Multicriteria Planning Framework for Regional Intersection Improvement Using Telematics Data of Connected Vehicles

Swastik Khadka, S.M.ASCE[1]; and Pengfei "Taylor" Li, Ph.D., P.E., M.ASCE[2]

Abstract: This paper presents a novel approach to intersection improvement planning utilizing telematics data from connected vehicles to generate performance measures for mobility, safety, and emissions. Congestion, crashes, and emissions are three major issues in urban areas, particularly at intersections, and agencies often struggle to prioritize improvement plans because of a lack of objective data. Traditional infrastructure sensors provide limited information at selected locations, but it is not feasible to deploy them at all intersections. The use of telematics data from connected vehicles provides a high granularity of information on driving events and trajectories that can be used in conjunction with vehicle emission modeling to efficiently generate performance measures for all intersections. In a case study of over 300 intersections in Arlington, Texas, the *Pareto front* method was used to evaluate and rank intersections based on multiple criteria. Intersections falling on the Pareto front were identified as having at least one outstanding (poor) performance measure and were required to be given priority for improvement. The results were cross-validated with historical crash reports and the judgments of city traffic engineers, demonstrating the effectiveness of the proposed framework in generating objective and reliable intersection performance measures. This approach has the potential to significantly improve intersection safety, mobility, and environmental impact, and can serve as a valuable decision-support tool for transportation agencies. **DOI:** 10.1061/JUPDDM.UPENG-4705.

Author keywords: Connected vehicles data; Pareto front; Multiobjective optimization; Congestion management; Traffic safety improvement.

Introduction

An efficient and safe transportation system is critical to social welfare and economic development in modern society. Transportation management is considered a core function of most federal and local governments. However, it is becoming increasingly complex because of the rapid growth of travel demand and diversified transportation modes. Among all the congestion, crashes, and vehicle-caused pollution in metropolitan areas, a large portion occurs at intersections. As such, most agencies closely monitor traffic performance at intersections and conduct intersection improvement regularly. Intersection improvement cannot be performed for all intersections at the same time but must be planned over multiple years owing to staffing and cost concerns. This practice creates a need for prioritizing those in need of immediate improvement according to certain data-driven performance measures. Agencies face two challenges in performance evaluation: (1) insufficient sensor deployment at and resulting data from all intersections, because it is financially infeasible; (2) insufficient safety and air quality information at intersections, because the infrastructure sensors typically only connect mobility data (e.g., delays, queue lengths). In-vehicle GPS data are the most widely used source of information regarding vehicle movements. These data are obtained through GPS devices installed in vehicles, which record the longitude, latitude, speed, and direction of the vehicles. They are valuable for accurately mapping the vehicle's position within a network. Over time, as technology advances, in-vehicle GPS devices have become more sophisticated and capable of collecting additional data through vehicles' Controller Area Network bus (CAN bus). One such example is the telematics data from connected vehicles (CV), also known as connected vehicle data in other literature. This type of data includes traces and driving events that occur while vehicles are in motion. Unlike traditional in-vehicle GPS data, CV data encompass both movement data and driving event data. Similar to traditional data, CV data include information such as longitude, latitude, date, time, and direction. However, they go beyond that and include driving events such as harsh braking, harsh acceleration, driving above the speed limit or below the speed limit, vehicle emissions, and even data on whether the seat belt was latched. With the emerging telematics data of connected vehicles, the aforementioned two challenges are being resolved. Such information has been by default collected via in-vehicle sensors by most automobile manufacturers while some manufacturers have decided to distribute part of the collected data, such as vehicle trajectories, for external users to exploit more business values. The published telematics data were reported to cover almost all roads and be reliably available during most times of the day. They represent a 2%–6% penetration rate of all vehicles (Khadka et al. 2023).

The objective of this paper is to use emerging telematics data to explore a multicriteria planning framework for regional intersection improvement. Through telematics data reduction, aggregation, and integration with a vehicle emission model, which we call the Motor Vehicle Emission Simulator (MOVES) Lite, we design an efficient workflow to evaluate the performance measures in traffic mobility, safety, and vehicle emissions for each intersection with the scope. While crossing an intersection, the trajectories of connected vehicles can reveal the number of vehicles' stops, an indicator of traffic mobility; the abnormal driving events such as *hard brakes* indicate

[1]Graduate Research Assistant, Dept. of Civil Engineering, Univ. of Texas at Arlington, Arlington, TX 76019. ORCID: https://orcid.org/0000-0002-5323-3081. Email: swastik.khadka@mavs.uta.edu

[2]Assistant Professor, Dept. of Civil Engineering, Univ. of Texas at Arlington, Arlington, TX 76019 (corresponding author). ORCID: https://orcid.org/0000-0002-3833-5354. Email: taylor.li@uta.edu

Note. This manuscript was submitted on May 12, 2023; approved on November 2, 2023; published online on December 27, 2023. Discussion period open until May 27, 2024; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Urban Planning and Development***, © ASCE, ISSN 0733-9488.

the likelihood of crashes; applying vehicles' instantaneous speeds to the MOVES Lite (Zhou et al. 2015), vehicles' emissions like CO_2 can be estimated to indicate the level of vehicle-caused air pollutions. All intersections are then evaluated using a multicriteria framework. The contribution of this paper is that it exploits the potential of telematics data to provide a more comprehensive evaluation framework for intersection improvement planning. The rest of this paper is structured as follows. The relevant literature is reviewed and summarized; then the data processing procedure is described in detail, followed by the development of the methodology for a multicriteria evaluation. At last, we conducted a case study using the City of Arlington in Texas to provide a recommended list of prioritized intersections for improvement according to their performances in mobility, safety, and emissions. The results were cross-validated with the crash records published by the Texas Department of Transportation and verified by the city traffic professionals.

Literature Review

Road congestion analysis is vital for transportation agencies to swiftly address congestion, enhancing traffic flow. Ongoing research seeks congestion prevention solutions, with some studies utilizing big data or artificial intelligence (AI) for congestion prediction, while others assess vehicle numbers or real-time intersection images. These techniques usually involve analyzing speed, density, travel time, and delay. Traffic congestion characterization often categorizes traffic into smooth, normal, and congested states. The reviewed literature in this paper focus on the previous studies on traffic congestion mitigation or ranking, which can be classified into two main categories: parametric and nonparametric methods.

Parametric Approach

Parametric approaches in methodologies are characterized by their reliance on explicit assumptions regarding the statistical and distributional properties of the data to be analyzed. This reliance on fixed parameters and assumptions can be a limitation, as the assumptions may not always align with the true nature of the data. Parametric methods offer simplicity and reduced complexity, but accuracy may suffer if assumptions are violated. Linear regression is a widely used example of a parametric approach for prediction and estimation. Lee et al. (2015) employed big data processing to establish correlations between traffic congestion and weather. Daily weather was correlated with congestion via multiple linear regression. The approach encompassed full regression modeling, variable removal, and residual analysis. The final model achieved 84.6% accuracy, surpassing observations. Porikli and Xiaokun (2004) introduced a low-latency, unsupervised congestion estimation algorithm using moving picture experts group (MPEG) video. Gaussian Mixture Hidden Markov Models (GM-HMMs) identified congestion phases via compressed domain congestion signal extraction. Traffic was categorized as follows: empty, open flow, mild congestion, heavy congestion, and stopped (Porikli and Xiaokun 2004).

Nonparametric Approach

Compared with the parametric approach, the nonparametric approach is a statistical method that does not make assumptions about the characteristics of the data given to the model. Because it does not rely on any predetermined assumptions or specific parameters, this type of model is not affected by the problem of predictions or estimations deviating from the assumed conditions. Some commonly used nonparametric approaches include machine learning and decision trees. Lu and Cao (2003) employed fuzzy logic to evaluate traffic congestion using key traffic parameters. They used a neuro-fuzzy inference system to process simulation data, categorizing results from smooth flow to congestion. Similarly, Krause et al. (1996) employed a comparable classification method to identify congestion on variable-road-sign multilane roads. Variants of fuzzy logic, like adaptive neuro-fuzzy inference, demonstrated higher accuracy and broader applicability (Olayode et al. 2021). Thianniwet et al. (2009) introduced a decision tree model, using GPS, camera, and survey data, to categorize traffic as light, heavy, or jammed with 91% accuracy. To enhance accuracy, advanced methods like the Bayesian Network prediction approach by Liu et al. (2014) were developed, considering directional dependencies for stochastic traffic congestion assessment. Some studies employed machine learning on images for traffic analysis. Gao et al. (2021) used a CNN to define congestion with weather impact. Huang et al. (2020) predicted congestion from drone images using volume and speed. Kurniawan et al. (2018) achieved 89.5% accuracy in classifying road conditions with CCTV images. Shen and Chen (2009) used roadside sensor data with an learning vector quantization neural network (LVQ-NN) model to predict congestion accurately, while Gong and Fan (2018) employed movable probe vehicle data to enhance reliability in ranking congested freeway bottlenecks beyond fixed sensors. This emphasis on probe vehicle data continued with Tran Quang and Hoon Bae (2021), who innovatively integrated gradient descent, pooling, and probe vehicle data for 5-min intervals. This approach outperformed comparable models and facilitated real-time visualization of evolving urban traffic patterns.

Various neural network models have been employed for road congestion prediction. ShirMohammadi and Esmaeilpour (2020) introduced an artificial neural network (ANN) using speed data, outperforming traditional methods. Similarly, Mondal and Rehena (2019) used an ANN to classify congestion based on vehicle speed and density. The recurrent neural network (RNN) approach by Zhou et al. (2019) with attention outperformed standalone RNN in predicting regional congestion. Additionally, diverse methods have emerged: Shenghua et al. (2020) introduced a two-step random forest-based congestion prediction system, while Elleuch et al. (2017) combined decision trees and neural networks, accommodating real-time incidents for improved prediction. These approaches collectively demonstrate the significance of accurately estimating and predicting congestion ranking on roadways. Some researchers have used complex models as key tools for traffic congestion prediction. Zhang et al. (2019) introduced a deep neural network (DNN) capturing network correlations from video data, while Sun et al. (2019) favored long short-term memory (LSTM), gated recurrent units (GRU), and RNN over standard models for short- and long-term forecasts. These studies highlight the importance of sophisticated methods in addressing congestion challenges.

In addition to these mathematical and statistically driven methods, another popular approach in optimization problems has gained traction. Especially in the field of engineering, the concept of the Pareto front and its efficient solutions are being used widely. It allows us to handle many different goals at the same time, without having to dive deep into all the specifics of each parameter. This approach has also been used in transportation problems such as choosing routes (Xunxue et al. 2007), making fair ramp decisions (Meng and Khoo 2010), improving signal control (Jiao et al. 2016), and optimizing traffic signals (Stevanovic et al. 2013). The aforementioned studies on transportation show that using the Pareto front method has clear advantages for solving these kinds of problems. Some additional literature is reviewed and Table 1 provides a summary of the selected literature.

Table 1. Summary of selected literature

Source	Methodology	Input data source/type	Application	Efficiency
Tran Quang and Hoon Bae (2021)	A hybrid deep convolutional neural network model	GPS probe vehicle and QGIS data	Prediction of traffic congestion index	Achieved higher R^2 value and outperforms other algorithms
Porikli and Xiaokun (2004)	Markov model	The model was trained based on MPEG video data	Traffic congestion estimation	The precision rate is over 95%
Lu and Cao (2003)	Fuzzy inference model	Input data were collected from a simulation result	Congestion evaluation from traffic flow	This model defines a new index name LOC (level of congestion)
Krause et al. (1996)	Fuzzy inference model	Dedicated short range communications (DSRC) detectors for speed data	Traffic management system for multilane highway	—
Pongpaibool et al. (2007)	Adaptive fuzzy inference model	The traffic data from vehicle recognition and tracking software	Traffic evaluation system (three-level congestion)	Manually tuned fuzzy logic achieves 88.79% accuracy, whereas the adaptive version achieves only 75.43% accuracy
Thianniwet et al. (2009)	Decision tree model	Data were collected using GPS devices, a webcam, and an opinion survey.	Identify road traffic congestion level	Accuracy as high as 91.29%
Zhao et al. (2016)	ARIMA model	Video camera induction coil detector and speed tachymeter	Travel distance estimation (TDE)	The predicted results were more reliable
Cao et al. (2018)	GRU and interactive temporal recurrent convolution network (ITRCN)	Based on Yahoo! data set	Traffic flow prediction	Outperforms the conventional GRU and CNN method 0
Tian and Pan (2015)	Long short-term memory recurrent neural network (LSTM RNN)	Data set from Caltrans performance measurement system (PeMs)	Traffic state prediction	Both MAPE and RMSE are lowered the greatest
Gao et al. (2021)	Image-based traffic congestion estimation framework, with basic CNN	1,400 traffic images including 66,890 vehicles	Traffic congestion estimation	The proposed framework can perform traffic congestion computation and estimation directly

Methodology

According to previous studies, CV data are highly accurate and ubiquitous (Comert 2013). As such, the CV data alone can be used to generate comprehensive performance measures for regional traffic signal evaluation. Nonetheless, the challenges lie in how to make the CV data *actionable*. The raw CV data cover all kinds of roads, from interstates to local drivers, and can easily be too big to be processed with traditional tools like spreadsheets. Another issue is that CV data were collected everywhere, including parking lots and private driveways, while only those CV data collected on the public roads are relevant to traffic management. Therefore, the CV data must be preprocessed, including for data reduction and map-matching. The preprocessed data can be aggregated and further analyzed with

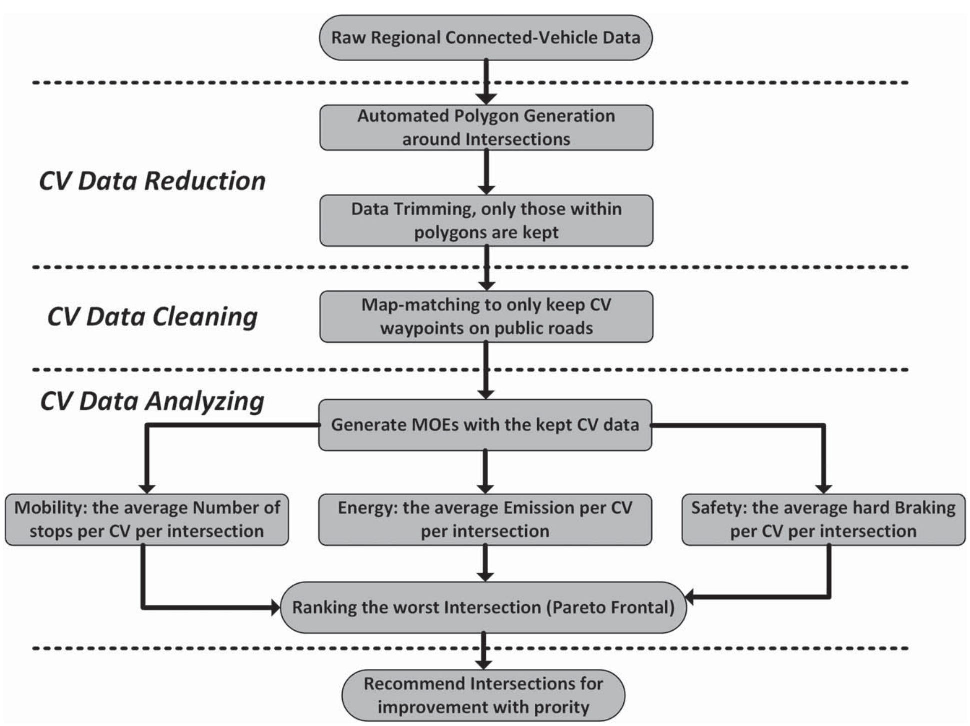


Fig. 1. Overall methodology to find the congested intersection.

traditional methods. To address these issues, we developed the multiobjective framework for traffic signal improvement planning following three sequential steps: (1) CV data reduction; (2) CV data map-matching; and (3) multiobjective ranking of signalized intersections.

Agencies usually procure the CV data for the entire region (statewide or citywide) while most projects focus on a few intersections/corridors. Therefore, it is necessary to quickly downsize the regional CV data to the target scope. Second, the reduced CV data will cover public roads, private roads, and parking facilities. It is necessary to further filter those data out of the public roads for the context. Third, multiple performance measures with the remaining CV data are estimated at each intersection and the intersections are cross-compared and ranked. The highly ranked intersections will be recommended for improvement as a priority. Fig. 1 shows the workflow of the proposed framework.

CV Data Processing

Automated Polygon Generation around Intersections

There are two ways to obtain the list of intersections in a region, from the geographic information system (GIS) managed by public agencies and from crowdsourced map engines, such as OpenStreetMap (OSM) (Haklay and Weber 2008). In this study, we used an open-source tool, referred to as *osm2gmns* (Lu and Zhou 2023) to retrieve the geolocations of all signalized intersections within the jurisdiction of the City of Arlington in Texas. The geolocation of each intersection's centroid was retrieved in the form of WGS84 coordinates: (latitude, longitude). Each intersection's scope was then extended from the centroid into a square geofence with an edge length of 305 m (e.g., 153 m upstream of the intersection centroid on all inbound roads). The intersection scope was selected this way to guarantee all the stops and hard brakes could be well captured when they were in queues. The queue lengths at intersections in Arlington, Texas, were almost shorter than 500 ft most of the time. Conversely, the intersection scope cannot be too large either because it may double counts stops for those closely spaced intersections. As illustrated in Fig. 2(a), if an intersection scope is too small, then some vehicles' controlled stops and hard brakes due to traffic signal operations may be ignored. If an intersection's scope is too large, then vehicles' stops and hard brakes may be overestimated. As illustrated in Fig. 2(b), V2's stops and hard brakes at the closely spaced upstream intersection may be counted by the downstream intersection because of the excessively large intersection scope. As such, the length of 500 ft is subject to changes according to local traffic conditions. The scope of each intersection at a region should be determined according to the prevailing spacing between intersections and peak-hour queue lengths.

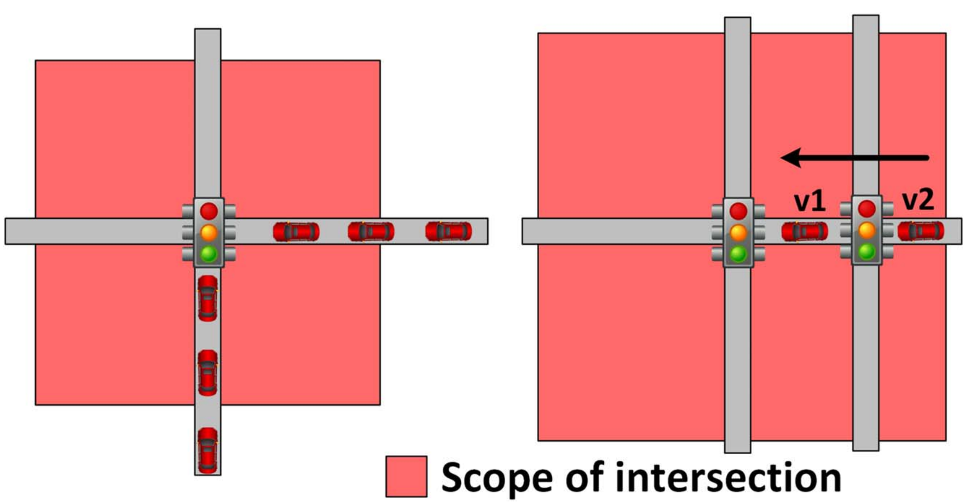
Data Trimming

In this step, it is necessary to trim the original data set covering the entire region or city and only keep data related to intersections. This step is to eliminate irrelevant waypoints and significantly reduce the total data size to be more manageable. Since millions or even billions of rows of the entire CV data need to be scanned, the trimming algorithm must be based on certain high-performance computing techniques. Khadka et al. (2022) described an efficient method to trim the data based on an open-source, off-the-shelf computer vision library. We suggest readers refer to that document for more details on data trimming.

Map-Matching to Further Clean the CV Data

After the data trimming, the remaining CV data must be further cleaned through map-matching. The map-matching technique was originally developed to mitigate the large positioning errors of certain early onboard global positioning system (GPS) receivers. As illustrated in Fig. 3, the reported waypoints may randomly be off the road. The assumption that drivers always adhere to traffic lanes while driving is not always accurate. Recent research has shed light on the impact of road layouts and traffic conditions on the dispersion of vehicle paths at intersections (Zhao et al. 2023). In that case, it is necessary to correct a wrongful waypoint by comparing it with its precedent waypoints and historically trace to match the wrong waypoint to the correct roadway link. The map-matching technique is critical to ensuring a stable and valid path in navigation systems.

Thanks to the advancement of GPS technologies, the CV data's positioning error can be almost ignored today. Nevertheless, the map-matching effort is still needed in this context because the CV data from the real world cover all available public and private roadways and park spaces. As illustrated in Fig. 4, the intersection scope covers both public roads and private spaces. The trimmed



(a) (b)

Fig. 2. (a) Data bias due to the smaller scope of intersections; and (b) data bias due to the larger scope of intersections.

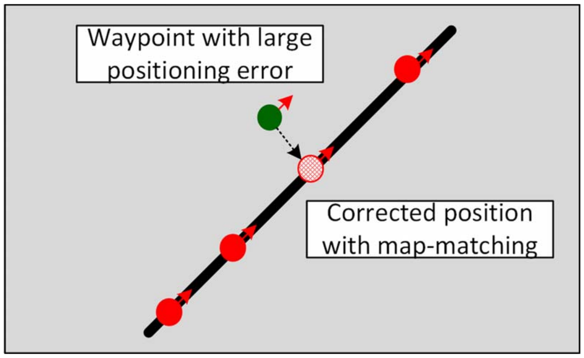


Fig. 3. Illustration of position correcting with map-matching.

CV data contains waypoints on both public roads and private driveways. Those waypoints collected within the irrelevant private spaces must be filtered out to avoid bias.

Multiobjective Intersection Ranking Framework

A multiobjective intersection performance ranking framework is developed in this paper. Based on the characteristics of CV data, three categories of performance measures are developed: (1) mobility performance: the average number of CV stops at each intersection; (2) safety performance: the average number of CV *hard brakes* at each intersection; and (3) energy performance: the average CO_2 emission of CVs, which is derived according to CVs' motions at intersections and the MOVES Lite emission model (Zhou et al. 2015). Previous research has provided evidence supporting the significance of the number of stops as a key factor in assessing the quality of service in coordinated systems. These stops, directly observed by drivers, enable the evaluation of signalized intersections' mobility aspects (Teply and Evans 1989). Furthermore, studies have indicated a correlation between a higher number of hard brakes and an increased occurrence of near misses, emphasizing the critical role of hard brakes in road safety (Subirats et al. 2010). Lastly, the rising concern over increasing CO_2 emissions, primarily from automobiles, has prompted the exploration of alternative solutions and the optimization of energy performance to address environmental safety (Change 2007). In the case study, a total of 383 signalized intersections in the City of Arlington, Texas were investigated. The aforementioned three performance measures were generated at all intersections. The three performance measures from the CV data were derived as follows.

Estimating the Average Number of Stops at Intersections

The average number of vehicle stops at intersections indicates the performance of traffic signal systems. More stops mean larger congestion, but it is necessary to point out that certain instantaneous stops (e.g., hard brakes) are not caused by traffic signal systems and should be separated from control-caused stops. To address

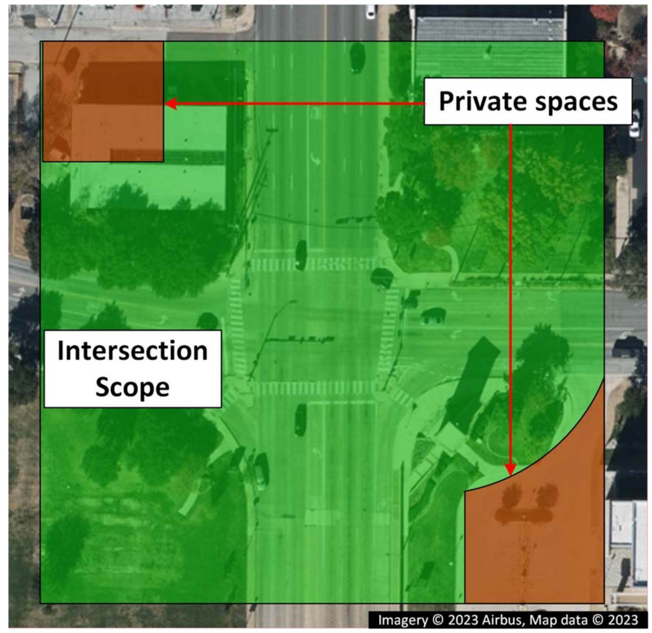


Fig. 4. Intersection scope with private driveways (The Cooper Street at the UTA Blvd, Arlington, Texas). (Image © 2023 Airbus, Map Data © 2023.)

these issues, we designed the following algorithm to tally CV stops at each intersection.

Algorithm 1. CV control stops capturing at intersections
For each CV trajectory at an intersection, examine the speed of each waypoint. If an instantaneous speed is slower than 8 KPH per hour (MPH) or 7.33 feet per second (ft/s), then examine the instantaneous speeds of the following waypoints. If a vehicle's low speed (less than 5 MPH) was kept equal to or longer than 6 s, then a control stop is identified.

A control stop is considered finished once the instantaneous speed becomes higher than 8 KPH. This algorithm can exclude the frequent stop-and-go behaviors while the queuing vehicles slightly reduce their separations from their front vehicles. Such maneuvers typically do not last longer than 6 s.

After the total stops of all CV trajectories are tallied at an intersection, the average number of stops is normalized by dividing the total number of stops by the total number of connecting vehicles passing that intersection. The following formula can be employed to compute the average number of stops:

$$S_{Ai} = \frac{\sum_{j=0}^{N} S_j}{N} \tag{1}$$

where S_{Ai} = average number of stops per vehicle at intersection i; N = total number of vehicles; and S_j = the number of stops for jth vehicle

Table 2. Average emission rate for zero-age passenger cars

Operating mode	Energy (kJ/h)	CO_2 (g/h)	NO_x (g/h)	CO (g/h)	HC (g/h)
0	49,206	3,536	0.05	2.37	0.04
1	45,521	3,271	0.01	4.06	0.00
…	…	…	…	…	…
40	641,649	46,113	14.34	407.60	2.73

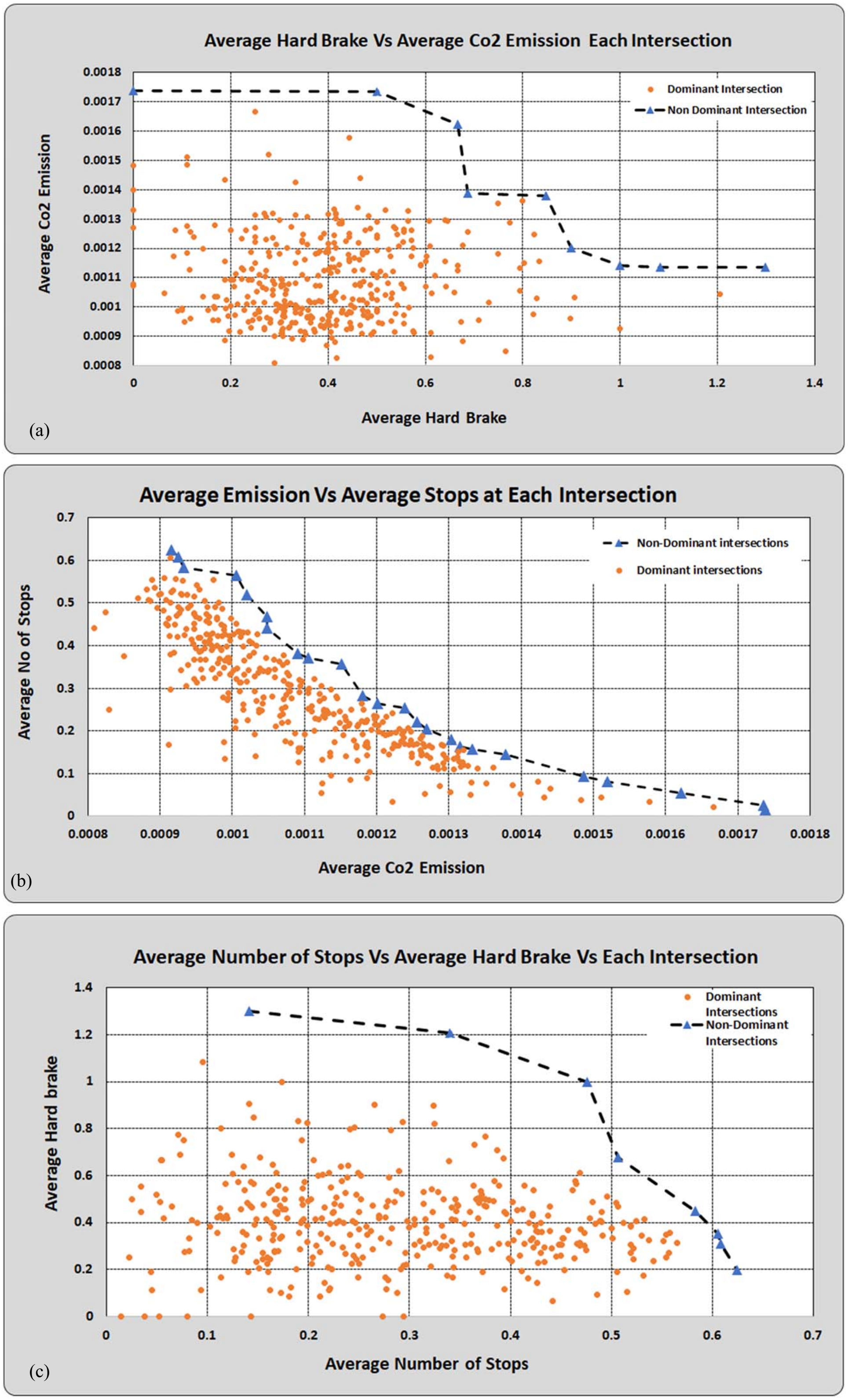


Fig. 5. (a) 2D Pareto fronts with the pairs of average hard brake versus average CO_2 emission; (b) average CO_2 emission versus average stops; and (c) average number of stops versus average hard brake based on the daily CV data set.

Discussion

Another popular performance indicator is the average vehicle delay at intersections. However, we found two hurdles in practice. The OSM does not provide reliable links to free-flow speeds. It would be necessary to collect that information for public agencies and it may not be available. Furthermore, the estimation of delay for connected vehicles at intersections is affected by the turning maneuvers they perform. The predetermined reduced turning speeds commonly used are often arbitrary and different from the

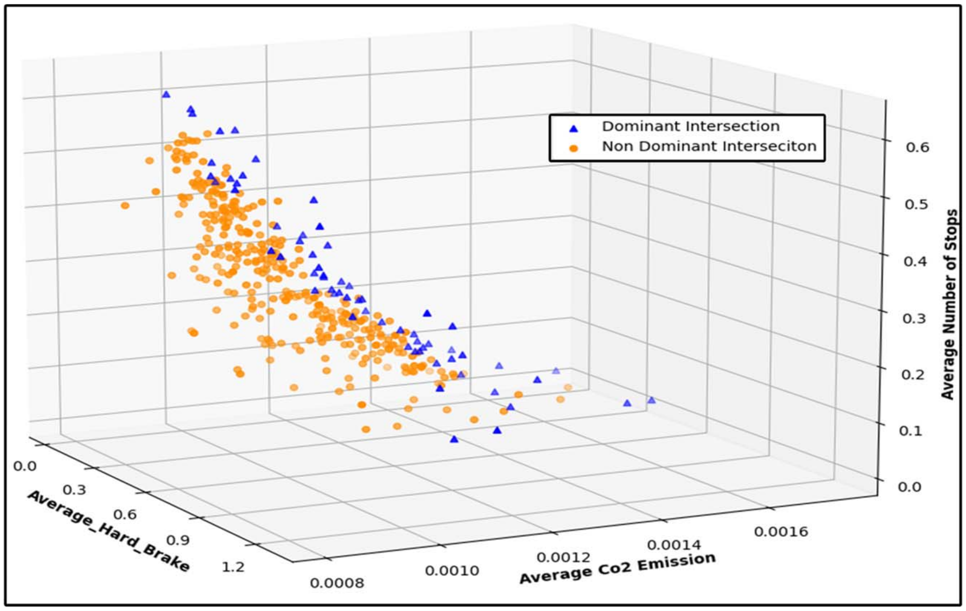


Fig. 6. 3D Pareto front with three performance measures based on the daily CV data set.

field observation. These facts can have an adverse impact on delay estimation. For instance, many drivers tend to compensate for delays by speeding up, resulting in most drivers consistently exceeding the speed limit by 8–16 KPH. Consequently, obtaining an accurate measurement of the true delay becomes challenging. Additionally, there is a lack of standardized benchmarks for turning maneuvers, making it difficult to accurately assess controlled delay by considering a specific turning movement speed. The high variability in speed reduction during turning movements further complicates the correct extraction of the control delay component from the overall delay. Considering these factors, it is more accurate to use vehicle stops as a representation of control delay at intersections. As such, we selected the vehicle stops over the vehicle delays.

Estimating the Average Number of Hard Brakes

The CV data contain the telematics data from vehicles such as hard brakes and hard accelerations. When a vehicle's speed change rate exceeds the threshold, hard braking and harsh acceleration events are generated and logged in the data set. Intuitively, if there are many hard brakes in certain areas, it would indicate excessive stop-and-go. Hence, near misses and crashes are more likely to occur. Therefore, we use the hard brakes as a proxy for crashes to represent the safety performance. The thresholds for hard brake events are programmatically incorporated into the vehicle's system, and an event is logged when the deceleration exceeds the designated threshold of 8.75 ft/s^2 (Wejo Data Service 2021). The hard braking events compiled for a specific intersection are calculated as follows:

$$\mathrm{HB}_{Ai} = \frac{\sum_{j=0}^{N} \mathrm{HB}_j}{N} \quad (2)$$

where HB_{A_i} = average number of hard brakes at intersection i; N = total number of vehicles; and HB_j = number of hard brakes of vehicle j.

Note that the hard brake events are also correlated with the control stops. One control stop may contain hard brake(s). Nonetheless, the directly reported hard brakes from the telematics data based on high-granular in-vehicle communication will be more reliable than those derived from the relatively low-granular vehicle waypoints.

Estimating the Average CO_2 Emission

Air quality is critical to maintaining clean and sustainable cities and intersections are where a large portion of air pollution by vehicles is generated. Therefore, it is important to evaluate the air quality performance at intersections to see whether the air quality performance should be addressed as a priority. We estimated the CO_2 emissions at intersections according to vehicle trajectories and motion-driven vehicle emission models. It would be necessary to include an environmental component in this process so that agencies can prioritize intersections that pose a significant environmental concern. Without loss of generality, carbon dioxide (CO_2) is a major greenhouse gas and can reflect the entire emission conditions at intersections. In 2004, the United States reported almost 33% of CO_2 emission, of which 80% came from the automobile industry (Change 2007). We compared variants of emission models to calculate the CO_2 emission from vehicles in the roadway network and developed the MOVES Lite model Zhou et al. (2015), which is based on MOVES, the state-of-the-art emissions modeling system developed by the US Environmental Protection Agency (USEPA) (USEPA 2010). The MOVES Lite is a light variant of the MOVES model. It uses the vehicle specific power (VSP)-to-operating mode conversation table and considers the average emission rates based on vehicle type and operating mode. To calculate VSP, Eq. (3) is used, adopted from the MOVES model (USEPA 2010), as follows:

$$\mathrm{VSP} = \left(\frac{A}{M}\right)^{*} v + \left(\frac{B}{M}\right)^{*} v^2 + \left(\frac{C}{M}\right)^{*} v^3 + (a + \sin(\varnothing))^{*} v \quad (3)$$

where, A = rolling term (t); B = rotating term [t/(m/s)]; C = drag term [t/(m/s^2)]; M = vehicle mass (t); v = vehicle speed (m/s); a = vehicle acceleration (m/s^2); and $\varnothing$ = road grade.

MOVES Lite first calculates the VSP for each vehicle using its corresponding operating parameters. With the combination of speed data and the calculated VSP, the appropriate operating modes are distinguished. Finally, using the following emission

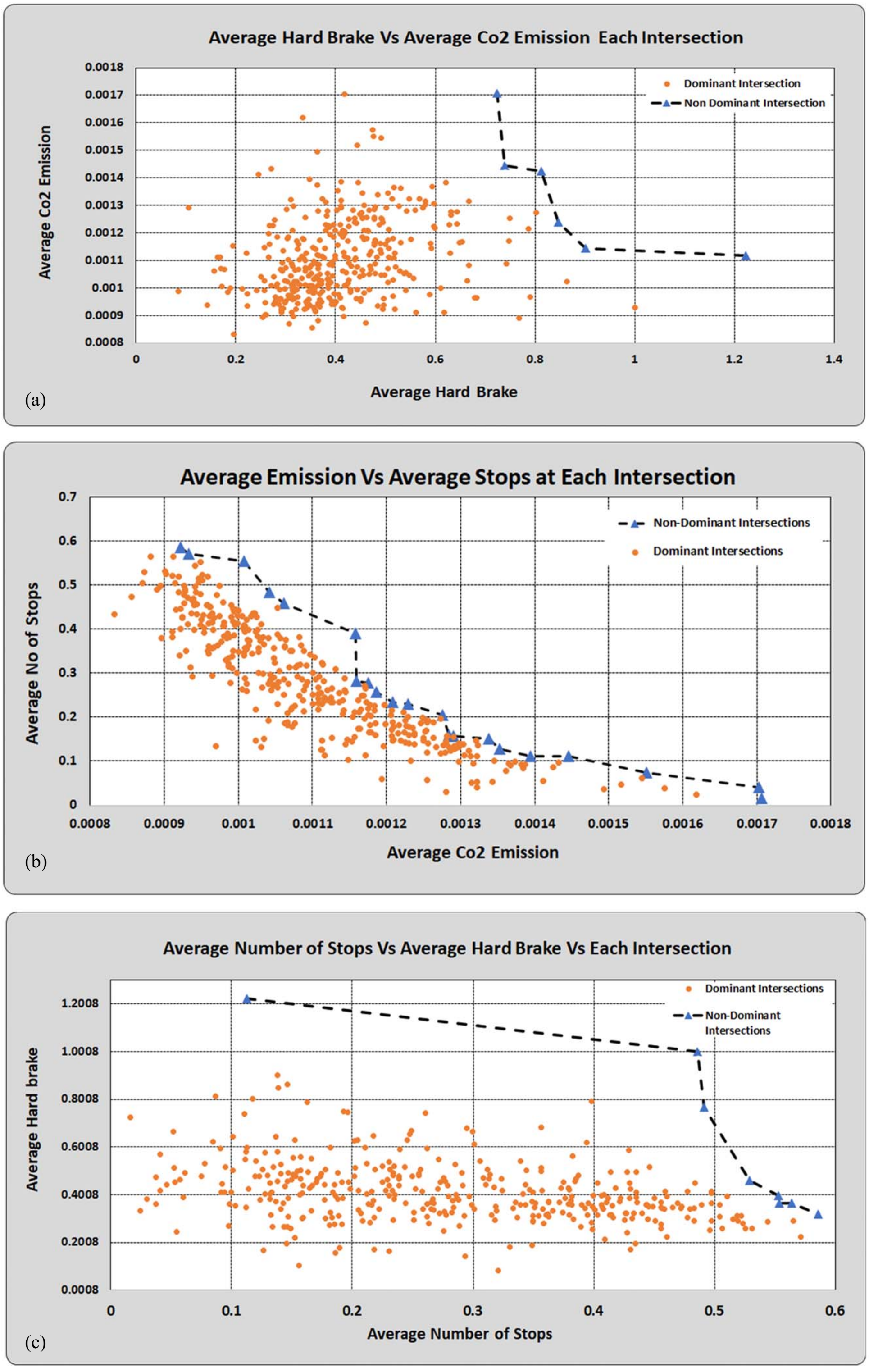


Fig. 7. (a) 2D Pareto fronts with the pairs of average hard brake versus average CO_2 emission; (b) average CO_2 emission versus average stops; and (c) average number of stops versus average hard brake based on the monthly CV data set.

Table 2, vehicle emission based on its operating mode, age, and type is obtained (Frey and Liu 2013).

The aforementioned emission model is applied to each CV trajectory where emissions are calculated every 3 s using the MOVES Lite model. Once the CO_2 emission has been determined, the total amount of CO_2 emissions is averaged over a given time (e.g., a day or a month) for each intersection and used as an emission performance indicator in the following multiobjective ranking process.

Identifying High-Priority Intersections with the Pareto Front Method

In terms of categories, this task falls into the category of multiobjective optimization. We employ the concept of the *Pareto front* to determine the best or worst points (i.e., intersections) between competing objectives. The Pareto front method is commonly used in engineering and planning to determine the optimal

J. Urban Plann. Dev., 2024, 150(1): 04023063

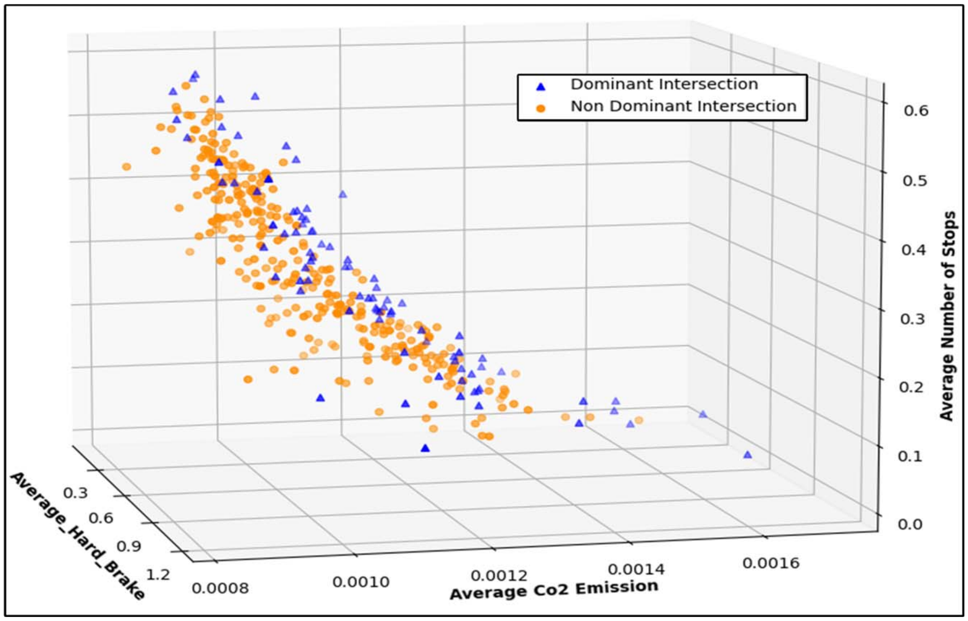


Fig. 8. 3D Pareto front with three performance measures based on the monthly CV data set.

solution(s) to a problem by comparing multiple objectives. The Pareto front is a set of solutions that are *Pareto efficient* where at least one objective is optimal while the other objectives are acceptable. In this study, we seek to identify the worst intersections (the opposite of best intersections) in a region by maximizing all three objectives: the number of stops, the number of hard brakes, and emissions using the Pareto front method. Intersections off the Pareto front will have at least one type of performance measure that is not as poor as those on the Pareto front and so they do not enter the high-priority intersection list.

The high-priority intersections are identified through three pairs of objectives. To compare intersection i, j with two performance measures $\boldsymbol{X}$, $\boldsymbol{Y}$, denoted as (X_i, Y_i)and (X_j, Y_j), if $X_i \geq X_j$ and $Y_i \geq Y_j$ (larger values mean worse), then intersection i with performance (X_i, Y_i) is said to dominate j with performance (X_j, Y_j). All intersections are compared in this way. If there are no intersections dominating intersection i, then i is a dominating intersection. The dominating intersections are identified in three rounds according to the three performance measure pairs: (*control stops versus hard brakes*), (*control stops versus* CO_2 *emission*), and (*hard brakes versus* CO_2 *emission*).

The three-objective Pareto front is also conducted using three performance measures altogether. The idea of comparing one intersection to the others is to check for dominance. It is similar to the aforementioned two-objective Pareto front method, but three performance measures should be compared.

Table 3. Intersections with high-priority need for improvement

Intersection ID	Location	Nearby area
2	S. Cooper St. and New Center Dr.	Interchange
17	E. Arkansas ln. and Forum Dr.	Industries and Warehouses
37	W. Tucker Blvd. and S. Bowen Rd.	Residential
100	New York Ave. and E I20 Frontage Rd.	Interchange
214	Green Oaks Blvd. and W I20 Frontage Rd.	Interchange
352	S. Watson and Davis St.	Interchange

The algorithm identifying dominating intersections using the Pareto front method is described as follows:

Algorithm 2. The algorithm to identify dominating intersections using the Pareto front

Denote two intersections' indices as i**, **j**;** X_i**, the average hard brakes;** Y_i**, the average CO_2 emission;** Z_i**, the average number of stops at intersection** $_i$. *The same notations apply to intersection* j. *Store (e.g., plot) the intersections as a list of dots with* X-, Y-*, and* Z-*coordinates in the performance planes or space.***:**

Initialize two empty lists for dominating and nondominating intersections, respectively.

Iterate over the length of the list created for *X*-coordinates.

For each *i* in all intersections

Set the initial state of intersection *i* with the coordinates of (X_i, Y_i, Z_i) **as "dominating"**

FOR any other intersection *j* with coordinates (X_j, Y_j, Z_j)

IF $X_i <= X_j$ **AND** $Y_i <= Y_j$ **AND** $Z_i <= Z_j$

Change the state of intersection *i* to "nondominating"

Break the FOR LOOP

End

End

Case Study: Prioritizing Intersections for Improvement Using the CV Data in Arlington, Texas

In the case study, we conducted a multiobjective intersection ranking with the CV data and the Pareto front method. Those intersections dominating all three performances were identified as high-priority intersections for improvement. Using the OSM platform, 381 signalized intersections were identified within the jurisdiction of Arlington, Texas. At each identified intersection, a square-shaped polygon was automatically drawn to cover around 500 ft back from the stop lines. After performing the data trimming, the

J. Urban Plann. Dev., 2024, 150(1): 04023063

Table 4. Quantitative comparison of intersections based on their objectives

Intersection ID	Average number of hard brake	Average CO_2 emission	Average number of stops	Dominating factor
2	0.90	0.001	0.14	Hard brake
17	0.65	0.001	0.25	CO_2 emissions
37	0.33	0.001	0.32	Number of stops
100	0.48	0.001	0.39	Number of stops
214	0.142	0.001	0.29	Number of stops
352	0.38	0.001	0.13	CO_2 emissions

CV data points were further screened out using the map-matching techniques and the roadway network obtained from the local agency. Each vehicle and its route information were verified by comparing the reported vehicle headings, or directions, and the link headings. The case study was divided into two parts: (1) Intersection prioritization using 1 day of CV data and (2) Intersection prioritization using 1 month of CV data. The two experiments aim to examine if the proposed framework can be used to help traffic operators identify the most problematic intersections using the daily incoming CV data as well as help the planners determine a strategic plan for intersections improvement in the region.

Identifying the High-Priority Intersections for Improvement Using Daily CV Data

We first processed the CV data collected on September 1, 2021, for the first experiment. The available CV data covers the Dallas–Fort Worth area in Texas, representing 2%–6% of all vehicles. The CV data set includes the movement data with which we derived the number of stops and the event data from which we retrieved the hard brakes. After the data trimming and map-matching, 368,236 vehicle trips were observed passing all intersections. At each intersection, the number of stops was estimated according to vehicles' speeds at intersections; the hard brake events were extracted from event data; and the emissions were estimated using the MOVES Lite model. In the original CV data set, the indicators used in the two-dimensional (2D) and three-dimensional (3D) Pareto front analysis were absolute. For instance, a vehicle may have recorded a total of three instances of hard braking throughout its entire trip. However, in the case study, the hard braking events for each vehicle were aggregated and then averaged for each intersection. As a result, the indicators used in the analysis became relative. It is important to note that all the indicators were averaged and calculated on a per-vehicle basis. From the event data set, we identified 193,510 hard brake events for September 1, 2021, within the city scope.

While conducting the Pareto front analysis, we first constructed three 2D Pareto fronts using two performance measures each time. Using Algorithm 2, the dominating and nondominating intersections were identified for each Pareto front. Fig. 5(a) compares the performance of average hard brakes and CO_2 emissions. Each dot represents an intersection. The dashed line and the dots on it indicate dominating intersections that at least have one performance worse than any other intersections. Those intersections on the front were problematic concerning either or both performance measures: hard brakes and average CO_2. In Fig. 5-(a), 9 out of 383 intersections were identified as dominating and they enter the list of high-priority intersections for improvement. We conducted another similar Pareto front using the pair of performance measures: the average CO_2 emissions and the average number of stops, as shown in Fig. 5(b). A total of 24 intersections out of 383 intersections were identified on the Pareto front and they entered the list of high-priority intersections for improvement. Fig. 5(c) shows the Pareto front for the average number of stops and hard brakes. Out of the 383 intersections studied, 8 intersections were identified as the dominating intersections.

In summary, there were a total of 32 unique intersections out of 383 intersections (i.e., 8%) in the three 2D Pareto fronts that were identified as dominating and they should be improved as a high priority.

Furthermore, we constructed a 3D Pareto front using Algorithm 2 and the three performance measures altogether. The 3D Pareto front enables us to find all the correlations among the three performance measures. Fig. 6 displays the plotted intersections according to their three performances. The triangular intersections are dominating. Fig. 6 has been positioned to the best angle to reveal results. Upon close examination of Fig. 6, 63 intersections were identified as dominating and should be improved as a priority, representing 16% of all intersections in the city.

Identifying the High-Priority Intersections for Improvement Using Monthly CV Data

In the second experiment, we used the entire CV data in September 2021 to conduct a similar analysis to the first experiment. The main difference between the two experiments is that the performance measures at each intersection were averaged from the 30 days of CV data regardless of the difference in the day of the week. The experiment is suitable to develop the intersection improvement plan based on historical CV data.

Fig. 7 reveals three 2D Pareto fronts, a total of 30 intersections were identified as dominating and entered the list of intersections with a high priority for improvement, representing 7.83% of all intersections. Fig. 8 reveals the 3D Pareto front, and 86 intersections

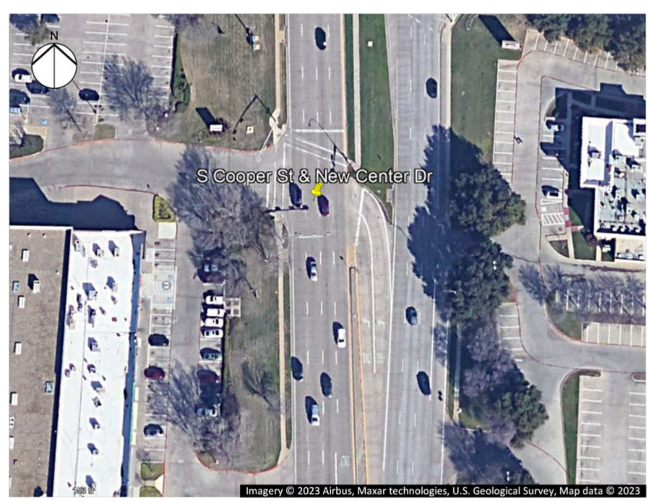


Fig. 9. Intersection (No. 2) layout of S cooper St. at the New Center Dr. (Image © 2023 Airbus, Maxar Technologies, US Geological Survey, Map Data © 2023.)

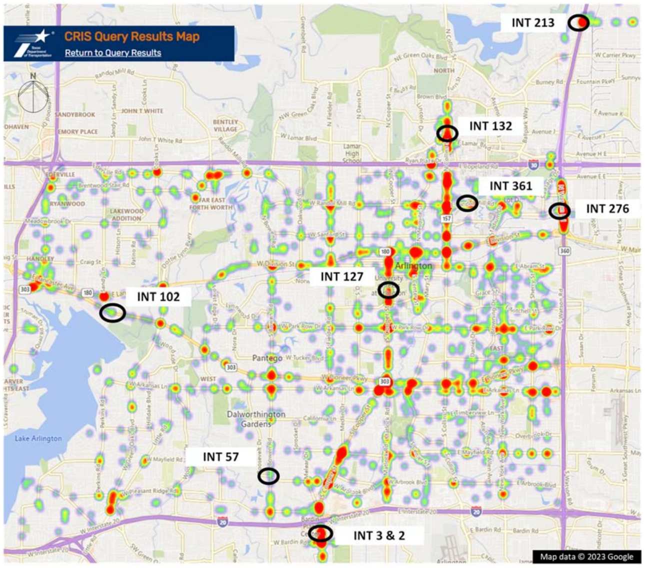


Fig. 10. Comprehensive crash record derived from the CRIS data set. (Map Data © 2023 Google.)

(22.5%) were identified as dominating and entered the list of intersections with a high priority for improvement.

Discussion

Two experiments represent two scenarios of using the CV data to develop a traffic signal improvement plan. For daily traffic operations, the first experiment shows how to use the CV data to identify problematic intersections every day. The second experiment demonstrates how to use the historical CV data to prioritize the intersections for improvement. We also compared the lists of identified high-priority intersections in two experiments. Most of them were on both lists. This finding suggests consistency in identifying the problematic intersections on a daily and monthly basis. The difference in the results between the two experiments can be caused by temporary traffic fluctuations at certain intersections, including big events, constructions, and malfunctioning traffic signals.

The 2D Pareto fronts can also imply certain solutions to the problematic intersections. For instance, if an intersection dominates the number of stops, then the solution may be to reoptimize the traffic signal systems. If the intersection dominates the number of hard brakes, it implies potential safety hazards and crashes.

In both experiments, it was also found that the number of dominating intersections in the 3D Pareto front was more than the total number of dominating intersections in the three 2D Pareto fronts. This finding implies that some problematic intersections can be identified only if all criteria are considered at the same time.

In practice, we would suggest using both 2D and 3D Pareto fronts as they each have unique advantages.

Results Validation

Validating the identified intersections with a high priority for improvement is important to ensure the proposed framework in this paper is reasonable. The list of high-priority intersections was cross-verified according to the historical crash data and local experiences. With the loss of generality, six intersections were selected from the 3D monthly Pareto front as listed in Table 3.

Table 4 offers a summary of the performances of six intersections.

As an example, upon a closer examination of Intersection 2's layout, as shown in Fig. 9, it appears that Intersection 2 has a skewed layout. The northbound vehicles must merge at a sharp angle. This layout may increase the possibility for vehicles to brake hard to make a left turn. Such intuitive observations can explain the high number of hard brakes there. Another example is Intersection 100, which has the dominating number of stops. This implies that this intersection may have a mobility issue that is likely caused by the signal timings.

Validation with the Crash Records Information System in Texas

Safety at intersections is one of the biggest concerns of agencies and we compared the high-priority intersections with the dominating hard brakes with the historical crash records retrieved from the Crash Records Information System (CRIS) in Texas in 2021 (TxDOT 2023). In the year 2021, there were a total of 6,943 recorded crashes in Arlington, Texas. However, due to limitations in the CRIS Query's map interface, only a maximum of 5,000 crashes could be displayed. As a result, a total of 4,941 crashes were visualized in the specified area of Arlington, Texas, as illustrated in Fig. 10.

According to the results of the second experiment, the mean value of average hard brakes among all intersections is 0.53 and there were 33 intersections where vehicles applied hard brakes more times than the mean value. Two dominating intersections with high hard brakes, the *Cooper St*. at *Americana* Dr. and *Cooper* St. at *New Center* Dr., were reported with severe crashes. Upon verification, a city official from Arlington, Texas, acknowledged that these two intersections are both three-leg intersections located near each other. The peculiar layout of these intersections may have resulted in an increased harsh braking, a fact supported by the analysis of the crash data obtained through the CRIS.

Table 5. Intersections with dominating hard brakes and the corresponding crash records

INT	ID	Average hard brake	Lat	Long	Number of crashes
Cooper at Americana Dr.	3	1.22	32.6729	−97.1344	15
HW 360 at Post and Paddock Rd.	213	1.00	32.7934	−97.0569	2
Cooper at New Center	2	0.90	32.6722	−97.1344	10
Random Mill at HW 360	276	0.86	32.7487	−97.0622	3
Cooper at Nedderman	127	0.85	32.7285	−97.1147	6
Pioneer at Oka Ridge	102	0.81	32.7248	−97.1964	2
Roosevelt at S Bowne Rd.	57	0.80	32.6859	−97.1495	2
Random Mill at AT&T	361	0.79	32.7502	−97.0862	9
Collins at Cantor	132	0.79	32.7693	−97.0964	5

According to available records, these two intersections collectively witnessed a total of 25 crashes in the year 2021. Table 5 presents the top nine intersections with the highest number of hard brakes and a cross-comparison with the crash records in the CRIS. Note that the high number of hard brakes implies high risks of various crashes at intersections and suggests that some preventive measures be prioritized. Additional research may be needed to further identify the correlation between crashes and hard brakes. In Table 5, other than the two intersections with higher crashes, various reasons can be identified for high hard brakes, including skewed layout, high truck ratios, limited vision, and excessively long signal cycles. Thus, all these intersections need to be paid attention to.

Conclusion and Future Work

In this paper, using the emerging CV data, we developed a multiobjective frame to prioritize intersections to help agencies decide their intersection improvement plan. The CV data are collected directly from recently manufactured vehicles' onboard units and therefore have ubiquitous coverages for all intersections within jurisdictions. Through state-of-the-art big data analytics techniques, we developed three categories of performance measures at each intersection: the average number of stops (mobility), the average number of hard brakes (safety), and the average CO_2 emissions (emission). Then we applied the Pareto front method to identify those dominating intersections that have at least one performance measure worse than any other intersections. Both 2D Pareto front and 3D Pareto front are constructed to identify the problematic intersections and possible causes. Through a case study in Arlington, Texas, we demonstrated two scenarios for intersection improvement planning: identifying problematic intersections using everyday CV data and scheduling intersection improvement plans using historical CV data. Both experiments showed reasonable and promising outcomes and all the identified dominating intersections can be justified with the local experiences. In the end, we further cross-checked the identified intersections dominating in safety concerns with the historical crash records as well as the comments by the local agencies.

Note that this data-driven framework only aims to find out those intersections that show outstanding (poor) performance(s). It mainly serves the agencies to locate those intersections in need for immediate improvement. However, this framework does not provide recommendations on what measures should be adopted at intersections and what their benefits will be. In addition, since many intersections in urban areas are closely spaced and correlated, the improvement at one intersection may have an impact on adjacent intersections and change their performance consequently. For instance, if the congestion at one intersection is reduced, it may attract more traffic from adjacent intersections, changing the overall pattern of traffic problems. As such, the intersection improvement planning may be performed iteratively to rerank all the intersections over time. In the meanwhile, predicting the changes in traffic pattern (e.g., travel demand, driving behaviors) will also be necessary at those targeted intersections to ensure the intended benefits. These efforts are out of the scope of this paper and will be studied in the future. Our future plans involve developing a comprehensive framework that combines CV data, infrastructure data, and traffic signal data. By incorporating big data, machine learning techniques, and travel demand modeling, we aim to extend the framework to identify and predict intersection performance for future scenarios. This includes the prediction of crash occurrences around intersections, allowing us to establish correlations between predictions and intersection performance based on CRIS data. Presently, we are using CRIS crash records to ensure the validity of the ranking. However, as we progress with advanced machine learning and travel demand modeling, we anticipate achieving a more precise and accurate ranking of congested intersections.

Data Availability Statement

Some or all data, models, or codes used during the study were provided by a third-party agency. Direct requests for these materials may be made to the provider as indicated in the Acknowledgments.

Acknowledgments

Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the official views or policies of the aforementioned organizations, nor do the contents constitute a standard, specification, or regulation of these organizations. The authors would like to thank Mr. Daniel Burnham, the City Engineer at the City of Arlington, for his comments on and verification of the identified problematic intersections. The connected vehicle data were distributed by Wejo Data Service, Inc.

References

Cao, X., Y. Zhong, Y. Zhou, J. Wang, C. Zhu, and W. Zhang. 2018. "Interactive temporal recurrent convolution network for traffic prediction in data centers." *IEEE Access* 6: 5276–5289. https://doi.org/10.1109/access.2017.2787696.

Change, O. C. 2007. *Intergovernmental panel on climate change*. Geneva, Switzerland: World Meteorological Organization.

Comert, G. 2013. "Effect of stop line detection in queue length estimation at traffic signals from probe vehicles data." *Eur. J. Oper. Res.* 226 (1): 67–76. https://doi.org/10.1016/j.ejor.2012.10.035.

Elleuch, W., A. Wali, and A. M. Alimi. 2017. "Intelligent traffic congestion prediction system based on ANN and decision tree using Big GPS traces." In *Proc., Intelligent Systems Design and Applications*, 478–487. Cham, Switzerland: Springer International.

Frey, H. C., and B. Liu. 2013. "Development and evaluation of simplified version of moves for coupling with traffic simulation model." In *92nd transportation research board annual meeting*. Washington, DC: Transportation research board (TRB).

Gao, Y., J. Li, Z. Xu, Z. Liu, X. Zhao, and J. Chen. 2021. "A novel image-based convolutional neural network approach for traffic congestion estimation." *Expert Syst. Appl.* 180: 115037. https://doi.org/10.1016/j.eswa.2021.115037.

Gong, L., and W. D. Fan. 2018. "Developing a systematic method for identifying and ranking freeway bottlenecks using vehicle probe data." *J. Transp. Eng. A Syst.* 144 (3): 04017083. https://doi.org/10.1061/JTEPBS.0000119.

Haklay, M., and P. Weber. 2008. "OpenStreetMap: User-generated street maps." *IEEE Pervasive Comput.* 7 (4): 12–18. https://doi.org/10.1109/MPRV.2008.80.

Huang, F. R., C. X. Wang, and C. M. Chao. 2020. "Traffic congestion level prediction based on recurrent neural networks." In *Proc., 2020 Int. Conf. on Artificial Intelligence in Information and Communication (ICAIIC)*, 248–252. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

Jiao, P., R. Li, and Z. Li. 2016. "Pareto front–based multi-objective real-time traffic signal control model for intersections using particle swarm optimization algorithm." *Adv. Mech. Eng.* 8 (8): 1687814016666042. https://doi.org/10.1177/1687814016666042.

Khadka, S., P. T. Li, and Q. Wang. 2022. "Developing novel performance measures for traffic congestion management and operational planning based on connected vehicle data." *J. Urban Plann. Dev.* 148 (2): 04022016. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000835.

Khadka, S., P. S. Wang, P. T. Li, and F. J. Torres. 2023. "A new framework for regional traffic volumes estimation with large-scale connected vehicle data and deep learning method." *J. Transp. Eng. A: Syst.* 149 (4): 04023015. https://doi.org/10.1061/JTEPBS.TEENG-7536.

Krause, B., C. V. Altrock, and M. Pozybill. 1996. "Intelligent highway by fuzzy logic: Congestion detection and traffic control on multi-lane roads with variable road signs." In Vol. 1833 of *Proc., IEEE 5th Int. Fuzzy Systems*, 1832–1837. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

Kurniawan, J., S. G. S. Syahra, C. K. Dewa, and G. Afiahayati. 2018. "Traffic congestion detection: Learning from CCTV monitoring images using convolutional neural network." *Procedia Comput. Sci.* 144: 291–297. https://doi.org/10.1016/j.procs.2018.10.530.

Lee, J., B. Hong, K. Lee, and Y.-J. Jang. 2015. "A prediction model of traffic congestion using weather data." In *Proc., 2015 IEEE Int. Conf. on Data Science and Data Intensive Systems*, 81–88. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

Liu, Y., X. Feng, Q. Wang, H. Zhang, and X. Wang. 2014. "Prediction of urban road congestion using a Bayesian network approach." *Procedia* 138: 671–678. https://doi.org/10.1016/j.sbspro.2014.07.259.

Lu, J., and L. Cao. 2023. "Congestion evaluation from traffic flow information based on fuzzy logic." In *Proc., 2003 IEEE Int. Conf. on Intelligent Transportation Systems*, 50–53. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

Lu, J., and X. S. Zhou. 2023. "Virtual track networks: A hierarchical modeling framework and open-source tools for simplified and efficient connected and automated mobility (CAM) system design based on general modeling network specification (GMNS)." *Transportation Research Part C: Emerging Technologies* 153: 104223.

Meng, Q., and H. L. Khoo. 2010. "A Pareto-optimization approach for a fair ramp metering." *Transp. Res. C Emerging Technol.* 18 (4): 489–506. https://doi.org/10.1016/j.trc.2009.10.001.

Mondal, M. A., and Z. Rehena. 2019. "Intelligent traffic congestion classification system using artificial neural network." In *Companion Proc., 2019 World Wide Web Conf.*, 110–116. San Francisco: Association for Computing Machinery.

Olayode, O., L. Tartibu, and M. Okwu. 2021."Application of adaptive neuro-fuzzy inference system model on traffic flow of vehicles at a signalized road intersections." In *Proc., ASME Int. Mechanical Engineering Congress and Exposition*, V009T009A015. New York: American Society of Mechanical Engineers (ASME).

Pongpaibool, P., P. Tangamchit, and K. Noodwong. 2007. "Evaluation of road traffic congestion using fuzzy techniques." In *Proc., TENCON 2007-2007 IEEE Region 10 Conf.* Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

Porikli, F., and L. Xiaokun. 2004. "Traffic congestion estimation using HMM models without vehicle tracking." In *Proc., IEEE Intelligent Vehicles Symposium 2004*, 188–193. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

Shen, X., and J. Chen. 2009. "Study on prediction of traffic congestion based on LVQ neural network." In *Proc., 2009 Int. Conf. on Measuring Technology and Mechatronics Automation*, 318–321. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

Shenghua, H., N. Zhihua, and H. Jiaxin. 2020. "Road traffic congestion prediction based on random forest and DBSCAN combined model." In *Proc., 2020 5th Int. Conf. on Smart Grid and Electrical Automation (ICSGEA)*, 323–326. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

ShirMohammadi, M. M., and M. Esmaeilpour. 2020. "The traffic congestion analysis using traffic congestion index and artificial neural network in main streets of electronic city (case study: Hamedan city)." *Program. Comput. Softw.* 46 (6): 433–442. https://doi.org/10.1134/S0361768820060079.

Stevanovic, A., J. Stevanovic, and C. Kergaye. 2013. "Optimization of traffic signal timings based on surrogate measures of safety." *Transp. Res. C Emerging Technol.* 32: 159–178. https://doi.org/10.1016/j.trc.2013.02.009.

Subirats, P., Y. Dupuis, E. Violette, D. Doucet, and G. Dupre. 2010. "A new tool to evaluate safety of crossroad." In *Proc., 4th Int. Symp. on Highway Geometric Design. Valence*, 2–5. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

Sun, S., J. Chen, and J. Sun. 2019. "Traffic congestion prediction based on GPS trajectory data." *Int. J. Distrib. Sens. Netw.* 15 (5): 1550147719847440.

Teply, S., and G. J. Evans. 1989. "Evaluation of the quality of signal progression by delay distributions." *Transp. Res. Rec.* 1225: 1–7.

Thianniwet, T., S. Phosaard, and W. Pattara-Atikom. 2009. "Classification of road traffic congestion levels from GPS data using a decision tree algorithm and sliding windows." In *Proc., World Congress on Engineering*, 1–3. London, UK: World Congress on Engineering (WCE).

Tian, Y., and L. Pan. 2015. "Predicting short-term traffic flow by long short-term memory recurrent neural network." In *Proc., Int. IEEE Conf. on Smart City/SocialCom/SustainCom (SmartCity)*. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

Tran Quang, D., and S. Hoon Bae. 2021. "A hybrid deep convolutional neural network approach for predicting the traffic congestion index." *Traffic Transp.* 33 (3): 373–385.

TxDOT. 2023. *Crash records information system*. Austin, TX: Texas Department of Transportation.

USEPA (US Environmental Protection Agency). 2010. *Motor vehicle emission simulator (MOVES) 2010 user guide*. Ann Arbor, MI: USEPA.

Wejo Data Service. 2021. "National Work Zone Awareness Week: Top 10 Questions from the Webinar." https://www.wejo.com/resources/nwzaw-top-10-questions-from-the-webinar#:~:text=Wejo%3A%20Hard%20braking%20or%20hard,8.75%20foot)%20per%20second%20squared.

Xunxue, C., L. Qinl, and T. Qing. 2007. "Genetic algorithm for Pareto optimum-based route selection." *J. Syst. Eng. Electron.* 18 (2): 360–368. https://doi.org/10.1016/S1004-4132(07)60099-1.

Zhang, S., Y. Yao, J. Hu, Y. Zhao, S. Li, and J. Hu. 2019. "Deep autoencoder neural networks for short-term traffic congestion prediction of transportation networks." *Sensors (Basel)* 19 (10): 2229. https://doi.org/10.3390/s19102229.

Zhao, Z., W. Chen, H. Yue, and Z. Liu. 2016. "A novel short-term traffic forecast model based on travel distance estimation and ARIMA." In *Proc., of Chinese Control and Decision Conf. (CCDC)*. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).

Zhao, J., V. L. Knoop, J. Sun, Z. Ma, and M. Wang. 2023. "Unprotected left-turn behavior model capturing path variations at intersections." *IEEE Trans. Intell. Transp. Syst.* 24 (9): 9016–9030. https://doi.org/10.1109/TITS.2023.3270962.

Zhou, X., P. Dong, J. Xing, and P. Sun. 2019. "Learning dynamic factors to improve the accuracy of bus arrival time prediction via a recurrent neural network." *Future Internet* 11 (12): 247. https://doi.org/10.3390/fi11120247.

Zhou, X., S. Tanvir, H. Lei, J. Taylor, B. Liu, N. M. Rouphail, and H. Frey. 2015. "Integrating a simplified emission estimation model and mesoscopic dynamic traffic simulator to efficiently evaluate emission impacts of traffic management strategies." *Transp. Res. D Transp. Environ.* 37: 123–136. https://doi.org/10.1016/j.trd.2015.04.013.

J. Urban Plann. Dev., 2024, 150(1): 04023063