Route based Freight Activity Metrics along the California State Highway System through a Pilot Multi Sensor Fusion System

December 2024

A Research Report from the Pacific Southwest Region University Transportation Center

Stephen G. Ritchie, University of California, Irvine

Andre Y.C. Tok, University of California, Irvine

Tanjeeb Ahmed, University of California, Irvine

Mingqi Yao, University of California, Irvine

Pacific Southwest Region UTC
University Transportation Center

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. PSR-23-03 TO 073	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle Route based Freight Activity Metrics along the California State Highway System through a Pilot Multi Sensor Fusion System		**5. Report Date** December 31, 2024 **6. Performing Organization Code** N/A
7. Author(s) Stephen G. Ritchie ORCID 0000-0001-7881-0415 Andre Y.C. Tok ORCID 0000-0002-0387-0170 Tanjeeb Ahmed Mingqi Yao		**8. Performing Organization Report No.PSR-23-03 TO 073**
9. Performing Organization Name and Address Institute of Transportation Studies University of California, Irvine 653 E Peltason Drive, Suite 4000 Irvine, CA 92697		**10. Work Unit No.** N/A **11. Contract or Grant No.** USDOT Grant 65A0674 TO 073
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590		**13. Type of Report and Period Covered** Final report (January 1, 2024 – December 31, 2024) **14. Sponsoring Agency Code** USDOT OST-R
15. Supplementary Notes		
16. Abstract This research project investigates the integration of point detector data to enhance vehicle tracking for heavy-duty trucks in Southern California. Leveraging a combination of automated license plate readers (ALPR) and inductive signature data, we aim to enhance the accuracy of vehicle monitoring, addressing limitations of traditional telematics data, such as sampling bias and data quality issues. Our research draws upon a dataset from sites deployed in the Freight Mobility Living Laboratory (FML2) testbed, enabling a comprehensive analysis of truck activity. Employing a Bayesian Logit model and principal component analysis, we developed an innovative framework aimed at improving truck tracking accuracy across key routes by matching inductive signatures. Findings indicate that integrating these advanced data sources leads to more accurate classifications and tracking of freight vehicles, ultimately contributing to better-informed planning and environmental sustainability efforts by Caltrans, the California Air Resources Board, and local agencies. By focusing on critical highways like I-710 and SR-60, we identify significant opportunities for optimizing freight mobility and addressing the increasing pressures on transportation infrastructure in urban settings. This research highlights the importance of innovative data integration techniques in developing effective traffic management strategies that meet the evolving demands of urban freight operations.		
17. Key Words Truck tracking, Inductive signature, ALPR, Bayesian logit, Freight activity monitoring, Multi sensor fusion		**18. Distribution Statement** No restrictions.

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 28	**22. Price** N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Contents

About the Pacific Southwest Region University Transportation Center

The Pacific Southwest Region University Transportation Center (UTC) is the Region 9 University Transportation Center funded under the US Department of Transportation's University Transportation Centers Program. Established in 2016, the Pacific Southwest Region UTC (PSR) is led by the University of Southern California and includes seven partners: Long Beach State University; University of California, Davis; University of California, Irvine; University of California, Los Angeles; University of Hawaii; Northern Arizona University; Pima Community College.

The Pacific Southwest Region UTC conducts an integrated, multidisciplinary program of research, education and technology transfer aimed at *improving the mobility of people and goods throughout the region.* Our program is organized around four themes: 1) technology to address transportation problems and improve mobility; 2) improving mobility for vulnerable populations; 3) Improving resilience and protecting the environment; and 4) managing mobility in high growth areas.

U.S. Department of Transportation (USDOT) Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

California Department of Transportation (CALTRANS) Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the United States Department of Transportation's University Transportation Centers program, in the interest of information exchange. The U.S. Government and the State of California assumes no liability for the contents or use thereof. Nor does the content necessarily reflect the official views or policies of the U.S. Government and the State of California. This report does not constitute a standard, specification, or regulation. This report does not constitute an endorsement by the California Department of Transportation (Caltrans) of any product described herein.

Disclosure

Principal Investigator, others, conducted this research titled, "Route based Freight Activity Metrics along the California State Highway System through a Pilot Multi Sensor Fusion System" at the Institute of Transportation Studies, University of California, Irvine. The research took place from March 14, 2024 to December 31, 2024 and was funded by a grant from the Caltrans in the amount of $105,000. The research was conducted as part of the Pacific Southwest Region University Transportation Center research program.

Acknowledgements

We would like to thank Caltrans Districts 7 and 8 for assisting with the fieldwork required to obtain data for this project.

Abstract

This research project investigates the integration of point detector data to enhance vehicle tracking for heavy-duty trucks in Southern California. Leveraging a combination of automated license plate readers (ALPR) and inductive signature data, we aim to enhance the accuracy of vehicle monitoring, addressing limitations of traditional telematics data, such as sampling bias and data quality issues. Our research draws upon a dataset from sites deployed in the Freight Mobility Living Laboratory (FML2) testbed, enabling a comprehensive analysis of truck activity. Employing a Bayesian Logit model and principal component analysis, we developed an innovative framework aimed at improving truck tracking accuracy across key routes by matching inductive signatures. Findings indicate that integrating these advanced data sources leads to more accurate classifications and tracking of freight vehicles, ultimately contributing to better-informed planning and environmental sustainability efforts by Caltrans, the California Air Resources Board, and local agencies. By focusing on critical highways like I-710 and SR-60, we identify significant opportunities for optimizing freight mobility and addressing the increasing pressures on transportation infrastructure in urban settings. This research highlights the importance of innovative data integration techniques in developing effective traffic management strategies that meet the evolving demands of urban freight operations.

Executive Summary

This research addresses the pressing need for improved traffic management strategies in response to the rapid growth of freight transportation in urban areas, particularly concerning heavy-duty trucks. As traffic congestion and freight operations become increasingly complex, this study seeks to enhance vehicle tracking through advanced sensor integration techniques. By employing automated license plate readers (ALPR) combined with inductive signature data, the research tackles existing challenges in vehicle monitoring, especially for long-distance tracking and real-time applications.

Our investigation builds upon the foundation of previous literature, which often struggles with the intricacies of varied traffic conditions and vehicle types. While significant progress has been made in vehicle classification and travel time estimation, this study focuses on creating a comprehensive analytical framework by integrating data sources from multiple sensor technologies. This methodology aims to improve the precision of vehicle classification-based tracking, thus informing freight planning and contributing to environmental sustainability efforts.

Focused specifically on Southern California's key highways, including I-710 and SR-60, this research highlights the diverse traffic patterns and heavy-duty vehicle operations in these regions. By developing a Bayesian Logit model to track trucks across three major routes, we provide a framework that enhances our understanding of freight movement and traffic dynamics.

The report reviews relevant literature on inductive signature data and weigh-in-motion (WIM) integration, showcasing advancements in vehicle classification methods, machine learning applications, and innovative tracking approaches. Among these, it emphasizes the effectiveness of integrating WIM data with inductive signatures to improve vehicle matching rates and tracking capabilities, even in congested traffic situations. We extend these methodologies by incorporating ALPR data, which serve as temporal and spatial constraints to search for candidate vehicles. Furthermore, our methodology utilizes principal components analysis (PCA) to reduce the dimensionality of the features, while minimizing correlation among features and maximizing the variance in the data.

The results indicate that the 25th and 26th features, representing the majority of Component 1, have the highest influence on the match probability. The Bayesian logit model demonstrated considerable prediction accuracy, achieving about 87% accuracy for matched vehicles and 99% for unmatched vehicles within the training and test datasets. The model identified 304 matched vehicle trips across three routes between I-710 Willow and SR-60 East End, showing considerable variations both in the overall route assignment proportions and also in the average hourly proportions across the three routes.

Overall, this research contributes valuable insights to the ongoing discourse on freight mobility, network optimization, and environmental impacts of freight transportation. By bolstering the capabilities of freight monitoring systems, this study aspires to promote greater efficiency and sustainability in urban freight transportation, ultimately addressing the challenges posed by increasing vehicle traffic in urban areas.

Introduction

The rapid growth of freight transportation in urban areas has intensified the need for efficient and effective traffic management strategies. Understanding vehicle behavior, specifically for heavy-duty trucks, is crucial for maintaining traffic flow, reducing congestion, and enhancing freight operations. This study focuses on leveraging advanced data integration techniques to improve vehicle classification, tracking, and weight estimation. By utilizing automated license plate readers and inductive signature data, we aim to address existing gaps in vehicle monitoring, particularly in the context of long-distance tracking and real-time data applications.

While some level of vehicle tracking can be achieved via telematics data such as GPS and cellular systems, the datasets from these sources nevertheless possess several drawbacks – such as sampling penetration, sampling bias, and data quality – that limit their benefits to state agencies.

In contrast, over 30 detector sites in Southern California are currently deployed with a combination of Inductive Signature and ALPR technologies as a part of UCI's Freight Mobility Living Laboratory (FML2) testbed, which offers a timely and prime platform for investigating the synergies of sensor fusion across these technologies to obtain better truck activity metrics such as truck tracking, classification, fleet identification that can benefit Caltrans, California Air Resources Board, California Energy Commission as well as Metropolitan Planning Organizations (MPOs) and local agencies.

This research builds on the established foundation in the literature on vehicle classification and detection using various technologies, recognizing that traditional methods often struggle with the complexities of dense traffic conditions and diverse vehicle types. While previous studies have made significant strides, our exploration emphasizes the integration of multiple data sources—specifically, automated license plate readers and inductive signatures—to create a robust framework for analyzing freight movement across Southern California's busy corridors. Through this approach, we aim to enhance the precision of vehicle classification and tracking, ultimately contributing to more informed freight planning and improved environmental sustainability.

The necessity for effective freight planning is amplified by the increasing pressure on transportation infrastructure and the need for environmentally conscious solutions. The integration of various data sources, including advanced econometric models, represents a significant advancement in our ability to classify and track vehicles accurately. This study explores these methodologies within the context of the Southern California regions, particularly along key highways such as I-710 and SR-60, where diverse traffic patterns and heavy-duty vehicle operations present both challenges and opportunities for enhanced analysis. Through this investigation, the goal is to contribute to the ongoing discussions about freight mobility, network optimization, and environmental impacts in transportation planning.

Our study builds upon the work of Hyun et al. (2017) and develops a Bayesian Logit model to track trucks across three routes between I-710 Willow (upstream) and SR-60 East End (downstream). In addition, we employ principal component analysis to capture the maximum variance of the influential features through the Bayesian Logit model. We use this model to calculate match probabilities based on the inductive signature of vehicle pairs and, thereby, track them across the three alternate routes between the upstream and downstream sites.

Literature Review

Inductive signature data, weigh-in-motion (WIM) data, and their integration have been extensively studied in previous literature, significantly contributing to freight and environmental planning by enabling improved vehicle classification, travel time estimation, and weight analysis. Most studies leverage WIM data as a core component, frequently integrated with inductive signatures or GPS data. Bayesian models and advanced machine learning approaches are consistently employed for vehicle classification and reidentification.

Meta and Cinsdikici (2010) proposed a vehicle classification algorithm utilizing single-loop inductive detectors. By applying discrete Fourier transform (DFT) to filter noise from raw data and principal component analysis (PCA) to extract features from the processed data, their downstream three-layer neural network (NN) model achieved a recognition rate of up to 94.21% under optimal configurations. While this method demonstrated robustness in classifying vehicle types even in congested traffic scenarios, its applicability to long-distance tracking remains unexplored.

Hernandez et al. (2016) focused their research on achieving high-resolution truck body classifications by integrating WIM data with inductive signature data. This approach provided detailed insights into freight operations and enhanced the precision of body and axle configuration models. Using a multiple classifier system (MCS), their method achieved over 85% accuracy for diverse body configurations. However, the study did not address longitudinal vehicle tracking directly.

To address the limitations in longitudinal tracking, Jeng and Chu (2014) developed an innovative approach for tracking heavy-duty vehicles on freeways by integrating WIM data with inductive signatures. Their RTREID-2MT method demonstrated effectiveness in tracking trucks over long distances across multiple stations. This method exhibited improved performance under congested conditions compared to previous Bayesian approaches, with accuracy rates exceeding 75% in free-flow traffic and 65.6% in congested scenarios. The integration of WIM and inductive signature data significantly improved vehicle matching rates while eliminating the need for site-specific calibration.

Moreover, Hyun (2016) proposed a network-wide truck tracking model that expanded the scale of vehicle tracking by combining Bayesian inference with advanced point detector data. This model successfully utilized inductive signatures and WIM attributes to achieve over 80% accuracy in vehicle reidentification across various locations. Additionally, it enhanced travel time estimations and demonstrated the potential to integrate truck body classification with travel pattern data, supporting improved freight planning.

Building upon prior research, Hyun et al. (2017) introduced a selective weighted Bayesian model (SWBM) to enhance long-distance truck tracking (up to 65 miles) using WIM and inductive signature data. Their method achieved an 81% correct matching rate for trucks tracked over a 26-mile corridor, effectively mitigating the impact of sensor measurement variations. However, the study also highlighted the limitations such as sensor calibration sensitivity and reduced accuracy in high-volume traffic conditions.

To address the spatial limitations of WIM sensors, Hernandez and Hyun (2020) proposed a methodology combining WIM and GPS data to estimate gross vehicle weight distributions at traffic count sites. This approach enhanced spatial resolution and improved weight estimation accuracy compared to traditional methods. However, the study's primary contribution was in refining weight distribution estimates rather than advancing vehicle tracking, as reliance on truck GPS data introduced limitations due to proprietary restrictions and incomplete fleet coverage.

Study Area

The study area comprises five sites in the Southern California region. Considering the location of the ALPR and inductive loop detectors, we designate I-710 Willow as the upstream site and SR-60 East End as the downstream site.

Study Site Setup

Two types of detector sites were used in this study: A pair of origin-destination control sites as well as three intermediate path coverage sites.

The origin-destination control sites comprise I-710 Willows and SR-60 East End, which are existing UCI Freight Mobility Living Laboratory (FML2) sites. These sites are equipped with Automated License Plate Readers (ALPRs) and inductive signature technology. The ALPR system at I-710 captures front vehicle license plates from the two slower lanes in the northbound direction, while the one at SR-60 captures plates in the two slower lanes in the eastbound direction (Figure 1). Both sites are also equipped with inductive loop signature technology across all lanes in both directions as a part of the UCI Truck Activity Monitoring System (TAMS), which provides real-time raw inductive loop signature data streams and detailed truck classification count data.

(a) I-710 @ Willows (b) SR-60 @ East End

Figure 1: ALPR Installations at the Origin-Destination Control Sites

The path coverage sites comprise SR-91 @ Carmenita, I-605 @ Slauson and SR-60 @ Albatross. SR-91 @ Carmenita and I-605 @ Slauson were previously deployed UCI TAMS sites and were re-initiated for this study. Another TAMS site on the SR-60 freeway (SR-60 @ Bixby) was originally planned for reinitialization, however the cabinet was found to be damaged, but most of the equipment was successfully relocated to a new location designated as SR-60 @Albatross (Figure 2).

(a) Damaged cabinet at SR-60 @ Bixby (b) Relocated setup at SR-60 @ Albatross

Figure 2: Relocation of equipment on SR-60 freeway

A fourth site on the SR-57 freeway (also a previously deployed TAMS site) was originally planned for this study near Diamond Bar exit. However, power at the traffic cabinet was down and no other locations were viable along the SR-57 freeway as a part of Route 91 (Figure 3).

Figure 3: No power at previously deployed SR-57 @ Diamond Bar TAMS site

Vehicles traveling through the control sites are generally expected to take either of the three routes listed below.

- Route 1 or "Route 91": Starting at I-710 Willow, through SR-91, through SR-57, through SR-60 and ending at SR-60 East End
- Route 2 or "Route 605": Starting at I-710 Willow, through I-605, through SR-60 and ending at SR-60 East End
- Route 3 or "Route 710": Starting at I-710 Willow, through SR-60, and ending at SR-60 East End.

To identify the route, three intermediate sites have been set up, which are SR-91 Carmenita, I-605 Slauson and SR-60 Albatross. Figure 4 shows the location of all the five sites and the three alternate routes.

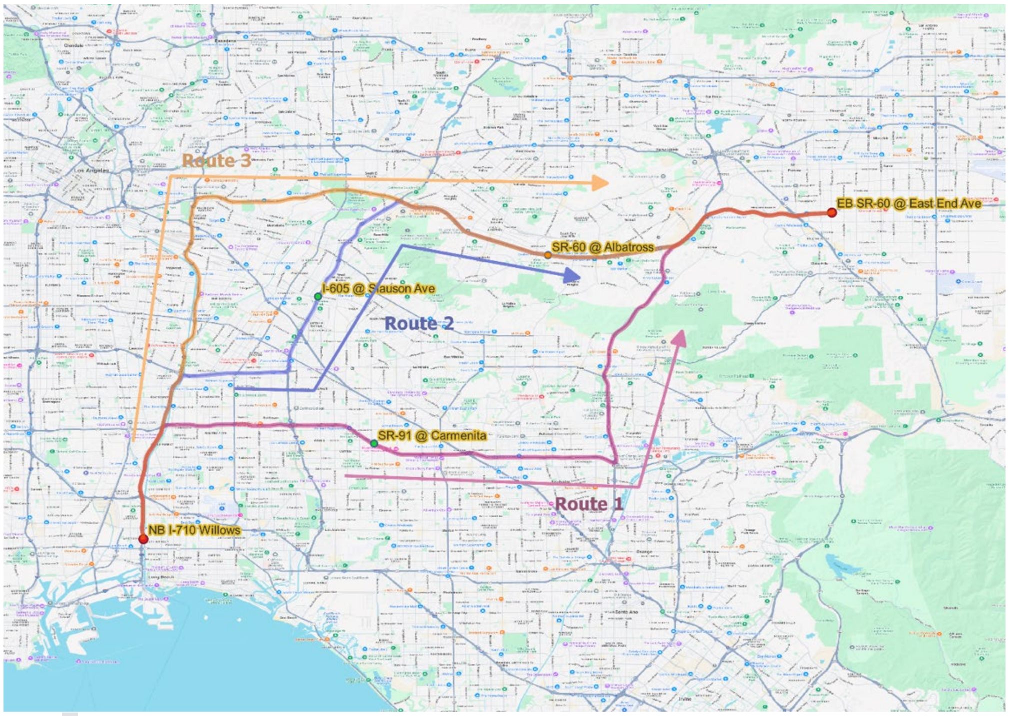


Figure 4: Location of Five Sites and the Three Alternate Routes in Study Network

The ALPR system provides data to facilitate license plate matching and confirm the origin-destination trip events.

Data

We used two sets of data in this project. For training the model data was obtained from the I-710 Willow and SR-60 East End sites for the month of February 2024. This dataset contains both ALPR and inductive loop detector data that are available from both of these sites. The second set of data was collected from 21 to 28 November 2024. We used this dataset to implement/test the model and track vehicles between the three alternative routes. Figure 5 and Figure 6 provide the average hourly volumes of trucks on weekdays and the respective average travel times observed in these two datasets. Both figures indicate that the most congested period is between noon to late afternoon, considering low truck volume and higher travel times.

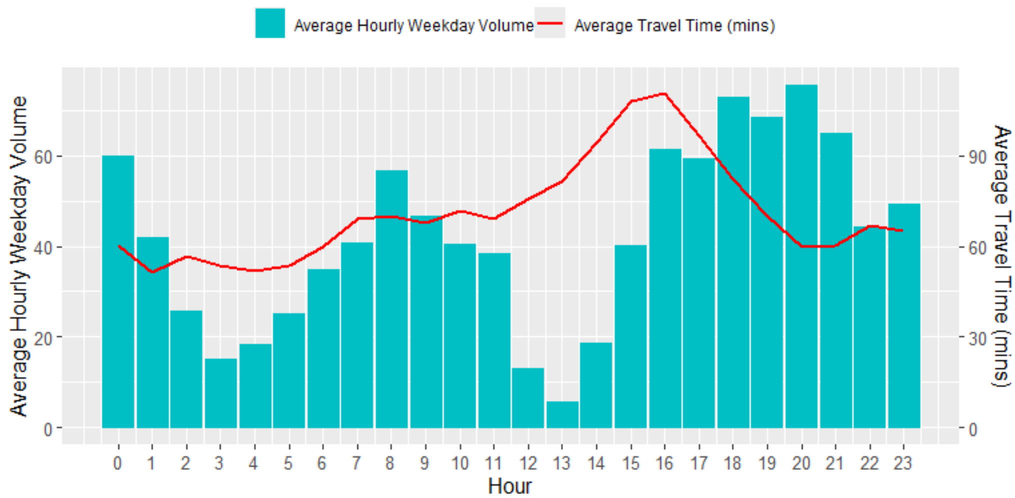


Figure 5: Hourly Distribution of Truck ALPR Volume and Average Travel Times in the Training Dataset

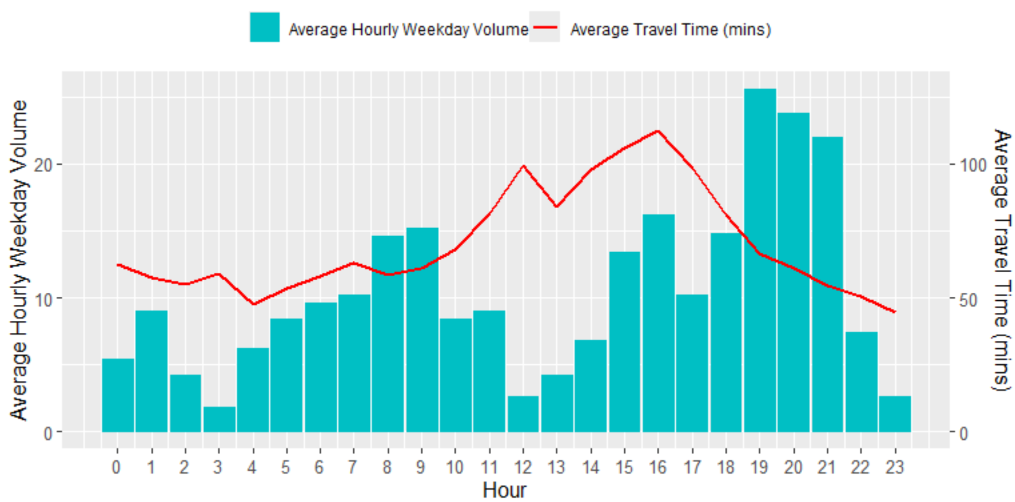


Figure 6: Hourly Distribution of Truck ALPR Volume and Average Travel Times in the Test Dataset

Vehicle Matching Model Development

In this section, we discuss the methodological approach of this project, including details on data collection, data pre-processing, and Bayesian model development for vehicle matching. Figure 7 illustrates the data flows and the output from each step of the methodology.

Data Collection

To develop the vehicle prediction model, we used the dataset from I-710 Willow and SR-91 East End sites as the upstream and downstream observations, respectively. Both of these sites are equipped with automated license plate recognition (ALPR) and inductive loop detector devices. Although there is only a few feet of distance between the two devices, we used a generous 10-second search window to capture all possible candidate vehicles at the loop detector for each license plate read by the ALPR. We matched the plates identified at the upstream ALPR site with the plates at the downstream ALPR site and obtained the corresponding inductive loop signature data from the downstream loop detector using the same 10-second threshold. Additionally, to enhance the model's capability in identifying matches from a large pool of unmatched vehicles, we used a 10-minute window for each ALPR record to establish signature candidates for building mismatched vehicle datasets. At the end, we have two datasets, one for I-710 Willow and another for SR-91 East End, containing daily observations of vehicles during the month of February 2024.

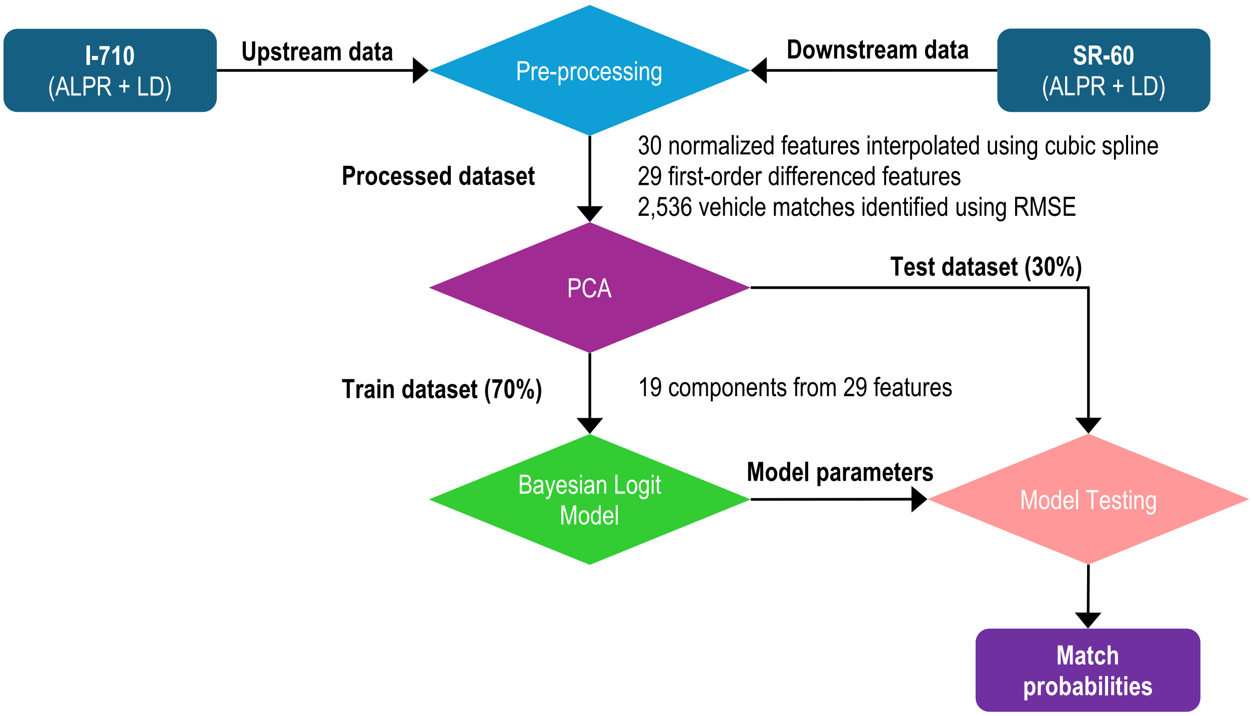


Figure 7: Flowchart Illustrating Data Processing and Modeling Process

Data Pre-Processing

The collected data was processed through several steps to make it suitable for developing Bayesian Logit Regression. In the first step, all vehicles observed in lanes other than the two rightmost lanes were removed. This filtering enhances the likelihood of aligning the ALPR data with the loop detector observations as the cameras were only configured to identify vehicles in the rightmost two lanes.

In the second step, only vehicles identified as belonging to FHWA classes 8, 9, and 10 are retained in the datasets. This class of vehicle represents tractor-trailers, which have very close signature profiles due to similar body configurations. Therefore, training the model to distinguish between these vehicles would allow for a higher sensitivity of the model parameters to the features that are crucial in identifying a match.

In the third step, both the signature and time values of each vehicle are normalized. Normalization of inductance value is required to remove the discrepancies in the electromagnetic sensitivity across different inductive loop detectors. On the other hand, the time axis is highly dependent on the speed of the vehicle. Hence, to adjust for the differences in speed, normalization of the time values is required.

In the fourth step, we interpolated 30 inductance values or features extracted from each vehicle's signature profile using a cubic spline. The inductive loop detector records signature values at 1-millisecond intervals. Considering the longer body length and slower speeds of the truck, the recorded number of inductance values can run into thousands. Using this large number of values to match vehicles can be computationally challenging and is less likely to provide any benefits since many of these values will be correlated. Hence, this step is important to extract a smaller number of potentially uncorrelated features that facilitate a computationally efficient and consistent matching process. At the end of this step, the resulting datasets for upstream and downstream vehicles consist of 30 interpolated normalized inductance values at the same normalized time intervals. Figure 8 illustrates the conversion of raw signature to normalized signature data and the 30 interpolated features.

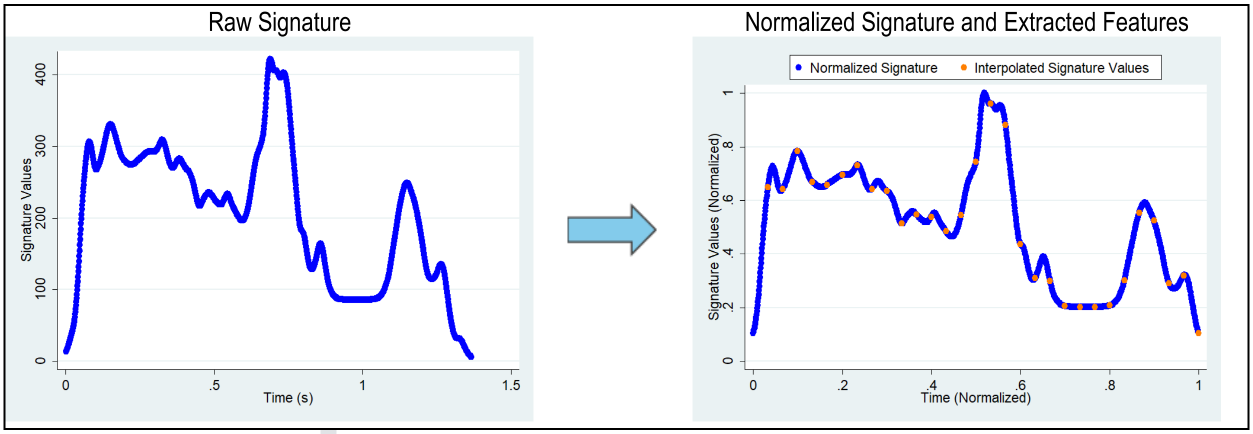


Figure 8: Signature Normalization and Interpolation Process

In the fifth step of data pre-processing, vehicle matches are identified between the two sites by calculating the root mean squared error (RMSE) of the interpolated features for each pair of candidate matches. To further reduce the impact of correlated features, we subtracted each inductance value from its adjacent value on the right to create 29 first-order differenced features. The calculation of RMSE is based on these 29 features, and the process identified 2,536 vehicle matches. A downstream vehicle is considered as a match with an upstream vehicle when the pair has the least RMSE value among all candidate matches with the upstream vehicle. Figure 9, shows the distribution of RMSE for the vehicles identified as matched and unmatched in the dataset.

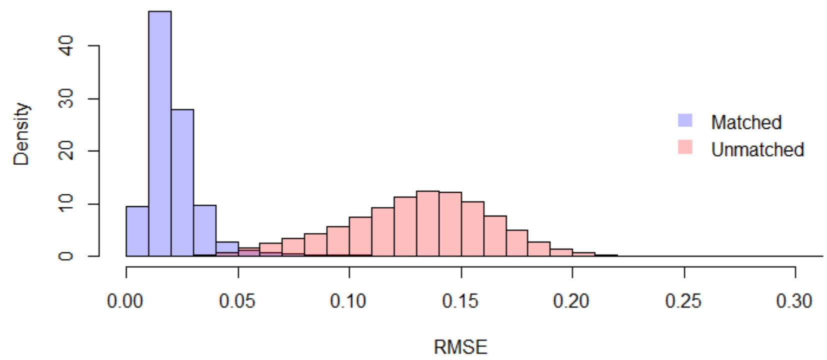


Figure 9: Distribution of RMSE for Matched and Unmatched Vehicles

As expected, the matched RMSE distribution is tighter with a lower mean than the unmatched RMSE, which has a significantly higher mean with a much wider distribution. This indicates the suitability of the dataset for training the model as it can make a large distinction between matched and unmatched vehicles.

Finally, in the sixth step, we performed a principal component analysis (PCA) on the 29 features. The primary aim of this step is to minimize the correlation between features by creating components that maximize the variance in the data (Palo et al., 2021). This is essential for the reliable estimation of the Bayesian model, which operates under the assumption of uncorrelated independent variables. Additionally, PCA serves to reduce feature dimensions through orthogonal transformation while preserving most of the variance and information within the dataset. As a result, it contributes to a better model fit and enhances computational efficiency. The resulting dataset contains each feature's principal component score, which is the projected value of the respective feature onto the principal component.

We then split the dataset into 70:30 ratios for model training and testing purposes, respectively. The training dataset contains 74,787 vehicle pairs, including 1,775 matched pairs, whereas the testing dataset contains 31,981 vehicle pairs, including 761 matched pairs. The two datasets are mutually exclusive in terms of vehicle ID, such that each dataset contains all the matched and unmatched pairs for the corresponding vehicles.

Bayesian Logit Regression

We used the prepared dataset to develop a Bayesian Logit model in order to distinguish between matched and unmatched vehicle pairs with a degree of certainty. Using a probabilistic model also allows for overcoming the measurement errors commonly seen in field data by estimating match through joint probabilities of the features. The logistic regression is suitable for modeling binomial responses, such as matched or unmatched vehicle pairs in our case. Eqn. 1 – Eqn. 3 provides the mathematical formulation of a Bayesian logit regression following the concepts and notation presented in Ncube (2023). Eqn. 1 presents the logit function used to model the binomial responses against the 19 components representing the 29 features.

$$\text{logit}\pi_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_{19} X_{i19} \tag{1}$$

where,

π_i = probability of match for vehicle pair i
α = intercept term
β = coefficients for the variables representing 19 features
X = 19 components representing 29 features.

In the Bayesian approach, the likelihood for a binomial model can be expressed by Eqn. 2.

$$\binom{n}{y} \pi^y (1-\pi)^{n-y} \tag{2}$$

where,

n = number of simulations
π = $g^{-1}(\mu)$ = probability of match for vehicle pair
μ = $\alpha + X^T\beta$

After incorporating logit link function of Eqn. 1 into Eqn. 2, the likelihood function transforms to Eqn. 3.

$$\binom{n}{y} (\frac{e^{\mu}}{1+e^{\mu}})^y (\frac{e^{\mu}}{1+e^{\mu}})^{n-y} \tag{3}$$

In the context of a Bayesian model, we need to define the prior distribution for the model parameters (i.e., intercept and coefficients for the components). The joint posterior distribution of the parameters is proportional to the product of the prior distribution of the likelihood function. We select the Cauchy distribution as prior for all model parameters, as suggested by Gelman et al. (2008), for logistic regression. The Cauchy distribution is characterized by thicker tails and an undefined mean or variance.

Results

Principal Components Analysis

From the results of PCA, we generated a scree plot that shows the percentage of variance in the 29 features explained by each of the 29 components or dimensions. The cumulative variance adds up to 100% at the 29th component. We see that by the 19th component, the proportion of explained variance exceeds 92%, which is sufficient for using these components in signature match identification. Therefore, we selected these 19 components as the independent variables for the Bayesian Logit model.

Figure 10 also shows that the first component itself explains 21.6% of the variance, which is significantly higher than any other individual component. Therefore, this component represents the most important features in terms of their ability to distinguish between signatures. To see which features have the most contribution to this component, we generated a squared cosine or cos2 plot in Figure 11. A detailed concept of squared cosine in PCA can be found in Abdi and Williams (2010). From Figure 11, feature 25 and 26 are the top two significant contributors to component 1. These features correspond to the rear part of the vehicle.

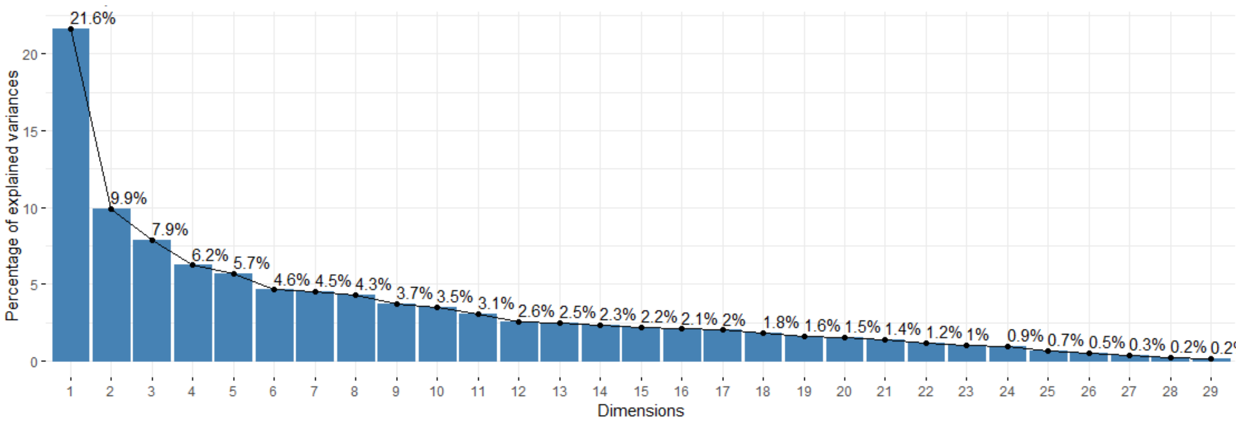


Figure 10: Scree Plot Showing the Percentage of Variance Explained by each Component

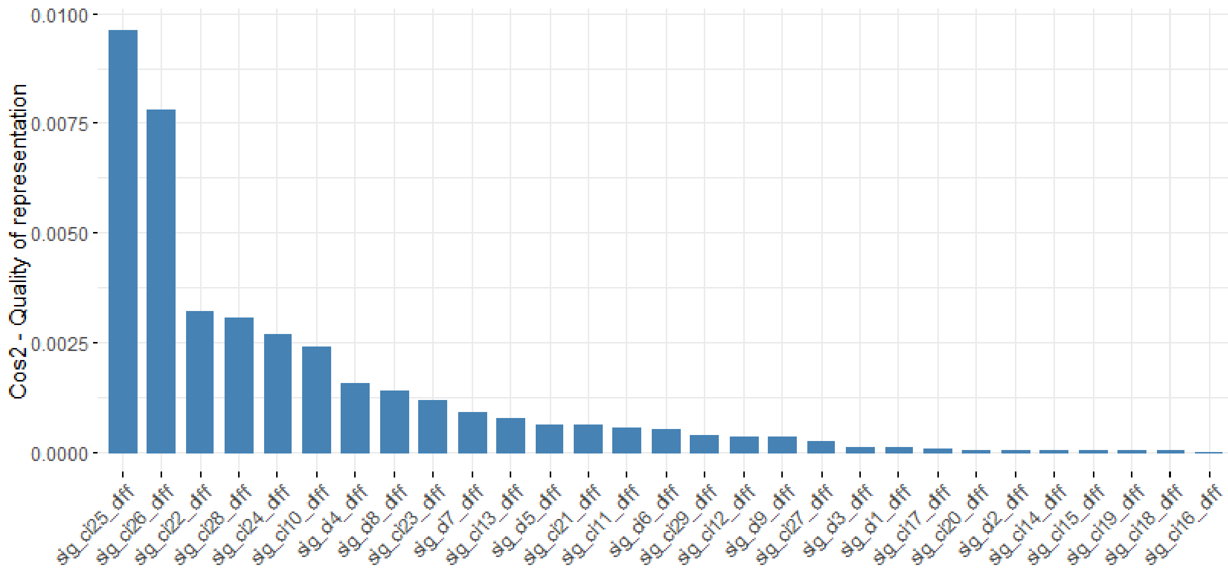


Figure 11: Cos2 Plot Showing the Quality of Representation of the Features in Component 1

Model Estimation Results

The Bayesian Logit model was estimated through the Markov Chain Monte Carlo (MCMC) simulations. After completing 4000 simulations, we checked the Monte Carlo Standard Error (MCSE) and the potential scale reduction (Rhat) to confirm the model goodness-of-fit. The MCSE indicates the error associated with the mean of estimated coefficients across the simulations. We obtained an MCSE value of zero at one decimal place, which indicates a good model fit. The Rhat is a converged diagnostic, which is equal to one when convergence is achieved (Roy, 2020). We obtained Rhat values of 1.0 for all of the coefficient estimates, suggesting the attainment of full convergence.

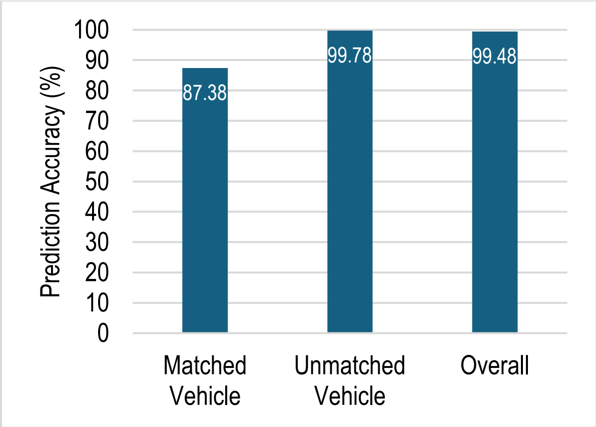
Table 1 provides the estimated coefficients from the Bayes Logit model for the intercept and the 19 components. Apart from the mean and standard deviation (S.D.) of the coefficients, the table also provides the 10th, 50th, and 90th percentiles for each of the distributions of estimated coefficients. The coefficients can be interpreted as the change in log odds by a unit change in each of the components. The sign on the coefficients indicates whether they contribute to the increase or decrease in the log odds. Since we used scores of principal components, rather than the actual difference in the inductance values, it is not possible to meaningfully interpret the relationship. Considering the magnitude of the coefficients, it appears that components 1 and 3 have a considerably higher effect on the log odds of a match than other components. On the other hand, component 17 has the least effect.

Table 1: Estimated Coefficients from the Bayesian Logit Model

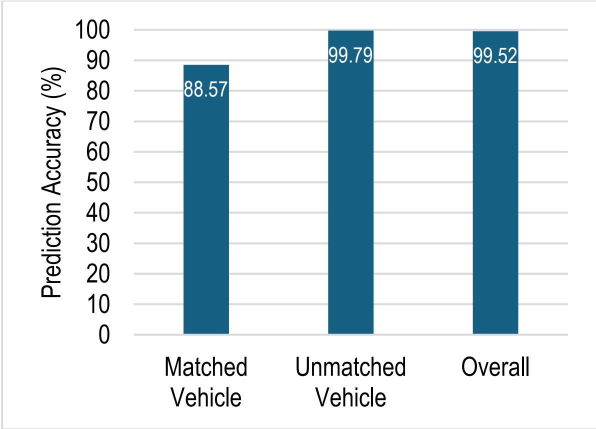
	Mean	S.D.	10%	50%	90%
(Intercept)	-14.3	0.4	-14.7	-14.3	-13.8
Component 1	-24.3	1.1	-25.7	-24.3	-23
Component 2	-14.9	1.4	-16.7	-14.9	-13.1
Component 3	-18.1	1.2	-19.7	-18.1	-16.5
Component 4	-11.2	1.9	-13.7	-11.3	-8.7
Component 5	1.9	1.4	0.2	1.7	3.8
Component 6	-4.7	1.9	-7.0	-4.7	-2.2
Component 7	-11.8	1.6	-13.8	-11.8	-9.7
Component 8	-3.2	1.8	-5.6	-3.1	-0.9
Component 9	-0.8	1.4	-2.6	-0.6	0.9
Component 10	1.0	1.5	-0.7	0.9	3.0
Component 11	1.4	1.4	-0.4	1.3	3.3
Component 12	-1.4	1.6	-3.5	-1.2	0.5
Component 13	-1.6	1.6	-3.7	-1.5	0.3
Component 14	10.4	1.9	8.0	10.4	13
Component 15	-0.7	1.6	-2.7	-0.7	1.1
Component 16	-3.6	2.0	-6.3	-3.6	-1.1
Component 17	0.5	1.6	-1.4	0.4	2.5
Component 18	0.8	1.5	-1.0	0.8	2.8
Component 19	-1.7	1.7	-4.0	-1.7	0.4

Model Validation

For model validation, we applied the estimated Bayesian Logit model to the training and testing datasets and calculated the probabilities of match from the predicted log odds. We identified a match if the probability exceeds 0.5 and all other pairs are considered as unmatched. The prediction accuracy is calculated as the proportion of cases in the model that identify a match or mismatch correctly. In Figure 12, we show the prediction accuracy for matched and unmatched vehicles separately as well as the overall prediction accuracy considering both correctly matched and unmatched vehicles. The result from the training dataset is shown in Figure 12 (a), and that of the testing dataset is shown in Figure 12 (b).



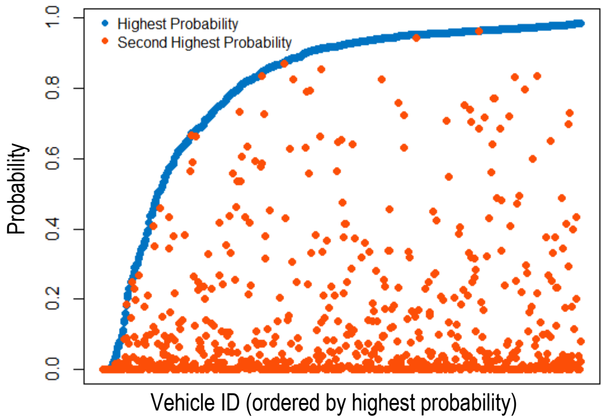
(a)



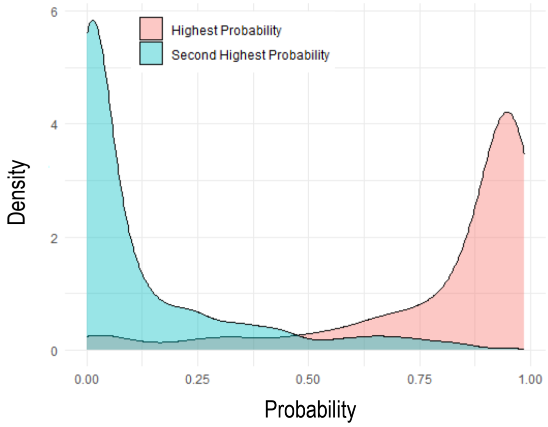
(b)

Figure 12: Prediction Accuracy of the Bayesian Logit Model

The predicted probabilities from the training dataset (Figure 12 (a)) show an 87.4% prediction accuracy for matched vehicles and 99.8% prediction accuracy for unmatched vehicles. Even with a testing dataset (Figure 12 (b)), we see similar and slightly higher prediction accuracy when the model is applied to the training dataset. To further verify the performance of the model in distinguishing between matched and unmatched vehicles, we compared the distribution of the highest and second-highest match probabilities (Figure 13). The objective is to determine how close are the second-highest probabilities to the highest probabilities, which designate a match. It is preferable to have a reasonably large difference between these two probabilities to avoid multiple close matches, which would reduce the reliability of the predictions.



(a)



(b)

Figure 13: Distribution of Highest and Second Highest Probabilities

Figure 13 (a) shows the scatterplot of the highest and second-highest probability with the x-axis ordered by the highest probability. We see that in only a few cases, the second-highest probability (red dots) is very close to the highest probability (blue dots), and most of the second-highest probabilities are less than 0.2, significantly farther away from the blue dots.

Figure 13 (b) shows the density plot of the highest and second-highest probabilities. This plot allows us to compare the means and the overlapping region. As expected from our analysis of the earlier plot, the means are quite distinct – the mean of the highest probabilities is close to one, and the mean of the second-highest probabilities is close to zero. We also see that the probability at which the two distributions intersect is close to 0.5. Hence, any match probability greater than 0.5 would indicate a significantly high chance of obtaining a correct match. The same argument can be made for probabilities less than 0.5 in the case of an unmatched vehicle.

Truck Tracking with Estimated Model

With the estimated and validated Bayesian Logit model, we tracked the trucks to identify the route they took between I-710 Willow and SR-60 Albatross. Considering the connectivity of the highway network, there are three possible routes between these upstream and downstream sites. Although these routes are discussed in the Study Area, we are presenting them again for reference.

- Route 1 or Route 91: Starting at I-710 Willow, through SR-91, through SR-57, through SR-60 and ending at SR-60 East End
- Route 2 or Route 605: Starting at I-710 Willow, through I-605, through SR-60 and ending at SR-60 East End
- Route 3 or Route 710: Starting at I-710 Willow, through SR-60, and ending at SR-60 East End.

We then applied the model to find matches between I-710 Willow and each of the intermediate sites, i.e., SR-91 Carmenita, I-605 Slauson, and SR-60 Albatross. We considered the following matching algorithm to identify routes for the vehicles observed both at sites I-710 Willow and SR-60 East End. To identify a preliminary match for routes, we used a 0.25 probability threshold considering the point of intersection in the probability density plots obtained from the three matches.

- Vehicles are assigned to Route 91 and Route 605 if it is observed at SR-91 Carmenita and I-605 Slauson, respectively, with at least 0.25 probability.
- Vehicles are assigned to Route 710 only if it is observed at SR-60 Albatross but not at I-605 with at least 0.25 probability.
- If multiple route assignments are found for a vehicle, it is assigned to the route with the highest probability.

Considering the aforementioned criteria, we were able to track 304 vehicle trips over the eight days. Among these vehicle trips, 48.7% are assigned to Route 91, 7.6% are assigned to Route 605, and 43.8% are assigned to Route 710. These values include the same vehicle observed over multiple days. Counting only unique vehicles, a total of 276 vehicles have been tracked out of the 727 vehicles identified between I-710 Willow and SR-60 East End using the ALPR dataset. Figure 14 presents the distribution of match probabilities in the three routes. It shows that all three routes have a wide range of match probabilities, with the majority of vehicles having a probability of less than 0.5. Among the three routes, Route 91 presents matches with the highest median probability of 0.42, whereas Route 710 presents matches with the lowest median probability of 0.35. Route 605 falls in between, with a median probability of 0.40. Considering the high prediction accuracy of the model with both train and test datasets, the overall low match probabilities in these routes are likely due to the variation in capturing the inductance by the loop detectors, which is an expected scenario when dealing with field data.

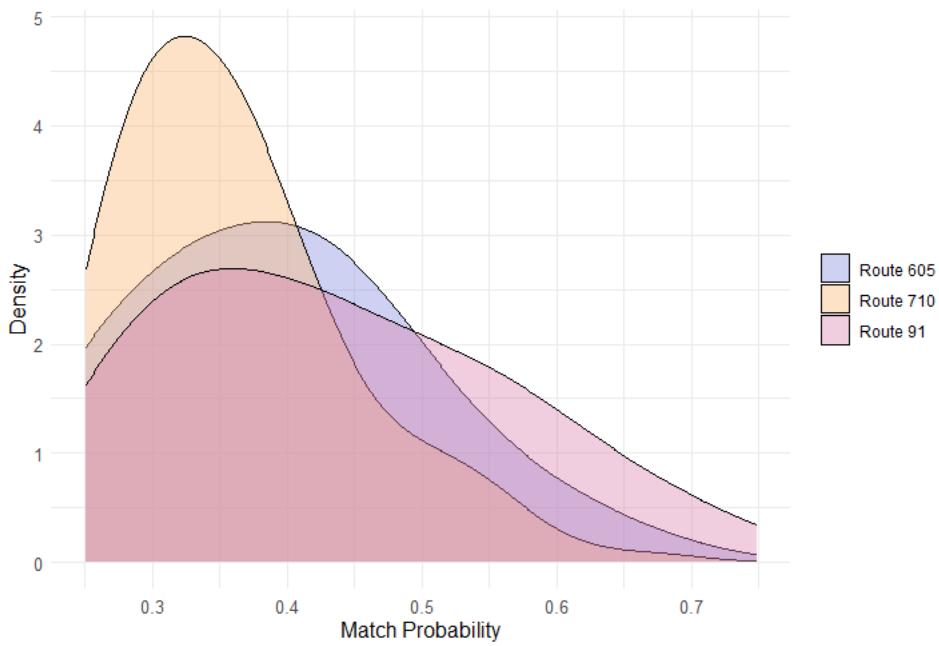


Figure 14: Distribution of Match Probability in the Three Routes

Figure 15 shows considerable differences in the hourly distribution of the 304 vehicle trips across the three routes during November 21-28, 2024. The *x*-axis denotes the hour when the vehicle passed the I-710 Willow site and the y-axis denotes the corresponding proportion of all the matched trips that passed through each of the three routes. For example, at the 4th hour, about 56%, 0%, and 43% of vehicle trips passed through Route 91, Route 605, and Route 710, respectively, between 4 am and 5 am. We did not find any matches for vehicles departing between 11 pm and midnight.

The distributions show more fluctuations in the choice of Route 91 and Route 710 compared to Route 605. Unlike Route 605, the other two routes have 100% assignment of trips, but at different times of the day. Route 91 has 100% assignment at 7 am, 10 am and 8 pm. On the other hand, Route 710 has 100% assignment between 12 to 1 am and at 11 am. Route 605 reaches the maximum assignment of about 25% at 4 pm.

Figure 15 clearly shows the high degree of preference for Route 91 and Route 710 throughout the day. If we look at the time segments when the proportions in these two routes surpass each other, we find that Route 91 has a high preference between 6 am and 10 am and between 4 pm and 11 pm. Similarly, Route 710 has a high preference between 12 am and 2 am and between 11 am and 1 pm. The highest preference for Route 605 is observed between 1 am and 4 am, and between 2 pm and 6 pm. Hence, Route 91 is preferred in the morning, evening, and night, Route 605 is chosen mainly in late night and late afternoon, and Route 710 is preferred in late night and mid-day.

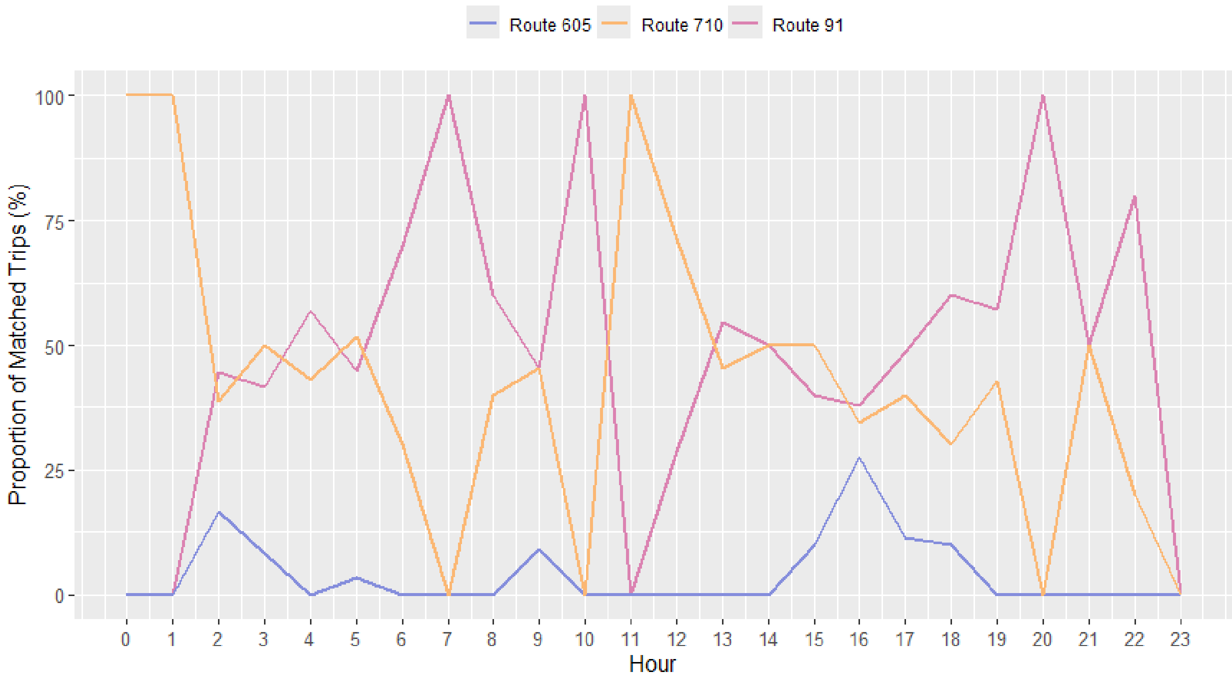


Figure 15: Hourly Distribution of Matched Trips Across Three Routes

Web Design Architecture for Truck Volume and Speed Visualization Dashboard

Design Overview

The web-based visualization dashboard is designed to:

- Display truck volume and speed dynamics across multiple road segments using a map.
- Provide user interaction through date selection and visualization content toggles.
- Animate hourly data with a progress bar and hour indicator.
- Handle edge cases like missing or zero data and dynamically adjust the visualization.

Dashboard Initialization

To avoid a CORS (Cross-Origin Resource Sharing) error when loading a local CSV file with D3 in your code, it is recommended to initialize the web-based visualization dashboard by starting a local development server using Python's built-in HTTP server. Follow these steps:

- Open a terminal in the folder containing both the HTML and CSV files;
- Run the command: python -m http.server or python3 -m http.server;
- Open a web browser and navigate to http://localhost:8000.

This will load the interface of the web-based dashboard, which should resemble the screenshot presented below:

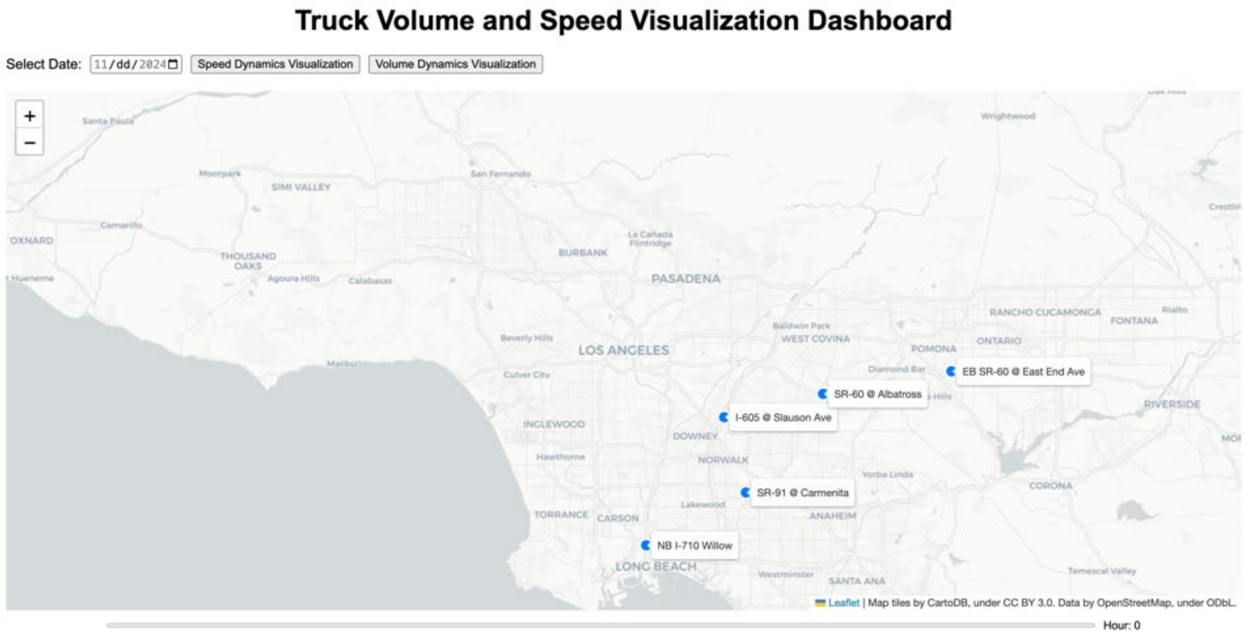


Figure 16: Screenshot Showing the Interface of the Web-Based Dashboard

Architecture Components

HTML Structure

The HTML defines the layout of the dashboard, including:

- A title for the page.
- Controls for user input (date picker and buttons).
- A map to display the visualizations.
- A progress bar and hour indicator for animation feedback.

JavaScript Logic

The JavaScript enables interactive and dynamic functionality, including:

- Loading and preparing the data.
- Handling animations for speed and volume visualization.
- Managing the map layers and dynamically updating visual elements.

Below are functional summaries for each of the key features:

- Data loading and initialization
  - Data is loaded from a local CSV file, which is derived by aggregating the matched vehicle data, using D3.js.
  - The date picker is constrained to align with the available data, ensuring that dates beyond the data range are unselectable (Figure 17).

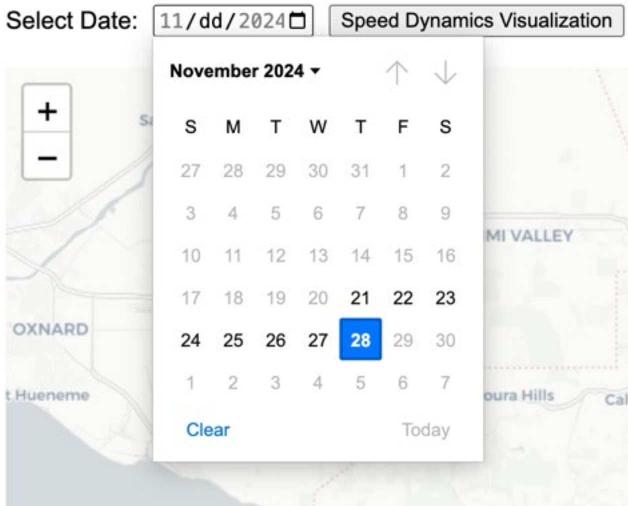


Figure 17: Screenshot of Date Picker

- A pop-up warning message is displayed if, upon clicking either visualization button, half of the selected data contains zero or missing values for either volume or speed across all corridors (Figure 18).

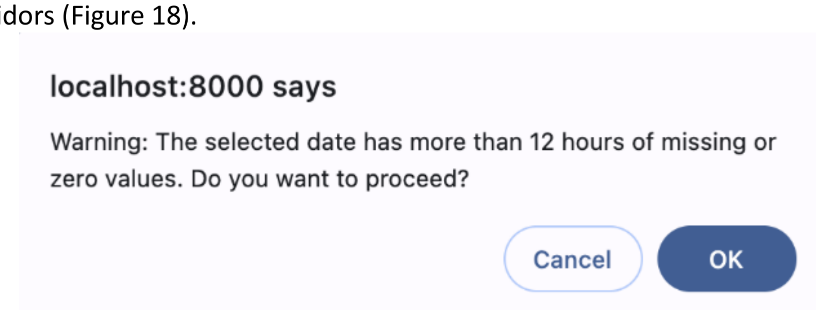


Figure 18: Screenshot of Error Message for Date with over 12 Hours Zero or Missing Data

- Base map and sensors
  - A light-colored Leaflet map is initialized to emphasize the animation.
  - All sensors are represented by filled dots, with their corresponding locations displayed adjacent to the dots
- Speed and volume animations
  - Speed is displayed as the hourly average, while volume is shown as hourly counts for each road segment.
  - Speed is represented using a red-orange-green color scheme with identical line widths, normalized by its maximum and non-zero minimum values (Figure 19). Similarly, volume is visualized with blue-colored lines of varying widths, also normalized by its maximum and non-zero minimum values (Figure 20).

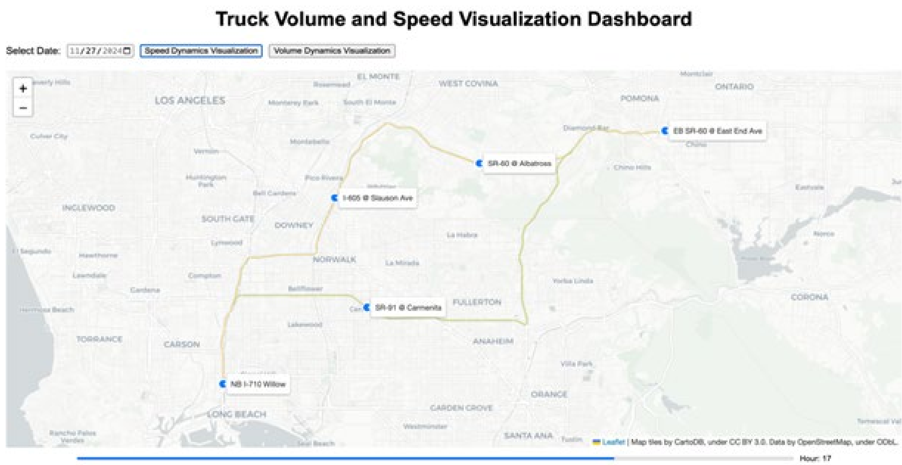


Figure 19: Representation of Speed Across Routes in the Dashboard

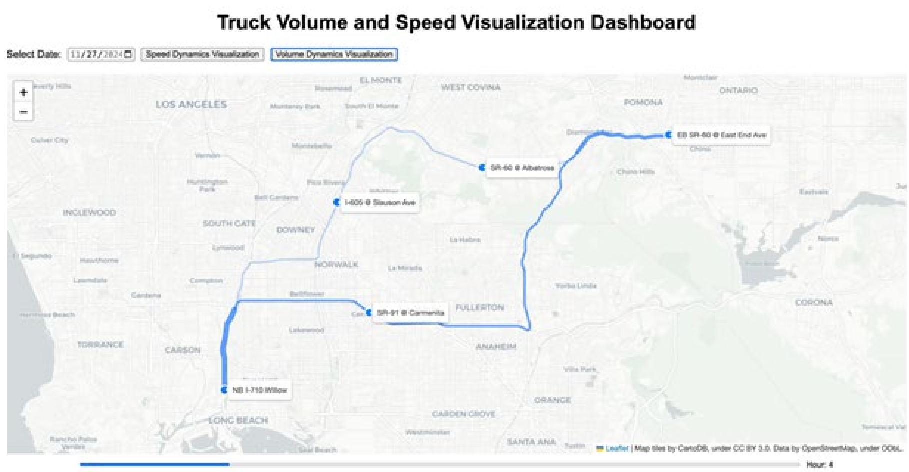


Figure 20: Representation of Volume Across Routes in the Dashboard

- For hours with zero or missing values on any road segment, no speed or volume animation will be displayed on the corresponding segment (Figure 21).

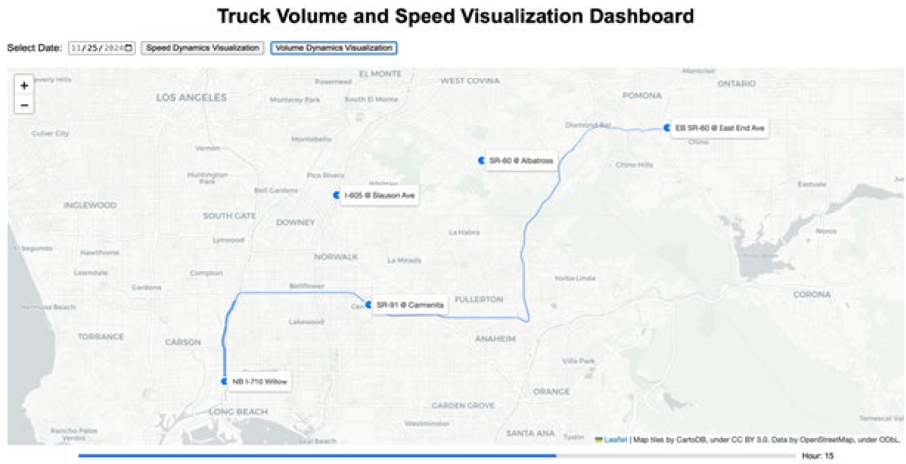


Figure 21: Representation of Missing Route Volume in the Dashboard

- Once the animation for all hours is completed, the map will be cleared, and no data will be displayed. To enhance the clarity of the animation, each hour is displayed for 1 second, with an additional 1-second pause before clearing the last hour (i.e., hour 23) display.
- The progress bar and hour indicator are updated dynamically to align with the animation timeline.

Conclusion

In this study, we aimed to develop and validate a Bayesian Logit model to evaluate the likelihood of a match between vehicles using signature data from inductive loop detectors. In addition, we used ALPR data to constrain the search window at the upstream and downstream sites, which helped to enhance the computational efficiency and reliability of the developed model. Our primary objectives included estimating the coefficients of the model, assessing the model's prediction accuracy, and tracking vehicle routes between selected sites.

The main findings from the analysis indicate that component 1, or the first-order difference between the 25th and 26th features, has the strongest influence on the log odds of a match, highlighting their critical roles in tracking vehicles. The model demonstrated considerable prediction accuracy, achieving 87.4% accuracy for matched vehicles and 99.8% for unmatched vehicles within the training dataset. The validation of the estimated model showed consistent performance across both training and testing datasets, suggesting robustness in its predictive capabilities. Moreover, our analysis of the relationship between the highest and second-highest match probabilities indicated a clear distinction between matches and mismatches, reinforcing the reliability of the model's predictions.

We matched 304 vehicle trips across three routes between I-710 Willow and SR-60 East End using the match probabilities from the estimated Bayesian Logit model. The results show considerable variations both in the overall route assignment proportions and also in the average hourly proportions across the three routes.

Despite these promising results, the study acknowledges several limitations. One significant limitation is due to the use of field data, which reduces accuracy in matches, especially for Route 710. Another limitation is the use of principal component scores, which may obscure the direct interpretation of the relationships with actual inductance values. Additionally, while the model effectively distinguishes between matched and unmatched vehicles, further investigations are necessary to evaluate its performance across diverse datasets and conditions. Moreover, the current framework is capable of identifying a single unique match per day, which limits analyzing route volumes as multiple trips by the same vehicle in a day cannot be distinguished.

In conclusion, while the Bayesian Logit model shows promise in accurately predicting vehicle matches, future work should focus on refining the model and exploring its applicability in broader scenarios considering more truck body configurations and multi-trips to enhance its effectiveness in real-world applications.

References

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley interdisciplinary reviews: computational statistics 2, 433–459. https://doi.org/10.1002/wics.101

Hernandez, S.V., Tok, A., Ritchie, S.G., 2016. Integration of Weigh-in-Motion (WIM) and inductive signature data for truck body classification. Transportation Research Part C: Emerging Technologies 68, 1–21. https://doi.org/doi.org/10.1016/j.trc.2016.03.003

Hyun, K., Tok, A., Ritchie, S.G., 2017. Long distance truck tracking from advanced point detectors using a selective weighted Bayesian model. Transportation Research Part C: Emerging Technologies 82, 24–42. https://doi.org/doi.org/10.1016/j.trc.2017.06.004

Kyung Hyun, 2016. Network-Wide Truck Tracking Using Advanced Point Detector Data (PhD Dissertation). University of California, Irvine.

Meta, S., Cinsdikici, M.G., 2010. Vehicle-Classification Algorithm Based on Component Analysis for Single-Loop Inductive Detector. IEEE Transactions on Vehicular Technology 59, 2795–2805. https://doi.org/doi.org/10.1109/TVT.2010.2049756

Ncube, B., 2023. Modeling Car Insurance Claim using Bayesian Logistic: Part 1.

Palo, H.K., Sahoo, S., Subudhi, A.K., 2021. Dimensionality Reduction Techniques:Principles, Benefits, and Limitations, in: Data Analytics in Bioinformatics Machine Learning and Genomic Technology, Feature Selection and Dimensionality Reduction. Scrivener Publishing LLC, pp. 79–107.

Roy, V., 2020. Convergence Diagnostics for Markov Chain Monte Carlo. Annual Review of Statistics and Its Application 7, 387–412.

Sarah Hernandez, Kyung Hyun, 2020. Fusion of weigh-in-motion and global positioning system data to estimate truck weight distributions at traffic count sites. Journal of Intelligent Transportation Systems 24, 201–215. https://doi.org/doi.org/10.1080/15472450.2019.1659793

Shin-Ting Jeng, Lianyu Chu, 2014. Tracking Heavy Vehicles Based on Weigh-In-Motion and Inductive Loop Signature Technologies. IEEE Transactions on Intelligent Transportation Systems 6, 632–641. https://doi.org/doi.org/10.1109/TITS.2014.2333003

Data Management Plan

https://www.metrans.org/assets/upload/PSR_DMP_Instructions.pdf

Products of Research
We collected ALPR and inductive signature data for the sites located at I-710 Willow and SR-60 East End for the period spanning the month of February 2024 and November 21-28, 2024. Additionally, we collected inductive signature data for the sites located at SR-91 Carmenita, I-605 Slauson and SR-60 Albatross for the period November 21-28, 2024.

Data Format and Content
All the data files are .csv or comma-separated values format segmented by vehicle ID. For privacy issues, the license plate column is replaced with generate vehicle ID that uniquely identifies each vehicle. Each file contains the merged ALPR and inductive signature data and contains the following columns: vehicle ID, site ID, lane direction, lane number, epoch of vehicle observation, inductance values, tier 2 class, tier 3 class, FHWA class and ID (based on lane direction, lane number and epoch).

Data Access and Sharing
Describe how the general public can access the data.

Reuse and Redistribution
State the restrictions on how the data can be reused and redistributed by the general public.