

## Article

# Comprehensive Data Analysis Approach for Appropriate Scheduling of Signal Timing Plans

Nemanja Dobrota <sup>1,\*</sup> , Nikola Mitrovic <sup>2</sup>, Slavica Gavric <sup>1</sup> and Aleksandar Stevanovic <sup>1</sup> 

<sup>1</sup> Department of Civil and Environmental Engineering, Swanson School of Engineering, University of Pittsburgh, 3700 O'Hara Street, 341A Benedum Hall, Pittsburgh, PA 15261, USA; slg130@pitt.edu (S.G.); stevanovic@pitt.edu (A.S.)

<sup>2</sup> A&P Engineers a CHA Company, 8935 NW 35th Lane, Suite 200, Doral, FL 33172, USA; nmitrovic@apcte.com

\* Correspondence: ned47@pitt.edu

**Abstract:** Improperly scheduled signal timing plans are one of the main reasons for reduced efficiency of traffic signals at coordinated urban arterials. Recently, most urban arterial roads are equipped with intelligent transportation systems devices capable of reporting the collected data on high temporal and spatial resolution, which gives us the opportunity to overcome traditional signal timing planning flaws. Previous studies have proposed methods for scheduling signal timing plans based on small quantities of data combined with various optimization approaches that ultimately require domain expert intervention to fine-tune proposed solutions. Consequently, the signal timing plans scheduling problem is still being addressed without a comprehensive approach. In this study, we propose a novel data-driven procedure based on visual analytics principles to identify the dominant traffic profiles and appropriate scheduling of signal timing plans. The medium-resolution volume data collected over a one-year period on a real-world corridor consisting of 12 signalized intersections were used to validate the proposed methodology. Applied principles from the visual analytics domain allow for a better understanding of traffic characteristics and ultimately alleviate the development of appropriate signal timing schedules. The results show that the proposed method more reliably schedules signal timing plans when compared to current practice.

**Keywords:** signal timing plans; scheduling; traffic profiles; medium-resolution data; *k*-means clustering; visual analytics; coordinated arterials



**Citation:** Dobrota, N.; Mitrovic, N.; Gavric, S.; Stevanovic, A.

Comprehensive Data Analysis

Approach for Appropriate

Scheduling of Signal Timing Plans.

*Future Transp.* **2022**, *2*, 482–500.

<https://doi.org/10.3390/futuretransp2020027>

futuretransp2020027

Academic Editor: Jian Li

Received: 13 April 2022

Accepted: 23 May 2022

Published: 1 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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/>).

## 1. Introduction

Traffic conditions on urban arterial vary significantly over time, exhibiting variations in traffic volumes during various time intervals (e.g., within an hour, day, week, month, season, or a year) [1–3]. Developing an appropriate scheduling of multiple (time-of-day) signal timing plans is often seen as an efficient and cost-effective solution, in coping with such inevitable fluctuations in traffic demand [4–7]. To develop and schedule the multiple signal timing plans, most of the agencies continue to rely on the traffic data gathered for a few (or even single) representative days. Often, to reduce the signal retiming and data collection costs, agencies deploy the identical set of signal timing plans for each weekday while those plans developed for low/moderate traffic are in operations during the weekend [4]. Improperly scheduled signal timing plans (STPs) are one of the most important reasons for improper operations of traffic signals [4].

Since a few years ago, most of the urban arterial roads are equipped with intelligent transportation systems devices capable of reporting the collected data on high temporal and spatial resolutions. This valuable source of information can undoubtedly be widely used by agencies in solving practical traffic-related problems, such as signal re-timing applications and for evaluation of the existing breaking points. Numerous studies were conducted in the past to address the issue of determination of time-of-day breakpoints [8–13]. Overall,

previous studies either proposed methods based on small temporal and spatial resolution (thus preventing for appropriate inclusion of diurnal traffic fluctuations into signal timing plans), or they were based on various optimization approaches that serve to identify signal timing plan schedules in an automatic manner that prevent domain experts from interacting with the specifics of traffic data fluctuations, which is very valuable considering one's in-house expertise. Consequently, determination of the TOD breakpoint is still being addressed without a comprehensive approach.

With the rapid development in sensor technology, a field of data science emerged as a consequence of the increased availability of data and our inability to appropriately analyze them [14]. This emerging field is called visual analytics (VA), which integrates domain principles with visual representations of the new information in order to gain insight and provide adequate solutions [15]. Numerous applications of analyzing large datasets in transportation domains are found so far [16–18]. However, very little attention is given to urban arterial management with respect to scheduling time-of-day (TOD) signal timing plans, also known as TOD breakpoints.

The goal of this study is to provide easy-to-apply solutions for determination signal timing plan schedules in the data rich environment. In this study we propose a straightforward yet novel data-driven procedure, based on clustering and the VA approach, for identification of the dominant traffic profiles and selection of multiple TOD breaking points. Medium-resolution volume data (i.e., data collected at mid-block stations in an ongoing manner and aggregated at several minutes bins) were collected from a typical urban arterial road for the period of one year. Noteworthy, definitions of other traffic data resolutions can be found in a recent study [19]. Collected data were then analyzed from its temporal and spatial aspect by using *k*-means clustering and principles of VA (that mainly advocates domain-expert involvement during the data analysis process). Such a novel way of analyzing the data allows for interaction with data and insightful analysis (e.g., directional splits, weekly patterns, etc.). The proposed method is quite robust and easily applicable to most of the ITS equipped coordinated arterial roads.

The rest of the paper is structured as follows. First, a brief overview of relevant research is presented. Then, the authors present the methodology, utilized for achieving research goals. In the subsequent chapter, each of the methodological steps are described. In the following section, obtained results from proposed method are presented and discussed. Finally, the paper is concluded with concluding remarks and ideas for future research.

## 2. Literature Review

Time-of-day (TOD) traffic signal control is based on the principle where multiple TOD signal timing plans (STPs) are scheduled to operate during the day. TOD plans are characterized by a unique cycle length that is highly correlated with traffic volumes. The determination of characteristic periods during the day where certain signal timing plans should operate was primarily done using the 24 h volume plots from single intersection and engineering judgments [20]. In cases where data availability is higher, the analytical methods that serve for the determination of robust TOD breakpoints were not in place.

Since early 2000, the problem of optimal determination of TOD plans was examined in greater detail. Smith et al. used a statistical, hierarchical clustering method on four months of data from a single intersection [20]. A follow-up study served to validate the proposed approach [21]. It was found that such statistical methods result in isolated clusters, whose assignment to dominant ones (e.g., AM peak) should be performed manually. Therefore, Park et al. proposed an heuristic approach to identify homogeneous TOD breakpoints, which will result in less frequent signal timing plans changes [22]. The follow-up study contained an automated approach which considered the effects of volume and signal timing plans while determining TOD breakpoints using the data from an arterial containing three intersections [8]. Later, Park and Lee also accounted plans transition costs while applying the heuristic-greedy algorithm on a 24 h dataset [9]. Wang et al. applied *k*-means clustering on several hours of volume data on two intersections [23]. A small

aggregation interval (of 5 min) resulted in frequent transitions between plans [21]. In 2008, Wong and Woon proposed an iterative  $k$ -means clustering method on twelve hours of volume data obtained from microsimulation (i.e., MITSIMLab) [24]. Dong et al. utilized isomap and  $k$ -means clustering algorithms for one day of volume data collected at a single intersection [25].

In 2011, Ratrout proposed subtractive clustering-based  $k$ -means technique on three coordinated intersections for 24 h volume data [10]. Jun and Yang applied Kohonen neural network on five consecutive weekdays volume on a three-leg, signalized intersection [11]. Guo and Zhang clustered seven days traffic data from nine mid-block detectors per direction, separately [26]. A final TOD schedule was developed by arbitrarily combining different TOD patterns from both directions [26]. Hao and Dong performed a two-dimensional clustering analysis method based on 24 h volumes on the multiple spatially isolated intersections [12]. Therefore, the proposed method did not account for specifics of traffic operations at coordinated arterials. Wan et al., using the bisecting  $k$ -means, determined TOD breakpoints based on trajectory data collected for several days at one intersection [13].

Ma et al. solved the problem of TOD determination by time series partitioning using 24 h data from a single intersection [27]. Chen et al., in a comparative study, used different clustering methods (i.e.,  $k$ -means, hierarchical, and Fisher ordinal clustering) to determine TOD breakpoints [28]. Data collected for three days at one intersection were used. Recently, Wang et al. proposed clustering of data collected for five consecutive days to account for data continuity rather than aggregating data for individual days [29]. Data were collected for one intersection and aggregated for all movements. Interestingly, the problem of characteristic time-of-day periods was examined recently for planning purposes [30]. Song and Yang used clustering on offline traffic data to examine similarity and characteristics of traffic flow patterns on the city-wide network area [31]. The authors did not utilize examined data to estimate the quality of existing signal timing plan schedules [31].

Previous studies mainly relied on analytical reasoning of data, which hinders an analyst's understanding of specific temporal flow fluctuations in cases when a longer series of data are available. Furthermore, most of the studies examined small-scale networks where spatial characteristics of utilized datasets are not examined. One emerging scientific field that primarily serve to extract more information from data is visual analytics (VA). In the domain of transportation engineering, numerous applications of VA were found in the literature [16–18]. However, to our knowledge, no related work to arterial management with respect to TOD breakpoints determination was found. Furthermore, big data were used for scheduling problems and intelligent transportation systems in many studies [32–36]. In their study, Shi and Abdel-Aty suggest constant monitoring of traffic operations and safety by using big data [32]. Antoniou et al. attempted to integrate a set of sensors and historical data using a data hub to generate signal traffic plans [33]. Fusco et al. utilized big data in public transportation for short prediction models [34]. Günther et al., developed driving cycle for busses using big data approach [35]. In 2015, Vij and Shankari discussed whether big data is big enough [36].

Schreck et al. developed a framework on visual interactive clustering analysis of vehicle trajectory data [37]. Andrienko et al. proposed an approach for extracting meaningful clusters from large databases by combining clustering and classification, driven by a human analyst through an interactive visual interface [16]. In a review study, Andrienko et al. presented the current state of practice in the field of visual analytics for movement and transportation systems [38]. Riveiro et al. developed a framework for the detection of anomaly events (e.g., near-accidents events) based on multidimensional road datasets [17]. Markovic et al. demonstrated the current application of trajectory data from the perspective of transportation agencies [18].

As shown, many research efforts were in development, with various modifications to clustering methods, in order to automatically develop TOD breakpoints. Mostly, small-sized networks and small datasets were considered. Even when relatively larger datasets were considered, due to the analytical reasoning approach, many prevailing traffic conditions

were overlooked. For solutions created and validated, based on such limited datasets, it is questionable how they will accommodate yearly flow fluctuations. This study fills a gap in the utilization of relatively large datasets (for the domain of arterial operations), from most of the nowadays urban arterials that can be used in order to provide TOD schedules that are considering temporal and spatial variations of traffic flow.

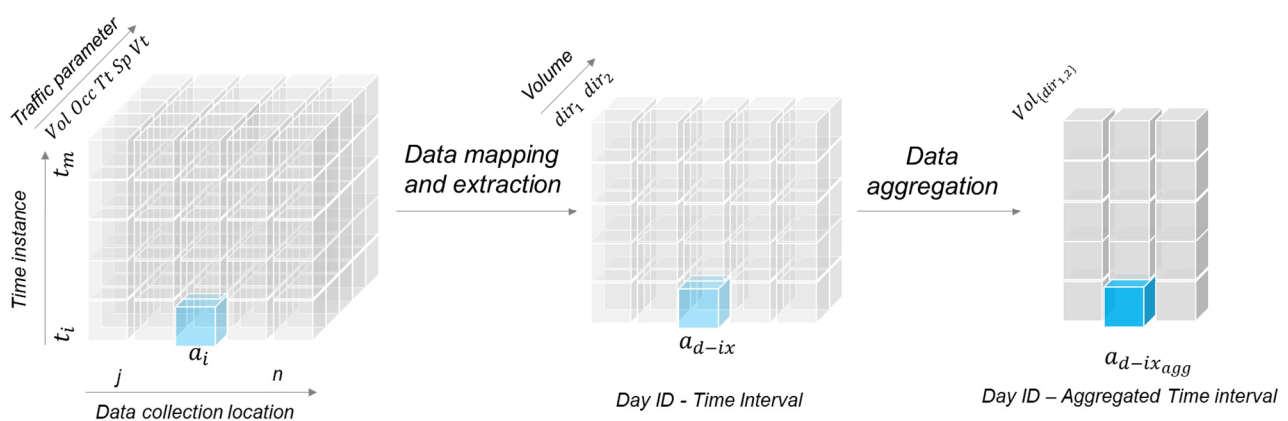
### 3. Methodology

Due to recent improvements in sensor technology, traffic data are nowadays collected with high temporal and spatial resolution and often classified in one of three groups, such as spatial event data, trajectories of moving objects, and spatial time series [16]. This study deals with the spatial time series, which are chronologically ordered sequences of aggregated values of time-variant (e.g., 1-min, 5-min intervals) thematic attributes (e.g., volume, speed, occupancy) associated with fixed spatial locations (e.g., mid-block, intersection stop-bar detectors). The key steps of the proposed methodology are as follows:

1. Data preprocessing
2. Revealing the traffic profiles:
  - 2.1. Determination of the appropriate number of clusters/traffic profiles
  - 2.2. Data clustering and visualization of temporal and spatial data components
3. Aggregating the results of clustering and visualizing them on a weekly level
4. Construction of the TOD breakpoints

#### 3.1. Data Preprocessing

Urban traffic detectors continuously collect various traffic parameters. In particular, volume (Vol), occupancy (Occ), travel time (Tt), speed (Sp), and vehicle type (Vt), as illustrated in traffic data tensor shown in Figure 1. For the purpose of signal timing plans scheduling, volume is the primary data type that is used. Therefore, volume data type is extracted, and data are mapped in the form of matrix  $A \in R^{m \times n}$ , with  $m$  unique time instances (records) and  $n$  facilities in the network.  $a_{i,j}$  refers to the observed traffic parameter at facility  $j$  and time instance  $i$ . Hence, the  $j$ th column of the matrix  $A$  contains data collected at a single location  $\{j\}_{j=1}^n$  in the network, while  $i$ th row of matrix  $A$  refers to the traffic data collected during the period  $t_i, t_i + T$ , where  $T$  is the sampling interval (e.g., 1 min).



**Figure 1.** Data preprocessing.

For each row  $i$  we assign a day-in-year, (e.g.,  $diy = \{1, 2, \dots, 365\}$ ) and an indexed time instance  $ix$  (e.g.,  $ix = \{1, 2, \dots, 1440\}$  for  $T = 1$  min), which refers to the instance-in-day index. For instance,  $diy = 50$  and  $ix = 60$  refers to the time instance at 1 a.m. on 19 February. Intuitively, variables  $d$  and  $ix$  are inferred from the time instant of  $i$  (i.e.,  $diy, ix = f(i)$  and  $\max(diy) \times \max(ix) = m$ ).

To examine spatial and temporal flow characteristic, we aggregate the collected data using common aggregation interval ( $T_{agg} = 15$  min). This would lead to the vertical shrinking in data and to the new data matrix  $A^{agg} \in \mathbf{R}^{m_{agg} \times n}$ , where  $m_{agg} = \frac{m}{T_{agg} \times T}$  and  $ix_{agg} = \{1, 2, \dots, 96\}$  for  $T_{agg} = 15$  min.

We aggregate the collected volume data to  $A^{agg} \in \mathbf{R}^{m_{agg} \times n}$  by examining the traffic information collected over predefined  $T_{agg}$  (such that  $m_{agg} = m(T_{agg})^{-1}$ ) for consecutive time intervals  $ix_{agg}$  for the same day  $d_{ij}$  and the same location  $j$ . We write this as:

$$a_{d-ix_{agg},j}^{agg} = \sum_{ix}^{ix_{agg}} a_{d-ix,j}, \text{ where } ix_{agg} = \frac{ix}{T_{agg}} \quad (1)$$

### 3.2. Revealing Dominant Traffic Profiles

Each row vector  $\alpha_{*,j} = [a_{*,1}, a_{*,2}, \dots, a_{*,n}]$ , where  $*$  =  $d - ix_{agg}$ , of the aggregated matrix  $A^{agg}$  contains the aggregated traffic variable for the entire network (or rather corridor) and for an aggregated period ( $m_a \times T$ ). We refer to this row vector  $\alpha_{*,j}^{agg} \in \mathbf{R}^{1 \times n}$  as a (corridor) traffic profile at time instance  $j$ . In this paper, we aim to split  $\alpha_{*,j}^{agg}$  into  $k$  mutually exclusive groups (or clusters) in order to identify the dominant traffic profiles in the network and their appearance throughout the year. To do so we apply  $k$ -means clustering method [39], method that previously was used for similar transportation-related applications [40,41]. This unsupervised clustering method aims to partition a given set of  $m_{agg}$  observations (i.e.,  $\alpha_{*,1}^{agg}, \alpha_{*,2}^{agg}, \dots, \alpha_{*,m_{agg}}^{agg}$ ) into  $k$  clusters by minimizing the within-cluster sum of squares (WCSS):

$$WCSS^k = \underset{S}{\operatorname{argmin}} \sum_{i=1}^k \sum_{\alpha_{*,j}^{agg} \in S_i} \|\alpha_{*,j}^{agg} - \mu_i\|^2 \quad (2)$$

where  $\mu_i$  is the mean point of  $S_i$  or centroid of cluster  $i$  ( $i = 1, 2, \dots, k$ ).

For the  $k$ -means clustering, the number of clusters  $k$  need to be specified beforehand. The number of clusters depends on the specific domain application, and it can usually be inferred with the help of the “elbow” curve  $WCSS^k = f(k)$ , where  $f(k)$  is defined as in Equation (2) for  $k = 1, 2, \dots, n$ .

Another practical method used for the determination of optimal clusters number is the silhouette method [42]. Each cluster is represented by a so-called silhouette, which is based on a comparison of its tightness and separation. The concept of silhouette width involves the difference between the within-cluster tightness and separation from the rest. The silhouette width  $s(i)$  is defined for entity  $i \in I$  as:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

where  $a(i)$  is the average distance between  $i$  and all other entities of the cluster to which  $i$  belongs and  $b(i)$  is the minimum of the average distance between  $i$  and all entities in each other cluster. The silhouette coefficient ranges from  $-1$  to  $1$ , where a cluster number which is close to  $1$  represents the optimal number of clusters for a given dataset.

The last method that we explored in this paper relies on the practice that the development of the new signal timing plan is justified if the difference in design volumes (between a pair of the most similar traffic profiles) is at least 10%. For each  $k = 2, \dots, n$  we compute the difference in design volumes  $d_{i,j}$  between each pair of clusters  $i$  and  $j$  as:

$$d_{i,j}^k = \frac{\|\mu_i - \mu_j\|^2}{\max(\|\mu_i\|, \|\mu_j\|)}, \text{ where } i \neq j \quad (4)$$

Then, we map the minimum of all  $d_{i,j}^k$  to the investigated  $k$  (i.e.,  $f(k) = \min(d_{i,j}^k)$ ,  $\forall i, j$ , where  $i \neq j$ ). To this end, we select the appropriate number of clusters and assign each data point  $\alpha_{d-ix_{agg,j}}$  to the corresponding cluster  $i$ , where  $1 \leq i \leq k$  such that  $\|S_i - \alpha_{d-ix_{agg,j}}\| \leq \|S_j - \alpha_{d-ix_{agg,j}}\|$  for  $\forall j$  such that  $i \neq j$ .

### 3.3. Identification of the Dominant Traffic Profile for Each Time Instance within a Week

The traffic often exhibits repetitive patterns over the course of a week, month, or even a year. Considering that (i) a single set of signal timing plans is often deployed during the entire year; and (ii) the fact the traffic in south Florida is quite heavy throughout the year with no significant seasonal/monthly fluctuations, we decided to investigate the dominant traffic profiles on a weekly basis. A high percentage of the commuters (in the investigated area) supports our assumption that the weekly patterns are more dominant and relevant than monthly and yearly patterns.

So, we first divide a week into  $\frac{7 \times 24 \times 3600}{T[s]}$  mutually exclusive time intervals within a week, where  $W_m = \frac{24 \times 3600}{T[s]}$  intervals are assigned to each day  $\{(dw_i)\}_i = 1^7$  of a week. Then, for each day  $dw_i$  each of the  $W_m$  intervals within a day we make a frequency distribution of the assigned clusters, based on their occurrence on day  $dw_i$  and the investigated time period. This distribution of frequencies helps us identify the dominant cluster (or traffic profiles) for each (aggregated) time instance during a week.

### 3.4. Construction of the Breakpoints

Finally, to set the breakpoints within a week, we observe the revealed dominant profiles for each time interval. We first combine the adjacent time intervals that share the dominant traffic profile. Then, using the trial error method, we aim to minimize the number of instances for which the assigned profile (or cluster) is different from the most frequent profile.

## 4. Experimental Setup

To apply the proposed methodology, we select a real-world corridor in Fort Lauderdale (Florida), Sunrise Boulevard. Special focus is given to a 3.6-mile long section of corridor that encompasses 12 signalized intersections, from the intersection at NW 31st Ave to NE 15th Ave, due to the availability of a longer series of data. In particular, data were available from 2015 to 2021. However, due to the COVID-19 pandemic and due to a longer series of missing data during some years, we selected data from 2017, as those had the lowest rate of missing and erroneous instances. The examined three-lane corridor carries a high percentage of commuters daily and serves multiple users (rail-road crossings, exclusive pedestrian crosswalks, etc.). The corridor is well equipped with ITS devices used for the collection of various traffic data (e.g., volumes, travel times, speeds). Figure 2 shows the locations of five microwave vehicle detection stations (MVDS) used for data collection.

Fully actuated signal control operates in the field with the single set of TOD signal timing plans active during the entire year. Due to budgeting limitations, such practice is common in many transportation agencies [4]. To be more specific, STPs are scheduled to operate in the same fashion for three distinctive periods throughout the year: (i) Monday to Friday, (ii) Saturday, and (iii) Sunday, as illustrated in Figure 3. In the first period, four signal timing patterns (AM peak, Midday peak, PM peak, and “Free” pattern) are used to combat flow fluctuations, whereas during the weekend, two signal timing plans (Midday peak and “Free”) are used. “Free” pattern means that a change of signal status depends only on detector actuations (i.e., cycle length is defined during this period).

To investigate prevailing traffic conditions and propose modifications to current TOD schedules proposed, our methodology was applied for a one-year volume dataset, collected throughout 2017. Without a loss of generality, the proposed method can be applied to any representative yearly dataset.

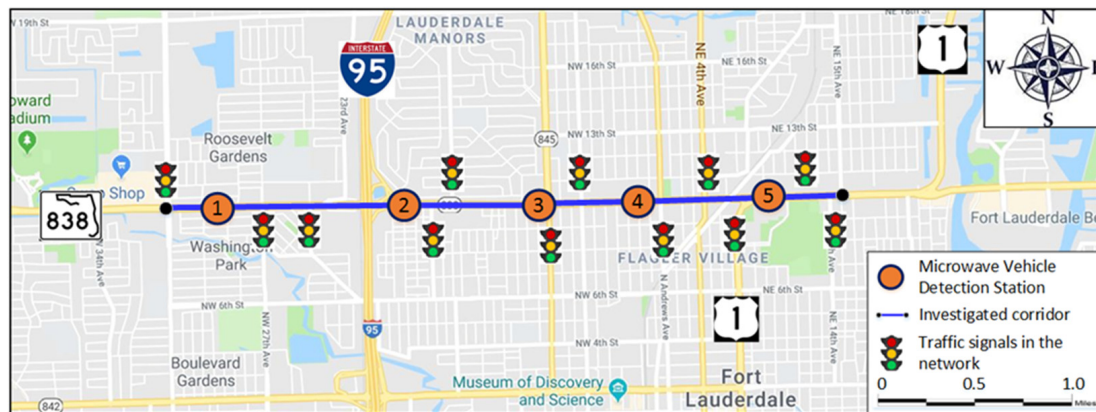


Figure 2. Study area.

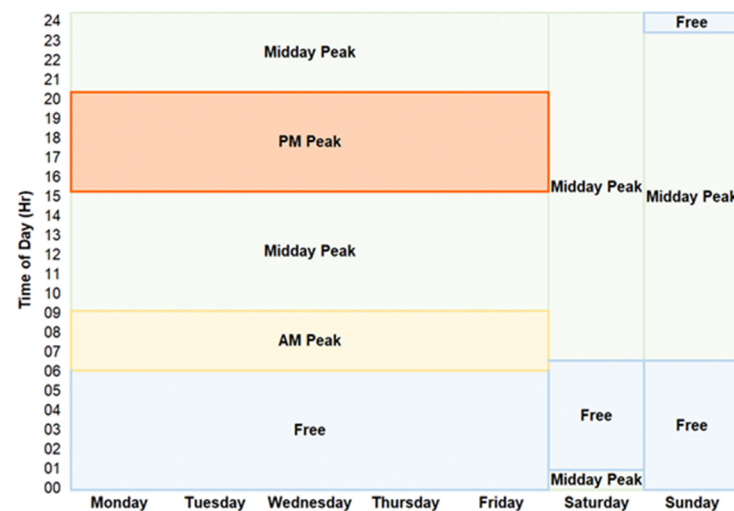


Figure 3. Currently deployed TOD signal timing plans.

## 5. Results and Discussion

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

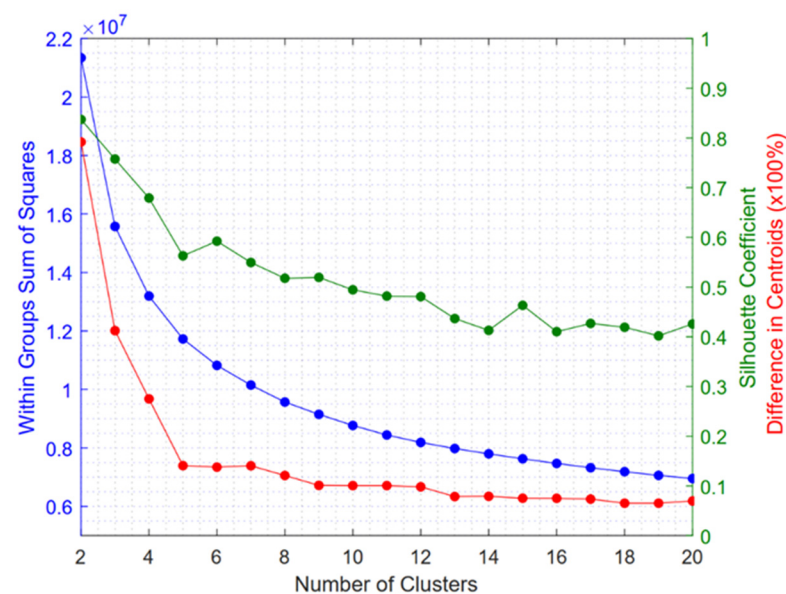
### 5.1. Revealing Underlying Patterns

#### 5.1.1. Identifying Appropriate Number of Clusters

The results of clustering analysis mainly depend on dataset characteristics and predefined parameters of the  $k$ -means algorithm, such as number of iterations, replications, and optimal number of clusters. Since the outcome of analysis mainly depends on a defined number of clusters, several methods were conducted to determine the optimal number of clusters. For a range of clusters  $k = 2, 3, \dots, 20$ , WCSS, SC and differences in centroid's values were computed and presented in Figure 4.

Based on the elbow curve method, due to the characteristics of the given dataset, no clear "elbow" can be spotted (as shown with the blue line chart in Figure 4). In such instances, a domain knowledge expert might suggest an appropriate value of  $k$ , based on the nature of the examined problem. In this particular case, we suggest a  $k$  value of 5 as the "elbow point". When values for the silhouette coefficient were plotted against the corresponding number of clusters (as illustrated with green line in Figure 4) based on the applied method, the recommended optimal number of clusters is two (the closest value of SC to 1). This result has a logical interpretation, since the dataset can be roughly

divided in two groups: day and night traffic. However, from the perspective of traffic signal operations throughout the day, this result has no practical application. Additionally, when differences between centroids were examined, by increasing the number of clusters, the difference significantly drops. By applying domain principles documented in previous studies, it is justified to develop signal timing plans when the difference in design volumes is at least 10% [43]. Based on such recommendations, for the observed dataset, it was found that the appropriate number of clusters is nine (as illustrated in Figure 4 with a red line). Since each of the methods for the determination of the appropriate number of clusters provided different recommendations, we conducted clustering for a range of values (i.e.,  $k = 5, \dots, 15$ ). Such an approach will allow us to examine the dataset in greater detail.

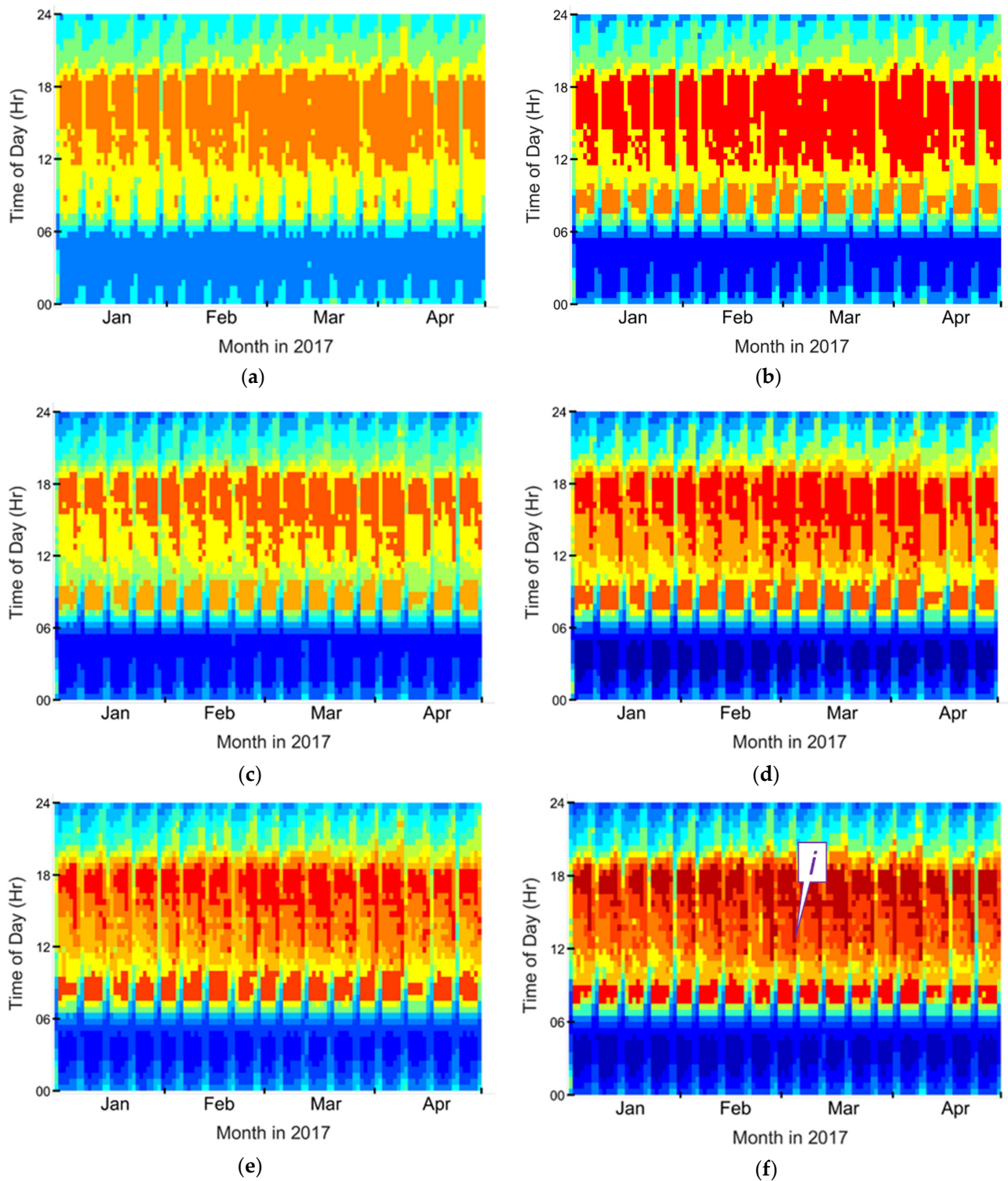


**Figure 4.** Determination of appropriate number of clusters.

#### 5.1.2. Clustering and Visualizing the Data

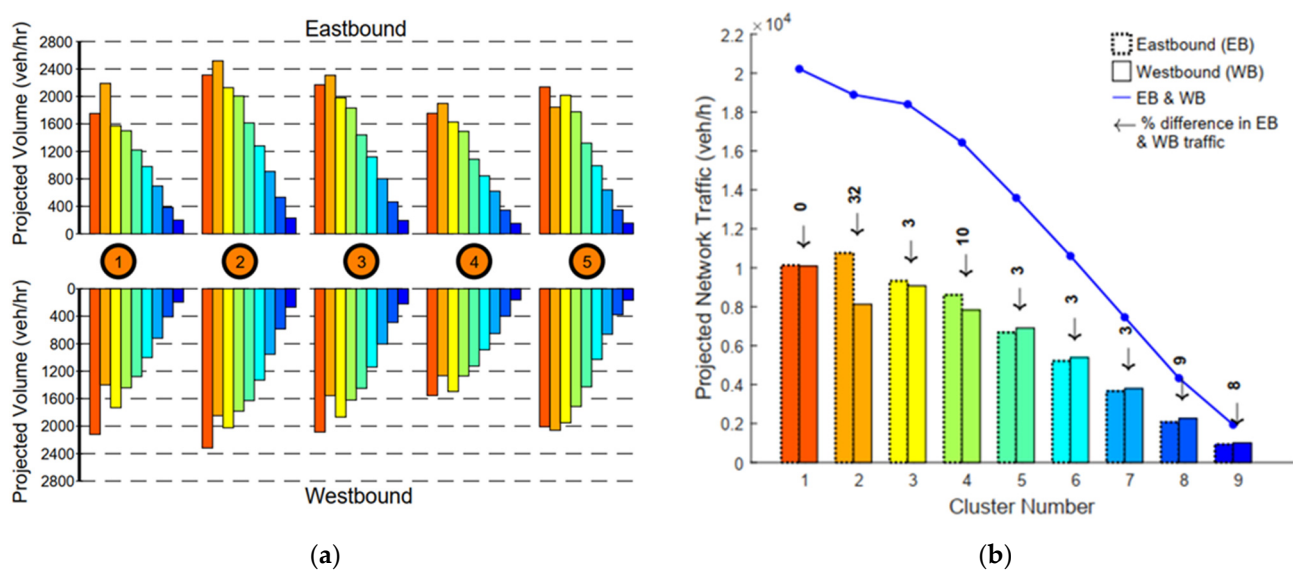
As highlighted in [34], the key to understanding the results obtained from cluster analysis is visualization. However, adequate visualization of feature vectors (i.e., points in multi-dimensional space of properties) is challenging [16]. Moreover, for representing both spatial and temporal aspects, there are no convenient visualization methods. Initially, we examined the temporal component of flow fluctuations. To visualize these features, we present an assigned cluster for each time instance in an examined period, in an XY grid, where, on the X and Y axis, days in a year, and time instances throughout the days are presented, respectively. In Figure 5, visualization of the first four months of 2017 is presented. Essentially, there are two important findings resulting from clustering analysis. First, with a low number of clusters, i.e.,  $k = 5$  (which can be selected arbitrarily based on WCSS value in Figure 3) some traffic specific patterns are overlooked (e.g., AM peak hour), and with a high number of clusters (i.e.,  $k = 15$ ) one can clearly identify some overlooked patterns (e.g., an earlier start of the PM peak period during Fridays (see callout *i* in Figure 5f)), increased traffic in late afternoon hours during Fridays and weekends, patterns of AM peak periods during weekdays. It needs to be noted that, due to the nature of clustering algorithms and the examined problem, a higher number of clusters can lead to the generation of isolated clusters, which may warrant frequent transitions of signal timing plans. Overall, such transitions can negatively impact the performance of signal control [44]. Based on the recommended appropriate number of clusters from the domain-knowledge method and a visual analysis of the performed clustering, we restrict our attention to further investigate the spatial component of the examined dataset for  $k = 9$ . Moreover, visual analysis of the temporal dataset component showed that a selected number of clusters allows one to distinguish specific events from the other (e.g., Friday PM

peak periods versus regular weekday Friday PM peak), and not to overlook regular events (e.g., AM peak hour).



**Figure 5.** Temporal characteristics of traffic profiles for a different number of clusters (for January (Jan), February (Feb), March (Mar) and April (Apr) months of 2017); (a) number of clusters-5; (b) number of clusters-7; (c) number of clusters-9; (d) number of clusters-11; (e) number of clusters-13; (f) number of clusters-15.

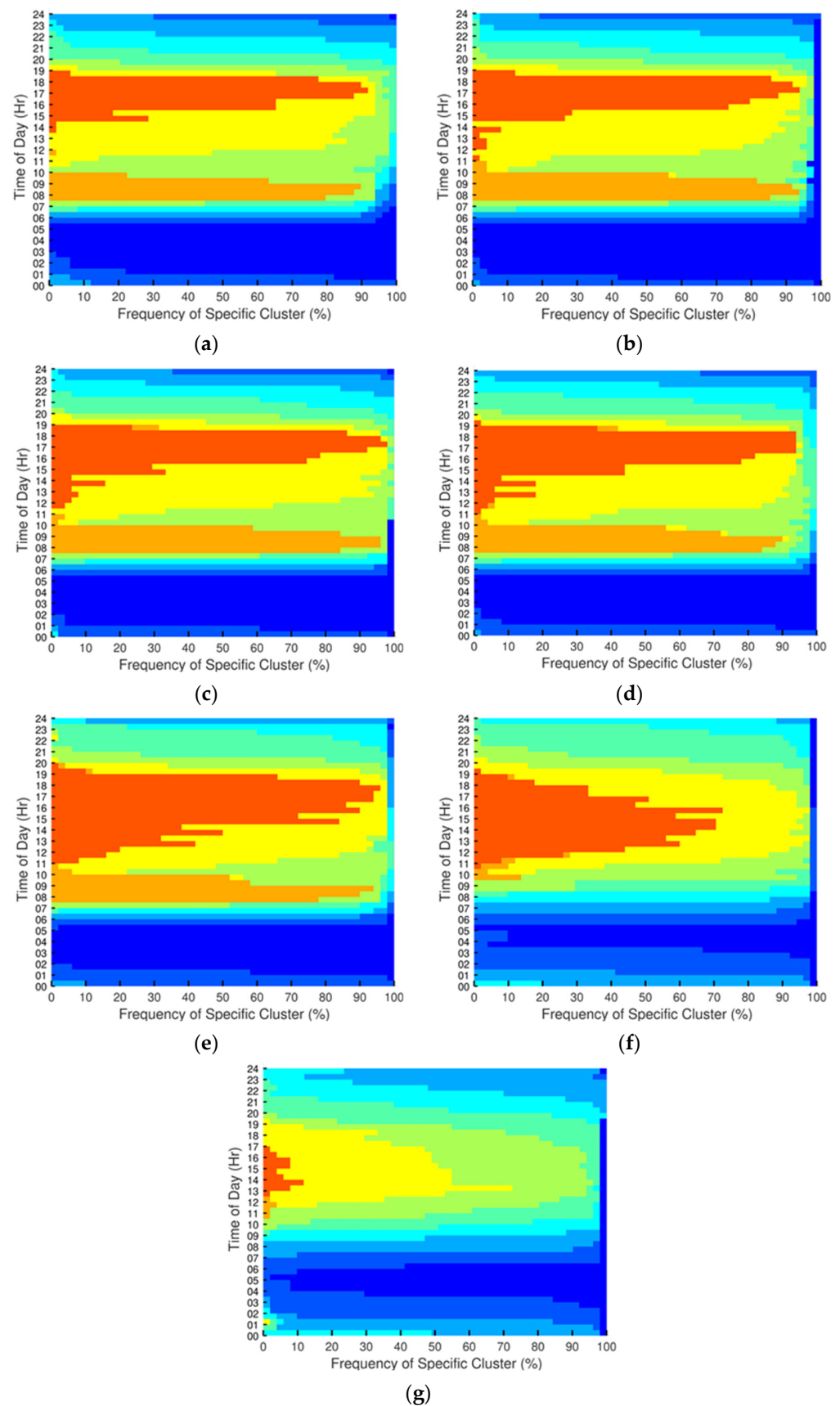
Although temporal analysis allows us to identify distinctive periods during the day, spatial analysis can show which directions of corridor carry more traffic within identified temporal clusters. The cluster center represents the projected volume value (determined, not actual) evenly distributed among all volume values within the same cluster. Clusters are defined in such a way where the preceding cluster, on average, carries higher volume values compared to the following cluster. However, since the feature vector contain 10 instances (data from each mid-block collection point per direction), it can be expected that cluster centers do not follow such trends at every data collection point. Figure 6a shows centroid volume values for five stations (annotated with numbers 1–5). It can be inferred that the orange cluster (cluster number 2), has a higher volume value in the eastbound direction than the dark orange cluster and vice versa. Therefore, it is hard to observe spatial flow fluctuations on a station-by-station basis. To provide a meaningful presentation of spatial flow characteristics, we developed Figure 6b, where we aggregated volume differences between each cluster center for all five data collection stations. In Figure 6b, it is evident that there is a significant difference in the eastbound traffic for the orange cluster (i.e., cluster 2). When examined, with the help of Figure 5c, it is evident that this cluster corresponds to the AM peak traffic, where eastbound traffic is more dominant. A similar trend can be seen in the clusters (from 5 to 9) that are mainly present within the night hours, where traffic is more dominant in the westbound direction (i.e., PM off-peak hour). Interestingly, for clusters that correspond to typical PM peak periods, it is found that volumes are evenly distributed per direction (annotated with a 0% difference in Figure 6b, see cluster 1, dark orange). Finally, clusters 3 and 4 correspond to transitional traffic (i.e., Midday period), which occurs between AM and PM peak periods, where we can see the high impact of the dominant EB direction.



**Figure 6.** Spatial characteristics of projected volume profiles; (a) projected volume on each data collection site; (b) directional projected volume differences.

### 5.2. Dominant Traffic Profiles within a Week

From the appropriate visualization of clustered data, we identified an appropriate number of clusters and were able to see how, spatially, traffic fluctuates. Considering that currently signal timing plan schedules (as illustrated in Figure 2) are developed for a week and scheduled throughout the year, we directed our analysis on a weekly level. We based our analysis on the frequency of each cluster occurring during a particular time (i.e., from 00 to 24 h) of a particular weekday (i.e., Monday–Sunday) aggregated using one year's worth of data. We illustrate such weekday cluster frequency distribution in Figure 7a–g.



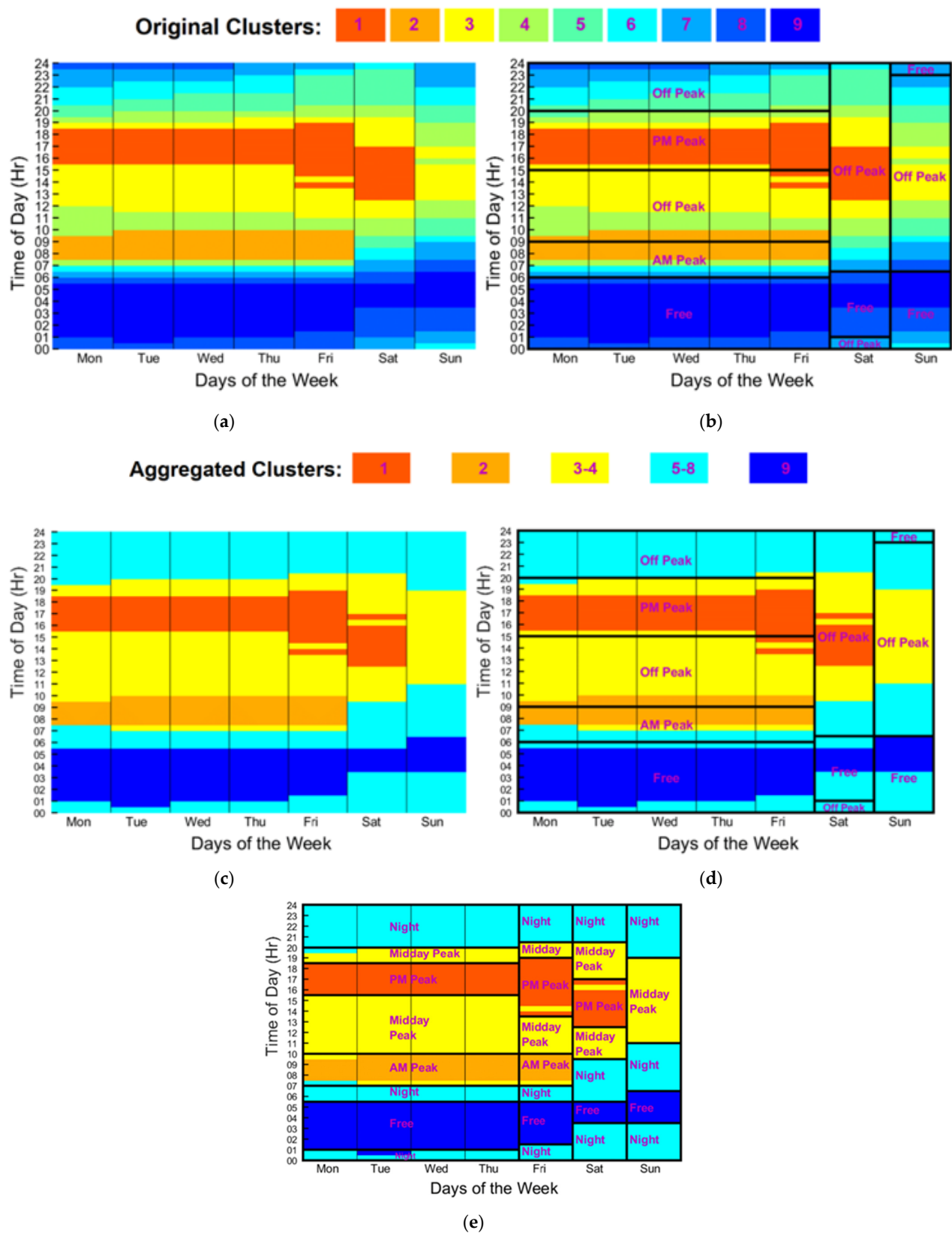
**Figure 7.** Spatial characteristics of projected volume profiles; (a) Monday; (b) Tuesday; (c) Wednesday; (d) Thursday; (e) Friday; (f) Saturday; (g) Sunday.

In general, it can be noted that starting from Monday (as shown in Figure 7a) until Thursday (as shown in Figure 7d) distinctive patterns of morning and afternoon peak periods (AM and PM) are present. The low midday traffic activity (as seen through the dark orange cluster) is noted on typical weekdays (Monday–Thursday). Due to generally increased transportation activities on the network during Fridays, it is evident that this day exhibits different patterns compared to the rest of the weekdays. On contrary to weekdays, weekend days are characterized by increased activities in the afternoon hours. When weekend days are compared, it can be noted that higher traffic activities occurred during Saturdays in the afternoon hours compared to Sundays (as illustrated in Figure 7f,g). Based on the similarities that particular days exhibit, we were able to note four distinctive week periods: (i) typical weekdays (Monday–Thursday), (ii) Friday, (iii) Saturday, and (iv) Sunday.

### 5.3. Development of TOD Breakpoints

Based on the cluster's frequency distributions for each day of the week, we were able to identify the most dominant clusters patterns throughout the week (as shown in Figure 8a). These patterns consist of the initially defined nine clusters (annotated as original clusters). We overlapped the current TOD schedule with the most dominant clusters to document the alignment with the current TOD schedule, illustrated in Figure 8b. Interestingly, it can be noted that in most of the cases, AM peak hour patterns are not appropriately scheduled, as seen through the increased traffic activity even after the defined period (i.e., 9 AM). The current pattern of Off-Peak traffic is used for a wide range of traffic profiles during midday and even night traffic. The same pattern is used for Saturday and Sunday traffic during most of the day. PM peak period is also scheduled in such a manner to cover a wide range of traffic profiles (i.e., clusters). Surprisingly, Saturday traffic, during the PM peak period, exhibits weekday patterns. Dominant clusters throughout the week are aggregated in five distinctive clusters that warrant the development of distinctive signal timing plans. We aggregate these clusters based on several characteristics: (i) significant differences in directional flow values (30%) (see Figure 6b, Cluster 2), for which distinctive signal timing plans should be developed and scheduled; (ii) clusters do not exhibit differences in any direction and volume values are high (see Figure 6b, Cluster 1), for which distinctive signal timing plans should be developed and scheduled; (iii) clusters that exhibit higher traffic volumes in one (see Figure 6b, Clusters 3 and 4) or other direction (see Figure 6b, Clusters 5–8), for which a set of distinctive timing plans should be developed; (iv) clusters that carry a low amount of traffic usually represent flows captured during late night times; these should be operated with signals running in “Free” mode. Based on such recommendations, aggregated profiles are developed and presented in Figure 8c. To show the alignment of proposed aggregated clusters with the current TOD schedule, Figure 8d is developed. Finally, to set the breakpoints within a week, we observe the revealed dominant profiles for each time interval. We first combine the adjacent time intervals that share the dominant traffic profile. Then, using the trial and error method, we aim to minimize the number of instances for which the assigned profile (or cluster) is different from the most frequent profile. Compared to the current schedule, we propose the development of an additional timing plan that will be used exclusively for night traffic, where higher activities are noted in the westbound direction of the examined corridor. Revision of current peak periods is necessary, with special attention paid to Fridays and Saturday traffic. The proposed schedule can be seen in Figure 8e.

It is noteworthy that the proposed method for the identification of the optimum number of clusters for the examined series of data, with five reported as the optimum number of clusters, or where the highest difference in centroids drop occurs (see Figure 4). Thus, the proposed method can be used for similar volume-based clustering problems in the transportation domain.



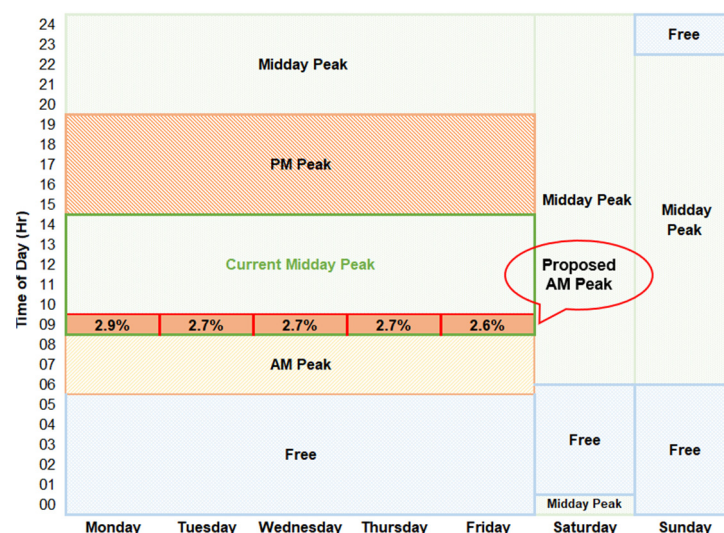
**Figure 8.** Spatial characteristics of projected volume profiles; (a) dominant cluster throughout the week; (b) aligning with current TOD schedule; (c) aggregated clusters throughout the week; (d) aligning with current TOD schedule; (e) proposed TOD schedule. Note: Monday (Mon), Tuesday (Tue), Wednesday (Wed), Thursday (Thu), Friday (Fri), Saturday (Sat), Sunday (Sun).

#### 5.4. Numerical Evaluation of the Proposed Approach

Evaluation of the proposed approach can be carried out either by applying plans in the field or simulating their operation in microsimulation [45–47]. Considering that both approaches would require a year worth of examination or extensive simulation modeling, in an attempt to evaluate the proposed approach, the authors conduct a numerical evaluation similar to a recent study [29]. To evaluate the proposed approach, we restrict our attention to weekdays (Monday–Friday) and the time period from 9 a.m. to 10 a.m., where the proposed approach suggests the deployment of the AM peak plan instead of the Midday (Off-peak) plan (as shown in Figure 8d,e). The reason behind such evaluation is that the data containing signal timing plans (i.e., cycle lengths and green splits) for other time periods (e.g., PM Peak, Night) were not available.

For evaluation, we estimate a vehicular delay that is a result of the deployment of both plans (AM peak vs. Off-peak) on the critical intersection (Sunrise Blvd.–NW 31st) of the examined corridor. The vehicular delay was selected as the main performance measure for the evaluation of the proposed signal timing schedules since it represents one of the fundamental and well-understood traffic signal performance measures [48]. We selected the evaluation of the critical intersection since, for coordination purposes, such an intersection drives coordination (overall corridor) cycle length and any changes in signal timing parameters (which result from applying other signal timing plans), highly affects the quality of the coordination of the examined corridor. In other words, if improvements are not obtained on the corridor’s critical intersection, resulting benefits from non-critical intersections from the coordination perspective can be irrelevant for mainline corridor travelers.

Figure 9 shows the change in the average delay per vehicle if the AM peak plan was used instead of the Midday Off-peak plan from 9 to 10 a.m. during weekdays. Detailed calculations are given in the Appendix A. As it can be inferred from Figure 9, the average vehicular delay is reduced by almost 3% for every one of the evaluated hours in each weekday.



**Figure 9.** Average vehicular delay reduction after the application of the signal timing plans suggested by the proposed method.

Although the reduction of the average delay per vehicle is indicative enough to show the efficiency of the proposed approach, we also estimated the amount of total delay (expressed in hours) by multiplying the average delays with the total volume (as outlined in Appendix A). For a period of one year, or 260 weekdays (which result from 5 weekdays during each of the 52 weeks), savings in total delay were estimated and they are presented in Table 1. From the results it can be concluded that the proposed approach reduces delay and efficiently supports the scheduling of signal timing plans. It is important to note that,

although signal timing related parameters were not available for other peak periods (PM Peak, Night), it is reasonable to expect that the proposed method would result in a similar magnitude of delay reduction.

**Table 1.** Total vehicular delay savings from 9 a.m. to 10 a.m. for all weekdays in a period of one year.

| Yearly Delay Savings [Hours] |          |            |           |         |
|------------------------------|----------|------------|-----------|---------|
| Mondays                      | Tuesdays | Wednesdays | Thursdays | Fridays |
| 106                          | 113      | 115        | 118       | 119     |

## 6. Conclusions

This study proposes a novel, visually interactive and easy to apply approach for the determination of signal timing plans schedules. The proposed method is a viable approach for transportation agencies to evaluate current and develop appropriate TOD breakpoints. The main motivation for this study lies in the fact that existing methods for scheduling signal timing plans (found in the literature) are developed and tested for relatively small datasets and network sizes and without proper consideration of real-world arterials. Consequently, transportation agencies are not fully utilizing the potential of longer series of data (a year's worth of data) that became available in the past couple of years. Even in cases in which sizable datasets were used in some of previous studies, the authors relied heavily on analytical reasoning of data, which hinders an analyst to reveal specific temporal and spatial traffic flow fluctuations. While performing this study, it is found that applying the clustering methods without consideration of several methods for the identification of an appropriate number of clusters as well as appropriate visualization of clustering analysis can lead to inconsistent and misleading results. One needs to be aware of both temporal and spatial components of data collected for longer time intervals.

Obtained results demonstrate how, compared to current operations, a different schedule should be developed for Fridays and Saturdays, which serves as a good indication that assumptions found in previous studies (where the same schedule operates from Monday through Friday) needs to be carefully examined. From the spatial aspect, it is found that clusters which contain higher directional traffic almost exclusively warrant distinctive timing plans (i.e., AM peak, Night patterns). In other cases, when these differences are negligible, but volume values are high, it is indicative that these periods require a distinctive traffic plan (i.e., PM peak). By performing an evaluation of the proposed schedule, it was found that scheduling AM peak signal timing plans instead of Midday peak plans, on average, close to 3% in average delay reduction can be achieved. Such savings are significant when the total number of hours saved during one year were estimated. Finally, this study contributes to the currently growing field of visual analytics, where the proposed approach for visualization of spatial time series is applicable for similar scale problems. A future step of research is related to the application and validation of the proposed approach in the field conditions.

Lastly, the development and scheduling of signal timing plans in the future might be considered with respect to environmental concerns [49–51], and on the basis of moving sensors data [52–57] that can overcome limitations of currently used fixed sensors [58].

**Author Contributions:** Conceptualization, N.D., N.M. and A.S.; methodology, N.D., N.M. and S.G.; formal analysis, N.D., N.M. and S.G.; writing—original draft preparation, N.D. and S.G.; writing—review and editing, N.D., N.M., S.G. and A.S.; visualization, N.D., N.M. and S.G.; supervision, A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partially supported by the Florida Department of Transportation under the grant, BDV27 TWO 462-02-Analysis of Traffic Demand Patterns and Signal Re-timing Strategies for ITS-data-rich Arterials.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to express their gratitude to Rodolfo Alfaro-Carcoba, Undergraduate Research Assistant from the Department of Electrical Engineering at the University of Nevada for his assistance during data processing efforts.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

The original traffic data were used to calculate average vehicle delay by using Webster's uniform delay equation (Webster, 1958) [59]. We estimated the average delay for eastbound and westbound traffic (main coordination movements), since data for side streets (northbound/southbound) were not available.

$$d = \frac{0.5C(1 - \lambda)^2}{1 - \lambda x} \quad (A1)$$

where:

C—cycle length (s)

$\lambda = \frac{g}{C}$ ; where g—green (s)

$x = \frac{v}{s}$ ; where v—volume (veh/hour); s—saturation flow rate (veh/h)

s = 1800 veh/h

The original field signal timing parameters are given in Table A1.

Traffic volumes for weekdays, Monday through Friday, are given in Table A2.

**Table A1.** Signal timing parameters.

|          | Cycle Length [s] | Average Green for Eastbound and Westbound Approach [s] |
|----------|------------------|--------------------------------------------------------|
| Off-peak | 180              | 82.5                                                   |
| AM peak  | 180              | 85.5                                                   |

Average vehicular delay was calculated by using Equation (A1) and results are presented in Table A3.

**Table A2.** Traffic volumes for weekday 9 a.m. to 10 a.m.

| Average Traffic Volume for Eastbound and Westbound Approach [veh/h/lanes] |         |           |          |        |
|---------------------------------------------------------------------------|---------|-----------|----------|--------|
| Monday                                                                    | Tuesday | Wednesday | Thursday | Friday |
| 1328                                                                      | 1427    | 1444      | 1482     | 1500   |

**Table A3.** Average vehicular delay from 9 a.m. to 10 a.m. for weekdays.

|          | Average Delay [s/veh] |      |      |      |      |
|----------|-----------------------|------|------|------|------|
|          | Mon                   | Tue  | Wed  | Thu  | Fri  |
| Off-peak | 39.3                  | 40.9 | 41.2 | 41.8 | 42.2 |
| AM peak  | 38.2                  | 39.8 | 40.1 | 40.7 | 41.1 |

After calculating the average vehicular delay, the reduction in delay was calculated by using Equation (A2).

$$\text{Delay reduction (\%)} = \frac{d_{\text{off-peak}} - d_{\text{am peak}}}{d_{\text{off-peak}}} \cdot 100\% \quad (A2)$$

Total delay for one hour is calculated as:

$$d_{total} = (d_{off-peak} - d_{am\ peak}) \cdot v \cdot 260 / 3600 \text{ (hours)} \quad (A3)$$

where

260—number of weekdays in a year (52 weeks \* 5 days).

## References

1. Stevanovic, A.; Transportation Research Board. *Adaptive Traffic Control Systems: Domestic and Foreign State of Practice*; Academic Press: Washington, DC, USA, 2010.
2. Dobrota, N.; Stevanovic, A.; Mitrovic, N. Development of assessment tool and overview of adaptive traffic control deployments in the US. *Transp. Res. Rec.* **2020**, *2674*, 464–480. [\[CrossRef\]](#)
3. Stevanovic, A.; Dobrota, N.; Mitrovic, N. *NCHRP 20-07/Task 414: Benefits of Adaptive Traffic Control Deployments—A Review of Evaluation Studies*; Technical Report, Final Report; Transportation Research Board of the National Academies: Washington, DC, USA, 2019.
4. Gordon, R.; Braud, C. *Traffic Signal Operations and Maintenance Staffing Guidelines*; Federal Highway Administration: Washington, DC, USA, 2009.
5. Shayeb, S.A.; Dobrota, N.; Stevanovic, A.; Mitrovic, N. Assessment of arterial signal timings based on various operational policies and optimization tools. *Transp. Res. Rec.* **2021**, *2675*, 195–210. [\[CrossRef\]](#)
6. Gavric, S.; Sarazhinsky, D.; Stevanovic, A.; Dobrota, N. Evaluation of Pedestrian Timing Treatments for Coordinated Signalized Intersections. In Proceedings of the Transportation Research Board 101st Annual Meeting, Washington, DC, USA, 9–13 January 2022.
7. Haq, M.T.; Molan, A.M.; Ksaibati, K. Evaluating the Operational Efficiency of Two Versions of Super Diverging Diamond Interchange Design: A Case Study in Denver, Colorado. *Transp. Res. Rec.* **2021**, *2676*, 747–762. [\[CrossRef\]](#)
8. Park, B.; Santra, P.; Yun, I.; Lee, D.-H. Optimization of Time-of-Day Breakpoints for Better Traffic Signal Control. *Transp. Res. Rec.* **2004**, *1867*, 217–223. [\[CrossRef\]](#)
9. Park, B.; Lee, J. A Procedure for Determining Time-of-Day Break Points for Coordinated Actuated Traffic Signal Systems. *KSCE J. Civ. Eng.* **2008**, *12*, 37–44. [\[CrossRef\]](#)
10. Ratrouf, N.T. Subtractive Clustering-Based K-Means Technique for Determining Optimum Time-of-Day Breakpoints. *J. Comput. Civ. Eng.* **2011**, *25*, 380–387. [\[CrossRef\]](#)
11. Jun, Y.; Yang, Y. Using Kohonen Cluster to Identify Time-of-Day Break Points of Intersection. *Lect. Notes Electr. Eng.* **2013**, *236*, 889–896. [\[CrossRef\]](#)
12. Hao, W.; Dong, C. Traffic Signal Segmentation Algorithm Based on Twodimensional Clustering of Traffic Volume and Vector Angles. In Proceedings of the 2017 International Smart Cities Conference (ISC2), Wuxi, China, 14–17 September 2017; pp. 1–5.
13. Wan, L.; Yu, C.; Wang, L.; Ma, W. Identification of Time-of-Day Breakpoints Based on Trajectory Data of Probe Vehicles. *Transp. Res. Rec.* **2019**, *2673*, 538–547. [\[CrossRef\]](#)
14. Keim, D.A.; Mansmann, F.; Schneidewind, J.; Thomas, J.; Ziegler, H. Visual Analytics: Scope and Challenges. In *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics*; Simoff, S.J., Böhlen, M.H., Mazeika, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 76–90. ISBN 978-3-540-71080-6.
15. Thomas, J.J.; Cook, K.A. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*; IEEE: New York, NY, USA, 2005.
16. Andrienko, G.; Andrienko, N.; Rinzivillo, S.; Nanni, M.; Pedreschi, D.; Giannotti, F. Interactive Visual Clustering of Large Collections of Trajectories. In Proceedings of the 2009 IEEE Symposium on Visual Analytics Science and Technology, Atlantic, NJ, USA, 12–13 October 2009; pp. 3–10. [\[CrossRef\]](#)
17. Riveiro, M.; Lebram, M.; Elmer, M. Anomaly Detection for Road Traffic: A Visual Analytics Framework. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 2260–2270. [\[CrossRef\]](#)
18. Marković, N.; Sekuła, P.; Vander Laan, Z.; Andrienko, G.; Andrienko, N. Applications of Trajectory Data From the Perspective of a Road Transportation Agency: Literature Review and Maryland Case Study. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 1858–1869. [\[CrossRef\]](#)
19. Dobrota, N.; Stevanovic, A.; Mitrovic, N. Modifying Signal Retiming Procedures and Policies by Utilizing High-Fidelity Modeling with Medium-Resolution Traffic Data. *Transp. Res. Rec.* **2022**, *2676*, 660–684. [\[CrossRef\]](#)
20. Smith, B.L.; Scherer, W.T.; Hauser, T.A. Data-Mining Tools for the Support of Signal-Timing Plan Development. *Transp. Res. Rec.* **2001**, *1768*, 141–147. [\[CrossRef\]](#)
21. Smith, B.L.; Scherer, W.T.; Hauser, T.A.; Park, B.B. Data-Driven Methodology for Signal Timing Plan Development: A Computational Approach. *Comput.-Aided Civ. Infrastruct. Eng.* **2002**, *17*, 387–395. [\[CrossRef\]](#)
22. Park, B.; Lee, D.-H.; Yun, I. Enhancement of Time of Day Based Traffic Signal Control. In Proceedings of the 2003 IEEE International Conference on Systems, Man and Cybernetics, Conference Theme—System Security and Assurance (Cat. No.03CH37483), Washington, DC, USA, 8 October 2003; Volume 4, pp. 3619–3624. [\[CrossRef\]](#)

23. Wang, X.; Cottrell, W.; Mu, S. Using K-Means Clustering to Identify Time-of-Day Break Points for Traffic Signal Timing Plans. In Proceedings of the 2005 IEEE Intelligent Transportation Systems, Vienna, Austria, 16 September 2005; pp. 586–591.
24. Wong, Y.K.; Woon, W.L. An Iterative Approach to Enhanced Traffic Signal Optimization. *Expert Syst. Appl. Int. J.* **2008**, *34*, 2885–2890. [\[CrossRef\]](#)
25. Dong, C.; Su, Y.; Liu, X. Research on TOD Based on Isomap and K-Means Clustering Algorithm. In Proceedings of the 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery, Tianjin, China, 14–16 August 2009; Volume 1, pp. 515–519. [\[CrossRef\]](#)
26. Guo, R.; Zhang, Y. Identifying Time-of-Day Breakpoints Based on Nonintrusive Data Collection Platforms. *J. Intell. Transp. Syst.* **2014**, *18*, 164–174. [\[CrossRef\]](#)
27. Ma, D.; Li, W.; Song, X.; Wang, Y.; Zhang, W. Time-of-day Breakpoints Optimisation through Recursive Time Series Partitioning. *IET Intell. Transp. Syst.* **2019**, *13*, 683–692. [\[CrossRef\]](#)
28. Chen, P.; Zheng, N.; Sun, W.; Wang, Y. Fine-Tuning Time-of-Day Partitions for Signal Timing Plan Development: Revisiting Clustering Approaches. *Transp. Transp. Sci.* **2019**, *15*, 1195–1213. [\[CrossRef\]](#)
29. Wang, G.; Qin, W.; Wang, Y. Cyclic Weighted K-Means Method with Application to Time-of-Day Interval Partition. *Sustainability* **2021**, *13*, 4796. [\[CrossRef\]](#)
30. Saracoglu, A.; Ozen, H.; Apaydin, M.S.; Maltas, A. A New Approach to Determine Traffic Peak Periods to Utilize in Transportation Planning. *Arab. J. Sci. Eng.* **2021**, *46*, 10409–10418. [\[CrossRef\]](#)
31. Song, R.; Yang, H. Clustering and Understanding Traffic Flow Patterns of Large Scale Urban Roads. In Proceedings of the 2021 International Conference on Control, Automation and Information Sciences (ICCAIS), Xi'an, China, 14–17 October 2021; pp. 246–250. [\[CrossRef\]](#)
32. Shi, Q.; Abdel-Aty, M. Big Data Applications in Real-Time Traffic Operation and Safety Monitoring and Improvement on Urban Expressways. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 380–394. [\[CrossRef\]](#)
33. Antoniou, G.; Batsakis, S.; Davies, J.; Duke, A.; McCluskey, T.L.; Peytchev, E.; Tachmazidis, I.; Vallati, M. Enabling the Use of a Planning Agent for Urban Traffic Management via Enriched and Integrated Urban Data. *Transp. Res. Part C Emerg. Technol.* **2019**, *98*, 284–297. [\[CrossRef\]](#)
34. Fusco, G.; Colombaroni, C.; Isaenko, N. Short-Term Speed Predictions Exploiting Big Data on Large Urban Road Networks. *Transp. Res. Part C Emerg. Technol.* **2016**, *73*, 183–201. [\[CrossRef\]](#)
35. Günther, R.; Wenzel, T.; Wegner, M.; Rettig, R. Big Data Driven Dynamic Driving Cycle Development for Busses in Urban Public Transportation. *Transp. Res. Part Transp. Environ.* **2017**, *51*, 276–289. [\[CrossRef\]](#)
36. Vij, A.; Shankari, K. When Is Big Data Big Enough? Implications of Using GPS-Based Surveys for Travel Demand Analysis. *Transp. Res. Part C Emerg. Technol.* **2015**, *56*, 446–462. [\[CrossRef\]](#)
37. Schreck, T.; Bernard, J.; Tekusova, T.; Kohlhammer, J. Visual Cluster Analysis of Trajectory Data with Interactive Kohonen Maps. In Proceedings of the 2008 IEEE Symposium on Visual Analytics Science and Technology, Columbus, OH, USA, 19–24 October 2008; pp. 3–10. [\[CrossRef\]](#)
38. Andrienko, G.; Andrienko, N.; Chen, W.; Maciejewski, R.; Zhao, Y. Visual Analytics of Mobility and Transportation: State of the Art and Further Research Directions. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 2232–2249. [\[CrossRef\]](#)
39. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A K-Means Clustering Algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1979**, *28*, 100–108. [\[CrossRef\]](#)
40. Ardalan, T.; Liu, D.; Kaisar, E. Truck Tonnage Estimation Using Weigh-In-Motion (WIM) Data in Florida. In Proceedings of the Transportation Research Board 99th Annual Meeting, Washington, DC, USA, 12–16 January 2020.
41. Alshayeb, S.; Stevanovic, A.; Effinger, J.R. Investigating impacts of various operational conditions on fuel consumption and stop penalty at signalized intersections. *Int. J. Trans. Sci. and Technol.* **2021**. [\[CrossRef\]](#)
42. Kodinariya, T.; Makwana, P. Review on Determining of Cluster in K-Means Clustering. *Int. J. Adv. Res. Comput. Sci. Manag. Stud.* **2013**, *1*, 90–95.
43. Stevanovic, A. Assessing Deterioration of Pretimed, Actuated-Coordinated, and SCOOT Control Regimes in Simulation Environment. Ph.D. Thesis, University of Utah, Salt Lake City, UT, USA, January 2006.
44. Shelby, S.G.; Bullock, D.M.; Gettman, D. Transition Methods in Traffic Signal Control. *Transp. Res. Rec.* **2006**, *1978*, 130–140. [\[CrossRef\]](#)
45. Iqbal, M.S.; Ardalan, T.; Hadi, M.; Kaisar, E.I. Developing Guidelines for Implementing Transit Signal Priority (TSP) and Freight Signal Priority (FSP) Using Simulation Modeling and Decision Tree Algorithm. In Proceedings of the Transportation Research Board 100th Annual Meeting, Washington, DC, USA, 5–29 January 2021. [\[CrossRef\]](#)
46. Arafat, M.; Nafis, S.R.; Sadeghvaziri, E.; Tousif, F. A data-driven approach to calibrate microsimulation models based on the degree of saturation at signalized intersections. *Trans. Res. Int. Persp.* **2020**, *8*, 100231. [\[CrossRef\]](#)
47. Haq, M.T.; Farid, A.; Ksaibati, K. Estimating Passing Sight Distances for Overtaking Truck Platoons—Calibration and Validation Using VISSIM. *Int. J. Trans. Sci. Technol.* **2021**, *in press*. [\[CrossRef\]](#)
48. Dobrota, N.; Stevanovic, A.; Mitrovic, N. A Novel Model to Jointly Estimate Delay and Arrival Patterns by Using High-Resolution Signal and Detection Data. *Transp. Transp. Sci.* **2022**, *2022*, 1–33. [\[CrossRef\]](#)
49. Stevanovic, A.; Shayeb, S.A.; Patra, S.S. Fuel consumption intersection control Performance Index. *Transp. Res. Rec.* **2021**, *2675*, 690–702. [\[CrossRef\]](#)

50. Alshayeb, S.; Stevanovic, A.; Dobrota, N. Impact of Various Operating Conditions on Simulated Emissions-Based Stop Penalty at Signalized Intersections. *Sustainability* **2021**, *13*, 10037. [[CrossRef](#)]
51. Haq, M.T.; Molan, A.M.; Ksaibati, K. Surrogate safety assessment of super DDI design: A case study in Denver, Colorado. *J. Trans. Saf. Sec.* **2022**, *14*. [[CrossRef](#)]
52. Azadi, F.; Mitrovic, N.; Stevanovic, A. Impact of Shared Lanes on Performance of the Combined Flexible Lane Assignment and Reservation-based Intersection Control. *Transp. Res. Rec.* **2021**. [[CrossRef](#)]
53. Cvijovic, Z.; Zlatkovic, M.; Stevanovic, A.; Song, Y. Conditional Transit Signal Priority for Connected Transit Vehicles. *Transp. Res. Rec.* **2021**, 2675. [[CrossRef](#)]
54. Cvijovic, Z.; Zlatkovic, M.; Stevanovic, A.; Song, Y. Multi-Level Conditional Transit Signal Priority in Connected Vehicle Environments. *J. R. Tr. Eng.* **2021**, *67*, 1–12. [[CrossRef](#)]
55. Niels, T.; Mitrovic, N.; Dobrota, N.; Bogenberger, K.; Stevanovic, A.; Bertini, R. Simulation-based evaluation of a new integrated intersection control scheme for connected automated vehicles and pedestrians. *Transp. Res. Rec.* **2020**, *2674*, 779–793. [[CrossRef](#)]
56. Arafat, M.; Hadi, M.; Raihan, M.A.; Iqbal, M.S.; Tariq, M.T. Benefits of connected vehicle signalized left-turn assist: Simulation-based study. *Transp. Eng.* **2021**, *4*, 100065. [[CrossRef](#)]
57. Arafat, M.; Hadi, M.; Hunsanon, T.; Amine, K. Stop sign gap assist application in a connected vehicle simulation environment. *Transp. Res. Rec.* **2021**, *2675*, 1127–1135. [[CrossRef](#)]
58. Alshayeb, S.; Stevanovic, A.; Mitrovic, N.; Dimitrijevic, B. Impact of Accurate Detection of Freeway Traffic Conditions on the Dynamic Pricing: A Case Study of I-95 Express Lanes. *Sensors* **2021**, *21*, 5997. [[CrossRef](#)] [[PubMed](#)]
59. Webster, F.V. *Traffic Signal Settings*; Road Research Technique Paper No. 39; Road Research Laboratory: London, UK, 1958. Available online: [https://www.scirp.org/\(S\(lz5mqp453edsnp55rrgjt55.\)\)/reference/referencespapers.aspx?referenceid=809303](https://www.scirp.org/(S(lz5mqp453edsnp55rrgjt55.))/reference/referencespapers.aspx?referenceid=809303) (accessed on 17 May 2022).