Developing Novel Performance Measures for Traffic Congestion Management and Operational Planning Based on Connected Vehicle Data

Swastik Khadka, S.M.ASCE[1]; Pengfei "Taylor" Li, Ph.D., P.Eng., M.ASCE[2]; and Qichao Wang, Ph.D.[3]

Abstract: In this paper, the authors present their efforts in exploring a new type of traffic data, referred to as internet-connected vehicle (ICV) data, for traffic congestion management and operational planning. Most currently manufactured vehicles contain onboard GPS and cellular modules, and they constantly connect to automobile manufacturers' clouds via cellular networks and upload their status. Some automobile manufacturers have recently redistributed the nonpersonal part of such data, such as geolocation, to third-party organizations for innovative applications. Compared with the traditional vehicle GPS data, the ICV data contain high-resolution GPS waypoints accompanied with the vehicles' abnormal moving events (e.g., hard braking). The ICV data also have huge potential in congestion management and operational planning. They explore to identify and analyze traffic congestion on both freeways and arterials using the ICV data. The ICV data adopted for this research are redistributed by Wejo Data Service, representing 10%–15% of all moving vehicles in the Dallas–Fort Worth (DFW) area in Texas. Through one case study for a freeway segment and one for an arterial segment, new traffic performance metrics based on the characteristics of ICV data have been presented. The highlights of these efforts are as follows: (I) queue length and propagation at freeway bottlenecks can be directly measured based on where and when most internet-connected vehicles slow down and join the queue; (II) an internet-connected vehicle's actual delay time on arterials can be directly measured according to its slow movement percentage, without assuming the nondelay travel speed; and (III) the ICV data set are also combined with the high-resolution traffic signal events to generate a *ground-truth* time-space diagram (TSD) on arterials—a common visualization of arterial signal performance for transportation planning and operations. **DOI: 10.1061/(ASCE)UP.1943-5444.0000835**. © *2022 American Society of Civil Engineers*.

Author keywords: Vehicle trajectories; Connected vehicle; Freeway queue management; Arterial traffic signal performance.

Introduction

The era of mobile computing enables ubiquitous smartphones, in-vehicle internet-of-things (IoT) devices, and connected and automated vehicles. Vehicles and infrastructure are connected in multiple ways. One approach is through crowdsourcing. The mobile devices are generating rich traffic data sets of new types, which will be supplemental to the infrastructure sensors maintained by public agencies to better plan the traffic congestion management heavily driven by data. Collecting and maintaining infrastructure sensors at a large scale requires a huge amount of construction and recurrent operational costs. Furthermore, infrastructure sensors are fixed-spot ones, and therefore they cannot provide a full spectrum of traffic conditions. Most agencies today still use the traffic data collected via fixed-spot infrastructure data for performance monitoring, such as inductive loops, road-side cameras, and radar sensors (Xu et al. 2016). Some agencies also adopt automated vehicle identification techniques, such as Toll Tag ID (Turner et al. 2000) and Wi-Fi/Bluetooth MAC address matching (Li and Souleyrette 2016; Araghi et al. 2016). While these techniques are proven to be effective, bias is inevitable because the agencies have to empirically select sensors' locations and use a point-to-point vehicle ID matching mechanism to represent the entire highway segments. In reality, congested segments and bottlenecks may occur at any place. Infrastructure detectors can only report the traffic conditions at certain fixed spots. As a result, the corresponding traffic performance monitoring is inevitably biased. One straightforward solution is to deploy infrastructure sensors more intensively to narrow the link segments while this option may be not cost-effective to agencies. On arterials, while intersections are explicit bottlenecks, hidden bottlenecks often exist at midblock driveways or two-way left-turn lanes. Such hidden bottlenecks can hardly be identified with traditional traffic analytics or fixed-spot detector data.

To overcome these challenges and to provide new cost-effective solutions, we exploit the potential of emerging internet-connected vehicle (ICV) data in congestion management. The ICV data are passively crowdsourced and collected by automobile manufacturers. Most currently manufactured vehicles constantly connect to automobile manufacturer's clouds and upload their real-time status (e.g., the GMC's *OnStar*, an add-on service based on subscriptions). Some automobile manufacturers have recently decided to redistribute such data sets (after removing the private info) to third-party organizations for innovative mobility applications. Compared with the traditional vehicle GPS data set, the ICV data contain high-fidelity waypoint locations accompanied with abnormal events

[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

[3]Postdoctoral Researcher, Computational Science Center National Renewable Energy Laboratory 15013 Denver West Parkway, Golden, CO 80401. ORCID: https://orcid.org/0000-0002-0863-4564. Email: qichao.wang@nrel.gov

Note. This manuscript was submitted on May 18, 2021; approved on December 21, 2021; published online on February 22, 2022. Discussion period open until July 22, 2022; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Urban Planning and Development***, © ASCE, ISSN 0733-9488.

(e.g., hard braking) while vehicles are moving. After some preliminary experiments, we conclude that the ICV data are highly accurate and hold great promise in congestion management. The ICV data are based on vehicle trips. Each trip will be allocated with a unique ID. For each waypoint of a trip, the vehicle's instantaneous latitude, longitude, current timestamp, speed, and heading are provided. Abnormal vehicle maneuvers (e.g., hard braking) are also collected from onboard units. The ICV data for this study are procured from Wejo Data Service which is licensed by General Motors. The penetration of the ICV data we studied represents 10%–15% of all moving vehicles in the Dallas–Fort Worth (DFW) area, Texas.

The paper is organized as follows. First, literature on traffic data collection and applications is reviewed; then a scalable framework for processing the ICV data is presented. Second, two case studies are conducted to find the hidden highway bottlenecks, accompanied by new performance metrics based on the ICV data sets. Third, we present new traffic performance metrics on arterials based on the ICV data and high-resolution traffic control log data.

Literature Review

Congestion management on highways focuses on bottleneck identification and mitigation. Traditional methods to discover bottlenecks mostly are based on infrastructure sensors. With fixed-spot detectors, Chen et al. design a statistical method to identify the possible appearance of bottlenecks on freeways by comparing the speed differences between the upstream and downstream locations. If the speed difference is more than 20 mi/h (32 km/h), then a bottleneck is identified (Chen et al. 2004). Other than the speed reduction between locations, bottlenecks on freeways can also be identified by the duration of speed reduction at certain locations (Banks 2009), by the reduction of traffic volumes (Bertini 2003), and by occupancy changes (Hall and Agyemang-Duah 1991; Zhang and Levinson 2004). The advantage of the fixed-spot detectors is that they can almost capture 100% of vehicle presence (if lane-by-lane detection is configured) and therefore reported results are accurate. However, the selection of sensor locations is very important, and hidden bottlenecks may be difficult to find.

Transportation agencies also explore to capture and match a small portion of vehicle *signatures* to collect segment travel time samples. The vehicle *signatures* refer to unique in-vehicle IDs. Through various roadside sniffing devices deployed at different locations, those unique vehicle IDs can be captured and rematched. The time difference between road-side devices is considered as a segment (i.e., space) travel time sample. Available vehicle signatures include, but are not limited to, probe vehicles with known IDs (Hofleitner et al. 2012), toll tags series numbers (Turner et al. 2000), license plate numbers (Bertini et al. 2005; Xu et al. 2011), cellphone locations (Qiu et al. 2009), vehicle's optical image identification (Kuroiwa et al. 2006), vehicles' magnetic signature matching (Charbonnier et al. 2012; Kavaler et al. 2011; Kwong et al. 2009; Sanchez et al. 2011a, b), and Bluetooth MAC address matching (Bakula et al. 2012; Barcelo et al. 2010; Brennan et al. 2010; Haghani et al. 2010; Hainen et al. 2011; Quayle et al. 2010; Richardson et al. 2011). Segment travel time is a direct indicator of road congestion and is considered superior to the estimated travel time by the fixed-spot detectors. Nonetheless, there are also challenges in leveraging between screening the sample outliers and sample sizes. For instance, in the commonly used Bluetooth-based travel time estimation method, the sample size is usually less than 5%. To increase the capturing rate, high-gain antennas (with a sensing radius of 1,000 or more feet) must be used, resulting in large measurement errors. In the meanwhile, vehicles are not tracked between two locations and their abnormal behaviors (e.g., pullover) cannot be tracked. As a result, the sample outliers must be screened. Other techniques also have their drawbacks. Many research efforts have been dedicated to addressing these issues. For example, a framework including multiple heuristic steps to process the Bluetooth-based travel time samples were designed by Haghani et al. (2010). In their framework, a set of 24-h historical travel time samples are selected to infer the travel speed distributions. A moving standard deviation is designed by Quayle et al. (2010) to screen the Bluetooth-based travel time outlier samples. Li and Souleyrette (2016) propose a Kalman filter framework to estimate the time-varying travel time with collected Bluetooth travel time samples. These techniques fall into the category of the *passive sensing* technique. Roadside capturing devices are typically needed at different locations for travel time estimations.

After entering the era of mobile computing, most vehicles become *probe vehicles* via drivers' smartphone apps and onboard units. These vehicles share their information regularly. The information can be collected through drivers' smartphones, in-vehicle global positioning system (GPS) receivers, and many more. Most new vehicles today are equipped with in-vehicle GPS receivers for location services (e.g., General Motor's OnStar service). Such services require vehicles to constantly share their locations and therefore the collected location data are essentially crowdsourced, and the nonpersonal part of such data sets is a novel traffic data set with massive sample sizes and broad coverages. The commercial probe vehicle data are typically provided in two forms: (I) dynamic link travel times/link travel speeds according to aggregated GPS trajectories; (II) individual vehicle trips containing all the trip waypoints at small time intervals (3–15 s). Both types of GPS probe data are currently used in congestion management. For example, Gong and Fan (2018) design a framework to identify freeway bottlenecks using the large-scale link travel time aggregated from vehicle GPS trajectories. They combine both probe data sets and traffic management center data to identify bottlenecks. Zhao et al. (2013) use commercial vehicle GPS trajectories with a time interval of 15 s or longer to evaluate traffic progression in Washington. Waddell et al. (2020a, b) investigate to estimate traffic signal performance along arterials using the vehicle GPS trajectories and automated traffic signal performance metrics (ATSPM). Deng et al. (2018) use GPS trajectories to model intersections.

Congestion management based on large-scale connected vehicle trajectories and events data is a relatively new concept. Some researchers have recently used such data sets to explore the congestion management on intersection operations (Hunter et al. 2021b; Saldivar-Carranza et al. 2021), work zone management (Desai et al. 2021; Sakhare et al. 2021), and impacts from connected vehicle penetration (Hunter et al. 2021a).

Framework for Processing the Raw ICV Data

A big challenge of using the ICV data for transportation management is how to reduce the ICV data size to a manageable level for specific purposes. Due to the rich information and high resolution, even a few weeks of ICV data set will be hundreds of gigabytes or even terabytes of text files. They are beyond the capability of traditional data processing tools. To overcome these challenges, we propose a scalable data processing framework to filter out irrelevant information from the raw ICV data including out-of-scope waypoints. Also, we develop efficient map-matching

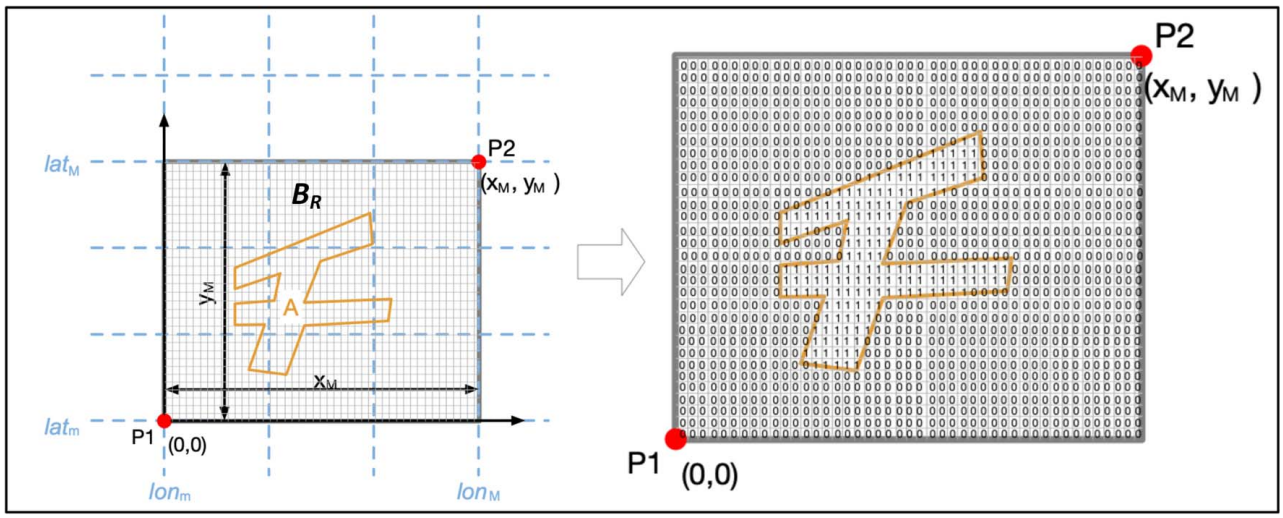

Fig. 1. Illustration of data reduction with the OpenCV library.

algorithms to map vehicle GPS traces onto road links. Fig. 1 demonstrates the proposed ICV data processing framework.

Data Reduction for the Scope of Interest

The first core process in this framework is to reduce the data size and screen out irrelevant data sets. The ICV data are often delivered in-vehicle traces covering the entire area while the scope of congestion management only includes much smaller areas along the highways whose shapes are likely irregular. Checking whether geolocation points are within an irregular polygon at a large scale is computing intensive. To address this issue, we adopt an efficient approach proposed by Li and Li (2010). The data-reduction method due to Li and Li is to first convert vector Geography Information System (GIS) maps into raster maps and use each pixel's value to represent a link ID. In light of this idea, we first rasterize a broader bounding rectangle on a GIS map. We only focus on the waypoints that are within the bounding rectangle. We then set all the pixels within the polygon of interest as one value that is different from those pixels out of the polygon. Each waypoint within the bounding rectangle will be assigned a corresponding pixel value. At last, all waypoints are grouped and filtered according to their assigned values. More details are provided as follows.

The congestion management is typically carried out for a city or a region within which the unit distances of longitude are viewed as constant (i.e., the city or region can be viewed as a flat plane). Therefore, we can safely map the vehicle GPS waypoints from the WGS84 coordinate system (latitude–longitude coordinates) into the local coordinates in a customized rectangular coordinate system.

As shown in Fig. 1, the broader bounding rectangle, denoted as B_R, is first defined to cover all the road segments of interest. The irregular area of interest is denoted as A. The bottom-left corner of B_R is the origin with coordinates (0, 0) in the local rectangle coordinate system, denoted as P_1. From an online map engine (e.g., Google Map), we can then find $P_1's$ latitude–longitude coordinates, denoted as (lon_m, lat_m). Second, the latitude–longitude coordinates of the top-right corner are denoted as P_2, which is identified as (lon_M, lat_M). Third, the rectangular B_R is further divided into $x_M \times y_M$ smaller cells. Using the cells as the units, then the local coordinates for P_1 and P_2 are (0, 0) and (x_M, y_M), $x_M \in N^+$, $y_M \in N^+$. A larger x_M or y_M means a small cell size (i.e., higher resolution) along with the x or y directions. The aforementioned conversion function from latitude–longitude coordinates(lon, lat) to local coordinates (x, y) in B_R can be formulated as follows:

$$(x, y) = F(\text{lon}, \text{lat}) = \left(\frac{x_M \times (\text{lon} - \text{lon}_m)}{\text{lon}_M - \text{lon}_m}, \frac{y_M \times (\text{lat} - \text{lat}_m)}{\text{lat}_M - \text{lat}_m} \right) \quad (1)$$

For each latitude–longitude point, (lon, lat), (1) will convert it into a local point, (x, y), in B_R.

With (1), the coordinates of polygon A' vertices, denoted as $\{(\text{lon}_i^v, \text{lat}_i^v),\ i \in \{0, 1, 2, \ldots\}\}$ can be converted into the local rectangle coordinate system as follows:

$$(\hat{x}_i, \hat{y}_i) = F(\text{lon}_i^v, \text{lat}_i^v) \quad (2)$$

To identify if a waypoint falls into A, the waypoint must be compared with all vertices of polygon A. This process, if carried out in sequence, will be time-consuming but can greatly speed up with advanced computing tools. In this context, because the proposed data-reduction process is similar to processing an image comprised of pixels, we adapt a computer-vision library, referred to as the *OpenCV* (Version 4.5.4), to perform this process. We formulated this problem as an image processing problem.

We constructed a binary matrix, $\boldsymbol{M}$, with dimensions of $x_M \times y_M$. Each element represents a cell defined previously in analogy of *image pixels* of B_R and A, and can be indexed by the row number i and column number j. Each element in $\boldsymbol{M}$ is initialized as 0 and those elements within A are then adjusted to 1. A function, M: $\mathbb{R} \times \mathbb{R} \rightarrow \{0, 1\}$, can be constructed from the matrix $\boldsymbol{M}$. For a point, (x, y), within the B_R, the value of $M(x, y)$ is the value of the element in matrix $\boldsymbol{M}$ indexed by row $x + 0.5$ and column $y + 0.5$. For the points outside of B_R, the value is 0. This is described by the following equation:

$$M(x, y) = \begin{cases} (\boldsymbol{M}_{x+0.5}, \boldsymbol{M}_{y+0.5}), & x \in [0, x_M), y \in [0, y_M) \\ 0, & \text{o.w.} \end{cases} \quad (3)$$

A latitude–longitude waypoint (lon, lat) will first be converted into the local coordinates (x, y) in B_R. We used the OpenCV library to mark down the corresponding latitude–longitude waypoints that are within A, that is, $\{(\text{lon}, \text{lat}) | M(F(\text{lon}, \text{lat})) = 1,\ \forall(\text{lon}, \text{lat})\}$. The aforementioned data-reduction process inspects each waypoint independently, allowing for parallelization. The GPS waypoints in this study are also delivered in multiple text files for each hour.

Map Matching of Vehicle GPS Waypoints to Time-Dependent Road Link Sequences

Traffic congestion is typically evaluated based on road segments. Therefore, it is critical to map the vehicle trajectories onto road links and convert vehicle trips to a time-dependent link sequence for quantitative analysis. For the real-world problems with the ICV data sets, the map-matching task often requires matching millions of vehicle GPS waypoints to tens of thousands of road links. There are plenty of commercial GIS packages including the map-matching functions. However, after exploring most of the existing tools, we conclude that most tools are not suitable for processing such a magnitude of vehicle trajectories targeted in this paper even though the outputs are accurate. As such, the map-matching algorithm must be designed in such a way that it can significantly reduce the matching space and can parallelize the tasks into multiple CPU cores or computer clusters. To meet these goals, we design the map-matching algorithm as follows:

- Step 1: Grid the road network
  The purpose of gridding the road network into cells is to exclude those road links irrelevant to a GPS waypoint. Fig. 2(a) demonstrates the grid road network in the Dallas–Fort Worth region. If a waypoint falls into Grid X, the algorithm does not need to compare those links in Grid ***Y***.
- Step 2: Identify the list of crossing grids for road links
  Regional highway networks are represented by nodes and links. After gridding the road network, certain long links (i, j), represented by two end nodes i, j, may cross more intermediate grids than just two grids. To ensure all crossing grids of a long link will be recorded, we extend the scope of long links. As illustrated in Fig. 2(b), link (i, j) crosses three grids (upper left, central and bottom right). We extend the list of crossing grids for (i, j) to all the grids within the shadow. This operation will guarantee any waypoint can always match its link even if the road network is sparse. The shadowed area in Fig. 2(b) is an extended rectangle on the gridded road network to cover multiple grids (upper left, central and bottom right). The rectangle uses the upper left and bottom right grids as diagonal corners.
- Step 3: Match GPS waypoints to links
  While matching a waypoint (e.g., Grid ***X***) to the road links, only those road links passing Grid ***X*** will be selected for matching and the link in the same grid with the shortest conjugate distance from the waypoint will be identified as the matched link (Fig. 1). After the map matching, each vehicle trip can be converted into a series of dynamic link sequences comprised of a link list and a time list. For instance, $(l_1, l_2, \ldots, l_n) + (t_1, t_2, \ldots, t_n)$ represents that a vehicle enters a link l_1 at t_1, leaves l_1 (i.e., enter l_2) at t_2, leaves l_2 (enters l_3) at t_3, and so on. At this time, it is ready to analyze the link traffic performance based on the converted ICV data sets.

Remarks: Although Step 2 ensures that a waypoint will match the correct link, it also introduces unnecessary matching efforts. For instance, in Fig. 2, a waypoint in those grids, which are within the rectangle but are not crossed by the trajectories, will have to calculate its conjugate distance to an irrelevant link (i, j). Nonetheless, the long links crossing multiple grids are rare in practical road networks. Therefore, Step 2 will not significantly increase the computing efforts.

Framework's scalability: Gridding the road network will dramatically reduce the number of links to be matched for each waypoint. Step 1 and Step 2 will jointly guarantee that the map-matching results will not deteriorate if the grid size is further reduced. Since the map matching for one waypoint is independent of other waypoints, the map-matching process for multiple waypoints can be parallelized on multiple CPU cores or computer clusters. Therefore, the proposed map-matching method is scalable, and it is suitable for large-scale problems.

Mismatching issue: A small portion of waypoints may be equally away from two links. For instance, while a vehicle is passing an overpass on freeways, its waypoints may be equally away from the freeway link and the overpass link. To avoid mismatching, it is necessary to examine the predecessors and successors of this waypoint within the same trip to ensure the correct links. Comparing the vehicle's heading (direction) with the road direction can also help remove the mismatching issue.

ICV-Data-Based Performance Metrics in the Time-Space Diagram

The time-space diagram (TSD) is a popular tool to reveal time-dependent traffic performance on roads. Because the ICV data ubiquitously cover all the locations with relatively large sample sizes, we can divide a road link into smaller segments to reveal the time-dependent traffic performance according to the ICVs' waypoints at different locations over time. The new performance metrics are designed according to the features of ICV data sets: time-dependent link speed map integrated with slow vehicle movements and degree of speed harmonization (DSH).

Time-Dependent Link Speed Map Integrated with Slow Vehicle Movements

Time-Dependent Travel Speed

The time-dependent segment speed is calculated as the average speed of waypoints reported within each segment and time.

$$\bar{v}_{t,l} = \sum_{i=1,\ldots,n} v^i_{t,l} \tag{4}$$

where $\bar{v}_{t,l}$ = average travel time on segment l during time-of-day period t; and $v^i_{t,l} (i = 1, 2, \ldots n)$ = reported speed in a waypoint within the segment and time-of-day period.

Vehicle Slow Movements and Queue Length

The slow movements are identified when a vehicle's speed has reduced below a threshold. When an internet-connected vehicle joins queue end at a bottleneck, its speed will significantly reduce. Therefore, we can estimate the propagation of queue lengths at bottlenecks according to when an internet-connected vehicle begins to report slow movements. As illustrated in Fig. 3, the start of a stable slow movement represents the instantaneous queue end at bottlenecks.

The proposed performance metric is a combination of travel speed and vehicle slow movements. This performance metric will be demonstrated in Case study I.

Degree of Speed Harmonization on Road Segments

While the travel speed is a straightforward indicator of traffic mobility, it also averages out the heterogeneity and fluctuation of individual vehicles. As shown in Fig. 4, Vehicle 1 accelerates and decelerates multiple times between location i and location i, whereas Vehicle 2 only accelerates once. If we use the average speed to represent the two vehicles' maneuvers, then Vehicle 1's stop-and-go pattern will be missing.

Because vehicles' stop-and-go characteristics directly reflect both traffic safety hazards and congestion, we propose the

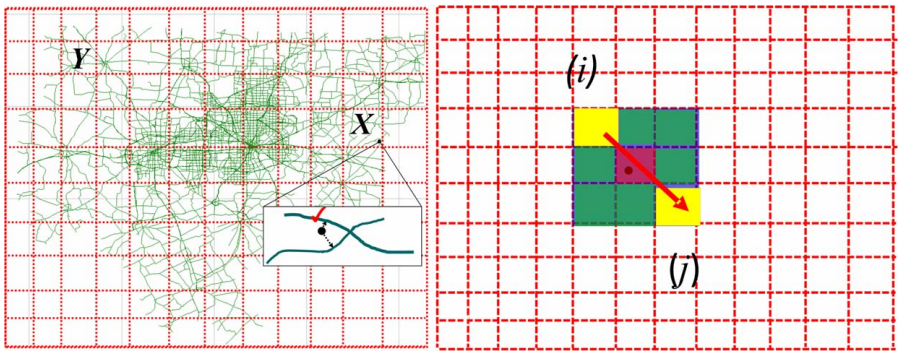

Fig. 2. Gridding road networks and map matching.

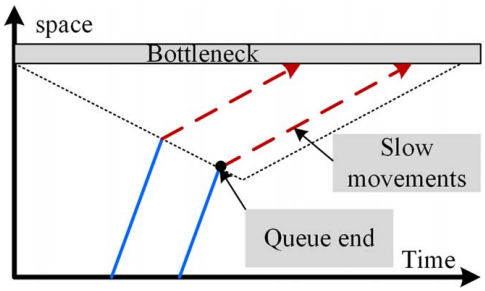

Fig. 3. Identifying queue end with vehicle slow movements at bottlenecks.

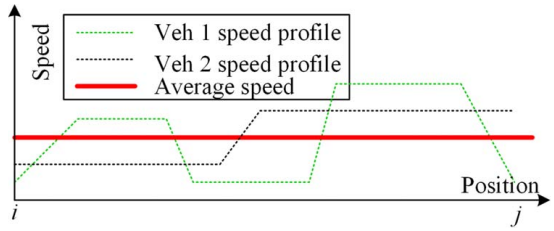

Fig. 4. Speed files of two heterogeneous vehicles with the same average link speed.

following performance metrics to reflect the speed harmonization on road segments.

Denote the collection of speeds of a trip p on a road link as $\{v_p^i\}$, $i = 1, 2, \ldots$ then nonnegative speed change value and the average speed change of p are defined as $\{d_p^i\} = \{|(v_p^i - v_p^{i-1})|\}$, $i = 2, 3, \ldots$ and $\bar{d}_p = \left(\sum_{i=1,2,\ldots n} d_p^i\right)/n$, respectively. The DSH on link l is calculated as follows:

$$\mathrm{DSH}_l = \frac{\left(\sum_{p=1,2,\ldots m} \bar{d}_p\right)}{m} \quad \text{for } \forall\, l \tag{5}$$

Eq. (5) shows the DSH of all vehicles on l during the study period. It is similar to the definition of *coefficient of variation* in statistics indicating the normalized deviation of a vehicle's reported speeds. DSH is related to traffic safety and vehicle emissions. The larger the dimensionless DSH of a cell is, the more likely vehicles frequently accelerate and decelerate, implying traffic safety hazards and vehicle emission issues.

Case Study I: Using the ICV Data to Identify Queue Propagation at Freeway Bottlenecks

The queue length at freeway bottlenecks is traditionally estimated according to traffic counts at upstream and downstream locations. For example, Lawson et al. (1997) proposed an approach to estimate the queue length based on the vehicle cumulative counts, also known as *A–D* curves; Newell (1993) proposed a so-called three-detector method to estimate the queue length at freeway bottlenecks. However, these methods of queue length estimation are sensitive to the selected parameters (e.g., vehicle speed while moving in the queue). Using the ICV data, we can directly measure the queue length and propagation at freeway bottlenecks. Through synthesizing multiple days of ICV data, we use the performance metrics proposed in the "ICV-Data-Based Performance Metrics in the Time-Space Diagram" section to identify the spatiotemporal characteristics of queue lengths at bottlenecks.

The selected freeway segment on I-20 interstate highway is one of the major corridors for travelers within the City of Arlington, Texas. There are also six ramps along the I-20 in Arlington, potentially causing vehicles to slow down. Following the framework proposed in the "Framework for Processing the Raw ICV Data" section we first process the raw ICV data and only reserve relevant vehicle trips. A related vehicle trip is identified if at least one of its waypoints is located within the study scope and has a correct heading. In addition, a vehicle trip is broken into two or more if a time interval between two consecutive waypoints is longer than 2 min. In that case, it means the same vehicle may likely have taken off the mainline and then taken on the ramp again. Therefore, the same vehicle trip should be viewed as two separate trips. In the preliminary data screening, a total of 36,345 vehicle trips were retrieved from June 1, 2020 to June 7, 2020. Using the processed vehicle trips, we then construct a 24-h time-space diagram with multiple days of vehicle trips to identify spatiotemporal bottlenecks. The speed limit on I-20 in Arlington is 70 mi/h (113 km/h). If a waypoint speed is lower than 40 mi/h (65 km/h)—a local empirical threshold—then the vehicle's waypoint is labeled as a *slow movement*, suggesting that this vehicle is in a moving queue. Because the ICV waypoints cover the entire freeway segment, we further divide the freeway into several subsegments with a 0.5-mi length to inspect the traffic performance metrics at each subsegment. Furthermore, we divide 1 day into 24 h.

Fig. 5 reveals a high correlation between travel speed and slow trajectories. For instance, we can identify low-speed cells, such as those in Areas I, II, III, and IV. We can also identify the queue propagation in those areas, too. The colored cells and slow trajectories jointly visualize the spatiotemporal characteristics of freeway bottlenecks. In particular, the East Bound (EB) queue length in Area III, developed

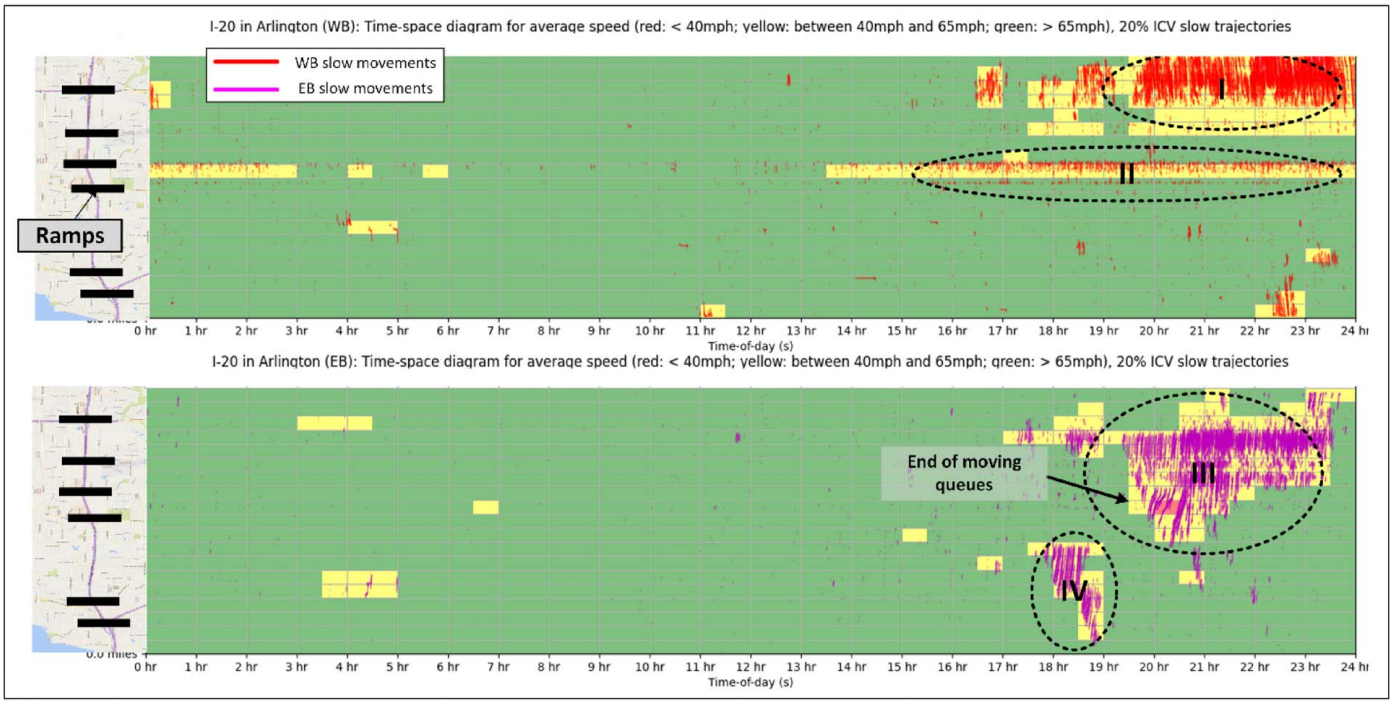

Fig. 5. Time-space diagram of dynamic travel speed integrated with slow vehicle movements. (Map data © 2021 Google.)

at a ramp, seems to have lasted a few hours, and the queue propagated back to upstream ramps. The queue length reached the maximal between 8 p.m. and 9 p.m. and then disappeared after 11: 30 p.m. Note that the ICV data for this case study were collected during the lockdown period of Texas. Therefore, there were no obvious morning and evening peak hours. The cause for the bottlenecks was road construction and work zone management. The local agencies decided to speed up road construction projects during the pandemic because the traffic volumes dramatically decreased.

Fig. 6 reveals the DSH values in each cell. A three-grey-level code is used to represent the DSH in each cell. We can tell that the DSH has a low correlation with travel speed when compared with the slow trajectories. It seems that vehicles' DSH reduces after they slow down and join the queue. This makes sense because vehicles have less flexibility of speed changes due to the close spaces with other vehicles.

Case Study II: Using the ICV Data and Traffic Control Data to Identify Arterial Congestion

Intersections are natural bottlenecks on arterials. Compared with freeways, the agencies typically collect various types of data via fixed-spot infrastructure detectors and traffic signal control events. Data-driven traffic control performance metrics are one of the focuses of the *Every Day Counts Initiatives* program overseen by Federal Highway Administration (Wagner 2014). More recently, researchers begin to explore integrating vehicle trajectories with infrastructure traffic data to create novel traffic performance metrics. The most common method to reveal the arterial traffic bottleneck is the TSD. The TSD reveals the performance of traffic signal control across multiple intersections and the performance metrics include green bandwidth, maximum queue length, and so on. Fig. 7 is an arterial time-space diagram based on traffic signal events (signal control records at each intersection) and vehicle trajectories (curves between intersections).

The high-resolution traffic signal events are footprints of traffic control operations. Whenever a traffic controller changes its status (e.g., capture an approaching vehicle, a green phase start or ending, and so on), this control event will be timestamped and archived. We can use the control events to construct the TSD using the control events. Constructing TSDs ideally needs both high-resolution traffic signal event data and synchronous vehicle trajectories with sufficient penetration rate, the latter of which are not available until recently. In this case study, we explore combining the ICV data with the high-resolution traffic control log data and design new arterial traffic performance metrics based on ICV data. The traditional TSD is solely based on infrastructure traffic data sets. Therefore, bias is likely introduced to reveal congestion under complex traffic conditions with queue spillbacks. The traditional TSD also assumes constant moving speeds between intersections unless vehicles join queues. This assumption may also be questionable because vehicles may slow down to enter a driveway, slowing down the following vehicles. Or vehicles may slow down and move in a two-way left-turn lane. These maneuvers cannot be captured by the infrastructure traffic sensors and so the TSD may convey misleading information about arterial traffic performance.

The selected arterial contains five consecutive intersections of Cooper Street in the City of Arlington, Texas. All five intersections can archive high-resolution traffic signal events data. The speed limit on Cooper Street is 40 mi/h (65 km/h). Because the mainline green duration varies from cycle to cycle under actuated traffic signal control, it will be more accurate if the high-resolution traffic signal events and ICV waypoints are synchronized. By contrast, the freeway congestion can be evaluated solely with the ICV data set. In North America, most arterials are controlled by actuated traffic signal systems. A common mechanism in the actuated traffic signal system is *early return to green*,

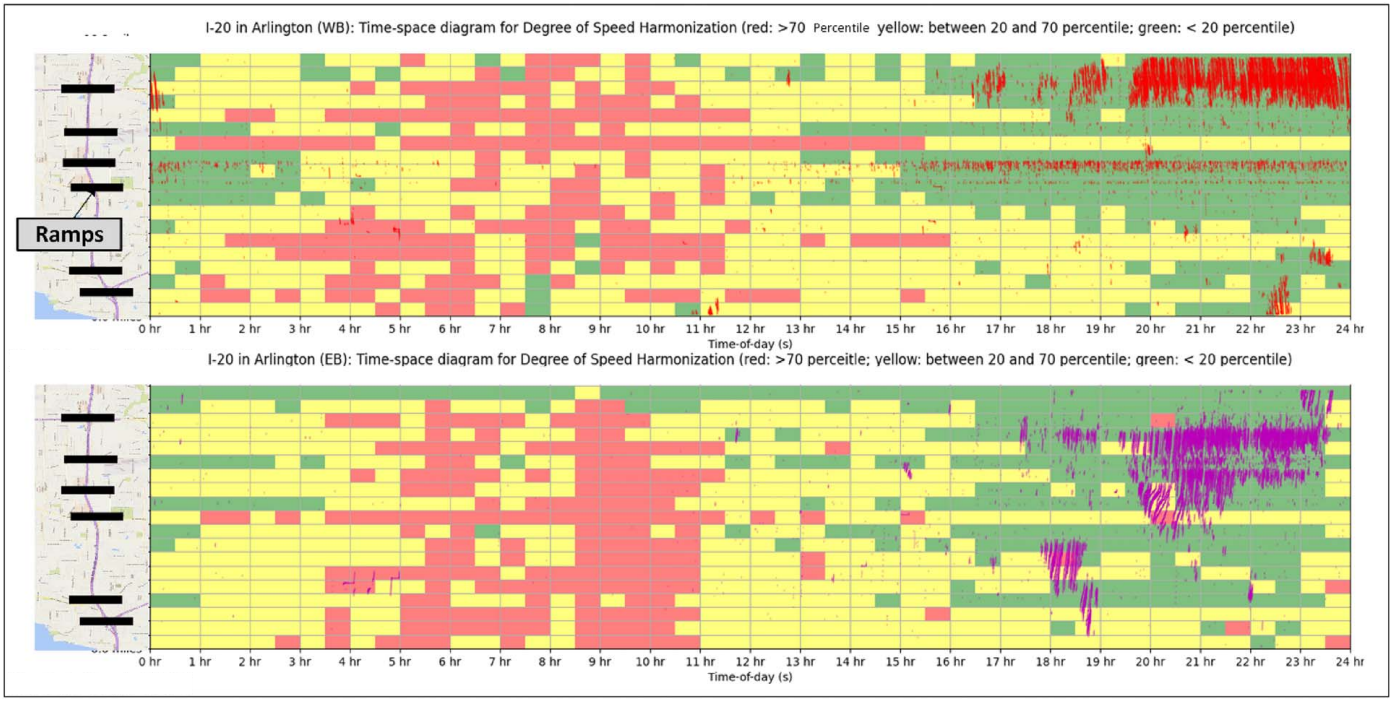

Fig. 6. Time-space characteristics of degree of speed harmonization (DSH).docx. (Map data © 2021 Google.)

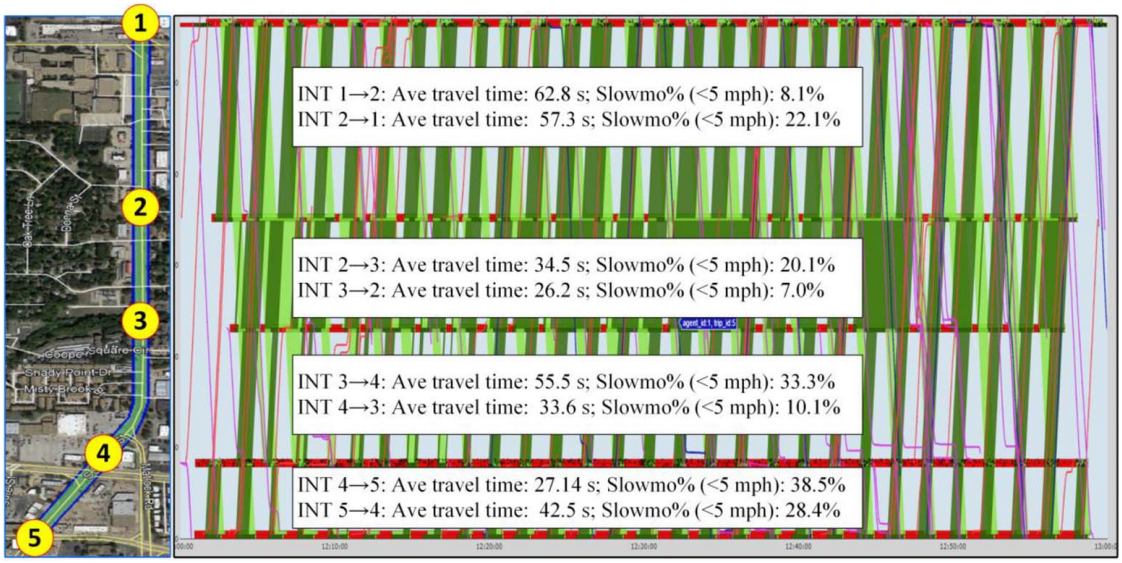

Fig. 7. Arterial time-space diagram and mobile-sensor-based performance metrics. (Image © Google, Maxar Technologies, U.S. Geological Survey, USDA Farm Service Agency © 2021.)

returning any unused green time allocated for minor approaches back to the mainline. As a result, the green duration on the arterial mainline may vary from cycle to cycle. To evaluate the high-fidelity delays, queue lengths, travel speed, and many more given, it is necessary to avoid averaging out the cycle-to-cycle differences. To meet these goals, we design an arterial time-space diagram based on the ICV trajectories and traffic signal events across multiple intersections. Other than the common travel speed or arrivals-on-green percentage described in other literature, we design a new performance metric between two consecutive intersections: the *percentage of slow movements* of vehicle trajectories. Using all the waypoints between two intersections during the study period, the slow movement percentage is calculated as follows:

$$p_s = \left(\frac{\sum_{i=1,..,m} v'_i}{\sum_{i=1,..,n} v_i} \right) \tag{6}$$

where p_s = slow movement percentage; v'_i = waypoint speed which is lower than a threshold [e.g., 5 mi/h (8 km/h)]; v_i = waypoint speeds; m = number of slow speeds; and n = number of total waypoints.

Compared with the link travel speed, an advantage of the slow movement percentage is that it can reflect the vehicles' actual

J. Urban Plann. Dev., 2022, 148(2): 04022016

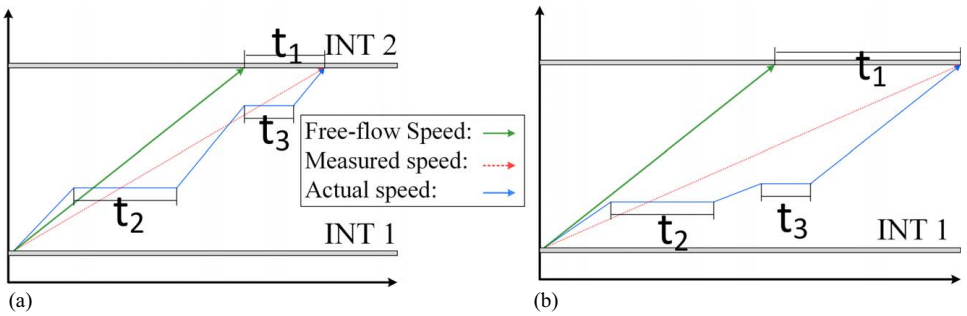

Fig. 8. Using slow movement percentages between intersections to calculate vehicles' actual delay time [when speed < 5 mi/h (8 km/h)].

queuing time between intersections. The link delay is defined as the actual travel time minus free-flow travel time. In practice, vehicles' moving speeds are seen as different from the speed limit. Therefore, the averaged link delay time may misestimate the actual queuing time. In Fig. 8(a), t_1 is the reported link delay while the vehicle's actual delay time is $t_2 + t_3$ and it is greater than t_1 because the vehicle's desired speed is faster than the free-flow speed. Fig 8(b) demonstrates the other possibility, $t_2 + t_3 < t_1$ because the vehicle's moving speed is lower than the speed limit. The driving experience, energy consumption, and emission are more sensitive to low-speed maneuvers. Identifying the actual queuing time accurately is important to evaluate the traffic signal performance and vehicle energy consumptions on arterials. To address this issue, the slow movement percentage is designed to capture the real delay time between intersections, which is critical to estimate the control delay given a traffic signal timing. The aggregated average travel time and slow movement percentages [speed < 5 mi/h (8 km/h)] between intersections are shown in Fig. 7, which is generated according to the ICV data (trajectories in the TSD) and high-resolution traffic signal events (mainline traffic control status and green band) from 12 p.m. to 1 p.m. on December 29, 2020. The actual delay time is calculated as average travel time multiplied by slow movement percentage.

Conclusions

In this paper, we present our efforts in exploring and exploiting the potential ICV data set. The ICV data are passively crowdsourced from the onboard positioning and communication modules in recently manufactured vehicles. The raw data from the distributor, the Wejo Data Service are well organized but have super big volumes. Therefore, we first design efficient algorithms to screen the data set and match the vehicle waypoints to the maps. We further design two new traffic performance metrics and their visualizations in the time-space diagrams based on the characteristics of ICV data sets. One case study is conducted on Interstate I-20 in Arlington, Texas, and the other case study is conducted on Cooper Street in Arlington, Texas, in conjunction with high-resolution traffic control events from the infrastructure. Cooper Street is a major arterial in Arlington. The features of the proposed traffic performance metrics include that: (I) we can identify the queue propagation at freeway bottlenecks, based on where and when most internet-connected vehicles began to considerably slow down and joined the queue; and (II) we can identify a vehicle's actual queuing time between intersections based on its slow movement percentage. The new performance metric does not need to assume a non-delay travel time. More accurate estimation of vehicles' queuing time is critical for traffic signal timing design and the estimation of vehicles' energy consumption. Based on the aforementioned findings, we conclude that the ICV data set has a huge potential in enhancing traffic congestion management planning both on freeways and arterials.

In the future, we will further seek how these newly proposed concepts can be implemented in an online manner and coupled with existing freeway/arterial ATMS and DSS systems that most public agencies are using.

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

This research was partially supported by the project "Exploring a Novel Public–Private–Partnership Data Sharing Policy through a collaborative Arterial Traffic Management System" sponsored by the USDOT UTC Center, Center for Transportation Equity, Decisions, and Dollars (CTEDD). The proposed map-matching method was inspired by the discussion with Dr. Simon Zhou from Arizona State University. 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 above organizations, nor do the contents constitute a standard, specification, or regulation of these organizations.

References

Araghi, B. N., R. Krishnan, and H. Lahrmann. 2016. "Mode-specific travel time estimation using Bluetooth technology." *J. Intell. Transp. Syst.* 20 (3): 219–228. https://doi.org/10.1080/15472450.2015.1052906.

Bakula, C., W. H. Schneider, and J. Roth. 2012. "Probabilistic model based on the effective range and vehicle speed to determine Bluetooth MAC address matches from roadside traffic monitoring." *J. Transp. Eng.* 138 (1): 43–49. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000284.

Banks, J. H. 2009. "Automated analysis of cumulative flow and speed curves." *Transp. Res. Rec.* 2124 (1): 28–35. https://doi.org/10.3141/2124-03.

Barcelo, J., L. Montero, L. Marques, and C. Carmona. 2010. "Travel time forecasting and dynamic origin-destination estimation for freeways based on Bluetooth traffic monitoring." *Transp. Res. Rec.* 2175 (1): 19–27. https://doi.org/10.3141/2175-03.

Bertini, R. L. 2003. "Toward the systematic diagnosis of freeway bottleneck activation." In *Proc., 2003 IEEE Int. Conf. on Intelligent Transportation Systems*, 442–447. New York: IEEE.

Bertini, R. L., M. Lasky, and C. M. Monsere. 2005. "Validating predicted rural corridor travel times from an automated license plate recognition system: Oregon's frontier project." In *Proc., Intelligent Transportation Systems*, 296–301. New York: IEEE.

Brennan, T. M., J. M. Ernst, C. M. Day, D. M. Bullock, J. V. Krogmeier, and M. Martchouk. 2010. "Influence of vertical sensor placement on data collection efficiency from Bluetooth MAC address collection devices." *J. Transp. Eng.* 136 (12): 1104–1109. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000178.

Charbonnier, S., A. C. Pitton, and A. Vassilev. 2012. "Vehicle re-identification with a single magnetic sensor." In *Proc., IEEE Int. Instrumentation and Measurement Technology Conference*, 380–385. New York: IEEE Computer Society.

Chen, C., A. Skabardonis, and P. Varaiya. 2004. "Systematic identification of freeway bottlenecks." *Transp. Res. Rec.* 1867 (1): 46–52. https://doi.org/10.3141/1867-06.

Deng, M., J. Huang, Y. Zhang, H. Liu, L. Tang, J. Tang, and X. Yang. 2018. "Generating urban road intersection models from low-frequency GPS trajectory data." *Int. J. Geog. Inf. Sci.* 32 (12): 2337–2361. https://doi.org/10.1080/13658816.2018.1510124.

Desai, J., H. Li, J. K. Mathew, Y. T. Cheng, A. Habib, and D. M. Bullock. 2021. "Correlating hard-braking activity with crash occurrences on interstate construction projects in Indiana." *J. Big Data Anal. Transp.* 3 (1): 27–41. https://doi.org/10.1007/s42421-020-00024-x.

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

Haghani, A., M. Hamedi, K. F. Sadabadi, S. Young, and P. Tarnoff. 2010. "Data collection of freeway travel time ground truth with Bluetooth sensors." *Transp. Res. Rec.* 2160 (1): 60–68. https://doi.org/10.3141/2160-07.

Hainen, A. M., J. S. Wasson, S. M. L. Hubbard, S. M. Remias, G. D. Farnsworth, and D. M. Bullock. 2011. "Estimating route choice and travel time reliability with field observations of Bluetooth probe vehicles." *Transp. Res. Rec.* 2256 (1): 43–50. https://doi.org/10.3141/2256-06.

Hall, F. L., and K. Agyemang-Duah. 1991. "Freeway capacity drop and the definition of capacity." *Transp. Res. Rec.* 1320: 91–98

Hofleitner, A., R. Herring, and A. Bayen. 2012. "Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning." *Transp. Res. Part B Methodol.* 46 (9): 1097–1122. https://doi.org/10.1016/j.trb.2012.03.006.

Hunter, M., J. K. Mathew, H. Li, and D. M. Bullock. 2021a. "Estimation of connected vehicle penetration on US roads in Indiana, Ohio, and Pennsylvania." *J. Transp. Technol.* 11 (4): 597–610. https://doi.org/10.4236/jtts.2021.114037.

Hunter, M., E. Saldivar-Carranza, J. Desai, J. K. Mathew, H. Li, and D. M. Bullock. 2021b. "A proactive approach to evaluating intersection safety using hard-braking data." *J. Big Data Anal. Transp.* 3: 81–94.

Kavaler, R., K. Kwong, A. Raman, P. Varaiya, and D. Xing. 2011. "Arterial performance measurement system with wireless magnetic sensors." In *Proc., 1st Int. Conf. on Transportation Information and Safety: Multimodal Approach to Sustained Transportation System Development—Information, Technology, Implementation*, 377–385. Reston, VA: ASCE.

Kuroiwa, H., T. Kawahara, S. Kamijo, and M. Sakauchi. 2006. "Vehicle matching between adjacent intersections by vehicle type classification." In *Proc., 2006 IEEE Int. Conf. on Systems, Man and Cybernetics*, 389–394. New York: IEEE.

Kwong, K., R. Kavaler, R. Rajagopal, and P. Varaiya. 2009. "Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors." *Transp. Res. Part C Emerging Technol.* 17 (6): 586–606. https://doi.org/10.1016/j.trc.2009.04.003.

Lawson, T. W., D. J. Lovell, and C. F. Daganzo. 1997. "Using input–output diagram to determine spatial and temporal extents of a queue upstream of a bottleneck." *Transp. Res. Rec.* 1572 (1): 140–147. https://doi.org/10.3141/1572-17.

Li, P., and R. R. Souleyrette. 2016. "A generic approach to estimate freeway traffic time using vehicle ID-matching technologies." *Comput.-Aided Civ. Infrastruct. Eng.* 31 (5): 351–365. https://doi.org/10.1111/mice.12159.

Li, Y., and Q. Li. 2010. "A fast algorithm for identifying candidate links for floating car map-matching: A vector to raster map conversion approach." *Ann. GIS* 16 (3): 177–184. https://doi.org/10.1080/19475683.2010.513153.

Newell, G. F. 1993. "A simplified theory of kinematic waves in highway traffic, Part I: General theory." *Transp. Res. Part B Methodol.* 27 (4): 281–287. https://doi.org/10.1016/0191-2615(93)90038-C.

Qiu, Z., P. Cheng, J. Jin, and B. Ran. 2009. "Cellular probe technology applied in advanced traveller information system." *World Rev. Intermodal Transp. Res.* 2 (2–3): 247–260. https://doi.org/10.1504/WRITR.2009.023310.

Quayle, S. M., P. Koonce, D. Depencier, and D. M. Bullock. 2010. "Arterial performance measures with media access control readers: Portland, Oregon, pilot study." *Transp. Res. Rec.* 2192 (1): 185–193. https://doi.org/10.3141/2192-18.

Richardson, J. K., B. L. Smith, M. D. Fontaine, and S. M. Turner. 2011. "Network stratification method by travel time variation." *Transp. Res. Rec.* 2256 (1): 1–9. https://doi.org/10.3141/2256-01.

Sakhare, R. S., J. C. Desai, J. Mahlberg, J. K. Mathew, W. Kim, H. Li, J. D. McGregor, and D. M. Bullock. 2021. "Evaluation of the impact of queue trucks with navigation alerts using connected vehicle data." *J. Transp. Technol.* 11 (4): 561–576. https://doi.org/10.4236/jtts.2021.114035.

Saldivar-Carranza, E. D., H. Li, and D. M. Bullock. 2021. "Diverging diamond interchange performance measures using connected vehicle data." *J. Transp. Technol.* 11 (4): 628–643. https://doi.org/10.4236/jtts.2021.114039.

Sanchez, R. O., C. Flores, R. Horowitz, R. Rajagopal, and P. Varaiya. 2011a. "Arterial travel time estimation based on vehicle re-identification using magnetic sensors: Performance analysis." In *Proc., 14th IEEE Int. Intelligent Transportation Systems Con.*, 997–1002. New York: IEEE.

Sanchez, R. O., C. Flores, R. Horowitz, R. Rajagopal, and P. Varaiya. 2011b"Vehicle re-identification using wireless magnetic sensors: Algorithm revision, modifications and performance analysis." In *Proc., 2011 IEEE Int. Conf. on Vehicular Electronics and Safety*, 226–231. Beijing: Institute of Electrical and Electronics Engineers (IEEE).

Turner, S., L. Albert, B. Gajewski, and W. Eisele. 2000. "Archived intelligent transportation system data quality: Preliminary analyses of San Antonio TransGuide data." *Transp. Res. Rec.* 1719 (1): 77–84. https://doi.org/10.3141/1719-10.

Waddell, J. M., S. M. Remias, and J. N. Kirsch. 2020a. "Characterizing traffic-signal performance and corridor reliability using crowd-sourced probe vehicle trajectories." *J. Transp. Eng. Part A. Syst.* 146 (7): 04020053. https://doi.org/10.1061/JTEPBS.0000378.

Waddell, J. M., S. M. Remias, J. N. Kirsch, and T. Trepanier. 2020b. "Utilizing low-ping frequency vehicle trajectory data to characterize delay at traffic signals." *J. Transp. Eng. Part A. Syst.* 146 (8): 04020069. https://doi.org/10.1061/JTEPBS.0000382.

Wagner, F. R. 2014. "Perspectives from the field: The Federal Highway Administration's 'Every day counts' initiative: Finding a path to better environmental outcomes within existing NEPA and agency regulations." *Environ. Pract.* 16 (4): 347–348. https://doi.org/10.1017/S1466046614000350.

Xu, C., P. Liu, B. Yang, and W. Wang. 2016. "Real-time estimation of secondary crash likelihood on freeways using high-resolution loop detector data." *Transp. Res. Part C Emerging Technol.* 71: 406–418. https://doi.org/10.1016/j.trc.2016.08.015.

Xu, T. D., L. J. Sun, Z. R. Peng, and Y. Hao. 2011. "Modelling drivers' en-route diversion behaviour under variable message sign messages using real detected traffic data." *Inst. Eng. Technol.* 5 (4): 294–301. https://doi.org/10.1049/iet-its.2011.0060.

Zhang, L., and D. Levinson. 2004. "Some properties of flows at freeway bottlenecks." *Transp. Res. Rec.* 1883 (1): 122–131. https://doi.org/10.3141/1883-14.

Zhao, W., E. McCormack, D. J. Dailey, and E. Scharnhorst. 2013. "Using truck probe GPS data to identify and rank roadway bottlenecks." *J. Transp. Eng.* 139 (1): 1–7. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000444.