


Article

Development of a Turning Movement Estimator Using CV Data

Somayeh Nazari Enjedani and Mandar Khanal * 

Department of Civil Engineering, Boise State University, Boise, ID 83725, USA

* Correspondence: mkhanal@boisestate.edu

Abstract: Turning movement (TM) data of vehicular traffic at intersections are a basic input for signal timing design. Existing methods of collecting TM data are time- and cost-intensive. Using connected vehicle (CV) data is an alternative method. Trajectories of vehicles through an intersection can be constructed using CV data. However, because of the low number of CVs in the traffic stream, it is imprecise to consider TM data from CVs as representative of the whole traffic flow. To address this issue, a Kalman filter (KF) for estimating TM rates at intersections based on CV data under low market penetration levels using commercially available connected vehicle data was developed in this study. This method is independent of intersection geometry or the presence of shared lanes. The algorithm was evaluated using data from an intersection in Salt Lake City, Utah. The manually collected TM counts at this intersection were compared with the raw CV data as well as the results obtained from the developed methodology. The comparison shows that while TM counts based on raw CV data show severe violations in accuracy, making them unreliable, the method developed in this research gives results that have much lower accuracy violations.

Keywords: turning movement counts; Kalman filtering; connected vehicle data; vehicle trajectory

1. Introduction

Traffic signal timing and efficiency analysis require vehicle TM count data at intersections [1]. In addition to signal timing applications, this information can be applied to dynamic traffic assignment and adaptive signal control [2]. It is not costly or complicated to estimate flows entering an intersection from its approaches using existing vehicle detection infrastructure [3]. Traditionally, traffic volumes have been obtained from fixed sensors such as loop detectors, microwave detectors, and video-imaging detectors [4]. While these devices can be used to measure approach volumes, they cannot measure TMs, which are the classification of an approach traffic stream into left, right, and through streams through the intersection [5]. At this time, manual counts are the most common and also the most labor-intensive method of collecting TM data.

The recent emergence of connected vehicles (CVs) opens up several possibilities for improving the operation of traffic signals [6]. With vehicle-to-infrastructure (V2I) communication, CVs can serve as mobile sensors by continually reporting their status to roadside equipment (RSE) or satellites. An alternative way of obtaining CV data is through commercial data services companies that provide data collected by various vehicle manufacturing companies. Such data provide vehicle movement data captured every 3 s from ignition on to ignition off for the full length of a journey. Vehicle trajectory data are also available at 3 m accuracy. Data services companies obtain such data from vehicle manufacturing companies and provide the data to customers. The data are available in near real time in regular batches (daily, weekly, or monthly) or as a historical one-off purchase. The research reported here used three months of historical data for selected months within the limits of Salt Lake City, Utah. Three months of vehicle movement data and three months of driving event data were purchased from a commercial data service company.

The necessity for fixed location detectors in the current signal systems may thus be greatly reduced or possibly completely eliminated with the increasing prevalence of CVs.



Citation: Nazari Enjedani, S.; Khanal, M. Development of a Turning Movement Estimator Using CV Data. *Future Transp.* **2023**, *3*, 349–367. <https://doi.org/10.3390/futuretransp3010021>

Received: 21 January 2023

Revised: 11 February 2023

Accepted: 27 February 2023

Published: 3 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Since CVs have not yet fully penetrated the vehicle market, we will need to use data fusion methods to obtain the most from the data we already have. CV data might be combined with data from traditional detectors to compute TMs with acceptable levels of accuracy.

It is possible to calculate TM percentages at an intersection for every time interval using CV data. These TM percentages can then be multiplied by the total approach volume obtained from other means, such as loop detectors, to estimate TM counts from that approach. However, such TM counts are not reliable because of the low penetration rate of CVs in the traffic stream. To address this issue, a Kalman filter (KF) methodology was developed in this study by utilizing historic data of TMs from CVs to obtain a reliable estimation of the TM rates for every time interval. Since CV data are available citywide, the methodology developed in this study can be used to estimate TMs at every intersection in the city for any desired time interval. The availability of such citywide concurrent data will also enable corridor- or network-level management of traffic operations. The methodology described here can be characterized as a TM estimator that can be used as an alternative to the traditional manual counting of TMs.

2. Material and Methods

2.1. Literature Review

Manual counting is the most basic method of obtaining TMs at an intersection, but this method is time-consuming and costly. Attempts to simplify this procedure first centered on utilizing an O-D matrix to solve the problem, but this strategy proved to be imprecise and unstable when utilized with an existing data collection system [7]. Other studies attempted to identify vehicle TMs using detector data [2,3,8–11]. However, the existence of shared lanes made vehicle tracking difficult, restricting the use of these methods to intersections without shared lanes.

Some researchers tried to improve the convenience and coverage of TM data collection by taking advantage of the existing video detection system at signalized intersections [12–15]. There were also some proposed approaches for intersections equipped with radar-based vehicle detection systems [16]. The main problem with these approaches is that they are useful only at intersections where video detection devices or radar-based devices have been installed. In addition, this process could be cost-prohibitive because it will require specialized hardware to be deployed at the intersections.

Meanwhile, others have tried to predict TMs at intersections using machine-learning-based models [17,18]. The issue with these studies and several other studies is that their algorithms require both inbound and outbound traffic volumes as input, whereas detectors often only gather traffic entering the intersection [19–22]. Overall, none of the suggested approaches has proved functional enough to be widely adopted as a substitute for manual counting by transportation planners and state authorities. Recently, some studies have attempted to use CV data to aid in estimating traffic volumes at intersections. Zheng and Liu attempted to estimate the volume of traffic at signalized intersections [23]. They defined the volume estimation problem as a maximum likelihood problem and solved it with an expectation maximum algorithm based on the theory of a time-dependent Poisson arrival process. The results of the Zheng and Liu study, which was the first attempt to estimate traffic volumes using sampled vehicle trajectories, can be very helpful in understanding how to extract traffic flow information from vehicle trajectory data. In their study, it appears that the volumes for various directions at a junction were calculated, although it is unclear how they found the split of the traffic volumes in each direction. In another study, Tang utilized sampled vehicle trajectories to estimate cycle-based traffic volume at signalized intersections using the tensor decomposition approach under low penetration rates [24]. Tang calculated cyclic volumes using CV data but did not estimate intersection TMs. A study by Saldivar-Carranza and another by Saldivar-Carranza, Li, and Bullock apply CV data to find TMs at intersections [25,26]. Individual vehicle TMs at signalized intersections are automatically identified using connected vehicle trajectory data. At an intersection, entry and exit trajectory directions are gathered to form movement clusters, which are then

analyzed using k-means to determine the number of clusters, and specific TMs are assigned to the centroids. The most serious obstacle to implementing this approach in real-world situations is the fact that the current penetration rate is below ten percent, which means that the findings based on only a limited sample of the real population may be unreliable. Utilizing a small sample to forecast the behavior of the entire population is undesirable; hence, a scientific approach to expanding the sample to the entire population is required.

2.2. Data Collection

Raw data used in this study come from CV data collected by a commercial data service company. The company has information for 9.1 trillion data points involving 44.4 billion journeys across a network of 10.7 million live vehicles from a supply base of over 50 million connected vehicles [27]. Another company collaborates with the data service company to provide the raw data from CVs in a user-friendly interface that makes it easier to process the raw data. The data used in this research were obtained for the months of March 2019, January 2021, and August 2021. These months were chosen because TM data from manual counts at three intersections in Salt Lake City were available from the Utah Department of Transportation (UDOT) for selected days during those months. The manually collected data were used as the ground truth in this study. The CV data have 179,070 journeys for March 2019, 2,353,877 journeys for January 2021, and 2,887,201 journeys for August 2021. These figures show the rapid growth of connected vehicles in Salt Lake City between March 2019 and August 2021.

It should be noted that the suggested method could not be used with the data for 2019 because there were not enough data points for the study month. Only the data from 2021 were used as the input for our algorithm. These restrictions forced us to evaluate our strategy only at one intersection, the 700 East and 900 South intersection in Salt Lake City. This was the only intersection with manual counts in 2021. Fortunately, three days of manual count data for this intersection were available: January 20, January 21, and August 24 of 2021.

2.3. Data Processing

The interface described above was used to capture CV information for the intersections. An interactive map and a control panel make up the interface. Waypoints generated by the CVs are shown on the map. Waypoints are generated once every three seconds when a CV is moving. A journey ID number is used to organize the waypoints according to the overall trip of which they are a part and allows for the collection of CV volumes. The waypoints carry many detailed pieces of information, such as the geographical location, timestamp, speed, acceleration, heading, information about the origin and destination of the trip, and many other features related to the particular waypoint.

Using the interface, it is possible to determine how CVs entering an approach at an intersection split into various directions. To do that, the entry and exit gates for every approach need to be defined first. Then, using a network tool available in the interface, it is possible to track the vehicles that enter an approach entry gate and exit from exit gates, which gives us TMs. Aggregating data from successive time intervals of 15 min, we can derive the TMs for every desired time interval that is a multiple of 15 min, for every desired date. This demonstrates the TM behavior of a small number of cars sending data to satellites but is not indicative of the movement of all traffic traveling through the intersection; however, this limited information can still be used to infer the behavior of the entire traffic flow.

For the analysis of a single intersection, a relevant part of the street network around the intersection must be defined. First, all links of the network that start or end inside a circle within a radius of around 300 feet from the center point of the intersection are considered. The region within this circle is believed to represent the area of influence due to the signalization. Second, another circle with a radius of 170 feet is defined around the same center. These radii are chosen based on the suggestion in [26]. Choosing the

influence area of an intersection in this manner also conforms to the recommendation of the *Highway Capacity Manual* [28], which suggests that the intersection influence area extends to a distance of 250 feet from the stop bar. Third, using these two circles, four polygons on four approaches are defined as the entry and exit gates for traffic flow into and out of the intersection. These circles and polygons are known as waypoint geometry. They are essentially geofences that record the entry and exit of vehicles inside these defined spaces. We need the geofences because grouping data into 15 min time intervals leads to journey IDs being counted multiple times. By defining this cross-shaped waypoint geometry, the query data would now have to follow the rules of the network and be inside this waypoint geometry during any given timeslot. This solves the problem of journey IDs being counted multiple times. Figure 1 illustrates the three waypoint geometries that help define entry and exit gates. Figure 2 shows the final defined gates.

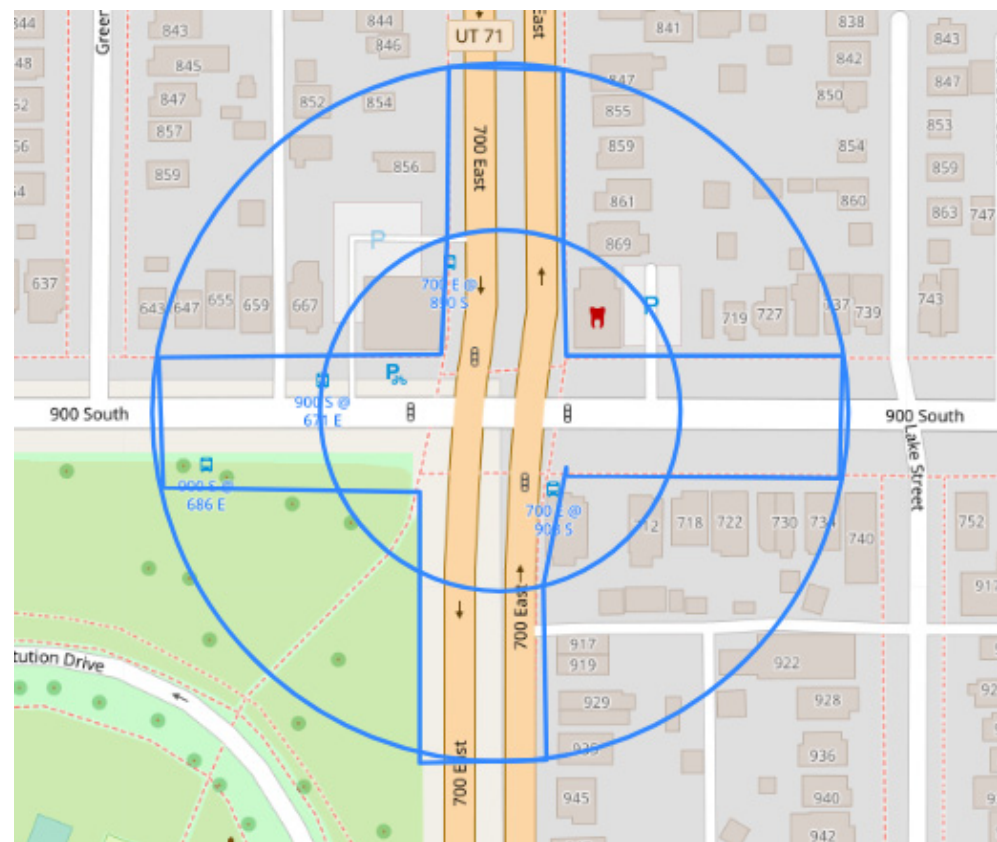


Figure 1. Geofences to Define Entry/Exit Gates.

The next step is to define the network to count TMs. Figure 3 depicts an example of such a network for northbound traffic. The four nodes in Figure 3 correspond to the four gates shown in Figure 2. This network is used to process a query that counts every journey ID that passes through both the entry gate (7184NB ENTRY) and one of the three exit gates (7184EB, 7184NB, and 7184WB) during the specified time intervals. The result of applying such a query on this network is shown in Figure 4. In this figure, the orange, green, and blue dots represent vehicles that are going through, turning right, or turning left, respectively. The existence of orange dots in segments other than northbound or green dots in southbound instead of eastbound segments can be a little confusing. This happens because vehicles are observed at different points in their journey and some drivers may have traveled through this intersection more than once during a trip; the system is picking up data points from different time periods during their trips. However, such occurrences do not affect the computation of TM percentages in the northbound direction. The TM data are exported to tables for further processing. Using these tables, it is easy to calculate the

turning percentages in each direction and for every desired time interval. However, due to the low CV penetration rates, the TM percentages computed in this manner are unlikely to accurately reflect the turning behavior of all vehicles passing through the intersection. The method described in this paper addresses this issue, and in the case study, the accuracy of the results obtained using this technique is compared to that of the results calculated directly from CV data.

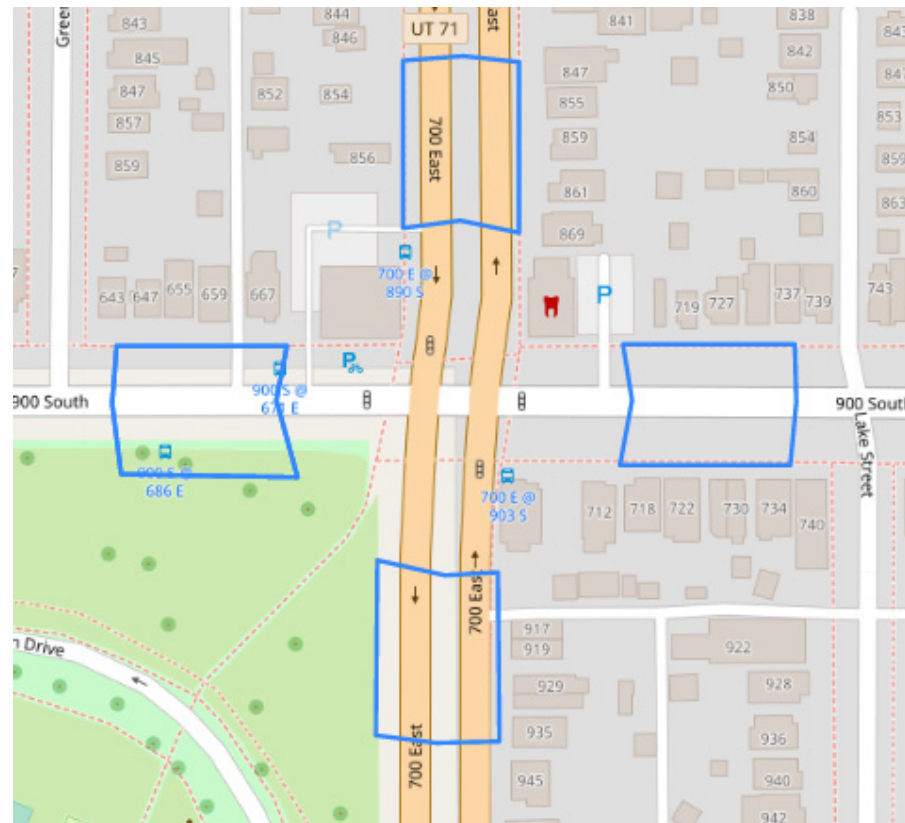


Figure 2. Entry and Exit Gates.

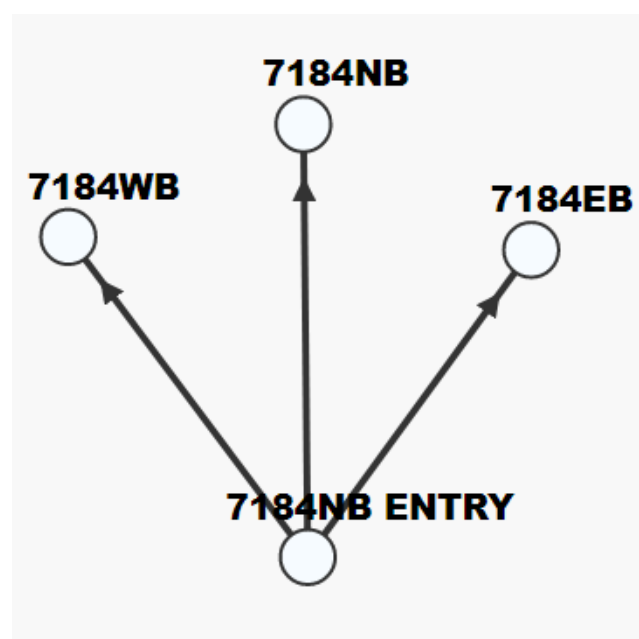


Figure 3. Network for Northbound Traffic.

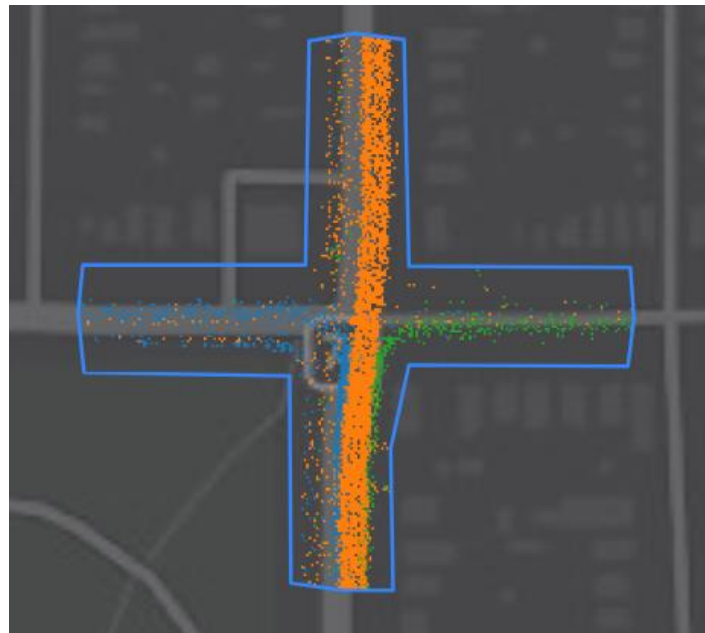


Figure 4. Turning Movement from DB4IoT.

2.4. Methodology

This study provides a method that uses CV data to estimate TM rates at intersections in a low-penetration CV environment. These estimated TM rates can then be applied to actual traffic counts derived from detector data to produce TM counts. This solution is not expensive since it does not require the installation of any hardware at the intersection, unlike many earlier attempts. It also functions for all intersection geometries, with or without shared lanes, and for intersections instrumented with any type of traffic detection system.

The TM percentages obtained from the CV data alone will likely not represent the turning behavior of all traffic flowing through the intersection because of the low penetration rates of CVs. At low penetration levels, the number of vehicles that send data to the data collection center will be small. The small sample size makes it unreasonable to make inferences about the entire population of vehicles using the network at any given time. In addition, these data are tied to the past, while we may also desire to predict turning counts for future time intervals. An alternative approach to address these concerns is necessary. A methodology based on Kalman filters (KFs) was chosen to fulfill this need. The KF approach was applied to historical CV data to estimate the turning rate for each subsequent time period.

A KF is a mathematical algorithm for obtaining a future estimate using historical data. The KF method was first proposed by Rudolf Kalman in 1960 [29]. A KF is used to estimate states based on linear dynamical systems in state-space format. The process model defines the evolution of the state from time k to time $k + 1$ [30]. In this study, we assume that k represents one day. According to the KF, in each step (one day in our case), the estimation of TM counts for a desired time interval is derived from a weighted sum of the prediction, which is computed as an intermediate step, and the present measurement, which is the CV TM counts for the desired day. A formula is used to calculate the prediction, which is based on all CV TM counts for all days leading up to the desired day. The Kalman gain is the metric that indicates the magnitude of the contribution of the measurement to the estimate. Each of the subsequent steps is described in detail below.

The “state” of the dynamical system is defined as an $(n \times 1)$ matrix and denoted by x_k . In this case, n equals 4, which includes three TM counts for a chosen time interval (left,

through, right) and the overall approach volume, which is the sum of the TM volumes. The vector x_k is shown as follows:

$$x_k = \begin{bmatrix} \text{Left – turning movements} \\ \text{Through movements} \\ \text{Right – turning movements} \\ \text{Approach volume} \end{bmatrix}$$

The evolution of the “state” is depicted by Equation (1). This is also known as the process model.

$$x_{k+1} = Ax_k + w_k \quad (1)$$

where

x_k : state variable for day k ; $(n \times 1)$ column vector (and similarly for day $k + 1$).

A : state transition matrix; $(n \times n)$ matrix.

w_k : state transition noise; $(n \times 1)$ column vector.

According to Equation (1), the “state” in each step is a function of the “state” in the previous step. The matrix A describes how the system changes over time. Further definition of the variables used in Equation (1) is given below.

A is the state transition matrix that is a user-defined matrix. In this study, the following value was chosen for the transition matrix A to reflect the linear relationship between the

TM and total approach volume of two successive days. $A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$ w_k is the process

noise vector, which is assumed to be a zero-mean Gaussian with a variance-covariance matrix Q , i.e., $w_k \sim N(0, Q)$. Q is an $(n \times n)$ diagonal matrix that will be used directly in the KF algorithm. Noise cannot be predicted but can be estimated statistically. We calculated Q based on the mean and variance of noise estimated from historical CV data. For more clarification, Equation (1) can be expressed as follows:

$$\begin{bmatrix} \text{Left – turning movements} \\ \text{Through movements} \\ \text{Right – turning movements} \\ \text{Approach volume} \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \text{Left – turning movements} \\ \text{Through movements} \\ \text{Right – turning movements} \\ \text{Approach volume} \end{bmatrix}_k + w_k$$

The process model is paired with the measurement model that describes the relationship between the state and the measurement at the current time step k [31]. The measurement is supposed to be TM counts coming from CV data for every time interval of each day (k). We applied this model for time intervals of one and two hours.

The measurement or observation is an $(m \times 1)$ vector. The measurement, which is derived from CV data, serves as the input to the KF algorithm. Indeed, the measurement is the number of connected vehicles that travel straight or make a left or right turn at the intersection at a given time each day. In our scenario, m equals 3, which corresponds to three TMs (left, through, and right) over a predefined time interval for an approach.

The measurement vector, z_k , is as follows:

$$z_k = \begin{bmatrix} \text{Left – turning movements} \\ \text{Through movements} \\ \text{Right – turning movements} \end{bmatrix}$$

The measurement model is shown in Equation (2).

$$z_k = Hx_k + v_k \quad (2)$$

z_k : measurement; $(m \times 1)$ column vector.

x_k : state variable for day k ; $(n \times 1)$ column vector.

H : state-to-measurement transition matrix; $(m \times n)$ matrix.

v_k : measurement noise; $(m \times 1)$ column vector.

The matrix H is also a user-defined transition matrix. It reflects the linear relationship between the measurement and the state variable. We defined matrix H as follows:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The measurement noise vector, v_k , is assumed to be zero-mean Gaussian with variance-covariance matrix R , i.e., $v_k \sim N(0, R)$. R is a $(m \times m)$ diagonal matrix, which will then be used in the KF algorithm; in our case, we used arbitrary values to define the matrix R . Equation (2) can be stated as follows for further clarity:

$$\begin{bmatrix} \text{Left - turning movements} \\ \text{Through movements} \\ \text{Right - turning movements} \end{bmatrix}_K = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \text{Left - turning movements} \\ \text{Through movements} \\ \text{Right - turning movements} \\ \text{Approach volume} \end{bmatrix}_k + v_k$$

The combination of the process model and the measurement model is called the system model, which is the foundation of the KF algorithm. The KF algorithm receives one input (the measurement, z_k) and returns one output (the estimate, \hat{x}_k). The notation “ \sim ” indicates that the variable to which it is applied is an estimation. Internal processing of the algorithm is performed through a four-step computation that is depicted in Figure 5. It is important to note that the superscript “ \sim ” shown on some variables in the algorithm means the variable bearing that symbol is a predicted value.

The four steps of KF consist of two separate processes: the process of prediction and the process of estimation. The first step is for the prediction, the second with calculating the Kalman gain, the third with the estimation, and the fourth with calculating the error covariance. Following is a description of these steps in greater detail.

The first step in the algorithm is the prediction. The two variables \hat{x}_k and P_k , which will be used in steps **II** through **IV**, are computed in this step. P is the covariance matrix of the estimation error. The formula in this prediction step has a very close relationship with the system model. In the prediction procedure, how the estimate \hat{x}_k will vary when time changes from K to $K + 1$ is predicted.

In step **II**, the Kalman gain (K_k) is computed. The Kalman gain plays an important role in estimating the TMs. It identifies the magnitude of the contribution of the measurement in estimation. This means that at every iteration, the Kalman gain describes the share of the measurement (the last data from CV) relative to the current-state estimate. With a high gain, the filter gives the most recent measurements more weight and responds to them more quickly.

In step **III**, an estimate for TMs is derived from a weighted sum of a measurement (the most recent data from CV) and the prediction (the aggregate of past data) from step **I**. The magnitude of the Kalman gain computed in step **II** represents the weight of the last CV data in the estimate. In step **IV**, the error covariance (P_k) is computed. Error covariance is a measure indicating the accuracy of the estimate. Normally, the decision to trust or discard the estimate computed in the previous step is made based on the review of the error covariance. If P_k is large, the error of the estimate is large, and, conversely, if P_k is small, the error of the estimate is small [31]. At the beginning of the computation, P is estimated, and in subsequent iterations, its value is estimated by the algorithm.

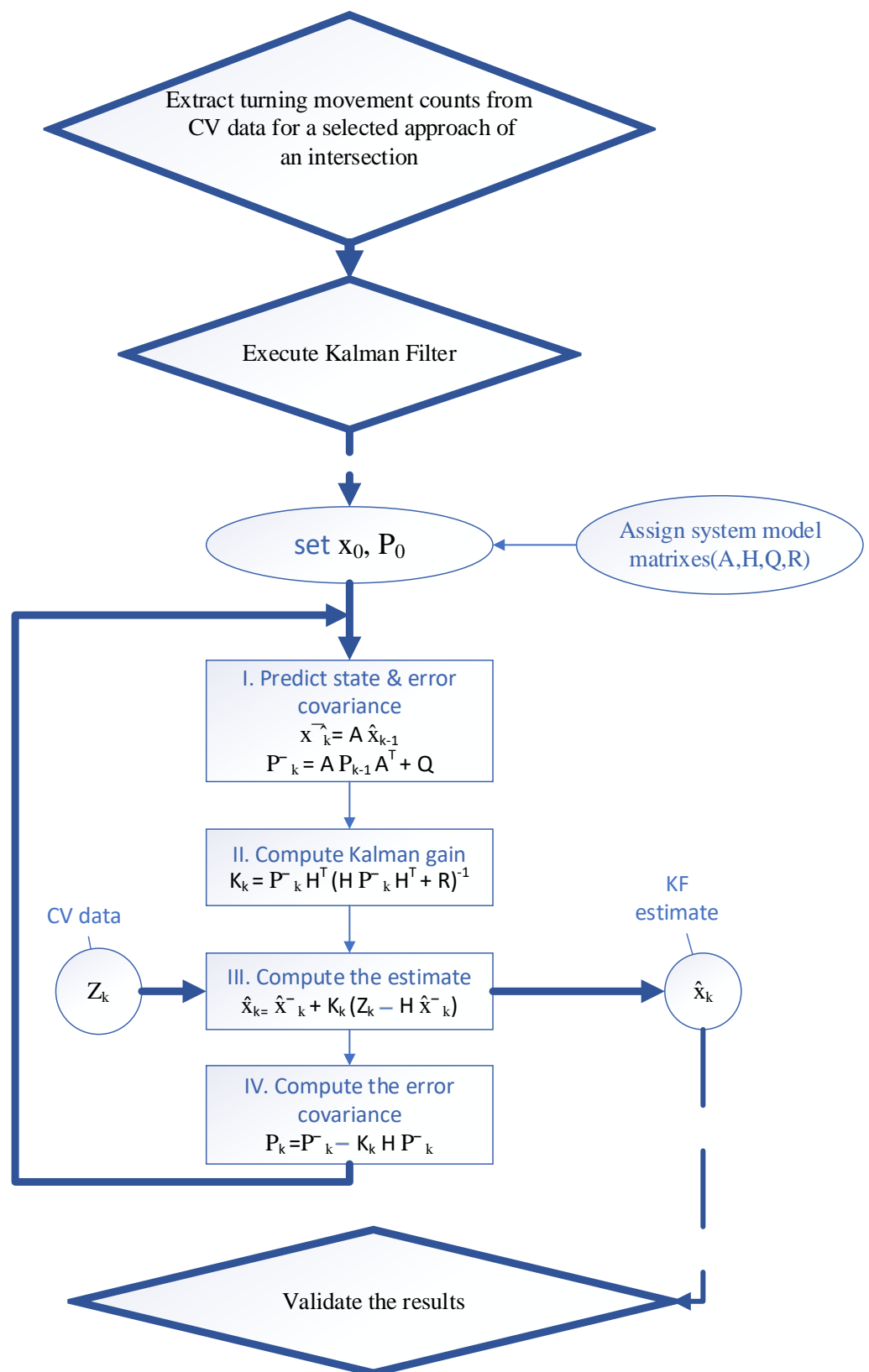


Figure 5. Kalman Filter Algorithm.

In summary, the KF algorithm estimates are obtained as the sum of the prediction (\hat{x}_k) and the current measurement (Z_k) with appropriate weightings. In the KF, the weighting (K_k) applied to obtain the estimate is not constant, but different at each iteration. The weighting is updated at each iteration by using the formula in step II of Figure 5 [30]. This characteristic of KF is what helps us to extract a more reasonable estimate for TM counts on the basis of CV data when the penetration rate is low compared to simply using CV data at each point.

Accurately defining the matrices Q and R is important, but doing so analytically has its limitations due to the multiplicity of error sources. In order to maximize the utilization of the noise information, these two matrices should be calibrated by trial and error [32]. In this study, the covariance matrix of x_k is calculated from CV data and is set to Q . Regarding R , for the first trial, an arbitrary value was set and then adjusted based on the variance of the CV data and the result from the first trial.

The bigger the Q , the bigger the Kalman gain, and the smaller the R , the bigger the Kalman gain. When the Kalman gain is large, so is the measurement's contribution. Because it is undesirable for each measurement to have an excessive influence, big values for Q and small values for R should not be employed. This concept stems from the fact that the penetration rate in this study is low (under 2%), making it impossible for every iteration to adequately introduce reality. Therefore, care was taken to reduce the impact of the most recent iteration and obtain an estimate with less variation while giving past data more weight.

All variables used in the algorithm and described above are summarized in Table 1. System model variables are defined by the user since these are design factors; all other variables are calculated or measured by the algorithm. Throughout the processes depicted in Figure 5, the algorithm calculates or measures all other variables.

Table 1. Variables Used in Kalman Filter.

External Input	z_k (Measurement)
Final output	\hat{x}_k (estimate)
System model	A, H, Q, R
For internal computation	\hat{x}_k, P_k, P_k, K_k

3. Results

A case study was conducted to assess the proposed estimation technique. In this case study, the TM counts at the intersection of 700 East and 900 South in Salt Lake City were investigated. We used CV data collected by the data service company as input to the proposed method. The CV data used were collected in January and August of 2021. The manual count data collected by UDOT for this intersection were used as the ground truth data. Manual count data were available for two hours around the peak hours of the evening of 20 January 2021 and two hours around the peak hours of the morning of 21 January 2021, and also from 6:00 a.m. to 8:00 p.m. on 24 August 2021. All three of these days are weekdays. Therefore, we were able to evaluate our method for weekdays. Evaluation for weekends could not be performed because of the lack of ground truth data. Figure 6 displays an image of the intersection downloaded from Google Maps. It is important to note that this technique is independent of the geometry of the intersection since it does not utilize any information derived from the geometry of the intersection. Instead, it obtains data from CVs and modifies them using a mathematical procedure. Consequently, it is not dependent on the geometry or the existence of shared lanes, unlike many prior techniques.

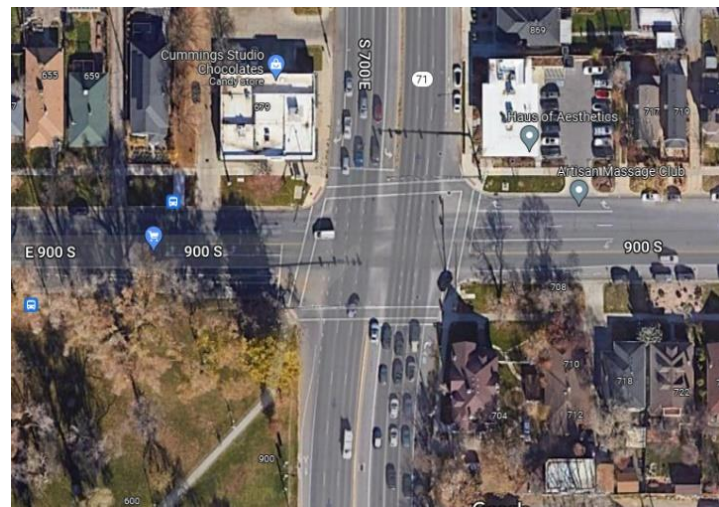


Figure 6. Intersection 700 East and 900 South, Salt Lake City. (Source: Google Maps).

Given the limited availability of ground truth data and the low penetration rate of CVs, the KF algorithm was evaluated at two levels. First, the algorithm was deployed and evaluated only during the morning and evening peak hours. Considering two different days of ground truth data during the morning and evening peak hours for each day as well as 12 distinct TMs for each intersection, there were 48 instances available for comparison in this scenario. It should be emphasized that the time interval in this case is two hours based on the peak hour reported by UDOT for the intersection. We picked this two-hour time period because, after examining the CV data, we determined that if we chose a time interval of one hour, we would not have enough CV data to run the algorithm for all directions. Thus, we opted for a wider range of time spanning the peak hours. Finally, for all weekdays (Monday through Friday) in August and January of 2021, a two-hour time interval in the morning from 7:30 to 09:30 and a two-hour time interval in the afternoon from 4:15 to 6:15 were chosen. This ensured that the algorithm could be run for all 48 cases. Second, the algorithm was applied and assessed hourly from 6:00 a.m. to 8:00 p.m. As mentioned earlier, we only have hourly ground truth data for the whole day of August 24; on the other dates, we only have two hours of TM counts around the peak hours. We anticipated having 168 instances since there are 14 time intervals in this scenario, and for each hour we have 12 different TMs, but due to the poor quality of the CV data during some hours, only 138 instances were available for analysis. For example, on some days of the month at 6:00 a.m., eastbound, there were no CV data available; hence, the algorithm was unable to finish its stages for that hour and direction. Future work may include the modification of the algorithm to overcome this issue. It may be feasible to complete the task by extrapolating the missing hour from days after or before the day with the missing data.

It is worth mentioning that comparing volumes from CV data with the UDOT manual counts, the CV penetration rate for the mentioned time intervals in January was found to be 1.23%; for August, it was 1.5%, for an average of 1.38%. However, one study in 2021 estimated the penetration rate on US roads in Indiana, Ohio, and Pennsylvania at around 4.5% [33]. This difference can be rooted in the fact that in this study, limited hours for three days at a specific intersection were used, which may not reflect the overall penetration rate of CV data in Salt Lake City.

For the intersection of interest, the CV data for TM counts and the total approach volumes were extracted first. The data were exported in CSV format. The CSV files were then imported into the statistical analysis software package, R, where a script for the KF algorithm was developed and run. The final outcomes obtained by running the algorithm for each time interval comprised the fraction of each TM volume as a percentage of the overall approach volume for the targeted time period, separately for the TMs coming directly from CV data and the ones generated by the KF algorithm. We also calculated the

fraction of each TM for time intervals of interest from ground truth data and compared results from CV and KF with the ground truth data by calculating the residuals in each case.

Regarding the first level of analysis, our goal was to estimate all TMs from the four approaches to this intersection for the morning and evening peak hours. The algorithm used CV data for two hours around the peak hours of one day for every iteration; each iteration is equivalent to one weekday.

Figures 7–9 show KF plots for through movements, left turns, and right turns for northbound traffic from 4:15 to 6:15 p.m. in August 2021. In these plots, dots indicate CV data (the measurement), and the solid line represents the estimated TM as a percentage of the total approach volume calculated by the KF algorithm. These figures show that the KF algorithm smoothens the estimation of the TM percentages by removing noise.

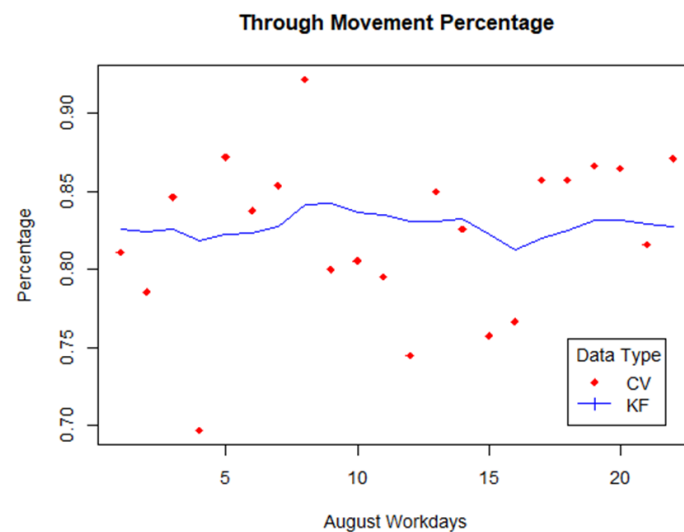


Figure 7. Northbound Through Movement 4:15 to 6:15 p.m. August, 2021.

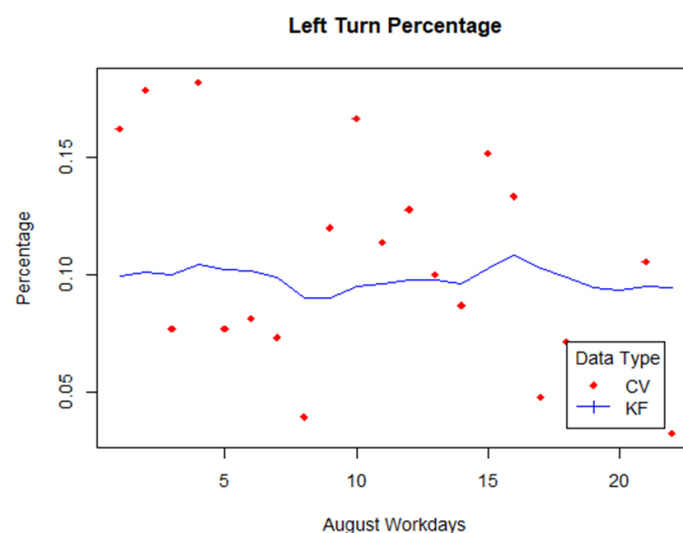


Figure 8. Northbound Left Turns 4:15 to 6:15 p.m. August, 2021.

Figures 10–13 show the TM percentages obtained from CVs, KF estimation, and manual counts for 24 August 2021 between 7:30 and 9:30 a.m., 24 August 2021 between 4:15 and 6:15 p.m., 21 January 2021 between 7:30 and 9:30 a.m., and 20 January 2021 between 4:15 and 6:15 p.m. for all TMs for the study intersection. Table 2 provides a summary of Figures 10–13. It illustrates the residuals of comparing KF estimation with ground truth data and the residuals of comparing CV results with ground truth data for all 48 cases. The

first column of this table displays the mean, standard deviation, minimum, 25%, 50%, and 75% quantiles, and maximum residuals and the overall root mean square error (RMSE) of the KF algorithm in comparison with the ground truth for the 48 examples used in this study. The second column displays the same information for CV results and ground truth data. These data make it quite evident that, on average, better outcomes are obtained through KF rather than raw CVs. As the data in the table show, the mean of the residuals for CV is around 5.89%, while it is around 2.15% for KF. This indicates that overall, the KF method gives better results. Another important observation is that the residual range for CV is between 0.15% and 24.5%, while this range in KF is much narrower, from 0.1 to 9.4%. It shows fewer variations in KF results that prevent large errors. The 75% quantile reveals that the residuals were close to 3% for 75% of KF results, while this figure is 9.6% for CVs. We can also see the overall RMSE for KF estimation and CV results. KF's overall RMSE was 3.07% as opposed to 8.31% for CVs, which again confirms the better accuracy of the method used in this study in comparison to using raw CV data.

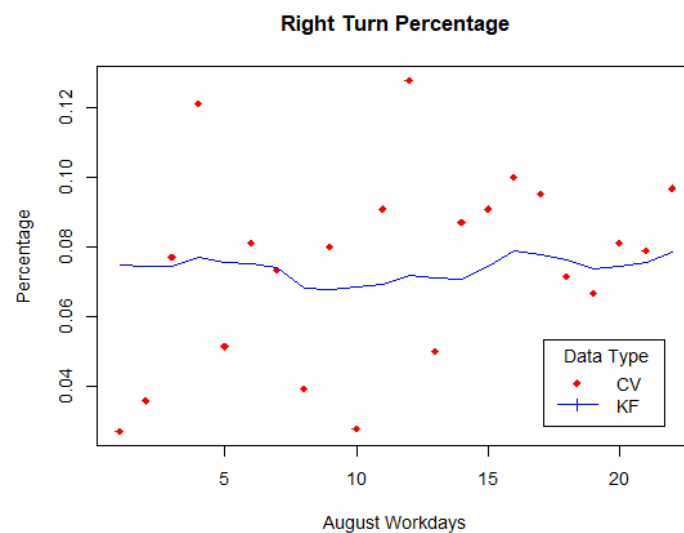


Figure 9. Northbound Right Turns 4:15 to 6:15 p.m. August, 2021.

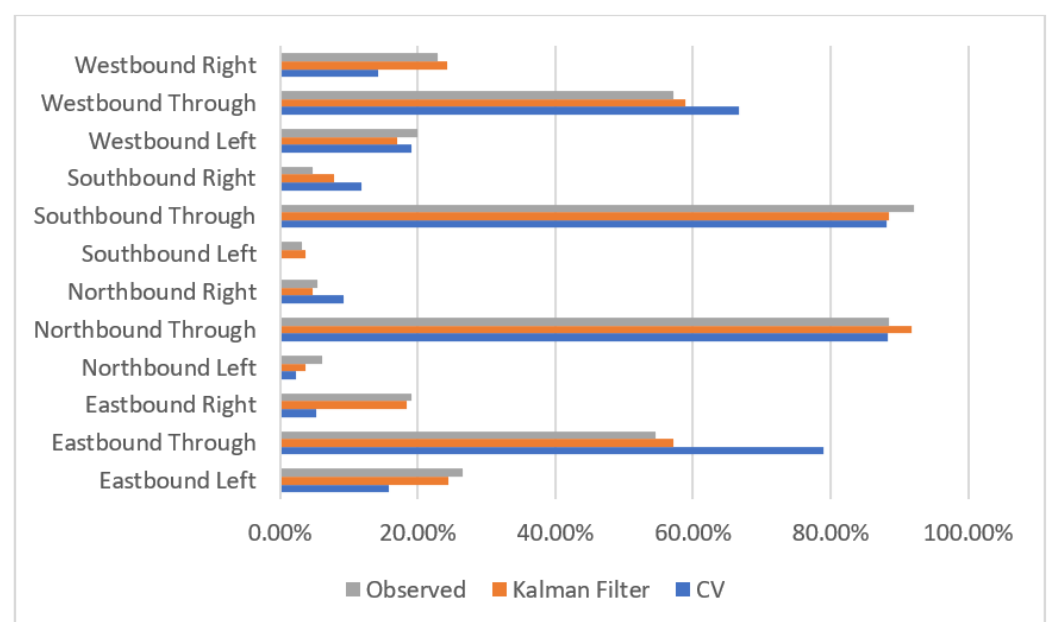


Figure 10. Comparison of CV, KF, and Observed TM Percentages for 7:30–9:30 a.m., 24 August 2021.

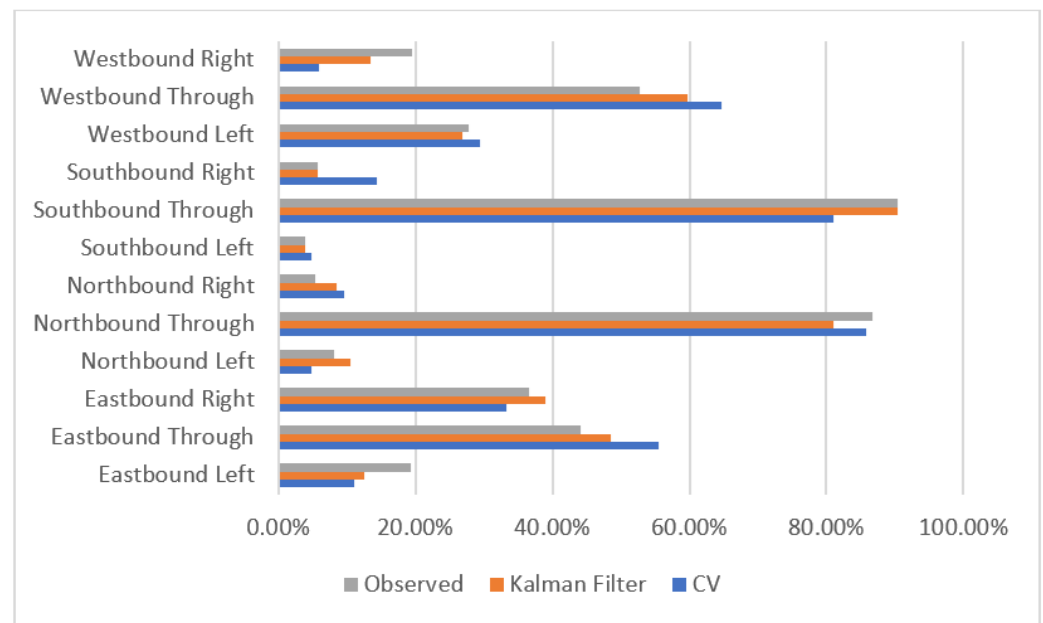


Figure 11. Comparison of CV, KF, and Observed TM Percentages for 4:15–6:15 p.m. 24 August 2021.

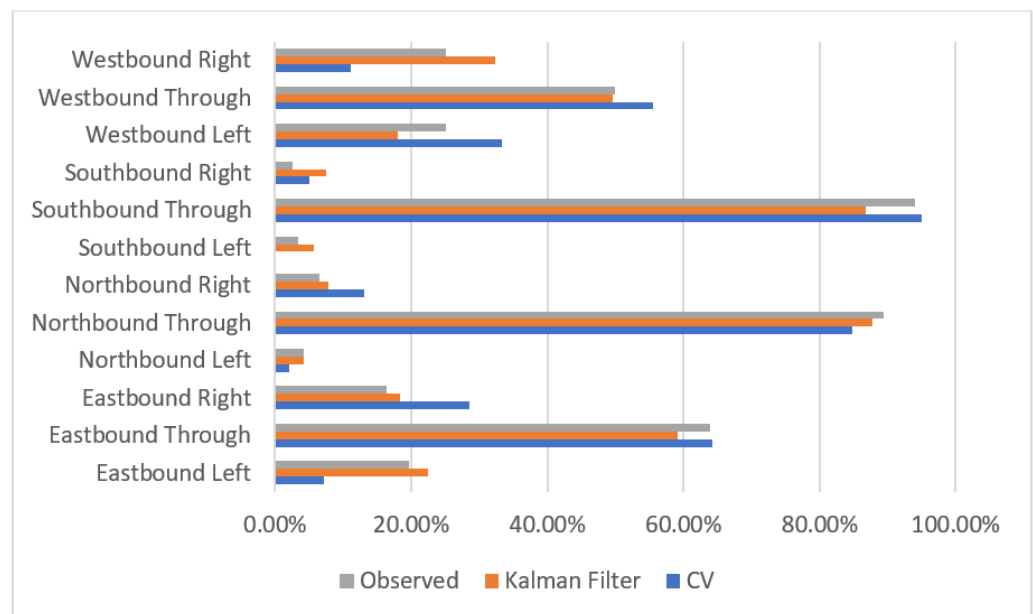


Figure 12. Comparison of CV, KF, and Observed TM Percentages for 7:30–9:30 a.m. 21 January 2021.

Although KF yields higher overall outcomes, it should be noted that CV sometimes outperforms KF in certain situations. Table 3 shows the residuals from CV and KF estimations and the ground truth data for 7 out of the 48 cases in which CV outperforms KF. Only 14.5% of cases fall in this category. However, this small proportion of cases with high accuracy results for CVs may not prove that the raw CV results provide accurate estimates for TMs. KF estimates may be more accurate due to their smaller residual variance.

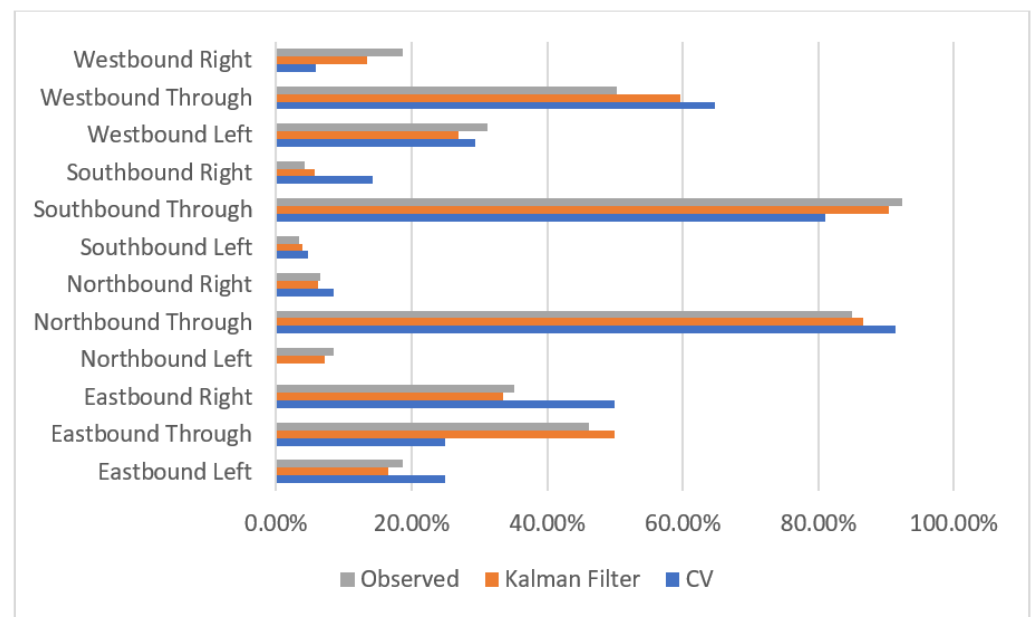


Figure 13. Comparison of CV, KF, and Observed TM Percentages for 4:15–6:15 p.m. 20 January 2021.

Table 2. Residuals of CV and KF Predictions vs. Ground Truth (Peak-Hour Case).

Metric	KF vs. Ground Truth %	CV vs. Ground Truth %
Number of Cases	48	48
Mean	2.15	5.89
Standard Deviation	2.23	5.93
Min	0.10	0.15
25% Quantile	0.38	1.05
50% Quantile	1.60	3.80
75% Quantile	3.05	9.65
Max	9.40	24.50
RMSE	3.07	8.31

Table 3. Residuals of Instances with Superior KF Outcomes (Peak-Hour Case).

Metric	KF vs. Ground Truth %	CV vs. Ground Truth %
Number of Cases	7	7
Mean	3.91	0.97
Standard Deviation	2.11	8.44
Min	0.30	0.15
25% Quantile	3.10	0.28
50% Quantile	4.30	0.90
75% Quantile	4.70	1.40
Max	7.20	2.40

It should be noted that the results shown in Tables 2 and 3 were obtained using only one month of historical CV data for each time period. In other words, for January, only

the CV data for that month were utilized, and the same was done for August. To find out whether this result would change if a wider range of CV data were used, the TM percentages were also calculated using CV data from the past two months and compared to the manual counts. In Table 4, the lowest, highest, and average RMSE values for one and two months of historical data are shown. The residuals for the KF results range from 0.04% to 7%, with an RMSE of 3.06%. This shows that using more historical data yields slightly better KF outcomes. However, despite a penetration rate of less than 2%, findings from only one month of historical data are also satisfactory. On the other hand, to estimate TM percentages for smaller time intervals, a wider range of CVs may be needed, especially under low CV penetration rates. This is something that needs to be explored in future studies.

Table 4. Comparison of Min and Max Residuals and Overall RMSE with One and Two Months of Historical Data.

Source of Data	Min Residual %	Max Residual %	RMSE %
Two months	0.04	7.00	3.06
One month	0.1	9.40	3.07

Regarding the second level of evaluation, the goal was to estimate TMs from the four approaches at the study intersection on 24 August 2021, from 6:00 a.m. to 8:00 p.m. At every iteration, the algorithm was processed with CV data for one weekday. As mentioned earlier, not enough data were available for shorter time intervals in all directions for all the desired hours. As the intention was to use hourly data, time intervals that lacked full CV data were ignored, as mentioned earlier. In the end, there were 138 examples to compare to real-world data.

Table 5 summarizes the residuals of the predictions using CV data and the KF algorithm in comparison to the ground truth data. These results also confirm that, on average, better outcomes are obtained through KF rather than raw CVs. The mean of the residuals for CV is around 10.31%, while it is around 2.68% for KF. This shows that the KF algorithm produces better outcomes overall. Another important note is that the residual range for CV is between 0.1% and 57.3%, while this range in KF is much narrower, from 0.0 to 13.6%. There is a smaller variation in the KF results which precludes large errors. The 75% quantile reveals that in CV, the residuals were less than 3.7% for 75% of KF results, while this number is 14.6% for CV, which is much higher. Overall the RMSE was 3.70% for KF, while it was 14.3% for CV.

Table 5. Residuals of CV and KF Predictions vs. Ground Truth (Hourly Case).

Metric	KF vs. Ground Truth %	CV vs. Ground Truth %
Number of Cases	138	138
Mean	2.68	10.31
Standard Deviation	2.57	10.03
Min	0.00	0.10
25% Quantile	0.90	3.13
50% Quantile	2.10	6.85
75% Quantile	3.70	14.60
Max	13.60	57.30
RMSE	3.70	14.30

As previously stated, although KF produces better results most of the time, CV sometimes surpasses KF. An example for the hourly scenario is shown in Table 6. This happened in 20 out of 138 instances. This is 14.5% of all cases. As the table shows, 75% of the errors using the KF method for these cases are less than 4.7%; as such, this should not be a significant concern. The overall findings of this research indicate that the KF method consistently produces outcomes that are superior to those of CV.

Table 6. Residuals of Instances with Superior KF Outcomes (Hourly Case).

Metric	KF vs. Ground Truth %	CV vs. Ground Truth %
Number of Cases	20	20
Mean	4.12	2.34
Standard Deviation	3.13	2.88
Min	0.40	0.10
25% Quantile	1.95	0.78
50% Quantile	3.95	1.20
75% Quantile	4.78	2.43
Max	12.00	10.40

4. Discussion

Traffic signal systems may undergo a paradigm shift as a result of CV technology's quick development. The information from CVs offers an opportunity to scale down or possibly do away with the necessity of traditional traffic detectors, enabling a wide range of real-time traffic analysis in the long run. However, in the near future, data from CVs may be especially valuable in providing offline performance measures for traffic signal systems and adjusting signal operation on a periodic basis, such as every one to two months. This possibility is particularly advantageous for enhancing signal operation.

In this study, we developed a KF technique to calculate the distribution of traffic volumes for each approach at intersections using data from private automobiles' navigation systems. Traffic volume distribution is an essential input for regulating traffic lights. One of the objectives of the suggested strategy was to cope with low CV penetration rates, for example, less than 2% for the intersection in Salt Lake City, Utah, that was employed in this research. Because CV technology is still in its early phases, the CV penetration rate is not high enough to make the behavior of connected vehicles representative of the whole traffic behavior. With the methodology developed in this study, the percentages of TMs obtained directly from CV data can be modified to make them highly accurate. These adjusted TM rates can then be multiplied by the overall approach traffic volume measured by traditional detectors or other counting devices to obtain the intersection TM volumes. The developed method was evaluated by comparing the TMs that were derived directly from CV data and those derived using the proposed KF method to the TMs that were manually collected using two scenarios: a peak-hour case and an hourly case. The findings demonstrated that, on average, KF produces better outcomes than CVs. Concerning peak-hour cases, the residuals of CVs against ground truth data varied from 0.15% to 24.50%, with an average of 5.89%, while the residuals of the KF against ground truth data were in a much tighter range from 0.10 to 9.40%, with an average of 2.15%. KF's overall RMSE was 3.07% as opposed to 8.31% for CVs in this case. In the hourly scenario, the residuals of the CVs versus ground truth data ranged between 0.10% and 57.30%, with an average of 10.31%, while the residuals of the KF versus ground truth data ranged from 0.00% to 13.60%, with an average of 2.68. The RMSE for KF was 3.70% compared to 14.30% for CVs. Both scenarios show fewer variations in KF results that prevent large errors, besides the smaller RMSE confirming the accuracy of the KF technique.

The recommended method may serve as a starting point for adjusting traffic signals using CV data. This approach may be utilized as an alternative to the traditional manual method of collecting TM counts at intersections, which is expensive and labor-intensive. This method is not expensive since it does not involve the installation of any hardware at the intersection, nor is it restricted to intersections with a certain sort of geometry or traffic detector system, unlike many previous efforts. The detector counts are still required for the approach volumes, but the labor and money required for traditional manual counts to determine TM volumes will not be required. Although data from CVs are currently not free, this study demonstrates that even one month of CV data can produce satisfactory outcomes.

This study can be extended in several ways. Updating the algorithm for turning movement estimations for shorter time intervals, such as every 15 min or even cycle-by-cycle estimation, can be one of the ways to extend this study.

Author Contributions: Conceptualization, M.K.; methodology, M.K.; software, S.N.E.; validation, S.N.E. and M.K.; formal analysis, S.N.E.; investigation, S.N.E.; resources, M.K.; data curation, S.N.E.; writing—original draft preparation, S.N.E.; writing—review and editing, M.K.; visualization, S.N.E.; supervision, M.K.; project administration, M.K.; funding acquisition, M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded partially by Subaward No. UWSC9924 from the University of Washington to Boise State University from the U.S. Department of Transportation award to the University of Washington.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used in this research can be obtained by contacting the corresponding author at mkhanal@boisestate.edu.

Acknowledgments: The authors are grateful for the funding received from the PacTrans Region 10 University Transportation Center that made the procurement of the CV data used in this study possible. They are also grateful to the Department of Civil Engineering at Boise State University for their support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Noyce, D.A.; Bill, A.R.; Chitturi, M.v.; Santiago-Chaparro, K.R. Turning Movement Counts on Shared Lanes: Prototype Development and Analysis Procedures Final Report for NCHRP IDEA Project 198. 2019. Available online: <https://trid.trb.org/view/1652249> (accessed on 15 June 2022).
2. Xu, K.; Yi, P.; Shao, C.; Mao, J. Development and Testing of an Automatic Turning Movement Identification System at Signalized Intersections. *J. Transp. Technol.* **2013**, *3*, 241–246. [CrossRef]
3. Hauer, E.; Pagitsas, E.; Shin, B.T. Estimation of Turning Flows from Automatic Counts. *Transp. Res. Rec.* **1981**, *795*, 1–7.
4. Vigos, G.; Papageorgiou, M. A Simplified Estimation Scheme for the Number of Vehicles in Signalized Links. *IEEE Trans. Intell. Transp. Syst.* **2010**, *11*, 312–321. [CrossRef]
5. Karapetrovic, J.; Martin, P.T. Estimation of Intersection Turning Movement Flows with the TMERT3 Model Version: Sensitivity to a Widespread Detector Failure. *Int. J. Traffic Transp. Eng.* **2021**, *11*, 442–453. [CrossRef]
6. Qi, H.; Dai, R.; Tang, Q.; Hu, X. Quasi-Real Time Estimation of Turning Movement Spillover Events Based on Partial Connected Vehicle Data. *Transp. Res. Part C Emerg. Technol.* **2020**, *120*, 102824. [CrossRef]
7. Nihan, N.L.; Davis, G.A. Application of Prediction-Error Minimization and Maximum Likelihood to Estimate Intersection O-D Matrices from Traffic Counts. *Transp. Sci.* **1989**, *23*, 77–90. [CrossRef]
8. Maher, M.J. Estimating the Turning Flows at a Junction: A Comparison of Three Models. *Transp. Res. Board* **1984**, *25*, 19–22.
9. Mahmoud, N.; Abdel-Aty, M.; Cai, Q.; Yuan, J. Predicting Cycle-Level Traffic Movements at Signalized Intersections Using Machine Learning Models. *Transp. Res. Part C Emerg. Technol.* **2021**, *124*, 102930. [CrossRef]
10. Virkler, M.R.; Kumar, N.R. System to Identify Turning Movements at Signalized Intersections. *J. Transp. Eng.* **1998**, *124*, 607–609. [CrossRef]
11. Noyce, D.; Chittori, M.; Santiago-Chaparro, K.; Bill, A.R. Automated Turning Movement Counts for Shared Lanes Using Existing Vehicle Detection Infrastructure Final Report for NCHRP IDEA Project 177. 2016. Available online: <https://trid.trb.org/view/1422700> (accessed on 15 June 2022).

12. Shirazi, M.S.; Morris, B. Vision-Based Turning Movement Counting at Intersections by Cooperating Zone and Trajectory Comparison Modules. In Proceedings of the 2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014, Qingdao, China, 8–11 October 2014; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2014; pp. 3100–3105. [\[CrossRef\]](#)
13. Yi, P.; Zhang, S. *Development and Field Testing of an Automatic Turning Movement Identification System*; State Job Number: 135141; The Ohio Department of Transportation, Office of Statewide Planning & Research: Columbus, OH, USA, 2017.
14. Shirazi, M.S.; Morris, B.T. Vision-Based Turning Movement Monitoring: Count, Speed & Waiting Time Estimation. *IEEE Intell. Transp. Syst. Mag.* **2016**, *8*, 23–34. [\[CrossRef\]](#)
15. Bélisle, F.; Saunier, N.; Bilodeau, G.A.; le Digabel, S. Optimized Video Tracking for Automated Vehicle Turning Movement Counts. *Transp. Res. Rec.* **2017**, *2645*, 104–112. [\[CrossRef\]](#)
16. Santiago-Chaparro, K.R.; Chitturi, M.; Bill, A.; Noyce, D.A. Automated Turning Movement Counts for Shared Lanes: Leveraging Vehicle Detection Data. *Transp. Res. Rec.* **2016**, *2558*, 30–40. [\[CrossRef\]](#)
17. Ghanim, M.S.; Shaaban, K. Estimating Turning Movements at Signalized Intersections Using Artificial Neural Networks. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 1828–1836. [\[CrossRef\]](#)
18. Shaaban, K.; Hamdi, A.; Ghanim, M.; Shaban, K.B. Machine Learning-Based Multi-Target Regression to Effectively Predict Turning Movements at Signalized Intersections. *Int. J. Transp. Sci. Technol.* **2022**, *12*, 245–257. [\[CrossRef\]](#)
19. Mechler, A.M.; Machemehl, R.B.; Lee, C.E.; Rpo, G. *The Design of an Automated Traffic Counting System with Turning Movement*; Center for Transportation Research, The University of Texas at Austin: Austin, TX, USA, 1986.
20. Gholami, A.; Tian, Z. Using Stop Bar Detector Information to Determine Turning Movement Proportions in Shared Lanes. *J. Adv. Transp.* **2016**, *50*, 802–817. [\[CrossRef\]](#)
21. Ghods, A.H.; Fu, L. Real-Time Estimation of Turning Movement Counts at Signalized Intersections Using Signal Phase Information. *Transp. Res. Part C Emerg. Technol.* **2014**, *47*, 128–138. [\[CrossRef\]](#)
22. Karapetrovic, J.; Martin, P.T. Estimating Intersection Turning Movement Flows with a NETFLO Algorithm: Weight Constraint Calibration. *Adv. Transp. Stud.* **2020**, *52*, 73–88. [\[CrossRef\]](#)
23. Zheng, J.; Liu, H.X. Estimating Traffic Volumes for Signalized Intersections Using Connected Vehicle Data. *Transp. Res. Part C Emerg. Technol.* **2017**, *79*, 347–362. [\[CrossRef\]](#)
24. Tang, K.; Tan, C.; Cao, Y.; Yao, J.; Sun, J. A Tensor Decomposition Method for Cycle-Based Traffic Volume Estimation Using Sampled Vehicle Trajectories. *Transp. Res. Part C Emerg. Technol.* **2020**, *118*, 102739. [\[CrossRef\]](#)
25. Carranza, S. *Scalable Operational Traffic Signal Performance Measures from Vehicle Trajectory Data*; Purdue University: West Lafayette, IN, USA, 2021.
26. Saldivar-Carranza, E.D.; Li, H.; Bullock, D.M. Identifying Vehicle Turning Movements at Intersections from Trajectory Data. In Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 19–22 September 2021; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2021; Volume 2021, pp. 4043–4050. [\[CrossRef\]](#)
27. Available online: <http://www.wejo.com> (accessed on 18 July 2022).
28. National Research Council (U.S.); Transportation Research Board. *Highway Capacity Manual*; Transportation Research Board, National Research Council: Washington, DC, USA, 2000.
29. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [\[CrossRef\]](#)
30. van Lint, H.; Djukic, T. Applications of Kalman Filtering in Traffic Management and Control. In *2012 TutORials in Operations Research*; INFORMS: Catonsville, MD, USA, 2012; pp. 59–91. [\[CrossRef\]](#)
31. Kim, P. *Kalman Filter for Beginners: With MATLAB Examples*; A-JIN Publishing Company: Seoul, Republic of Korea, 2011.
32. Bishop, G.; Welch, G. *An Introduction to the Kalman Filter*; SIGGRAPH ACM, Inc.: Los Angeles, CA, USA, 2001.
33. Hunter, M.; Mathew, J.K.; Li, H.; Bullock, D.M. Estimation of Connected Vehicle Penetration on US Roads in Indiana, Ohio, and Pennsylvania. *J. Transp. Technol.* **2021**, *11*, 597–610. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.