Identifying Stop-and-Go Congestion with Data-Driven Traffic Reconstruction

Shreyaa Raghavan*, **Edgar Ramirez Sanchez***, **Cathy Wu**
Massachusetts Institute of Technology
Cambridge, MA 02139
`{shreyaar, edgarrs, cathywu}@mit.edu`

Abstract

Identifying stop-and-go events (SAGs) in traffic flow presents an important avenue for advancing data-driven research for climate change mitigation and sustainability, owing to their substantial impact on carbon emissions, travel time, fuel consumption, and roadway safety. In fact, SAGs are estimated to account for 33-50% of highway driving externalities. However, insufficient attention has been paid to precisely quantifying where, when, and how often these SAGs take place—necessary for downstream decision-making, such as intervention design and policy analysis. A key challenge is that the data available to researchers and governments are typically sparse and aggregated to a granularity that obscures SAGs. To overcome such data limitations, this study thus explores the use of traffic reconstruction techniques for SAG identification. In particular, we introduce a kernel-based method for identifying spatiotemporal features in traffic and leverage bootstrapping to quantify the uncertainty of the reconstruction process. Experimental results on California highway data demonstrate the promise of the method for capturing SAGs. This work contributes to a foundation for data-driven decision-making to advance the sustainability of traffic systems.

1 Introduction

Transportation is the most significant contributor to greenhouse gas (GHG) emissions, accounting for 28 percent of the total in the U.S.. Traffic congestion on its own is estimated to cost the average driver around 2000 USD in large cities and, to the United States alone, 305 billion USD in 2017 (1). "Stop-and-go"(SAG) traffic, a common phenomenon in congested traffic, happens when cars traveling on a highway periodically form waves that propagate down the highway, leading cars to constantly accelerate and brake unnecessarily. These SAG waves or events lead to up to 67% higher fuel consumption, longer travel times, higher GHG emissions levels, and could be a safety hazard (2). Concerningly, studies have found that SAGs can be responsible for 33-50% of highway driving externalities, which are external costs such as carbon emissions, highway accidents, etc (3; 4). To mitigate climate change and improve air quality in human settlements, making road transportation more environmentally sustainable and reducing SAG congestion should be a priority.

However, there is hope because SAGs are preventable, implying that a considerable portion of these adverse effects and externalities of highway driving are avoidable. Multiple studies have proven the potential of different technologies to alleviate SAGs using Autonomous Vehicles (AVs) (5; 6; 7; 8) , reinforcement learning (9), and variable speed limits (10), among others. Moreover, several studies have provided insights into how these events form, propagate, and fade (5; 6; 7; 8). It is clear that in theory, we can make proactive efforts to enable and support these interventions if we are able to identify when and where these SAGs occur.

In particular, machine learning could play a pivotal role, as it has the capability to identify and predict a broad range of events through techniques like image processing, kernel methods, and anomaly detection. However, a major limitation of these models is that they require rich datasets to capture the nuanced dynamics of vehicles, and there is very limited rich traffic data available. In contrast, the generally available data, of which there is a large amount, is typically too sparse or noisy to capture SAGs. In the US, for example, the highway dataset with the largest coverage is the Caltrans Performance Measurement System (PeMS), but even in this best-case scenario, sensors are spaced thousands of feet away and the data is both temporally and spatially sparse. The existing data, in its current state, is not enough to capture the minute behaviors of stop-and-go events, especially at a large scale.

In response, this paper proposes the use of traffic reconstruction as a way to fill in the gaps and generate rich trajectory information to enable the deployment of complex ML techniques. We apply this framework to a county in California and devise a kernel-based SAG identification method with bootstrapping to provide robust insights into traffic behavior.

2   Related Work

The study of stop-and-go events (SAGs) in traffic has a longstanding history, dating back to the 1960s (11). While much of the literature has traditionally focused on fundamental analyses, there has been a discernible shift toward identification methods (12), in part driven by challenges in obtaining sufficient data for large-scale identification. On one hand, theoretical studies encompass dynamic wave modeling, control mechanisms, instabilities, driving models, and other aspects that shed light on triggers, formation, propagation, and dissipation of these waves (13; 6; 14; 15; 16; 17; 18; 19; 20; 21; 5; 7; 8).

On the other hand, empirical studies focus on identification through various techniques, such as time series and signal processing methods, to identify SAG occurrences in space and time from the observed traffic or vehicle-specific data (22; 23; 24; 25; 12; 26; 27; 28). These approaches typically rely on rich trajectory data that only exists for very few stretches of highway, and only some of them have sufficient resolution to incorporate spatiotemporal features. In contrast, most studies use data from fixed and sparse sensors and identify SAGs using stationary wave processing techniques. In particular, wavelet transformation has been a popular approach for SAG identification (12; 20), matching a Mexican hat profile to the signal to identify where this behavior occurs.

Finally, while research links stop-and-go waves to negative effects such as increased fuel emissions, safety risks, and driver distress (27; 18; 29; 7), SAG quantification has not been conducted at scale. A unique example in quantifying the number of SAGs is a study estimating that 20% of traffic jams in the Netherlands could be stop-and-go waves (30). The significance of this estimation is underscored by its inclusion as a priority in the national plan of Intelligent Transportation Technologies (ITS) Netherlands. Yet, this is an anomaly, as most countries, including the United States, do not treat stop-and-go mitigation as a priority due to a lack of quantitative evidence.

Unlike previous literature, our work makes use of both spatiotemporal features in traffic data and broadly available data, which is sparse in nature.

3   Traffic Feature Identification Framework

To leverage available yet sparse data, we propose a technique to augment raw traffic data with physics-based traffic simulations to build a richer, more continuous data representation, with the goal of enabling ML pipelines for stop-and-go identification. Our raw data source is the Caltrans Performance Measurement System (PeMS) dataset, which consists of roughly 39,000 induction loop detectors (electromagnetic highway sensors) that capture traffic flow, occupancy, and speed aggregated every 30 seconds on most freeways in major metropolitan areas of California. The key concern with PeMS is that the dataset is too sparse to draw conclusions about traffic events or have a complete understanding of vehicle trajectories. As a result, we seek to fill in these data gaps via traffic reconstruction while using PeMS as our ground truth data.
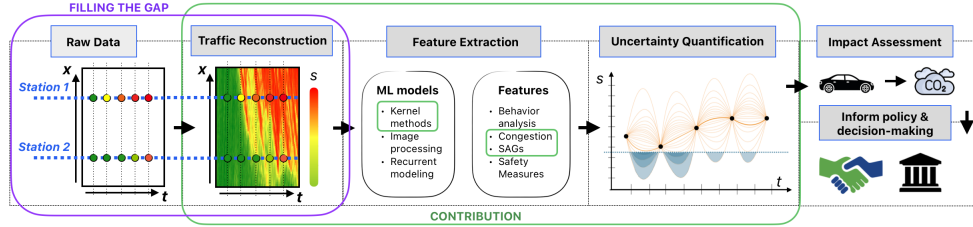
Figure 1: Pipeline for traffic feature identification. Our contribution is boxed in green.

3.1 Traffic Reconstruction

We leverage a traffic reconstruction model built as part of a safety analysis framework, which simulates traffic scenarios at a fine-grain level to build individual vehicle trajectories. It does so by combining micro-level dynamics of driving behavior on California highways, captured by highway video data, with macro-level data obtained by induction loop detectors.

At the large scale, we represent the entire road network as a graph $G = (E, V)$. The goal is to recover the optimal flow through every edge e at time t such that the routes of the vehicles align closely to the ground-truth traffic data while still being realistic. We model this flow estimation problem as a maximum flow problem. At the smaller scale, we create a driver model that relies on the widely used "Intelligent Driver Model" (IDM) (31). In our traffic reconstruction framework, the driver model is calibrated to digital video footage of California highways to better capture behaviors commonly seen in that region (32). Both the driver model and the estimated vehicle routes are then combined in Simulation of Urban MObility (SUMO), a traffic simulator, to extrapolate detailed trajectories of traffic. The outcome is a rich representation of an entire highway network, that has information about every vehicle at every point in space and time. A more comprehensive overview of this traffic reconstruction method is currently in preparation.

3.2 Uncertainty Quantification

With a continuous, time-space representation of traffic, we can now extract salient features. However, each reconstruction of traffic may have deviations, so relying on a single reconstruction for predictions could inflate or erase certain behaviors. This prompts us to question: given a single simulation, how certain are we in an observed event or behavior? Instead, we want a robust measure of confidence in our identification of features. To accomplish this, we use a technique known as bootstrapping and leverage our traffic reconstruction as a distribution sampler, which allows us to estimate the distribution of speeds for a point in time and space (33). We replicate the traffic reconstruction k times and build a distribution of average vehicle speed at position x and time t: $p_s(s|t, x)$. In each replication, we randomly sample (with replacement) stochastic elements of the reconstruction, such as speed deviations, offsets in departure times, and vehicle driving dynamics.

To provide a measure of certainty, we want to know the probability that an event of interest, A, occurs at (t, x). We define A as a function of speeds, $S_{t,x}$, so $p_A(A|t, x) = p_s(f(S_{t,x}, ..., \nu)|t, x)$. Hyperparameters of f are denoted as ν. In the next section, we show how to leverage this formulation for robustly identifying stop-and-go events.

4 Application to Stop-and-Go Event Identification

Using the individual trajectories from the traffic reconstruction, we build a rich time-space representation of traffic that encodes meaningful features. This representation, known as a time-space diagram, is commonly used in the transportation community to visualize vehicle behaviors as they move through time (x-axis) and space (y-axis). The color of the diagram encodes the aggregated speed of the vehicles across all lanes.

In our study, we reconstruct 100 variations of traffic for the entire Los Angeles County from 11:15 to 11:30 a.m.. We focus our analysis below on a 0.5 km segment of the 110-N freeway.
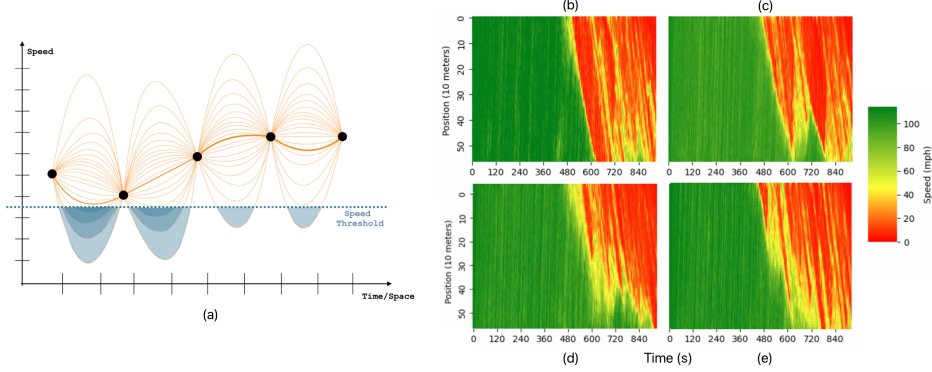
3

Figure 2: (a) Schematic of possible reconstructed trajectories that fill the gaps between ground-truth readings. These trajectories vary in each replication. (b-e) 4 varying time-space diagrams out of 100.

A stop-and-go event is defined by an alternating wave of deceleration and acceleration, or high speeds and low speeds. Despite the minimal work done on identification specifically, extensive research has focused on explaining and characterizing this oscillating pattern using fundamental traffic models, such as the Lighthill–Whitham–Richards (LWR) model (34; 35). We know, for example, that in a time-space representation, they appear as waves with a negative slope, as seen by the red slants in Figure 2(b-e) (13). Additionally, empirical studies have recorded concrete properties of SAGs: for example, during rush hour on US-101, the average width of a SAG was recorded to be 7 min with an average amplitude of 15 km/hr (13). We capture these known behaviors with a parameterized spatiotemporal kernel.

We take into account both the dip in speed over time and the backward movement over space by designing a 2-D kernel based on the Sobel operator for diagonal edge detection (36). However, the behavior of a SAG depends on many factors such as the time of day or road conditions, so not all SAGs appear the same on a time-space diagram. To capture SAGs with different oscillation periods, we parameterize the width of the kernel. It is important to note that other properties of the kernel, such as the slope of propagation or the magnitude of the dip (which corresponds to the amplitude of the wave) can also vary depending on the setting. Here, we only specify the width, i.e. the wave period, as a parameter.

An example of the kernel K with width $= 4$ seconds is as follows:

$$K(4) = \begin{bmatrix} 0 & -1 & -1 & -1 & -1 & 0 & 2 & 2 \\ 2 & 0 & -1 & -1 & -1 & -1 & 0 & 2 \\ 2 & 2 & 0 & -1 & -1 & -1 & -1 & 0 \end{bmatrix}$$

We convolve this kernel $K(w)$ across neighborhoods of speeds, $S_{t,x}$, in the time-space diagram and normalize these values from -1 to 1. $S_{t,x}(m,n) = \{s_{t+i,x+j} : \forall i \in [-m,m], j \in [-n,n]\}$ is a neighborhood of speeds surrounding $s_{t,x}$. We define this convolution as the function $f_{SAG}(S_{t,x}, w|t,x) = norm_{(-1,1)}(K(w) * S_{t,x}(w,1))$, which we also refer to as the "Kernel Activation Value". The Kernel Activation Value provides a measure of how strongly the behavior at (t,x) matches the kernel.

Then, we define $C_{t,x}$ as the event that a SAG occurs at (t,x), which is 1 when $f_{SAG}(S_{t,x}, w|t,x) \geq \epsilon$ and 0 otherwise. This ϵ controls how steep the stop-and-go behavior should be in order to be considered a SAG event. For our study, we use a threshold of $\epsilon = 0.30$. We consider ϵ as a hyperparameter, as it can be tuned to accommodate different settings. We build an indicator variable $C_{t,x}^j$ defined by the occurrence of a SAG in trial i at (t,x). Then, we estimate the probability of a SAG occurring at (t,x) by using the $C_{t,x}^j$ as samples of the probability distribution:

$$p(C_{t,x} = 1|t,x) = p_s(f_{SAG}(S_{t,x}, w|t,x) \geq \epsilon|t,x) = \frac{1}{100} \sum_{i=1}^{100} C_{t,x}^i(w, \epsilon)$$

Using the continuous time-space representation and our kernel-based method, we are able to detect and visualize stop-and-go traffic in time and space, as shown by the blue diagonal lines in the first

4

row of Figure 3. The probabilities in the second row of Figure 3 create hotspots that we know are more prone to SAGs. As a result, we can quantify how certain we are that a SAG detected in a single trial is not simply variance or noise. This probabilistic analysis allows us to extrapolate the observed traffic patterns in data to the real world and provides a measure of uncertainty, which is crucial for real-world decision-making and intervention.
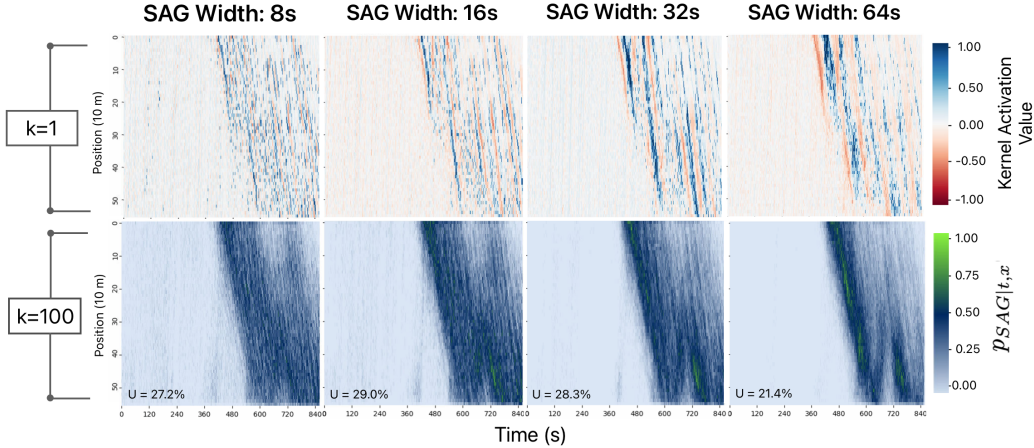

Figure 3: f_{SAG} for kernels of different widths applied to a single trial (same trial as Figure 2(b)) are shown in the first row. Values classified as SAG ($> \epsilon$) are boosted for visual reference. The second row shows the probability (combined across all 100 simulations) of a SAG occurring. U represents the percentage of the heatmap for which the detector is uncertain, i.e. $0.25 < p_{SAG} < 0.75$.

When considering the impact of stop-and-go events, each one has a social cost (i.e., an aggregate measure of air quality, GHG emission, travel time, and safety measures) associated with it. In Figure 3, a greater likelihood of a SAG at any point indicates a greater chance of incurring that social cost. While a proper social cost estimation requires further research, we could interpret this representation as a map of where avoidable congestion and avoidable social costs are most likely present.

5 Conclusion

Stop-and-go wave mitigation is a significant opportunity for making highway driving more sustainable. This paper presents a novel traffic reconstruction framework, offering a rich spatio-temporal representation that effectively and robustly detects traffic behaviors, particularly stop-and-go events. Our visual representation enables the creation of comprehensive traffic datasets, while the success of our kernel method in capturing SAG patterns demonstrates the potential for more complex ML methods to leverage this data. With local SAG identification, policy-makers can pinpoint when and where SAGs are likely to occur in their communities, empowering the design and deployment of targeted interventions. Through SAG identification at scale, we can quantify how much of the total highway congestion is avoidable. This would concretely motivate the need for smart infrastructure and other technology-based interventions and is a crucial step toward convincing decision-makers to take action at a national level.

Future research directions include the development of sequential ML tools for time-series analysis, extending to other features such as safety measures, causal analysis via individual trajectories, behavior analysis, and prediction through supervised learning. A key goal in the future will be to validate the traffic reconstruction more quantitatively. We plan to compare reconstructed trajectories to the available ground truth by testing the pipeline on a richer dataset, such as a vehicle trajectory dataset or highway camera footage. Finally, scaling the SAG identification process for the full California highway system and for longer periods of time is a promising step towards at-scale SAG intervention and more sustainable roadways.

References

[1] E. P. Agency, *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2021.* Jul 2023.

[2] R. E. Stern, S. Cui, M. L. Delle Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, H. Pohlmann, F. Wu, B. Piccoli, *et al.*, "Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 205–221, 2018.

[3] K. Goldmann and G. Sieg, "Quantifying the phantom jam externality: the case of an autobahn section in germany," *European transport research review*, vol. 13, no. 1, pp. 1–15, 2021.

[4] K. Goldmann and G. Sieg, "Economic implications of phantom traffic jams: evidence from traffic experiments," *Transportation Letters*, vol. 12, no. 6, pp. 386–390, 2020.

[5] C. Balzotti and E. Iacomini, "Stop-and-go waves: A microscopic and a macroscopic description," in *Mathematical Descriptions of Traffic Flow: Micro, Macro and Kinetic Models*, pp. 63–78, Springer, 2021.

[6] J. A. Laval and L. Leclercq, "A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1928, pp. 4519–4541, 2010.

[7] S. Oh and H. Yeo, "Impact of stop-and-go waves and lane changes on discharge rate in recovery flow," *Transportation Research Part B: Methodological*, vol. 77, pp. 88–102, 2015.

[8] A. Portz and A. Seyfried, "Analyzing stop-and-go waves by experiment and modeling," in *Pedestrian and Evacuation Dynamics*, pp. 577–586, Springer, 2011.

[9] A. R. Kreidieh, C. Wu, and A. M. Bayen, "Dissipating stop-and-go waves in closed and open networks via deep reinforcement learning," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1475–1480, IEEE, 2018.

[10] M. Wang, W. Daamen, S. P. Hoogendoorn, and B. Van Arem, "Connected variable speed limits control and car-following control with vehicle-infrastructure communication to resolve stop-and-go waves," *Journal of Intelligent Transportation Systems*, vol. 20, no. 6, pp. 559–572, 2016.

[11] P. I. Richards and P. I. Richards, "Shock waves on the highway," *Operations Research*, 1956.

[12] N. Saxena, N. Saxena, N. K. Saxena, Dixit, and S. Waller, "Techniques for identifying the occurrence of stop-and-go waves in traffic: a literature review," *null*, 2016.

[13] J. A. Laval and L. Leclercq, "A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic," *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1928, pp. 4519–4541, 2010.

[14] J. A. Laval, C. Toth, and Y. Zhou, "A parsimonious model for the formation of oscillations in car-following models," *Transportation Research Part B-methodological*, 2014.

[15] X. Li and Y. Ouyang, "Characterization of traffic oscillation propagation under nonlinear car-following laws," *null*, 2011.

[16] R. E. Wilson, "Mechanisms for spatio-temporal pattern formation in highway traffic models," *Philosophical Transactions of the Royal Society A*, 2008.

[17] D. Chen, S. Ahn, J. A. Laval, and Z. Zheng, "On the periodicity of traffic oscillations and capacity drop: The role of driver characteristics," *Transportation Research Part B-methodological*, 2014.

[18] X. Li, J. Cui, S. An, and M. Parsafard, "Stop-and-go traffic analysis: Theoretical properties, environmental impacts and oscillation mitigation," *Transportation Research Part B-methodological*, 2014.

[19] H. Yeo and A. Skabardonis, "Understanding stop-and-go traffic in view of asymmetric traffic theory," *null*, 2009.

[20] Z. Zheng, S. Ahn, D. Chen, and J. A. Laval, "Applications of wavelet transform for analysis of freeway traffic: Bottlenecks, transient traffic, and traffic oscillations," *Transportation Research Part B-methodological*, 2011.

[21] A. Corli and H. Fan, "Hysteresis and stop-and-go waves in traffic flows," *Mathematical Models and Methods in Applied Sciences*, 2019.

[22] Z. He, Y. Lv, L. Lu, L. Lu, W. Guan, W. Guan, and W. Guan, "Constructing spatiotemporal speed contour diagrams: using rectangular or non-rectangular parallelogram cells?," *Transportmetrica B-Transport Dynamics*, 2019.

[23] R. Jiang, C.-J. Jin, H. Zhang, H. Zhang, Y.-X. Huang, J. Tian, W. Wang, W. Wang, W. Wang, W. Wang, W. Wang, M.-B. Hu, H. Wang, H. Wang, H. Wang, H. Wang, and B. Jia, "Experimental and empirical investigations of traffic flow instability," *Transportation research procedia*, 2017.

[24] A. Abdi, M. Saffarzadeh, and A. Salehikalam, "Identifying and analyzing stop and go traffic based on asymmetric theory of driving behavior in acceleration and deceleration," *null*, 2016.

[25] Y. Lin, B. Ran, B. Ran, Y. Rui, Y. Zheng, Y. Zheng, P. pei Mao, P. pei Mao, J. Zhang, and J. Zhang, "Real-time detection of stop-and-go waves on freeways using probe trajectories," *CICTP 2019*, 2019.

[26] X. Li, X. Wang, and Y. Ouyang, "Prediction and field validation of traffic oscillation propagation under nonlinear car-following laws," *null*, 2012.

[27] F. Wu, R. E. Stern, S. Cui, M. L. D. Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, R. Haulcy, B. Piccoli, B. Seibold, J. Sprinkle, and D. B. Work, "Tracking vehicle trajectories and fuel rates in phantom traffic jams: Methodology and data," *Transportation Research Part C-emerging Technologies*, 2019.

[28] M. Mauch and M. J. Cassidy, "Freeway traffic oscillations: Observations and predictions," *null*, 2004.

[29] L. J. J. Wismans, L. Suijs, L. Kroll, L. Krol, E. van Berkum, and E. C. van Berkum, "In car advice to reduce negative effects of phantom jams," *null*, 2015.

[30] D. V. Noordegraaf, F. Faber, and D. Vukovic, "Filegolven binnen no time in beeld-excelleren in detecteren," in *Bijdrage Aan Het Colloquium Vervoersplanologisch Speurwerk*, 2011.

[31] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical review E*, vol. 62, no. 2, p. 1805, 2000.

[32] C. Zhang and L. Sun, "Bayesian calibration of the intelligent driver model," *arXiv preprint arXiv:2210.03571*, 2022.

[33] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. CRC press, 1994.

[34] A. Corli and H. Fan, "Hysteresis and stop-and-go waves in traffic flows," *Mathematical Models and Methods in Applied Sciences*, vol. 29, no. 14, pp. 2637–2678, 2019.

[35] X. Li and Y. Ouyang, "Characterization of traffic oscillation propagation under nonlinear car-following laws," *Transportation Research Part B-methodological - TRANSP RES PT B-METHOD*, vol. 17, 11 2011.

[36] N. Kanopoulos, N. Vasanthavada, and R. L. Baker, "Design of an image edge detection filter using the sobel operator," *IEEE Journal of solid-state circuits*, vol. 23, no. 2, pp. 358–367, 1988.