

Research Article

Traffic Management as a Service: The Traffic Flow Pattern Classification Problem

Carlos T. Calafate,¹ David Soler,² Juan-Carlos Cano,¹ and Pietro Manzoni¹

¹Department of Computer Engineering (DISCA), Universitat Politècnica de València, 46022 Valencia, Spain

²Institute for Pure and Applied Mathematics (IUMPA), Universitat Politècnica de València, 46022 Valencia, Spain

Correspondence should be addressed to Carlos T. Calafate; calafate@disca.upv.es

Received 30 July 2015; Revised 22 September 2015; Accepted 27 September 2015

Academic Editor: Shengbo Eben Li

Copyright © 2015 Carlos T. Calafate et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Intelligent Transportation System (ITS) technologies can be implemented to reduce both fuel consumption and the associated emission of greenhouse gases. However, such systems require intelligent and effective route planning solutions to reduce travel time and promote stable traveling speeds. To achieve such goal these systems should account for both estimated and real-time traffic congestion states, but obtaining reliable traffic congestion estimations for all the streets/avenues in a city for the different times of the day, for every day in a year, is a complex task. Modeling such a tremendous amount of data can be time-consuming and, additionally, centralized computation of optimal routes based on such time-dependencies has very high data processing requirements. In this paper we approach this problem through a heuristic to considerably reduce the modeling effort while maintaining the benefits of time-dependent traffic congestion modeling. In particular, we propose grouping streets by taking into account real traces describing the daily traffic pattern. The effectiveness of this heuristic is assessed for the city of Valencia, Spain, and the results obtained show that it is possible to reduce the required number of daily traffic flow patterns by a factor of 4210 while maintaining the essence of time-dependent modeling requirements.

1. Introduction

In densely populated urban areas, traffic-related problems, such as air quality, noise, vibration, and accidents, are critical issues for management authorities. In terms of solutions to make traffic flow more efficient or to reduce it, especially in downtowns, authorities develop initiatives to promote the use of public transportation, forbid access to the most polluting vehicles, alternate the days of downtown access according to the vehicles' plate number, charge drivers for access, and so forth. In addition to these initiatives, traffic engineers analyze the traffic flow in our cities taking into account important factors like the adequate street directions to minimize travel times, influence of traffic lights synchronization and placement in traffic congestion, fuel consumption and CO₂ emissions, traffic noise modeling [1–6], and so forth.

Particularly, in the field of fuel consumption and exhaust pollutant, Intelligent Transportation Systems (ITS) have recently emerged as a powerful ally in order to improve traffic flows [7]. Moreover, the massive adoption of smartphones

and the ever increasing efforts to achieve smartphone-vehicle integration [8, 9] pave the way towards novel traffic management solutions where real-time interaction between drivers and traffic management authorities becomes possible. Such interaction provides mutual benefits since traffic authorities are able to have real-time feedback about traffic congestion states at different parts of a city, while drivers are also able to have more information, aiding them in the decision process of finding the optimal route.

In this paper we present a novel platform for centralized traffic management in urban environments which attempts to avoid known problems associated with current route planning solutions based on fixed path costs. The proposed solution takes into account the historical data about traffic patterns in order to provide time-dependent route recommendations to drivers traveling through dense traffic areas. As a first approach to deploy this solution, we propose using existing traffic measurements based on induction loop detections [10] in order to obtain all the required time-dependent traffic flow models. We focus on the specific case of the city

of Valencia, Spain, to gain further insight into the problem. Based on the results obtained, we propose a heuristic to address the problem efficiently by grouping elements with a similar behavior, and we assess the effectiveness of the proposed heuristic in terms of the number of interpolation functions required. We show that it is possible to reduce the required number of interpolations functions describing daily traffic patterns by a factor of 4210, which significantly reduces the problem complexity.

The paper is organized as follows: in the next section we introduce some related works. In Section 3 we present the proposed traffic management platform. Section 4 describes the time-dependent traffic analysis problem and provides an overview of the traffic patterns for the city of Valencia, Spain. Section 5 describes the selected heuristic to the modeling problem, along with the results achieved. Section 6 then presents the overall aggregation gains, detailing the origin of those gains. Finally, in Section 7 we conclude the paper.

2. Related Works

After several decades of research, the existing traffic engineering literature is quite broad and extensive. Recently, some solutions have emerged that rely on mobile devices to monitor the traffic in real time, for example, the Mobile Millennium [11] project. Such information can be used for administrative purposes, for example, to visually analyze the traffic conditions, but, in addition, it can also be useful to optimize the routes taken by vehicles, as shown analytically by Kim et al. [12].

Among these proposals we can find TrafficView [13], which defines a framework to gather and disseminate information about the vehicles on the road. With such a system, drivers will be provided with road traffic information that helps driving in adverse situations such as foggy weather or finding an optimal route in a long trip. Work and Bayen [14] highlight the potential of mobile devices to provide real-time traffic information for the entire transportation network, providing some case studies. Claudel et al. [15] emphasize how mobile devices may allow obtaining more reliable estimations about the time required to traverse specific routes. Leontiadis et al. [16] propose an opportunistic traffic management system where vehicles share traffic information in an ad hoc manner, allowing them to dynamically reroute based on individually collected traffic information. Recently, solutions such as EcoTrec [2] introduced a VANET-based ecofriendly routing algorithm for vehicular traffic which considers road characteristics and traffic conditions to improve the fuel savings of vehicles, thereby reducing gas emissions.

Moreover, when attempting to solve the vehicle route planning problem in the most accurate way, we must take into account the traffic variability throughout the day, as well as other situations that take place in real life when driving a vehicle [17, 18]. For instance, it is quite clear that, on large metropolitan areas, the cost of traversing certain arteries, especially large avenues, heavily depends on the time of day, being critical at peak traffic hours [19]. However, it has been

proved that integrating time-dependencies in route optimization algorithms significantly increases their complexity [20, 21].

To tackle this increase of complexity, we present in this paper an approach to significantly reduce the amount of data that our platform will need to find the time-dependent shortest routes. Specifically, we detail how to aggregate large amounts of historical traffic flow data into the most meaningful set of information to properly describe traffic flow variations throughout the day on the different streets and avenues of a city.

To this aim, we will use a clustering technique. Cluster analysis is an unsupervised learning technique used for the classification of data. Data elements are partitioned into groups called clusters that represent proximate collections of data elements based on a distance or dissimilarity function. There exist two main clustering methods. The hierarchical methods basically start with each member of the set in a cluster of its own and fuse nearest clusters until there are k remaining. The partitioning methods start by building a set of k representative objects and cluster around those, iterating until (locally) optimal clustering is found. See, for example, the classical book by Kaufman and Rousseeuw [22] and Xu and Wunsch II [23].

Clustering techniques have been already used in the last years as part of ITS solutions in order to provide real insights into traffic management policies. For brevity, we only refer to some of these works. We recommend consulting Guardiola et al. [24] for further information on the topic.

For example, Wang et al. [25] present a dynamic traffic prediction model that deals with traffic flow data to convert them into traffic status. In this model, two data mining techniques, the clustering analysis and the classification analysis, are applied to historical traffic flow data. Caceres et al. [26] present a methodology for estimating traffic flows using road features as clustering variables, so that it can be applied to any road section, even without detector data. More recently, Yildirimoglu and Geroliminis [27] partition the historical data set from loop detectors on Californian freeways in clusters with similar characteristics based on the traffic patterns observed on the roadway. The building block of their methodology is the development of stochastic congestion maps, which identify the probability that a space-time domain is congested. Finally, Guardiola et al. [24] present a new methodology for analyzing the daily traffic flow profile using Functional Data Analysis. They claim that their methodology allows a maximum exploitation of the recorded historical data and results in the detection of changes in the flow pattern, which would otherwise be difficult to detect via classical statistical methods.

3. Traffic Management as a Service

Current vehicle navigation systems are typically based on locally stored static information from which routes are calculated. Among such systems we can find commercial applications like TomTom (<http://www.tomtom.com/>) or Garmin (<http://www.garmin.com/>). There are also free tools, like

Google Maps Navigator and OsmAnd (<http://osmand.net/>) that operate in a similar manner. The main drawbacks of navigation systems based on static information are the inability to adapt to traffic congestion states or unexpected events, like accidents or other problems on the road, which cause travel times to be much higher than expected.

More sophisticated route navigation solutions update route information in real time, based on reported traffic conditions. As an example, the TomTom navigation software has been enhanced to support client-server interaction in order to inform clients about alternative routes when atypical traffic delays are detected.

In this paper we will address the specific problem of traffic congestion in urban environments. Instead of accidents and other conditions causing atypical delays, we will focus on predicting daily traffic flow patterns for a specific urban environment, detailing how it is possible to reduce travel times based on historical information about the traffic density distribution throughout the day.

The proposed traffic management platform is named ABATIS: *Automatic Balancing of Traffic through the Integration of Smartphones with vehicles*. The main novelty of ABATIS as a route planning system is providing time-dependent route recommendations based on traffic congestion history. Specifically, it offers client-server interaction, where the route selection process is performed at the route server (see Figure 1) based on real-time information stored in the route database and historical data. The traffic analysis and visualization server allows making traffic congestion forecasts based on historical data while also allowing traffic management authorities to check the traffic conditions in real time.

Clients contribute to improving the route database information by providing real-time feedback about traffic congestion conditions, which allows maintaining both a real-time map of traffic fluidity in a city and accurate historical data of traffic behavior. This approach supports global traffic load balancing and event-based management (e.g., reducing traffic congestion in the route of an ambulance).

This strategy, although offering significantly better routes, has a higher cost since the estimated time for traversing each path segment will no longer be a fixed value based on segment length and speed limit, but instead it will vary dynamically along the day. In order to achieve time-dependent costs for the different streets and avenues in a city, ABATIS will use existing historical data about traffic logs in a city to estimate travel times. Since such logs provide per-hour congestion measurements for all induction loop detectors in a city for a whole year, they must be properly summarized and synthesized by the traffic analysis server to allow seamlessly integrating such information in the route server. Thus, in the remainder of the paper, we will focus on the traffic analysis component, proposing a heuristic able to reduce the complexity of the problem by converting huge amounts of historical data about traffic intensity into a small but representative set of daily patterns able to describe the expectable traffic behavior in the city along the day.

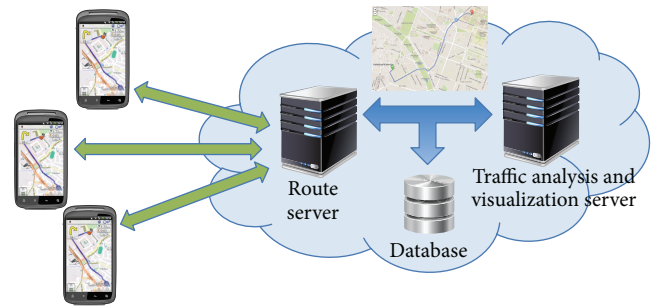


FIGURE 1: ABATIS traffic management architecture.

4. Flow Pattern Classification Problem

Attempting to model the daily traffic flow pattern of hundreds of streets/avenues for every day of the year would lead to hundreds of thousands of interpolation functions able to provide a smooth description of per-street traffic flow variations throughout the day, based on several million input values (assuming a per-hour granularity). Such modeling effort for a single city can be considered excessive and, in addition, causes route recommendation tasks at the server to have an extremely high computational cost. Nevertheless, when attempting to provide an accurate characterization of path segment costs in a specific urban environment, it quickly becomes clear that (i), from a yearly perspective, seasonal differences are expectable as, for example, more people use their vehicles during cold weather seasons than during the warm and hot seasons where, for example, bicycles or public transport can become a more attractive alternative; (ii), from a weekly perspective, labor days are characterized by mobility patterns and traffic congestion states that drastically differ from the behavior during weekends and holidays; (iii), from an hourly perspective, different hours of the day are associated with different congestion levels (e.g., day versus night); and finally (iv), from a spatial perspective, different streets/avenues have different traffic levels at any time of the day, requiring independent modeling.

Taking the aforementioned factors into consideration, in this section we will take an in-depth look into traffic behavior when focusing on a medium-size European city like Valencia, Spain, which is the third largest metropolitan area in Spain with about 1.77 million inhabitants. Detailed trace files containing the amount of traffic flowing in each of the streets/avenues each hour for a full year (2013) were provided to us by Valencia's City Hall Traffic Department, in particular, data concerning the 421 most relevant streets/avenues (those monitored by traffic services through induction loop detectors).

Our goal is to obtain insight into the traffic flow, detecting traffic patterns according to the day of the week, hour, and type of street. Based on the traffic patterns detected, we will propose a heuristic in order to simplify the number of models required while maintaining most of the time-dependent modeling effectiveness. Although we use the city of Valencia as the target of our analysis, the modeling methodology followed is quite general, being applicable to other cities as well.

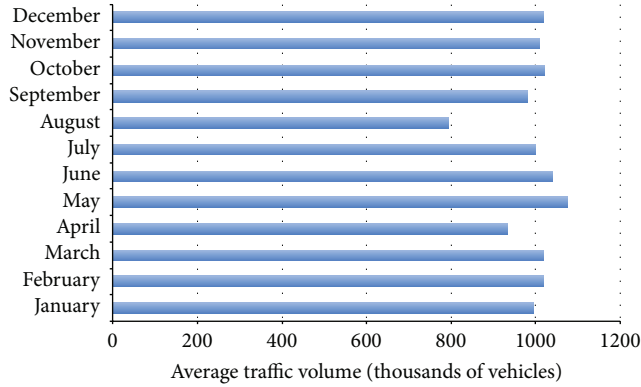


FIGURE 2: Average traffic volume in Valencia per month.

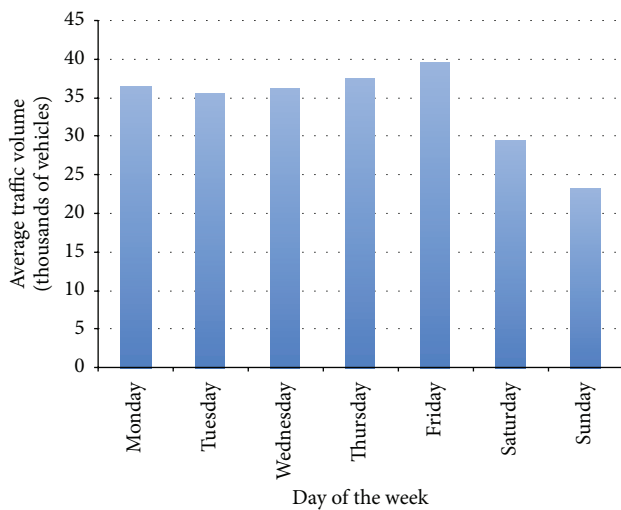


FIGURE 3: Average traffic volume in Valencia for the different days of the week.

We start by analyzing the monthly traffic, assessing whether we can detect significant seasonal differences. As shown in Figure 2, there are minor fluctuations in terms of overall traffic on a monthly basis. It quickly becomes evident that holiday periods, like August and also Easter (in April), have a clear and expectable impact on the overall traffic volume. For the remaining months of the year the values can be considered relatively similar, having a mean value of about 1 million vehicles.

For the analysis that follows we picked a month with an average overall traffic volume close to the mean; specifically, we selected November, which has no holiday periods. Focusing on the traffic pattern variation throughout the week, Figure 3 shows that there are very significant differences between the days of the week, especially between the weekend and weekdays. Also, we can observe an overall increasing trend from Monday to Friday, with Friday being the weekday with higher traffic volume.

In addition to the differences in terms of daily traffic volume, there are also clear differences in terms of the daily traffic pattern itself. For instance, Figure 4 shows that on

Mondays the traffic follows a typical pattern where the peak hour is between 8 and 9 a.m., when most people go to work. Another peak occurs between 2 and 3 p.m., which denotes mobility from people working in the afternoon. Finally, a last traffic peak is detected between 6 and 8 p.m., when workers return to their homes. Other weekdays follow a similar pattern.

A totally different pattern is detected, for example, on a Sunday. Compared to weekdays we find that (i) work-related traffic peaks are no longer present; (ii) the total traffic volume is significantly lower; and (iii) the peak hours differ. In particular, peak hours are now related to mobility towards food courts at lunch time (between 1 and 2 p.m.) and mobility from relax areas to homes (between 6 and 8 p.m.).

When focusing on the traffic distribution throughout a city, it is well known that main streets and avenues will experience a much higher traffic load than secondary and isolated ones. Discriminating between them is a relevant issue since some streets barely experience any traffic load increase during peak hours, meaning that travel times are not affected by congestion in the same way as the main arteries of the city.

To be able to discriminate between the streets of Valencia based on traffic flow, we first obtained the peak traffic intensity per street during November, and we then obtained the cumulative distribution for these values (see Figure 5).

We observe that 30.3% of all streets have a traffic intensity lower than 690 vehicles/hour during peak hours, which according to [28] means that these low traffic intensity streets will not experience traffic congestion even at peak hours, and so they can be discarded from our time-dependent modeling efforts. Additionally, we observe that the number of streets/avenues with very high traffic volumes (more than 10,000 vehicles during the peak hour) is rather limited (about 10%). Thus, the majority of the streets in a city will experience moderate traffic volumes, and the global peak hour behavior will not cause any noticeable effect on these streets. To confirm this observation, Figure 6 shows the traffic load per hour in two different streets for the same day. Notice that although both share quite similar values for peak traffic intensity, the daily traffic patterns significantly differ that the peaks in one pattern often match valleys in the other pattern.

Observing the daily traffic pattern in Figure 6(a), we find that it closely matches the traffic pattern of a typical Monday, as shown in Figure 4(a); on the contrary, Figure 6(b) shows a quite different traffic pattern. Hence, it becomes necessary to discriminate between the different streets based on their daily traffic pattern. To achieve this goal, we will apply a clustering technique in order to automatically classify streets according to their daily traffic pattern.

5. Clustering Heuristic

In this section we propose a heuristic to simplify traffic modeling for the city of Valencia by taking into consideration the results presented in the previous section.

The proposed heuristic aggregates into a single pattern all those daily traffic patterns having a common behavior. This is made possible by making the obtained time-dependent

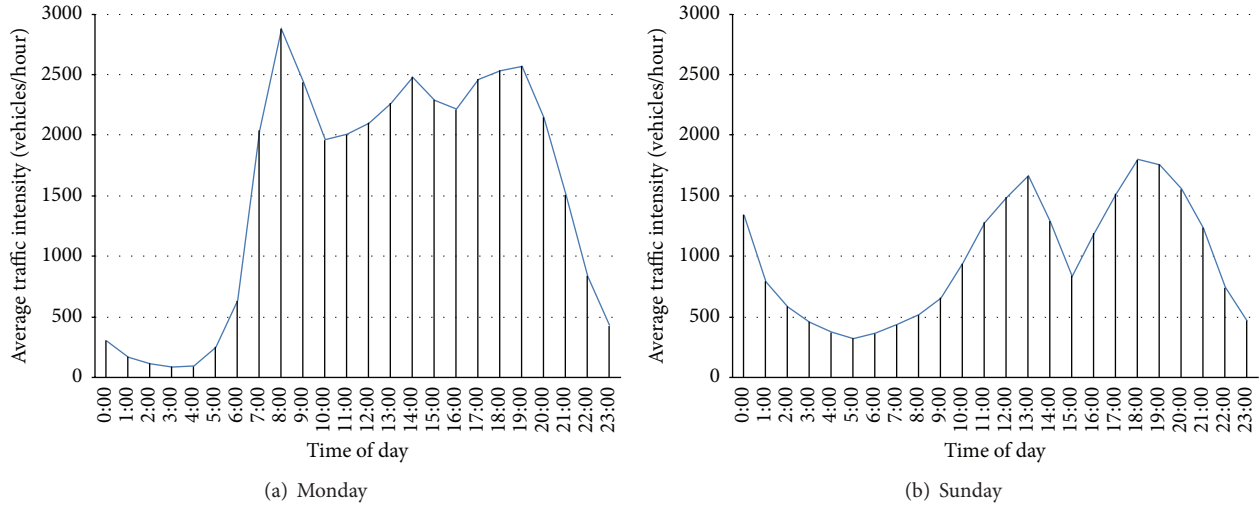


FIGURE 4: Average daily behavior for different days of the week.

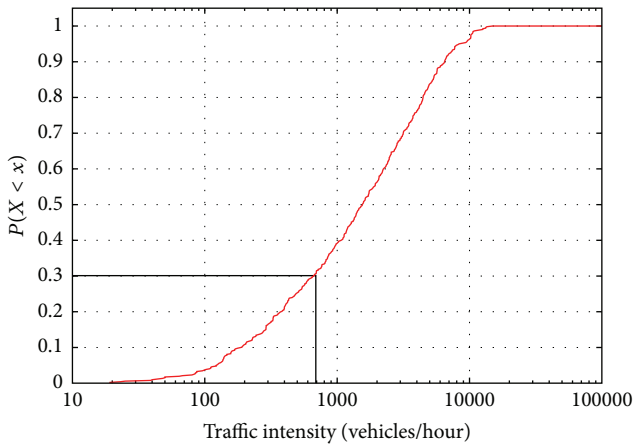


FIGURE 5: Cumulative distribution for traffic intensity using the monthly peak hours.

models independent of the actual number of vehicles in each street through normalization using the mean daily value.

To this aim, we use Mathematica 9.0.1 [29], which is a widely recognized tool to solve mathematical problems, especially in engineering. This tool provides function *FindClusters*, which returns the number of clusters as well as the elements on each cluster. This function has several options and suboptions. In fact, we can choose between a hierarchical method or a partitioning method. The partitioning method it uses is based on the Partitioning Around Medoids (PAM) algorithm [22], which seeks to find k representative objects called medoids from the data set such that the sum of the dissimilarities within a cluster are minimized. A medoid can be defined as that object of a cluster whose average dissimilarity to all the objects in the cluster is minimal. After finding the set of medoids, each object of the data set is assigned to the nearest medoid.

We have chosen the partitioning method of *FindClusters* for two reasons. The first one is that this method is the default

option, and the second and most important one is that the PAM algorithm is the one used by reference authors on the topic such as Guardiola et al. (see [24]), who claim that the choice of PAM is due in part to the large number of statistics it provides for thorough analysis of the resultant clusters.

At this point, we want to stress the fact that while [24] (and also [27]) try to cluster different days corresponding to the same section of a freeway, the aim of our procedure is quite different; particularly, we attempt to cluster different streets corresponding to the same day. Moreover, as far as we know, the clustering distance that we will use here has not been used in any previous paper on ITS.

Finally, note that although we have not made use of them, function *FindClusters* has suboptions in order, for instance, to fine-tune the number of clusters. Probably the best known suboption to do this is the silhouette statistic [22], but according to [23] there is no criterion providing evidence about its superiority compared to others in the general case of adjusting the number of clusters. In addition, notice that two properties that define a good heuristic and that we have taken into account to our aim are low time overhead and simplicity of its steps.

Below we describe the five steps followed to reduce the number of independent daily patterns to be modeled: (i) select the appropriate clustering metric, (ii) find the optimal number of clusters per day of the week, (iii) determine how representative mean days are, (iv) group days of the week with similar characteristics, and (v) group clusters with similar daily patterns.

5.1. Selection of a Clustering Metric for Per-Hour Street Behavior.

If for each street (or street segment) we have the number of cars that traverse it every hour, we can represent each street by a point $x = (x_1, x_2, \dots, x_{24})$ in \mathbb{R}^{24} , where x_i is the number of cars traversing the street at hour i . Suppose we have two streets $x = (x_1, x_2, \dots, x_{24})$ and $y = (y_1, y_2, \dots, y_{24})$. By default, the distance used to form clusters is the Euclidean

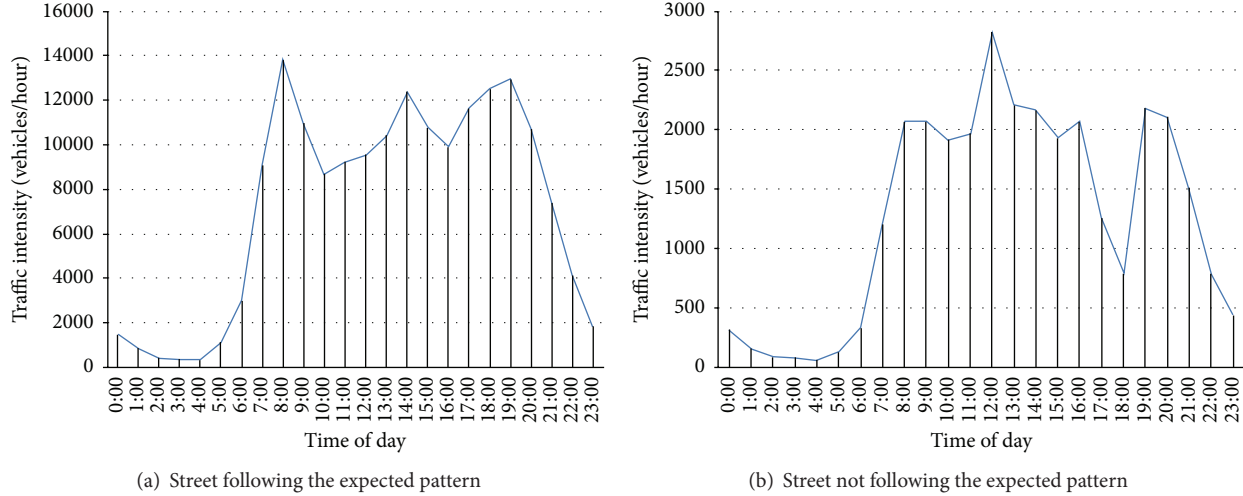


FIGURE 6: Daily traffic intensity pattern for streets with different characteristics.

distance, $\sqrt{\sum_{i=1}^{24} (x_i - y_i)^2}$. If the Euclidean distance between two points is relatively small, both streets will belong to the same cluster. However, if we attempt to classify streets taking into account the traffic variability as a function of the time of day, we believe that this distance is not adequate. Let us take a small illustrative example in this regard. Suppose that we only consider six consecutive hours for four different streets and that their respective points are $a = (12, 9, 10, 9, 8, 11)$, $b = (24, 20, 22, 20, 17, 21)$, $c = (6, 16, 20, 25, 17, 7)$, and $d = (15, 35, 44, 48, 34, 17)$.

Streets a and b have a similar behavior: the relative number of vehicles traversing them every hour is more or less the same, within certain bounds. Although the actual number of vehicles differs greatly from one street to another, both streets should be in the same group encompassing all those streets where there is little traffic variability, where vehicle speeds can be considered mostly constant over the considered period.

With respect to streets c and d , central hours are peak periods where we have about three times the traffic volume compared to edge values. Although the number of vehicles differs greatly from one street to another, they should belong to the same group characterized by a single peak corresponding to hours in the mid-range and with much lower values on the edges.

However, if we classify the four streets using the Euclidean distance, the result is quite predictable: $\{a, c\}$ and $\{b, d\}$. In this example the Euclidean distance has created two clusters grouping the two streets with less traffic and the two streets with high traffic volume. To address this problem, we believe that the distance metric that best fits our objective is the correlation distance, defined as $1 - |r_{xy}|$, where r_{xy} is the correlation coefficient:

$$r_{xy} = \frac{\sum_{i=1}^{24} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{24} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{24} (y_i - \bar{y})^2}}. \quad (1)$$

Recall that $|r_{xy}|$ is always less than or equal to 1 and that values close to 1 indicate that variables x and y have a direct linear relationship, meaning that the graphical representation of the 24 points (x_i, y_i) is approximately a straight line. Therefore, the higher the correlation between points x and y is, the closer to zero $1 - |r_{xy}|$ becomes, and so the probability of belonging to the same cluster will increase. If we classify the four streets according to correlation distance, the result obtained is the desired one: $\{a, b\}$ and $\{c, d\}$.

On the other hand, it is easy to see that the correlation distance is the same if we work with the coordinates (x_i, y_i) or with coordinates $(x_i / \sum_{j=1}^{24} x_j, y_i / \sum_{j=1}^{24} y_j)$, taking into account that, to compare streets considering traffic variability throughout the day, it also seems useful to compare the percentage of the daily traffic passing on every street for each hour. This way, it does not matter whether we compare both streets considering the number of cars per hour or the percentage of traffic per hour: the classification using the correlation distance will generate the same clusters. This is obviously not true when adopting Euclidean distances.

5.2. Finding the Optimal Number of Clusters for Each Day of the Week. Using the correlation distance defined previously, in this section we will determine the optimal number of clusters for the 292 streets in Valencia considered by the City Hall as representative in terms of traffic flow for every day of the week. Subsequently, to reduce the overall number of clusters, we will attempt to join the different days in a week whenever the same number of clusters are detected.

Therefore, for our analysis, we apply the *FindClusters* function to each of the 28 days of November studied enabling the correlation distance option. For each day, the function will cluster the 292 points in \mathbb{R}^{24} corresponding to the streets taken for our study.

In the analysis that follows we work with the percentage of vehicles traversing each street every hour with respect to the overall daily value. As referred in the previous section, the actual number of vehicles *per se* is not relevant to

TABLE 1: Number of clusters obtained and associated statistics.

	Mo	Tu	We	Th	Fr	Sa	Su
A: Week 1	3	3	1	2	1	2	3
B: Week 2	1	2	4	1	3	2	4
C: Week 3	5	3	1	4	2	2	3
D: Week 4	3	1	1	3	3	2	1
E: mean(A, B, C, D)	3	2.25	1.75	2.5	2.25	2	2.75
F: median(A, B, C, D)	3	2.5	1	2.5	2.5	2	3
G: average day	4	2	2	4	3	2	2
H: round(E) == G	False	True	True	False	False	True	False
I: mean(E, F, G)	3.3(3)	—	—	3	2.58(3)	—	2.58(3)
Number of clusters	3	2	2	3	3	2	3

TABLE 2: Percentages of matching for the different clusters compared to the average day clusters.

	Mo	Tu	We	Th	Fr	Sa	Su
Number of clusters	3	2	2	3	3	2	3
Week 1	83.11	92.31	84.42	30.86	70.15	91.98	73.72
	66.67	84.56	62.32	59.32	68.75	71.43	66.67
	81.33			56.99	90.43		35.82
Week 2	60.14	89.74	80.52	81.48	70.15	96.26	74.36
	55.07	58.09	59.42	43.22	60.64	81.90	69.56
	80.00			58.06	51.56		73.13
Week 3	62.84	84.62	80.52	58.02	70.15	88.77	51.28
	69.57	32.35	91.30	75.42	74.47	84.76	47.83
	84.00			31.18	35.94		89.55
Week 4	81.76	96.15	74.68	62.96	86.57	97.87	82.05
	88.41	84.56	82.61	74.58	65.96	53.30	59.42
	76.00			65.59	56.25		58.21
Average	73.63	78.68	77.05	58.56	69.18	86.21	66.70

our purposes, and the correlation distance metric adopted provides the same output on both cases.

Since our study period encompasses 4 weeks, we create an “average day” for each day of the week, which is calculated for each street by averaging the number of vehicles traversing it each hour. Such “average day” attempts to filter out the peculiarities of a specific day, obtaining a representative trend instead.

Table 1 shows the results obtained, where the last row shows the cluster allocation for each day of the week. To attain those values, we first apply function *FindClusters* to different weeks (A–D) and to the “average days” (G). In addition, we calculate the mean (E) and the median (F) for the cluster groups corresponding to the different weeks. If this mean value (E) is rounded to a number that matches the number of clusters for the average day (G), then we define such value as the number of clusters for that day of the week. Otherwise, we obtain the average of the mean (E), median (F), and average day (G) to obtain a value (I) that when rounded defines the number of clusters to be used. We find that the proposed

number of clusters matches the rounded mean (E) except for a minor change in one day.

5.3. Determining Cluster Matching on a Per-Day Basis. Once the number of clusters for each day of the week was defined, the next step was to validate that cluster elements for each day of the week resembled the cluster elements obtained for the average day. If a good degree of matching is obtained, then the conclusions associated with streets in that cluster are valid; otherwise, we could be considering that streets belong to a group with a specific behavior, when in fact their behavior significantly differs.

For our endeavor we apply the *FindClusters* function to the 35 days (28 real days plus 7 average days), but this time fixing the number of clusters defined a priori, as obtained in the previous section. Afterwards, for each of the four weeks under analysis, we compare the clusters obtained against the average day of the week, determining the percentage of streets that both clusters have in common. These results are presented in Table 2.

TABLE 3: Percentages of cluster matching for average days of the week with same number of assigned clusters. Valid combinations are shown in boldface.

Combinations	Degree of matching (%)	Average matching (%)
Monday-Thursday	57.43	48.29
	8.69	
	66.67	
Monday-Friday	77.70	68.84
	72.46	
	48.00	
Monday-Sunday	58.78	43.49
	27.54	
	28.00	
Thursday-Friday	37.04	59.25
	71.19	
	63.44	
Thursday-Sunday	30.86	41.78
	55.93	
	33.33	
Friday-Sunday	61.94	51.37
	36.17	
	51.56	
Tuesday-Wednesday	91.67	91.78
	91.91	
Tuesday-Saturday	71.15	58.56
	44.11	
Wednesday-Saturday	69.48	56.51
	42.03	

We find that the average degree of matching for all the days of the week is 72.71%. Globally, we find that this value is quite acceptable and that differences appearing on specific days are expectable since traffic patterns may suffer some changes depending on weather, special events, or other conditions.

5.4. Grouping Days of the Week with Similar Cluster Characteristics. The next step of our clustering procedure was to assess the feasibility of grouping those days of the week having the same number of clusters. With this purpose we tested all combinations and calculated the percentage of cluster matching for each pair of mean days of the week. The results are shown in Table 3.

All combinations show an average degree of matching below 70%, except for the Tuesday-Wednesday combination which is close to 92%. Thus, we agree that these two weekdays can be combined as if they were a single day since similar patterns are obtained in terms of traffic variability throughout the day. Data shown earlier in Figure 3 also emphasize this similitude.

To confirm that the grouping did not have a negative impact on the error associated with specific days, we now

proceed to compare the degree of matching for the different clusters against the average day, the crossed average day, and the proposed union of both days. These results are shown in Table 4.

We find that the differences between the three cases are quite low. Specifically, the impact of grouping these two days into one is of only 1.6%, which is quite acceptable. The results using cross averages also strengthen the point of unifying these two days. As a result, by accounting for the number of clusters of each average day and by merging Tuesday and Wednesday into a single day, we obtain a total of 16 different traffic patterns.

5.5. Grouping Clusters with Similar Daily Patterns. In this section we present the normalized traffic patterns corresponding to the 16 clusters created: 3 for Monday, 2 for Tuesday/Wednesday, 3 for Thursday, 3 for Friday, 2 for Saturday, and 3 for Sunday.

As shown in Figure 7, there are some pattern similitudes between the first weekdays (Monday versus Tuesday/Wednesday), between the last weekdays (Thursday versus Friday), and between weekend days (Saturday versus Sunday). However, this initial insight obtained visually must be confirmed through statistical evidence. With this purpose we picked the clusters for those days which visually show some similitude and calculated the correlation between the daily patterns associated with each cluster for relevant time ranges. The results of these analyses are presented in Table 5.

When comparing the daily pattern for the clusters of Monday against Tuesday/Wednesday (see Table 5(a)), we find that there is a high correlation (>92%) between the patterns corresponding to the first 2 clusters of each of these days. Thus, a single model will suffice when attempting to represent the daily pattern for these clusters that only a different model is required for Monday's Cluster number 3.

When comparing Thursday against Friday, we find that only Cluster number 2 for Thursday and Cluster number 1 for Friday present a high correlation (~94%).

Finally, when comparing Saturday against Sunday, we find that Cluster number 1 and Cluster number 3 present a good degree of matching (~94%), and these two clusters can also be represented through same daily pattern.

6. Generalization and Benefits of the Proposed Model

In this section we assess the benefits of our model in terms of the minimum number of patterns required to adequately describe traffic intensity throughout the day for the city of Valencia. Then, we detail how these different models obtained can be integrated in our traffic management platform to predict route costs. Finally we summarize our proposal by presenting the proposed heuristic in pseudocode format to allow generalizing the proposed procedure to any target city.

6.1. Aggregation Gains Achieved. Below we discuss the different aggregation techniques that integrate our heuristic and the previous analysis.

TABLE 4: Percentages of matching for the different clusters against the average day, the crossed average day, and the proposed union of both days.

	Original average days		Crossed average days		Union of average days	
	Tu	We	Tu	We	Tu	We
Week 1	92.31	84.42	90.26	86.54	86.83	86.23
	84.56	62.32	81.16	65.44	84.00	69.60
Week 2	89.74	80.52	89.61	78.21	85.63	77.25
	58.09	59.42	57.25	58.09	56.80	60.00
Week 3	84.62	80.52	86.36	78.85	82.04	72.46
	32.35	91.30	34.06	90.44	30.40	88.00
Week 4	96.15	74.68	95.45	73.08	92.22	68.26
	84.56	82.61	82.61	81.62	86.40	80.00
Average	78.68	77.05	77.82	76.71	77.14	75.34

TABLE 5: Correlation between clusters (period between 7 a.m. and 9 p.m.).

(a) Monday and Tuesday/Wednesday

		Tuesday/Wednesday	
		Cluster number 1	Cluster number 2
Monday	Cluster number 1	0.9221668	0.578729
	Cluster number 2	0.6229643	0.9422671
	Cluster number 3	0.5900097	0.7910942

(b) Thursday and Friday

		Friday		
		Cluster number 1	Cluster number 2	Cluster number 3
Thursday	Cluster number 1	0.6741969	0.2552095	0.7292981
	Cluster number 2	0.9393144	0.6691599	0.6666628
	Cluster number 3	0.7247197	0.7841533	0.8645128

(c) Saturday and Sunday

		Sunday		
		Cluster number 1	Cluster number 2	Cluster number 3
Saturday	Cluster number 1	0.8859214	0.8585393	0.9368844
	Cluster number 2	0.8948805	0.8840648	0.7977545

Yearly Analysis. The monthly behavior results shown before allow assuming that traffic volumes throughout the year are mostly constant, except for vacation periods like summer and festivities lasting for long periods (e.g., Easter), meaning that partitioning weeks into three groups (typical week, relevant holiday period, and summer holidays) seems appropriate.

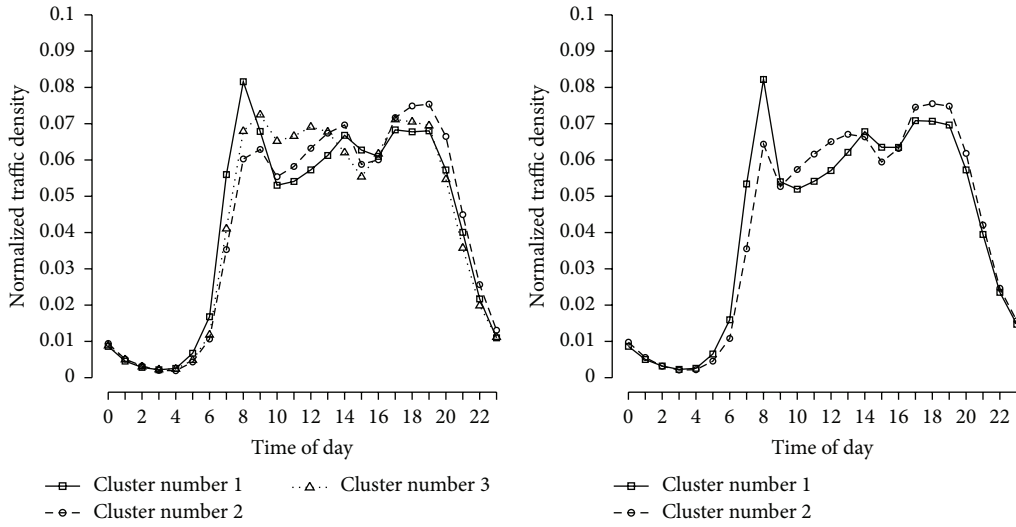
Monthly Analysis. Results have shown that, for the same type of period, data is consistent across weeks, which allows clustering the different days of a month in a single *average representative week*.

Traffic Intensity Analysis. Concerning traffic congestion for the different streets and avenues of a city, our heuristic assumes that only a subset of these streets/avenues actually face significant congestion problems deserving time-dependent modeling, while for the rest, the use of traditional fixed-cost approaches suffices. Based on the thresholds defined in [28] for class IV (urban) arterial types, we consider that only

those streets with a peak traffic value surpassing 690 vehicles per hour are actually experiencing congestion-related traffic delays. This way, the target number of streets/avenues can be reduced from 421 (total number of streets being monitored by traffic services) to 292 (number of streets with a relevant traffic load).

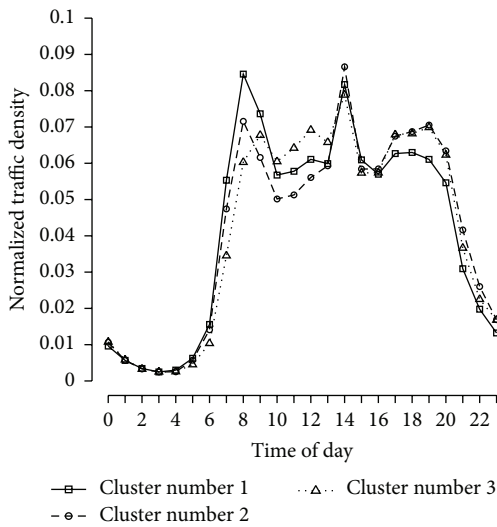
Clustering Analysis. Focusing on the street/avenue subset significantly affected by congestion, the clustering analysis showed that a small number of groups can be created, where for each group all streets/avenues follow very similar traffic congestion patterns. Thus, the target number of models required can be reduced from 292 per 7 days in a week to a total of 18, and this value can be further reduced to 16 by noticing the similarity between Tuesday and Wednesday.

Daily Pattern Analysis. An analysis of the daily patterns associated with the different clusters defined for the different days of the week has shown that some of these clusters have

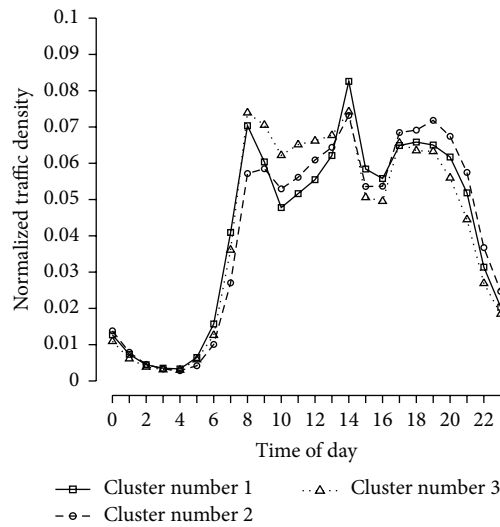


(a) Monday

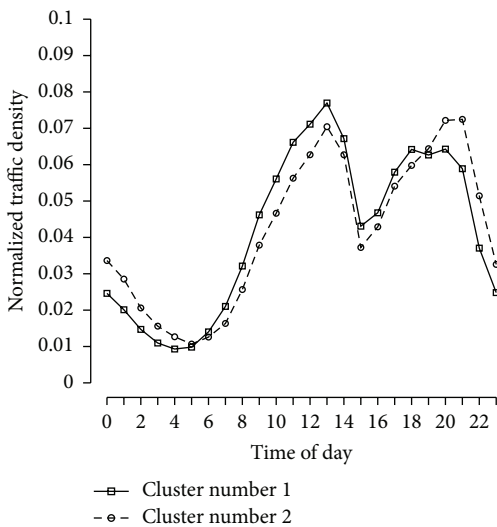
(b) Tuesday/Wednesday



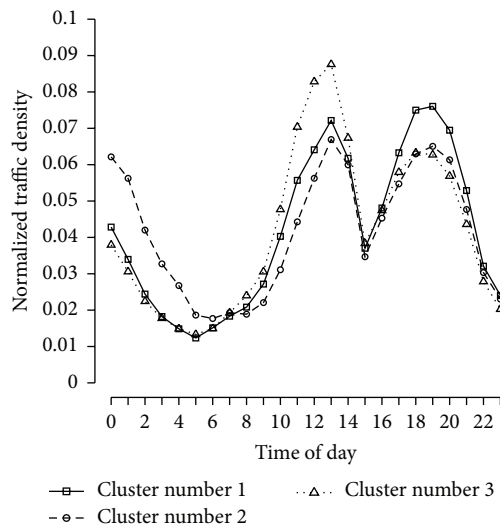
(c) Thursday



(d) Friday



(e) Saturday



(f) Sunday

FIGURE 7: Cluster description for the different average days considered.

TABLE 6: Benefits of the proposed heuristic in terms of aggregation gain.

Target heuristic	Number of elements	Aggregation gains	Independent modeling domains
Monthly patterns per year	12	12 : 3	3
Daily patterns per month		30 : 7	
Traffic intensity analysis		421 : 292	
Street clustering	421×30	$(292 \times 7) : 18$	12
Similar days clustering		18 : 16	
Daily pattern analysis		16 : 12	
Total	151,560	4210	36

a common behavior. This means, in general, that the same group of streets behaves similarly across different days, which allows slightly reducing the number of patterns from 16 to 12.

Based on aforementioned aggregation proposals for the city of Valencia, in Table 6 we detail the benefits obtained in terms of model simplification. As can be observed, street clustering is the key element when reducing the number of separate modeling domains required to characterize the traffic behavior throughout the year. In particular, aggregation based on the clustering analysis is the most critical one, allowing for substantially reducing the number of interpolation functions required. The second most relevant aggregation gain is associated with yearly and week behavior, based on segregating work periods from short/long holiday periods and by finding that we have the same behavior across the different weeks. Eliminating secondary streets that experience fluid traffic throughout the whole year also provides some contribution in terms of aggregation gain by eliminating the need for modeling their traffic throughout the day. Finally, the daily pattern analysis across clusters has further helped reducing the number of models required.

Overall, the proposed heuristic allows reducing the required number of interpolation functions for the city of Valencia by a factor of 4210 while maintaining the essence of time-dependent modeling requirements. Such a significant reduction certainly simplifies the integration of these models in our ABATIS platform and allows accelerating the associated calculations. This way, route decisions are taken in a centralized route server based on traffic states prediction throughout the day and for the different streets/avenues of a city, thus providing the most time-efficient routes.

6.2. Applicability of the Model in the Context of ABATIS. The relationship between traffic flow levels and average travel speed is a well-known topic in traffic flow theory [30]. As shown in Figure 8, this relationship can be closely approximated through a parabolic behavior represented through the following expression, obtained by interpolating points $(0, 0)$, $(s_f/2, v_m)$, and $(s_f, 0)$:

$$v(s) = \frac{4 \cdot v_m}{s_f^2} \cdot s \cdot (s_f - s). \quad (2)$$

As expected, average travel speed starts to decay when traffic density per lane increases beyond a certain threshold

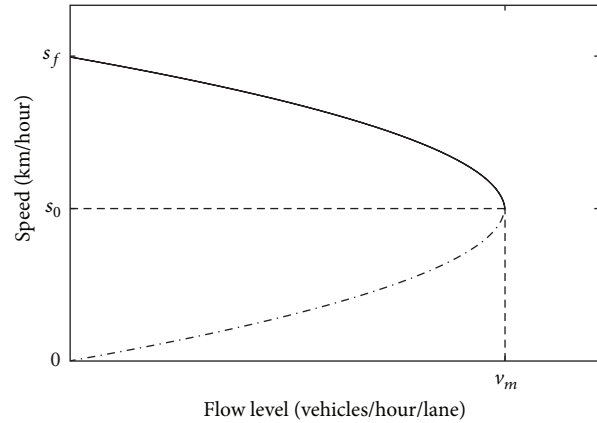


FIGURE 8: Relationship between vehicle flow level and vehicle speed.

and becomes close to zero when approaching the maximum road capacity.

Since our models required a normalization of traffic levels of each street in order to perform model aggregation for similar patterns, given a street and an instant of time a vehicle is expected to enter the street, we show below the four steps involved in calculating the travel time for that street starting at the given instant of time. Note that, for simplicity, we do not put to the variables the subindexes corresponding to the given street and instant of time.

- (i) Find the normalized traffic intensity (pattern) n at the time the vehicle is expected to enter the target street, using the daily pattern for the target street.
- (ii) Obtain the expected traffic flow level v for that street and instant of time by denormalizing the obtained value using the mean traffic volume \bar{v} for the target street:

$$v = n \cdot \bar{v}. \quad (3)$$

- (iii) Based on the average free-flow speed s_f and the maximum flow v_m for the target street (provided by authorities), the expected travel speed s_v can be obtained based on the predicted traffic flow level v . Specifically and taking the behaviour of Figure 8 corresponding

```

input: 3D array of traffic density per street, per hour, per day
output: pattern-dependant cluster classification
BEGIN
for each street in All_streets do {
  if (peak_traffic_intensity in P < 690 veh/h) then
    remove street from All_streets
  }
for each Week_day in WEEK_DAY do {
  average_Week_day = get_average_pattern(Week_day)
  clusters[] = FindClusters(Week_day, average_Week_day)
  mean_clusters = get_average(clusters[Week_day])
  median_clusters = get_median(clusters[Week_day])
  if (clusters[average_Week_day] == round(mean_clusters)) then
    num_clusters[week_day] = clusters[average_Week_day]
  else
    num_clusters[week_day] = round(get_average(mean_clusters,
      median_clusters, clusters[average_Week_day]))
  }
for all week_day pairs (w_i, w_j)
where num_clusters[w_i] == num_clusters[w_j] do {
  if (Matching(cluster_elements(w_i), cluster_elements(w_j)) > 90%)
  then pattern[w_i] = pattern[w_j]
  }
for all week_day pairs (w_i, w_j) with different pattern do {
  for all clusters c_i in w_i and c_j in w_j do {
    if (correlation(average_street(c_i), average_street(c_j)) > 0.9)
    then
      pattern[c_i] = pattern[c_j]
    }
  }
}
RETURN cluster pattern classification
END algorithm

```

ALGORITHM 1: Cluster patterns.

to below flow saturation levels as reference (solid line section), s_v can be approximated as follows:

$$s_v = \frac{s_f}{2} \cdot \left(1 + \sqrt{1 - \frac{v}{v_m}}\right). \quad (4)$$

- (iv) Calculate the travel time t for the target street with length L using the expected travel speed:

$$t = \frac{L}{s_v}. \quad (5)$$

Notice that, since the ABATIS platform is able to offer, among others, Traffic Management as a Service, it is able to serve optimal routes to clients. Currently, route costs are calculated using free-flow speeds. Thus, the proposed models can be integrated in the route calculation engine so that optimality conditions now account for the updated path costs using our predictive model. In addition, if the current status of the traffic flow is available in the future, it can be combined with the predicted value to further improve path cost accuracy.

6.3. Pseudocode for the Proposed Heuristic. Let P represent the time period under analysis and let *Week_day* represent

the set of days in P corresponding to a particular day of the week. *WEEK_DAY* is a superset containing all *Week_day* sets and *All_streets* represents the set containing all the streets for the target city.

Algorithm 1 shows the pseudocode that allows applying the proposed heuristic in a systematic manner, thereby making it applicable to any target city.

7. Conclusions

Traffic management has evolved substantially in the last decades. Nowadays, traffic engineers require effective solutions to help them improve the traffic flow in cities, while minimizing travel times and tackling traffic-related problems such as CO₂ emissions, noise, and accidents.

In this paper we define a procedure to obtain reliable traffic congestion estimations for all the streets/avenues in a city for the different times of the day and for every day in a year. Considering the modeling effort required, we proposed a heuristic that allows reducing the number of required interpolation functions characterizing daily traffic patterns.

By specifically addressing the city of Valencia, we made a detailed analysis of traffic behavior on the different streets/avenues of the city to determine (i) the behavior along the

year, (ii) which days of the week show a similar pattern, (iii) which streets/avenues experience more traffic congestion, and (iv) how streets can be grouped into clusters based on their daily traffic pattern. The results of our analysis show that it is possible to model the traffic behavior in the city by aggregating elements with a similar behavior in the same interpolation function. This way, we will be able to account for the travel time variations along the main paths of a city, providing users with both optimized and accurate travel plans, while reducing the modeling complexity.

As future work we will develop a smartphone application that interacts with the ABATIS platform in order to obtain the most efficient routes, and we will implement a route planning algorithm that allows selecting these best paths while accounting for time-dependencies, FIFO restrictions, turn penalties, and so forth.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was partially supported by Valencia's Traffic Management Department and by the "Ministerio de Economía y Competitividad, Programa Estatal de Investigación, Desarrollo e Innovación Orientada a los Retos de la Sociedad, Proyectos I+D+I 2014," Spain, under Grant TEC2014-52690-R.

References

- [1] S. Salcedo-Sanz, D. Manjarrés, Á. Pastor-Sánchez, J. Del Ser, J. A. Portilla-Figueras, and S. Gil-López, "One-way urban traffic reconfiguration using a multi-objective harmony search approach," *Expert Systems with Applications*, vol. 40, no. 9, pp. 3341–3350, 2013.
- [2] R. Doolan and G.-M. Muntean, "VANET-enabled eco-friendly road characteristics-aware routing for vehicular traffic," in *Proceedings of the IEEE 77th Vehicular Technology Conference (VTC Spring '13)*, pp. 1–5, Dresden, Germany, June 2013.
- [3] A. Corti, V. Manzoni, and S. M. Savaresi, "Simulation of the impact of traffic lights placement on vehicle's energy consumption and CO₂ emissions," in *Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems (ITSC '12)*, pp. 620–625, Anchorage, Alaska, USA, September 2012.
- [4] Y. Xu, Y. Zhang, and M. Liu, "Multiagent based decentralized traffic light control for large urban transportation system," *Mathematical Problems in Engineering*, vol. 2014, Article ID 104349, 13 pages, 2014.
- [5] F. Pop and C. Dobre, "An efficient pagerank approach for urban traffic optimization," *Mathematical Problems in Engineering*, vol. 2012, Article ID 465613, 9 pages, 2012.
- [6] V. Nedic, S. Cvetanovic, D. Despotovic, M. Despotovic, and S. Babic, "Data mining with various optimization methods," *Expert Systems with Applications*, vol. 41, no. 8, pp. 3993–3999, 2014.
- [7] M. K. Nasir, R. Md Noor, M. A. Kalam, and B. M. Masum, "Reduction of fuel consumption and exhaust pollutant using intelligent transport systems," *The Scientific World Journal*, vol. 2014, Article ID 836375, 13 pages, 2014.
- [8] G. Allì, L. Baresi, A. Bianchessi et al., "Green move: towards next generation sustainable smartphone-based vehicle sharing," in *Proceedings of the Sustainable Internet and ICT for Sustainability (SustainIT '12)*, pp. 1–5, IEEE, Pisa, Italy, October 2012.
- [9] J. Zaldivar, C. T. Calafate, J. C. Cano, and P. Manzoni, "Providing accident detection in vehicular networks through OBD-II devices and android-based smartphones," in *Proceedings of the 36th Annual IEEE Conference on Local Computer Networks (LCN '11)*, pp. 813–819, IEEE, Bonn, Germany, October 2011.
- [10] M. Westerman, R. Litjens, and J. Linnartz, "Integration of probe vehicle and induction loop data: estimation of travel times and automatic incident detection," Tech. Rep., 1996.
- [11] J. C. Herrera, D. B. Work, R. Herring, X. Ban, Q. Jacobson, and A. M. Bayen, "Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century field experiment," *Transportation Research, Part C: Emerging Technologies*, vol. 18, no. 4, pp. 568–583, 2010.
- [12] S. Kim, M. E. Lewis, and C. C. White III, "Optimal vehicle routing with real-time traffic information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 178–188, 2005.
- [13] T. Nadeem, S. Dashtinezhad, C. Liao, and L. Iftode, "Trafficview: traffic data dissemination using car-to-car communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 8, no. 3, pp. 6–19, 2004.
- [14] D. Work and A. Bayen, "Impacts of the mobile internet on transportation cyberphysical systems: traffic monitoring using smartphones," in *Proceedings of the National Workshop for Research on High-Confidence Transportation Cyber-Physical Systems: Automotive, Aviation, & Rail*, pp. 18–20, Washington, DC, USA, November 2008.
- [15] C. Claudel, A. Hofleitner, N. Mignerey, and A. Bayen, "Guaranteed bounds on highway travel times using probe and fixed data," in *Proceedings of the 88th TRB Annual Meeting Compendium of Papers*, Washington, DC, USA, 2009.
- [16] I. Leontiadis, G. Marfia, D. MacK, G. Pau, C. Mascolo, and M. Gerla, "On the effectiveness of an opportunistic traffic management system for vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1537–1548, 2011.
- [17] C. Qian, C.-Y. Chan, and K.-L. Yung, "Reaching a destination earlier by starting later: revisited," *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, no. 5, pp. 641–647, 2011.
- [18] O. Bräysy, E. Martínez, Y. Nagata, and D. Soler, "The mixed capacitated general routing problem with turn penalties," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12954–12966, 2011.
- [19] A. L. Kok, E. W. Hans, and J. M. J. Schutten, "Vehicle routing under time-dependent travel times: the impact of congestion avoidance," *Computers and Operations Research*, vol. 39, no. 5, pp. 910–918, 2012.
- [20] F. Dehne, M. T. Omran, and J.-R. Sack, "Shortest paths in time-dependent fifo networks," *Algorithmica*, vol. 62, no. 1-2, pp. 416–435, 2012.
- [21] D. Soler, J. Albiach, and E. Martínez, "A way to optimally solve a time-dependent vehicle routing problem with time windows," *Operations Research Letters*, vol. 37, no. 1, pp. 37–42, 2009.

- [22] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, New York, NY, USA, 1990.
- [23] R. Xu and D. Wunsch II, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.
- [24] I. G. Guardiola, T. Leon, and F. Mallor, "A functional approach to monitor and recognize patterns of daily traffic profiles," *Transportation Research, Part B: Methodological*, vol. 65, pp. 119–136, 2014.
- [25] Y. Wang, Y. Chen, M. Qin, and Y. Zhu, "Dynamic traffic prediction based on traffic flow mining," in *Proceedings of the 6th World Congress on Intelligent Control and Automation (WCICA 0'6)*, vol. 2, pp. 6078–6081, Dalian, China, June 2006.
- [26] N. Caceres, L. M. Romero, and F. G. Benitez, "Estimating traffic flow profiles according to a relative attractiveness factor," *Procedia—Social and Behavioral Sciences*, vol. 54, pp. 1115–1124, 2012.
- [27] M. Yildirimoglu and N. Geroliminis, "Experienced travel time prediction for congested freeways," *Transportation Research Part B: Methodological*, vol. 53, pp. 45–63, 2013.
- [28] Transportation Research Board, *Highway Capacity Manual*, Transportation Research Board, 2010.
- [29] Wolfram, Mathematica, 2015, <http://www.wolfram.com/mathematica/>.
- [30] Federal Highway Administration, "Revised monograph on traffic flow theory," Tech. Rep., Federal Highway Administration, Washington, DC, USA, 2001.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

