

ESTIMATING SOURCE OF TRAFFIC CONGESTION FROM VEHICLE TRAJECTORY DATA USING HIERARCHICAL BAYESIAN MODEL

K. HAYASHI^{a*}, and T. SEO^b

^a School of Environment and Society, Tokyo Institute of Technology, Japan

Email: hayashi.k.bi@m.titech.ac.jp

^b School of Environment and Society, Tokyo Institute of Technology, Japan

Email: seo.t.aa@m.titech.ac.jp

ABSTRACT

Traffic congestion is the result of complex interactions between vehicles. Therefore, understanding driving behavior is essential to understanding the mechanisms behind congestion. The fundamental diagram (FD), which describes the relationship between speed and spacing (or equivalently, flow and density), is crucial for understanding driving behavior and congestion. The FD typically depends on road and vehicle characteristics, such as road geometry and driver characteristics. However, it is not clear which factor is dominant under certain conditions, especially at complicated bottlenecks such as sags. This makes it difficult to model and alleviate traffic congestion. This study aims to identify the sources of congestion by estimating an accurate FD that distinguishes the effects of both vehicle and road characteristics. We propose a novel FD model based on a hierarchical Bayesian framework to capture the influence of these factors and their interactions. This model enables the estimation of a more accurate FD that accounts for the effects of both road and vehicle. We estimated the proposed model using actual vehicle trajectory dataset on highway. The results confirm that the model would represent the sources of congestion, including specific road features and specific vehicle types. For example, at a certain sag, most of the vehicles contribute to the capacity reduction, but the effects of few vehicles were significantly high.

Keywords: (up to five)

Traffic flow theory, Newell's car following model, Hierarchical Bayesian model, Complete vehicle trajectory data

1. INTRODUCTION

Traffic congestion arises from the intricate interplay between vehicles on the road. Thus, investigating driving behavior is important to unraveling the mechanisms that cause congestion. The fundamental diagram (FD) is a macro/mesoscopic law of traffic flow. It describes the critical relationship between speed and spacing (or, equivalently, traffic flow and density) depending on vehicle or road characteristics. The FD allows us to assess traffic capacity and the prevailing traffic conditions.

The FD is influenced by the specific road section and the heterogeneity of vehicles. Road features such as sags or curves can affect the FD. Furthermore, the vehicle characteristics influence the FD due to the inherent differences in driving behavior. However, it is not clear which factor is dominant under certain conditions, especially at complicated bottlenecks such as sags.

Various methods have been proposed to estimate and investigate empirical FDs from actual traffic data. Chiabaut et al. (2009) developed a technique for FD estimation using data from vehicle trajectories, enabling detailed FD estimation. Taylor et al. (2015) proposed a method for estimating the car following model using a dynamic time-warping approach, making it possible to estimate different FDs for each vehicle. Seo et al. (2019) proposed a method for FD estimation using probe vehicle data, enabling FD estimation over a wider area.

However, most existing methods for FD estimation do not capture the driver and road section dependent effects separately. To our knowledge, one exception is Yoshida et al. (2017), whose FD depended on both road sections and vehicles. A limitation of this method is its inability to perform a detailed analysis of individual sections, likely due to the inherent difficulty of estimating complex models.

Advancements in technology are overcoming this problem in estimating complicated models. Bayesian modeling offers a solution. For example, Zhang et al. (2024) analyzed driving behavior by estimating car following model parameters by hierarchical Bayesian modeling. Their approach caught the vehicle effect of each parameter through Bayesian estimation. In this way, the hierarchical Bayesian framework is effective for detailed analysis.

This study proposes a novel FD model based on a hierarchical Bayesian framework to capture the vehicle- and section-dependent effects in a distinguishable way. Then, we validate the proposed model by estimating it using actual vehicle trajectory data. Finally, we interpret the impact of congestion on different sections and vehicles based on the results.

2. METHOD

2.1 Model formulation

In this study, Greenshields (1935) FD represents the speed-spacing relation as:

$$v = u \left(1 - \frac{\delta}{s} \right) \quad (1)$$

where u is the free flow speed, δ is the jam spacing, v is the vehicle speed, and s is the spacing. Estimating FD depending on section and vehicle identifies estimating u and δ for each section on the road and each vehicle simultaneously.

The model incorporates section-specific and vehicle-specific free flow speed and jam spacing represented by $u^{i,j}$ and $\delta^{i,j}$, respectively, when j represents the section where the vehicle is located, and i represents the vehicle number. These parameters are decomposed into four components: an average $(\bar{u}, \bar{\delta})$ where the influence of sections and vehicles is removed, a section effect $(\Delta u_X^j, \Delta \delta_X^j)$ that represents the impact of road feature such as sags, a vehicle effect $(\Delta u_N^i, \Delta \delta_N^i)$ that captures the difference of driving across vehicles, and an interaction effect $(\Delta u_{NX}^{i,j}, \Delta \delta_{NX}^{i,j})$ that captures the combined influence of a specific section and vehicle. Notably, all these components are estimated simultaneously through hierarchical Bayesian modeling, allowing for their subsequent separation and individual analysis.

The following paragraphs show details of model formulation. Spacing and speed of vehicle i at section j are denoted as $s^{i,j,k}$ and $v^{i,j,k}$ where k is the data index in i of j . To account for noise, we assume $v^{i,j,k}$ follows a normal distribution (denoted by $N(\mu, \sigma^2)$, whose mean is μ , and the variance is σ^2). μ represents the expected speed from data. σ^2 means the total error of observation noise and behavioral noise. The probability distribution of the observed speed is shown as:

$$v^{i,j,k} \sim N \left(u^{i,j} \left(1 - \frac{\delta^{i,j}}{s^{i,j,k}} \right), \sigma^2 \right) \quad (2)$$

Half-Cauchy distribution (HC) as the prior for σ^2 due to its suitability for variance is used, according to Gelman (2006). HC denoted by $C^+(\beta)$, represents the positive side of the Cauchy distribution with a scale parameter β . σ^2 is distributed as:

$$\sigma \sim C^+(100) \quad (3)$$

where 100 represents weakly informative prior parameters. A weakly informative prior parameter has no meaning. A large value like 100 minimizes the influence of the prior on the posterior distribution, allowing the data to drive the estimation primarily. Weakly informative prior parameters do not need to have values as long as they are large enough.

$u^{i,j}$ is the free flow speed for each section and each section. $u^{i,j}$ is the linear sum as:

$$u^{i,j} = \bar{u} + \Delta u_X^j + \Delta u_N^i + \Delta u_{NX}^{i,j} \quad (4)$$

For estimation stability, Δu_X^j , Δu_N^i , and $\Delta u_{NX}^{i,j}$ are each subject to constraints such that the sum of the sections or vehicles is zero.

The first term in Eq (4) $\bar{u}$ means average free flow speed independent of section and vehicle. To account for uncertainty in this parameter, $\bar{u}$ is assigned a normal distribution prior, denoted as:

$$\bar{u} \sim N(18, 100) \quad (5)$$

where 18 and 100 represent weakly informative prior parameters.

The second term in Eq (4) Δu_X^j means the deviation of the free flow speed from average in j . To improve the efficiency of estimation, Δu_X^j is employed a Non-centered Parameterization according to Betancourt (2015). This reparameterization is achieved as:

$$\Delta u_X^j = A_X e_X^j \quad (6)$$

where A_X means the scale of deviation of the free flow speed by section and e_X^j means the ratio of deviation of the free flow speed by j . A_X is assigned HC, the preferred distribution. A_X is denoted as:

$$A_X \sim C^+(500) \quad (7)$$

where 500 represents a weakly informative prior parameter. e_X^j means captures the impact on free flow speed in j . To ensure stable parameter estimation during the hierarchical modeling process, the sum of e_X^j is constrained to be zero. This constraint is expressed as:

$$\sum_j e_X^j = 0 \quad (8)$$

Given this constraint and the requirement for e_X^j to handle noise, they are assigned normal distribution priors with a mean of zero and a variance of one. This is denoted as:

$$e_X^j \sim N(0, 1) \quad (9)$$

where 0 and 1 are standard normal distribution parameters for Non-centered Parameterization.

The third term in Eq (4) Δu_N^i means the deviation of the free flow speed from average for i . Δu_N^i is reparameterization by Non-centered Parameterization as:

$$\Delta u_N^i = A_N e_N^i \quad (10)$$

as in Eq (6) where A_N means the scale of deviation of the free flow speed by vehicle, and e_N^i means the ratio of deviation of the free flow speed by i . A_N is expressed as:

$$A_N \sim C^+(500) \quad (11)$$

as in Eq (7) where 500 represents a weakly informative prior parameter. e_N^i captures the impact on free-flow speed by i . e_N^i is expressed as in Eq (8) and (9) as:

$$\sum_i e_N^i = 0 \quad (12)$$

$$e_N^i \sim N(0, 1) \quad (13)$$

The fourth term in Eq (4) $\Delta u_{NX}^{i,j}$ means the deviation of the free-flow speed from average in j for i . $\Delta u_{NX}^{i,j}$ is reparameterization by Non-centered Parameterization as:

$$\Delta u_{NX}^{i,j} = A_{NX} e_{NX}^{i,j} \quad (14)$$

as in Eq (6) where A_{NX} means the scale of deviation of the free flow speed by interaction, and $e_{NX}^{i,j}$ means the ratio of deviation of the free flow speed by i and j . So, A_{NX} is expressed as:

$$A_{NX} \sim C^+(500) \quad (15)$$

as in Eq (7) where 500 represents a weakly informative prior parameter. $e_{NX}^{i,j}$ captures the impact on free-flow speed for i in j . As in Eq (8), the sum of $e_{NX}^{i,j}$ is constrained to be zero for j or i . $e_{NX}^{i,j}$ is expressed as in Eq (8) and (9) as:

$$\sum_i e_{NX}^{i,j} = 0 \quad \forall j \quad (16)$$

$$\sum_j e_{NX}^{i,j} = 0 \quad \forall i \quad (17)$$

$$e_{NX}^{i,j} \sim N(0, 1) \quad (18)$$

$\delta^{i,j}$ is modeled like the free flow speed given by Eq (5) to (18) as:

$$\delta^{i,j} = \bar{\delta} + \Delta\delta_X^j + \Delta\delta_N^i + \Delta\delta_{NX}^{i,j} \quad (19)$$

where the following prior distribution gives the average jam spacing:

$$\bar{\delta} = N(8,100) \quad (20)$$

where 8 and 100 represent weakly informative prior parameters.

The model structure exhibits a clear hierarchical organization, as depicted in Figure 1.

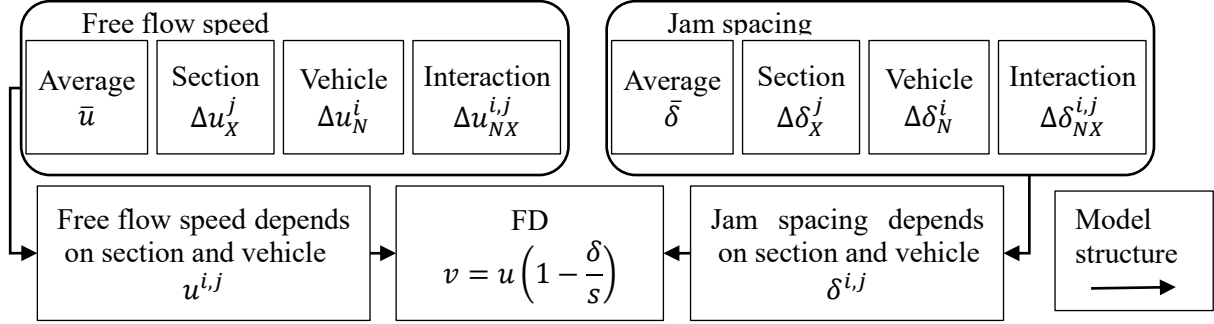


Figure 1. A diagram showing the hierarchical structure of the parameters in the model.

2.2 Discussion

The proposed model offers the distinct advantage of enabling the separation of the effects exerted by road sections and vehicles on the FD. This allows for the straightforward extraction of FD that considers only one of these effects from the estimated results. Furthermore, it facilitates a clear comparison of the influence of each factor.

The proposed model presents two key limitations: computational cost and the use of the Greenshields FD. The model estimates parameters for each vehicle within each road section. As a result, expanding the sample size becomes computationally expensive due to the increasing number of parameters. This makes it difficult to increase the number of sample vehicles thus the model is susceptible to biases in the data. The model utilizes the Greenshields FD, which may not fully capture real traffic dynamics. While a more complex FD could improve the model's accuracy, a trade-off between reproducibility and computational cost led to Greenshields FD being used. Nakanishi (2022) made the same choice.

3. CASE STUDY

3.1 Data

Our case study utilizes Zen Traffic Data (ZTD), the complete vehicle trajectory dataset from the Hanshin Expressway Company (2023). ZTD captures vehicle position, speed, and other for all vehicles within the target section at a frequency of 0.1 seconds. We focus on a specific zone within ZTD encompassing a 1,000 m stretch (from 2,500 m to 3,500 m) on the Hanshin Expressway Route 13 Higashi-Osaka Line (towards Nara) near the Morinomiya junction. This zone features lane change restrictions and sags, as illustrated in Figure 2. Percentages in Figure 2 indicate the slope of each section.

Since the FD represents steady state traffic conditions, a specific data selection process was employed to identify periods closest to this ideal state. To achieve this, the coefficient of variation (CV) of speed was calculated for each datum, considering a 2 s window before and after. Only data with a CV less than 0.1, indicative of minimal speed fluctuations, were retained for estimation.

To guarantee a balanced spacing during data reduction, we select vehicles. This involved selecting data from 100 vehicles, ensuring an equal amount for each of the first ten spacing classes (ranging from 0 m to 100 m with 10 m increments). This selection ensures the model is trained on a dataset that reflects

various traffic densities. Subsequently, the data was resampled to a frequency of 0.3 seconds.



Figure 2. Route alignment (Edited from Geospatial Information Authority of Japan, 2024)

3.2 Estimated result

The model estimation converged successfully. The Gelman-Rubin statistics and effective sample size for all parameters were checked. Following Kruschke (2017), Gelman-Rubin statistics below 1.1, and effective sample sizes exceeding 10,000 are thresholds for checking satisfactory convergence.

Figure 3 presents the estimated average FD. The superimposed 2D histogram represents the observed speed-spacing relation within the data, while the overlaid line signifies the estimated FD. Figure 3 indicates that the model's average driving behavior is obtained for the data.

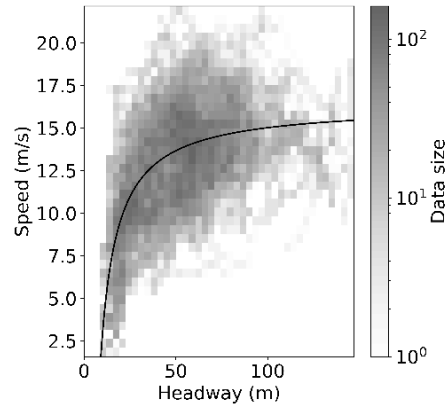
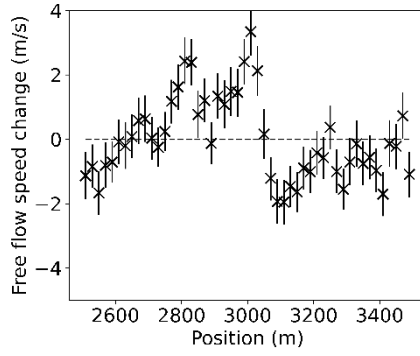


Figure 3. Average FD

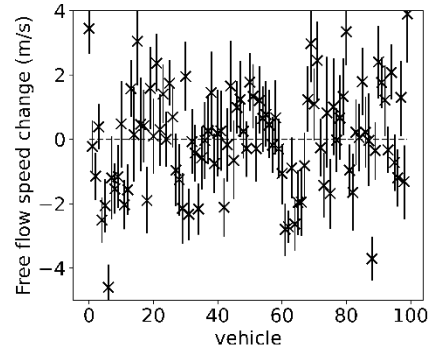
Figure 4 illustrates the variations in FD parameters (free flow speed and jam spacing) across both sections and vehicles. The upper figures depict changes in free flow speed, whereas the lower show variations in jam spacing. The left figures represent variations attributable to different sections, while the right side reflects variations associated with individual vehicles. The error bars in Figure 4 represent 96% confidence intervals. Figure 4 shows that sag etc. affects the free flow speed, and vehicle heterogeneity affects both FD parameters. The section effect is relatively continuous but changes significantly over 3000 m.

FD parameters, encompassing the independent effects of sections, vehicles, and their interactive effects, are presented in Figure 5. In Figure 5, free flow speed shows faint stripes, indicating the strong influence of interaction effect, while jam spacing shows clear stripes, indicating little interaction effect.

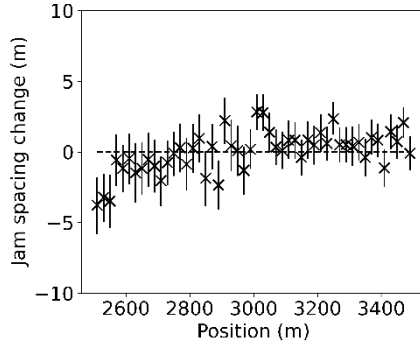
The FD allows for the estimation of traffic capacity. Utilizing the estimated $u^{i,j}$ and $\delta^{i,j}$ for a given section (i) and vehicle (j), the capacity can be calculated as $u^{i,j}/(4\delta^{i,j})$. This estimated capacity, based on both section and vehicle characteristics, represents “the capacity of a specific section when all vehicles being the same characteristics as the considered vehicle”.



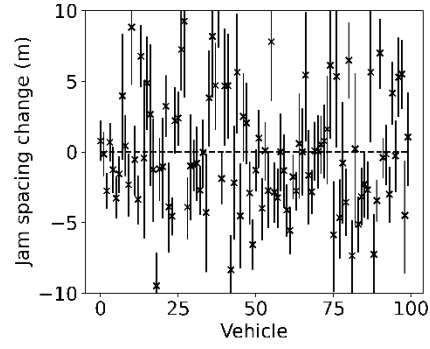
(a) Free flow speed change by section



(b) Free flow speed change by vehicle

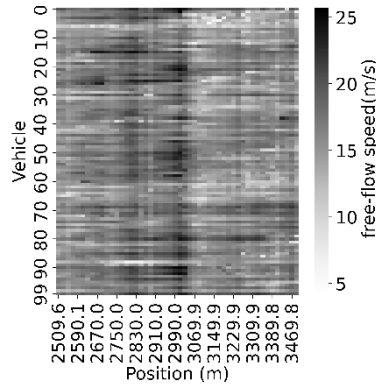


(c) Jam spacing change by section

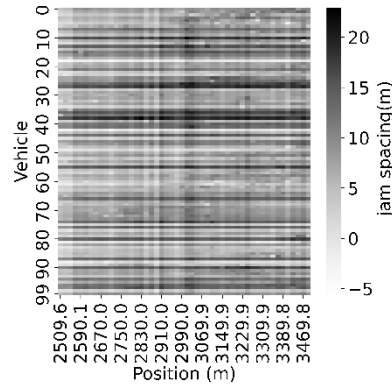


(d) Jam spacing change by vehicle

Figure 4. Parameter distribution by vehicle, and section



(a) Free flow speed



(b) Jam spacing

Figure 5. Parameter distribution

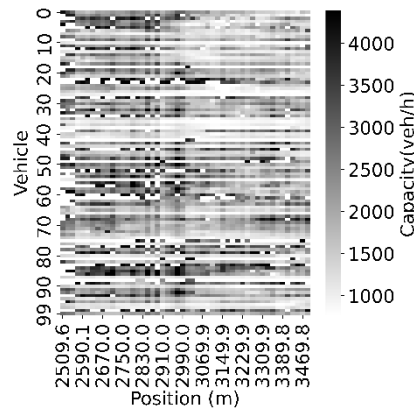


Figure 6. Traffic capacity distribution

Figure 6 visualizes the estimated capacity across sections and vehicles. Note that, sections, and vehicles with a jam spacing of 3 meters or less have been excluded from this figure. These data points are therefore excluded to ensure the capacity estimates remain reliable. According to the figure, both sections and vehicles impact capacity. This can be seen as vertical and horizontal stripes. Capacity varies greatly depending on location and vehicle and can vary by more than 3000 veh/h.

3.3 Discussion

Figures 2, 4(a), and 4(c) reveal that sections (e.g., sag and lane change prohibition area) have an impact on free flow speed. The free flow speed decreases at the sag and increases at the lane change prohibition area. Conversely, the effect of sections on jam spacing appears to be minimal. This suggests that driving safety awareness, which dictates jam spacing, is primarily determined by vehicle characteristics rather than road section. The large change in free flow speed around 3000 m may be associated with the apex of the road.

On the other hand, Figures 4(b) and 4(d) demonstrate vehicle characteristics' influence on free flow speed and jam spacing. This can be attributed to the role of vehicle-specific safety awareness in determining both the maximum speed a vehicle can safely travel and the minimum spacing for safety. Depending on the vehicle, the vehicle effect is as large as the section effect, suggesting the importance of vehicle effect.

By comparing Figures 4 and 5, we can gain insights into the interaction effects between sections and vehicles on the FD parameters. These interactions seem to affect free flow speed, as evidenced in Figures 4(a), and 5(a). In Figures 4(a), and 5(a), we observe that certain vehicles (e.g., Vehicle 15) exhibit more changes in free flow speed compared to others when transitioning between sections. This may suggest that interactions affect how much vehicles are influenced by the section they are traveling on.

In contrast, the interaction effect on jam spacing appears minimal, as indicated by the clear striped pattern in Figure 5(b). The few remaining effects, along with the observed slight interactions, likely represent the preceding vehicle's acceleration and deceleration behaviors. Notably, the selection process focused on near steady state speed, leading to the inclusion of data with sudden changes in spacing. When vehicle 86 had a large interaction, the vehicle in front suddenly slowed down. These data have a bigger interaction effect in jam spacing.

Figure 6 highlights the significant influence of both section and vehicle on traffic capacity. We observe decreases in capacity beyond the 3000 m, likely due to section-specific factors. Similarly, capacity decreases at sag. In contrast, lane change prohibition can be understood as a good measure to increase capacity. Additionally, Figure 6 reveals variations in vehicle-specific capacity. While some vehicles (e.g., vehicle 40) exhibit a low capacity with minimal dependence on the section, others (e.g., vehicle 85) demonstrate a high capacity. This may suggest the limitations of solely focusing on section-dependent or vehicle-dependent capacity or driving behavior estimation for effective congestion countermeasures.

4. CONCLUSION

This study has investigated the factors shaping driving behavior characteristics, particularly traffic capacity. We propose and estimate a model that incorporates the influence of both road sections and vehicle characteristics on the fundamental diagram (FD). The model uses hierarchical Bayesian modeling to represent the Greenshields FD for each specific section and vehicle. Complete vehicle trajectory data is then applied to estimate the model parameters. Following estimation, the results are interpreted to validate the influence of sections and vehicles on driving behavior and traffic capacity.

Our approach successfully has separated the effects of sections and vehicles, enabling the estimation of

free flow speed, jam spacing, and capacity for each section and vehicle. This allows for quantifying phenomena such as that both sections and vehicles impact free flow speed, whereas vehicle characteristics primarily influence jam spacing. Finally, using the estimated parameters, we can quantitatively describe the impact of both sections and vehicles on traffic capacity. As a result, the capacity depends on the section and the vehicle, which suggests a limit of solely focusing on section-dependent or vehicle-dependent capacity. It also becomes clear that the impact of the section varies from vehicle to vehicle. For example, some vehicles always contribute to congestion and others always contribute to relieving congestion in the same section.

A key limitation of this study lies in using the Greenshields FD, which may not always provide the most accurate representation of real-world traffic conditions. Utilizing a more accurate FD than the Greenshields one holds the potential for a deeper understanding of vehicle driving behavior. One such example of a highly accurate FD is Newell's (2002) triangular FD.

5. ACKNOWLEDGMENT

This research was supported by JSPS KAKENHI Grant Number 23H01524. The Hanshin Expressway provided Zen Traffic Data. The Tokyo Institute of Technology's supercomputer TSUBAME3.0 was used to estimate the model.

6. REFERENCES

- Betancourt, M. and Girolami, M. (2015) Hamiltonian Monte Carlo for hierarchical models. *Current trends in Bayesian methodology with applications*, Vol. 79, No. 30, pp. 2 – 4,
- Chiabaut, N., Buisson, C., and Leclercq, L. (2009) Fundamental Diagram Estimation Through Passing Rate Measurements in Congestion, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, no. 2, pp. 355–359
- Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper), *Bayesian Analysis*, Vol.1, pp. 515-534
- Geospatial Information Authority of Japan (2024) *GSI Tiles*, <https://maps.gsi.go.jp/vector>
- Greenshields, B. D., Bibbins, J. R., Channing, W. S., and Miller, H. H. (1935) A study of traffic capacity. *In Highway research board proceedings*, Vol. 14, pp. 448-477.
- Hanshin Expressway Co. (2023) *Zen Traffic Data*, <https://zen-traffic-data.net>
- Kruschke, J. K. (2017) *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, Academic Press New York
- Nakanishi, W. (2022) Sparse modelling for location dependent fundamental diagram with complete vehicle trajectory dataset, *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, Vol. 78, No. 1, pp. 24-33
- Newell, G.F. (2002) A simplified car-following theory: a lower order model. *Transportation Research Part B: Methodological*, Vol. 36, No.3, pp.195-205
- Seo, T., Kawasaki, Y., Kusakabe, T., and Asakura, Y. (2019) Fundamental diagram estimation by using trajectories of probe vehicles, *Transportation Research Part B: Methodological*, vol. 122, pp. 40–56
- Taylor, J., Zhou, X., Rouphail, N. M., and Porter, R. J. (2015) Method for investigating intradriver heterogeneity using vehicle trajectory data: A Dynamic Time Warping approach, *Transportation Research Part B: Methodological*, Vol. 73, pp. 59–80
- Yoshida, K., Nakanishi, W., and Asakura, Y. (2021) Analysis of car-following behavior in different time zones by estimating the parameters of Newell's model, *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, Vol. 76, No. 5, pp. I 1253-I 1261
- Zhang, C., Wang, W., and Sun, L. (2024) Calibrating car-following models via Bayesian dynamic regression, *Transportation Research Part C: Emerging Technologies*, pp. 104719