,





For the Underwood (1961) model,



Converting it to linear function,



Linear it by





For the Northwestern Model (1967), the formula is



Converting it to linear function,



Linear it by





