Vehicular Congestion Detection and Short-Term Forecasting: A New Model with Results

Gustavo Marfia and Marco Roccetti

Abstract

While vehicular congestion is very often defined in terms of aggregate parameters such as traffic volumes, and lane occupancies, based on recent findings the interpretation that receives most credit is that of a state of a road where traversing vehicles experience a delay exceeding the maximum value typically incurred under light or free-flow traffic conditions. We here propose a new definition, according to which a road is in a congested state (be it high or low) only when the likelihood of finding it in the same congested state is high in the near future. Based on this new definition, we devised an algorithm which, exploiting probe vehicles, for any given road: (a) identifies if it is congested or not, and; (b) provides the estimation that a current congested state will last for at least a given time interval. Unlike any other existing approach, an important advantage of ours is that it can be generally applied to any type of road, as it neither needs any a-priori knowledge nor require to estimate any road parameter (e.g., number of lanes, traffic light cycle, etc.). Further, it allows to perform short term traffic congestion forecasting for any given road. We present several field trials gathered on different urban roads whose empirical results confirm the validity of our approach.

Index Terms

Vehicular traffic congestion definition, traffic forecasting, intelligent transportation systems, vehicular software technology for congestion detection.

I. INTRODUCTION

Two main approaches have emerged with the aim of limiting the business and societal costs of vehicular congestion. The first approach amounts to provide aggregate traffic information (such as the intensity of traffic volumes and lane occupancy rates) to transportation authorities, which,
in turn, feed this information into Advanced Traffic Management Systems (ATMSs) to control traffic lights. The second approach is more pervasive and is based on the idea of gathering road traversal times from probe vehicles [1]–[7]. This information is then managed by Advanced Traveler’s Information Systems (ATISs) which, in turn, supply single drivers with a feedback on traffic and with suggestions on the best routes to a destination as a function of real-time traffic conditions (simplistic examples of ATIS are Personal Navigation Devices, PNDs [5]).

Although the importance of providing updated information on traffic conditions, on a per single vehicle basis, is widely understood, to this date, the most amount of research has been devoted to devise traffic estimation and forecast algorithms to be used by ATMSs. As such, these algorithms are mostly concerned with the problem of keeping under control specific locations (e.g., highways and principal arterial roads) subject to high volumes of traffic [8], [9]. In fact, roads where flow intensities are high, as in freeways, are easy to monitor, since they usually have a small number of interconnections with other roads while represent a small subset of the whole urban map.

It is, instead, hard to obtain a comprehensive, road by road, picture of urban traffic, since such roads are tightly interconnected and consequently subject to a high traffic variability. Therefore, traffic information is not widely available in this specific context, with only a few of the main transportation authorities of the most densely populated cities having the tools to monitor (a small subset of) urban roads. As an example, only cities of the magnitude of Los Angeles and Milan are currently provided with a pervasive monitoring infrastructure composed of induction loops and video cameras, which however do not cover the entire urban area [10].

But things may change as soon as wireless sensor technologies will enter the game and play a primary role. For example, recent market research forecasts that the 88% of PNDs mounted on vehicles in 2015 will integrate a GPS and a cellular connection [11]–[14], thus paving the way to the deployment of a mobile traffic sensing infrastructure comprised of vehicles which provide on-the-fly the time they spend to traverse a given portion of road. As a result, all this information sensed by a multitude of vehicles could be processed and put to good use to guide each single driver through the least congested path towards its destination.

It is hence not surprising that many researchers have recently shifted their attention to the design of mechanisms able to detect, as well as to forecast, the congestion state of any given segment of a road, even if not classifiable as a principal arterial way [15]–[18]. Most of such
schemes rely on the idea of collecting and processing the traversal time data from all the vehicles that pass through a given road. Obviously, vehicles should be equipped with a GPS receiver, a wireless communication interface and a software protocol needed to exchange data with a centralized entity. In turn, the centralized entity should process this data and subsequently distribute it to drivers, thus providing a useful aid for routing decisions.

The challenge, at this point, is to design a set of algorithms capable of detecting and forecasting traffic congestion, based on a pervasive traffic sensing infrastructure [16]. Obviously, the starting point of all these research initiatives is a good definition of what vehicular congestion is for any given road segment. Indeed, such definition has been available for long, and sounds as follows: congestion is the travel time or delay in excess of that normally incurred under light or free-flow travel conditions [19].

Although clear in theory, this definition has not found a successful algorithmic counterpart, as it does not provide an unambiguous method, independent of any parameter and applicable to any road, to find the traversal time value $T^*$, which distinguishes a congested from a non-congested road. To overcome this problem, we provide a brand new definition which identifies the state of congestion of a given road as a state that lasts for at least $S$ units of time and during which travel times or delays exceed the time $T^*$ normally incurred under light or free-flow travel conditions.

Following our definition, a road is congested when the traversal times of vehicles exceed $T^*$ and all subsequent vehicles which enter the road within a time $S$ keep exceeding the same value. The intuition behind this is given by observing that even when the input flow to a congested road suddenly drops, the inertia of the existing queue causes the road segment to be seen as congested also by those vehicles that enter the road within time $S$. From this observation we can logically draw that if a vehicle traverses a road segment when congested, a second vehicle will very likely experience a similar traversal time if it enters the road segment within $S$ units of time from the first vehicle. It is exactly this kind of consideration that allows one to understand how our congestion detection definition may also be used for estimating the duration of congestion states, thus providing a simple and effective tool useful for traffic forecasting.

From the above considerations, assuming $S$ known, an algorithm able to compute the congestion threshold $T^*$ for any given road is straightforward. The idea at the basis of such algorithm is as follows. Take a platoon of vehicles all entering a given road segment within $S$ units of time since the beginning of the platoon. We say that a road segment is in a congested state if
the number of pairs of subsequent vehicles, for both of which the traversal time exceeds $T^*$, is much greater than the number of pairs for which the traversal time of only the first vehicle of the pair exceeds $T^*$ (being this ratio $N : M$, for example). Conversely, a road segment is in a non-congested state if the amount of pairs of subsequent vehicles, for both of which the traversal time is below $T^*$, is much greater than the amount of pairs for which the traversal time of only the first vehicle of the pair is lower than $T^*$ (we can suppose this ratio is $K : H$). Assuming $N : M$, $K : H$ and $S$ known, it is as easy as pie to find $T^*$. The rationale is that we see congestion as that state where the number of vehicles in a platoon for which the traversal times are all stably high outnumber the number of vehicles for which their traversal time gracefully drop to a non-high value. Similarly, we deem a road segment as not congested when a much greater and steady number of vehicles with low travel times traverse that road with respect to the number of vehicles with high travel times.

The issue of quantifying the ratios N:M and H:K is crucial and should be left to the experience and sensibility of who is in charge of tuning our mechanism for managing traffic operations (typically the Traffic Operations Manager). However, a good choice can be based on the following consideration. A general reasoning can be conducted observing that human beings (and drivers as well) require a high reliability on the information they process to make their decisions. Specifically, drivers perceive a road as congested when a high rate of vehicles suffers from delays, which are above a congestion threshold. The point of how much high should be this rate is still open. Obviously, any value exceeding 60/70% matches that sense of “perceived congestion” drivers have in mind. Indeed, authors of [20] identify in the 80th percentile (more precisely the 80th-50th inter-quartile difference) a reasonable level of reliability, which drivers require on the traffic information they receive.

Now, we have not yet provided any precise recipe for determining $S$. It is worthwhile to mention that a reasonable assumption of $S$ is of great importance for our approach as it helps in determining the congestion threshold $T^*$. Not only, given the nature of our algorithm, $S$ gives us the means of providing a forecast of traffic. To this aim, we can consider a study from the American National Bureau of Transportation Statistics which can be of help, as it reveals that the average american driver spends 55 minutes a day behind the wheels [21]. Considering that the average daily traffic pattern for any driver is from home to work in the morning and back in the late afternoon, and assuming similar travel time values for both directions, we find that the
average one-way travel time equals 27.5 minutes. This indicates a plausible reference value for $S$, as an average driver will not spend more than such value stuck in traffic. As per a minimum value of $S$, a good choice could be that of 2 or 3 minutes; this being, on average, the traversal time of a typical urban road without traffic.

Once implemented, to confirm the validity of our traffic detection algorithm, we carried out real experiments amounting to 450 miles of travelled roads throughout different locations in the world. We report in this article results from those experiments, which have validated our approach.

To conclude this Section, we emphasize that our scheme differs from any other we are aware of on one important point: it computes a congestion threshold dynamically, as a function of the congestion duration $S$. We think that this design choice brings at least three advantages over all other alternative schemes:

- A road’s congestion threshold solely depends on probe vehicle traversal times, not requiring any other contextual information;
- Our scheme binds the detection of congestion (or no congestion) to the prediction of its duration, while other algorithms either ignore this relationship or separately address these two problems;
- Our algorithm takes into account that $S$ may change, because of sudden differences in a road’s capacity, traffic light cycles, or varying weather conditions, for example, and adapts accordingly.

The rest of this paper is organized as follows. In Section II we review the main schemes proposed as effective means to detect traffic congestion in urban roads. In Section III we provide both the intuition behind and the formal implementation of our algorithm. A description of the experimental scenario and with empirical results may be found in Section IV. We finally conclude with Section V.

II. RELATED WORK

A great amount of research has been carried out in the past years working on the problem of evaluating the performance of urban road networks by means of new congestion metrics. In most of such works, a congested state is distinguished from a non-congested state by analyzing the traversal times of the vehicles that flow through a given road segment and comparing them with
some fixed threshold [22]–[27]. In the following, we will describe how three relevant traffic detection algorithms work. The first two methods both rely on the use of probe vehicle data for the detection of congestion, while the third, is based on the Highway Capacity Manual delay formula for signalized intersections, and will serve, as we shall see in Section IV-A, as a benchmark for the validation of our results.

The methodology described in [23], is based on the use of fixed speed thresholds to determine whether a given road segment is congested or not. In such work, vehicle probes are periodically collected from a fleet of four thousand taxis operating in Shanghai and averaged out providing instantaneous traversal times and speeds at specific locations. In practice, traffic is classified according to the average speed experienced by a group of taxis that traverse a given road segment, as follows. If in urban contexts, taxis result moving at a speed that exceeds 30 km/h, hence traffic is classified as very smooth. If instead taxis’ speeds are between 25 and 30 km/h, traffic is smooth. Finally, if the average speed falls in one of the [16, 25), [11, 16) or [0, 11) km/h intervals, traffic is defined as medium, congested and very congested, respectively. Such methodology has these two important drawbacks. The first one amounts to the fact that an a-priori traffic classification based on predetermined values of speed is too rigid. In fact, a given road may exist where a speed of 20 km/h cannot be considered as a symptom of congestion simply because this is the maximum speed cars can reach, due, for example, to a specific traffic light cycle. The second problem is that this mechanism lacks the ability to predict how long a state of congestion will last.

The second method, termed SSTE (Surface Street Traffic Estimation), was specifically proposed to identify congestion on signalized road segments, which are road segments whose downstream intersection is managed by a traffic light [28]. Indeed, congested traversals are distinguished from non-congested traversals by analyzing the GPS traces collected by vehicles. Two different algorithms cooperate to this aim. The first one estimates the red light duration of a traffic light as the 95th percentile of the stopping duration of vehicles. The second one computes two thresholds. The first threshold is an average speed, computed as the road segment length divided by the sum of the 5th percentile of traversal times plus the red light duration. The second threshold is a space mean speed, computed as the 5th percentile of the spatial mean speed values that exceed the first threshold. While the meaning of the first threshold is clear, as a vehicle that experiences an average speed below this value traversed the road with a delay
that is above the free flow traversal time plus the red light duration, it is worth spending a few more words on the meaning of the second threshold. The space mean speed of a vehicle is the arithmetic mean of the instantaneous speed samples taken at fixed locations. Therefore, this second threshold differentiates the values given by those vehicles that traverse a road with a *stop and go* pattern, from the values of those vehicles that, instead, smoothly flow through the road. Summarizing, vehicles that exceed both thresholds are classified in a free flow state, whereas vehicles that fall below both thresholds are classified as congested. While this strategy is clear, as it identifies congestion as queueing, this algorithm falls short in two main aspects. The first is that this method does not provide any forecasting information, thus resulting questionable as to its utility. The second is that this method focuses on segments with signalized intersections, being not clear how it may be extended to more general cases.

Finally, an approach often used to distinguish congested from non-congested states is based on the HCM delay formula for signalized intersections \[8\], \[29\]–[31]. This formula computes the average traversal time \(\bar{T}_{HCM}\) of a vehicle as the sum of three values. The first, \(d_0\), is the average traversal time per vehicle in free flow conditions. The second, \(d_1\), is equal to the additional average delay per vehicle due to traffic light phases. Finally, \(d_2\), amounts to an additional average delay experienced by a vehicle because of congestion. While \(d_0\) can be simply obtained dividing the length of a road segment by its speed limit, \(d_1\) and \(d_2\) are functions of: the capacity of the road (determined by the number of lanes and the length of traffic light phases), the average amount of vehicles entering the road within a given time, and the expected duration of the given analysis conditions. In summary, the \(\bar{T}_{HCM}\) value may be computed on a per vehicle basis assuming that the input traffic volume of a road matches its capacity and that the analysis period is long enough. This value is finally exploited to differentiate congested from non-congested portions of roads. While this approach exploits the duration of the analysis with the aim of capturing the future state of a road segment, it has, indeed, a limit given by its inability to adapt to the capacity fluctuations of a road segment (being the capacity of a street subject to a number of modifications caused by obstructing vehicles and maintenance works, for example). Moreover, accurate results require accurate estimates of the parameters that influence the capacity of a road, which are not easy to obtain on a large-scale basis.

Summarizing, the above approaches either: (a) do not couple the congestion identification and forecasting problem as one, thus risking to identify as congested roads that will not last in that
state any longer in time, or (b) excessively rest upon statically chosen parameters, jeopardizing their adaptability to new and different road settings. We will show that our approach is successful in overcoming both of the previous problems, as it does not need the a-priori setting of any parameter. Hence, we are confident that it can represent a good candidate for traffic estimation and forecasting in pervasive urban traffic scenarios.

III. A NOVEL CONGESTION DETECTION MODEL

Before proceeding with a detailed explanation of our congestion detection algorithm, we briefly account for the general context where it should be exploited. In particular, we assume vehicles mounting an advanced PND integrated with a GPS receiver and a full-duplex communication device.

First of all, any given segment of a road must be put under observation for a duration of circa half a day/a day, as the idea is that, upon completion of that given road segment traversal, any vehicle sends to a centralized entity a message that includes an identifier of that road segment, its entry and exit times (i.e., the traversal time). As soon as the centralized entity has completed the observation activity and has collected a sufficient number of samples from probing vehicles, it has a clear picture of the congestion states characterizing that given segment of a road, as a function of the value of a given congestion threshold $T^*$ and of the value of the congestion duration $S$.

Afterwards (ten minutes later, or even two days after), another vehicle will traverse that road experiencing a travel time equal to $T$ and will transmit this information to the centralized entity. Depending on if $T > T^*$ or $T \leq T^*$, the centralized entity will inform all the subsequent vehicles that plan to traverse that road that they will incur, with a given likelihood (typically 80%), in a congested or not congested state of duration $S$.

Obviously, the initial observation phase required to tune the system may be performed only once in a given period, or repeated with a frequency to be obtained by the transportation authority on a per road basis.

Let us explain now how the algorithm needed to detect and forecast congestion works.

A. Congestion Detection and Forecasting Algorithm

We first begin by recalling the general definition given in Section I which is as follows:
Definition 3.1. A road segment $R$ is in a congested state if travel times or delays of traveling vehicles exceed the time $T^*$ normally incurred under light or free-flow travel conditions, and this congested state lasts for at least $S$ units of time.

Owing to this generic definition, it is possible to infer a further set of operational definitions from which a simple algorithm can be derived to identify when a road segment is in a state of high or low congestion. Let us start with a couple of definitions aimed at identifying two different types of sets of vehicles entering a road segment under diverse congestion conditions.

Definition 3.2 (High Congestion Vehicle Set). Take a platoon $P$ of vehicles entering a road segment $R$, with the first vehicle of the platoon entering $R$ at time $t_0$ and the last one entering $R$ no later than time $t_0 + S$. Now, $\text{HighCongestion}_{T^*}$ is defined as the set of all the pairs of vehicles, $(i, j)$ (with $i$ entering $R$ before $j$), of this platoon for which their traversal times, say $T^*_i$ and $T^*_j$, exceed both the congestion threshold $T^*_1$ (i.e., $(T^*_i > T^*_1) \land (T^*_j > T^*_1)$). We also define as $\text{Noise}_{1T^*}$ the set of all the pairs of vehicles, say $h, k$, for which the traversal time $T^*_h$ of only the first vehicle $h$ exceeds $T^*_1$ (i.e., $(T^*_h > T^*_1) \land (T^*_k \leq T^*_1)$).

In essence, $\text{HighCongestion}_{T^*}$ represents the amount of all those vehicles suffering from a stable situation of congestion. Indeed, all their traversal times lie above the congestion threshold $T^*_1$. Instead, $\text{Noise}_{1T^*}$ represents the set of those vehicles, a part of which are leaving the congestion state. In fact, the traversal times of those vehicles at the end of the platoon lie below the congestion threshold $T^*_1$.

Definition 3.3 (Low Congestion Vehicle Set). Take the same platoon $P$ of vehicles subject to the same conditions as before, $\text{NoCongestion}_{T^*}$ is defined to be the set of all the pairs of vehicles, say $(i, j)$ (with $i$ entering $R$ before $j$), of this platoon for which their traversal times, say $T^*_i$ and $T^*_j$, are both below the congestion threshold $T^*_2$ (i.e., $(T^*_i < T^*_2) \land (T^*_j < T^*_2)$). We also define as $\text{Noise}_{2T^*}$ the set of all the pairs of vehicles, say $h, k$, for which the traversal time $T^*_h$ of only the first vehicle $h$ is below $T^*_2$ (i.e., $(T^*_h < T^*_2) \land (T^*_k \geq T^*_2)$).

Similarly as before, $\text{NoCongestion}_{T^*}$ groups all the cars incurring in a stable situation of no congestion (i.e., all their traversal times are below $T^*_2$). In this case, instead, $\text{Noise}_{2T^*}$ represents a situation where the last cars in the platoon are entering in a new congestion state, as their
traversal times exceed $T_2^*$. 

Given the above sets, we now introduce the indicator functions of those sets, with the aim of providing a tool with which the number of vehicles of a platoon $P$ entering a road segment $R$ within a time period $S$, can be counted under different congestion conditions.

**Definition 3.4 (High Congestion State).** Let $1_{HighCongestion_{T_1^*}} : (P \times P) \rightarrow \{0, 1\}$ be defined as:

$$1_{HighCongestion_{T_1^*}}((i, j)) = \begin{cases} 
1 & (i, j) \in HighCongestion_{T_1^*} \\
0 & (i, j) \notin HighCongestion_{T_1^*}.
\end{cases}$$

Let $1_{Noise_{T_1^*}} : (P \times P) \rightarrow \{0, 1\}$ be defined as:

$$1_{Noise_{T_1^*}}((h, k)) = \begin{cases} 
1 & (h, k) \in Noise_{T_1^*} \\
0 & (h, k) \notin Noise_{T_1^*}.
\end{cases}$$

Similarly, for non-congested states:

**Definition 3.5 (No Congestion State).** Let $1_{NoCongestion_{T_2^*}} : (P \times P) \rightarrow \{0, 1\}$ be defined as:

$$1_{NoCongestion_{T_2^*}}((i, j)) = \begin{cases} 
1 & (i, j) \in NoCongestion_{T_2^*} \\
0 & (i, j) \notin NoCongestion_{T_2^*}.
\end{cases}$$

Let $1_{Noise_{T_2^*}} : (P \times P) \rightarrow \{0, 1\}$ be defined as:

$$1_{Noise_{T_2^*}}((h, k)) = \begin{cases} 
1 & (h, k) \in Noise_{T_2^*} \\
0 & (h, k) \notin Noise_{T_2^*}.
\end{cases}$$

At this point, we are able to count: a) the number of pairs of vehicles which suffer of high congestion versus the number of pairs of vehicles which are leaving a congested situation, and b) the number of pairs of vehicles which do not suffer of congestion versus the number of pairs of vehicles which are entering in a congestion state. This allows us to provide the following proposition on which our congestion detection algorithm is based:
Proposition 3.1: (Congestion). A given road segment R is congested during a period S if the following holds:

\[
\frac{\sum_{(i,j) \in P \times P} 1_{\text{HighCongestion}_T^1}(i, j)}{\sum_{(i,j) \in P \times P} 1_{\text{Noise1}_T^1}(i, j)} \geq \frac{N}{M},
\]

with \(N/M = 80\%/20\%\), as discussed in Section I [20].

The same can be drawn for a non-congested state, as follows:

Proposition 3.2: (No Congestion). A given road segment R is not congested during a period S if the following holds:

\[
\frac{\sum_{(i,j) \in P \times P} 1_{\text{NoCongestion}_T^2}(i, j)}{\sum_{(i,j) \in P \times P} 1_{\text{Noise2}_T^2}(i, j)} \geq \frac{H}{K},
\]

again, with \(H/K = 80\%/20\%\) [20].

The rationale of Proposition 3.1 is simply that if the number of cars in a stable situation of congestion largely exceeds the number of cars that are leaving a congestion state, then that state can be confirmed as a congested state. Conversely, Proposition 3.2 states that if the number of cars suffering of no congestion outnumbers the set of those entering in a congested situation, then that state can be confirmed as a non-congested state.

B. Efficient Implementation

Up to this point, we have devised a mathematical model which can be used to distinguish a congested state of a road segment \(R\) from a non-congested one, based on the computation of two congestion thresholds, namely \(T_1^*\) and \(T_2^*\). Before explaining how to compute the values of \(T_1^*\) and \(T_2^*\), we have to anticipate here that it is not only a matter of the choice of an adequate mathematical model, but a matter of its application to only those real cases where it can return an effective solution that meets drivers’ needs, in terms of knowledge about the fact if a given road is congested or not. We here appeal to a principle of reality, which says that it makes sense to apply our algorithm to all those roads where congested and non-congested states alternate, while it does not make sense to apply it when only one of the two mentioned states can be found on a road at any given time, because in such case there is no reason for using any existing model. In this latter case we would find paradoxical \(T_1^*\) and \(T_2^*\) values. Take for example this
case showing when our algorithm should NOT be used. Consider a road segment \( R \) for which we know for sure it is congested in a given period \( S \), with a certain platoon \( P \) of vehicles running on it. Take now Proposition 3.1 and apply it to the traversal times returned by those vehicles. Obviously, this proposition will be able to find a \( T_1^* \) value which is exceeded by the traversal times of all the vehicles in \( P \). Take then Proposition 3.2 and apply it to the set of the same traversal times as before. Obviously, again, a new value \( T_2^* \) can be obtained exceeding all those traversal times. In this way, we would have obtained the paradox of having \( T_2^* > T_1^* \) without any possibility of understanding if our road is in a congested state or not.

Indeed, this is not a problem of our algorithm, but wrong was its application. Instead, the right way to apply our procedure is that of working on a set of data which, for a given road \( R \), be sampled both from states of congestion and from states of no congestion. From an operational viewpoint, this means that, for our algorithm to work correctly, a given road must be put under observation for a period whose duration is long enough to capture both congested and non-congested traffic situations. Hence, if the traversal times are sampled correctly, applying both Propositions 3.1 and 3.2 returns two threshold values \( T_1^* \) and \( T_2^* \) ordered according to their natural way, that is \( T_2^* \leq T_1^* \).

Based on the above considerations, an efficient way to implement the statements of both Propositions 3.1 and 3.2 goes through two different steps. The first step amounts to searching the pair \((\bar{T}_1^*, \bar{T}_2^*)\) that maximizes: a) the number of pairs of vehicles whose traversal times are both larger than \( \bar{T}_1^* \) (congestion), b) the number of pairs of vehicles whose traversal times are both below \( \bar{T}_2^* \) (no congestion). Simultaneously, of minimal size should be: c) the amount of pairs of vehicles for which the traversal times of only the latter vehicle is smaller than \( \bar{T}_1^* \), d) the amount of pairs of vehicles for which the traversal time of only the latter vehicle is larger
than $\bar{T}^*_2$ (noisy conditions). All this can be obtained with the following formula:

$$\left( \bar{T}^*_1, \bar{T}^*_2 \right) = (T^*_1, T^*_2) \text{ s.t.}$$

$$\left\{ \max_{T^*_1, T^*_2} \sum_{(i,j) \in P \times P} \left[1_{\text{HighCongestion}_{T^*_1}}(i,j) + 1_{\text{NoCongestion}_{T^*_2}}(i,j) - 1_{\text{Noise1}_{T^*_1}}(i,j) - 1_{\text{Noise2}_{T^*_2}}(i,j) \right] \right\} = |\Delta|,$$

(3)

where $\Delta = \sum_{(l,m) \in P \times P} 1_{\text{HighCongestion}_{T^*_1}}(l,m) - \sum_{(s,t) \in P \times P} 1_{\text{NoCongestion}_{T^*_2}}(s,t)$ is a term accounting for a few noisy traversal time values which could have a negative effect on the computation of the congestion threshold. This is the problem of a situation where two well-defined and different clusters of traversal time values coexist with a few isolated samples (either very high or very low), which, in turn may have the effect of shifting the value of the congestion threshold. $\Delta$, therefore, ensures that the two clusters of samples contain the maximum number of points each by minimizing the difference between their sizes. In other words, subtracting $\Delta$ guarantees that if, for example, an isolated point lies along the x-y bisector right below (or also right above) all the other points plotted on a congestion graph, our mechanism returns a solution where the $\bar{T}^*_1$ and $\bar{T}^*_2$ values simply separate the two clusters, while it is excluded the possibility that a solution exists where $\bar{T}^*_1$ and $\bar{T}^*_2$ separate the isolated points from the union set that contains the two clusters.

The second step amounts to taking the just computed $\left( \bar{T}^*_1, \bar{T}^*_2 \right)$ values, respectively replacing them in Equations 1 and 2 and finally checking if the inequalities are satisfied. The motivation behind the execution of this second step is that Step 1 could complete giving us a percentage of pairs of vehicles in the state of congestion equal to $N$, with $N < 80\%$, thus resulting in a percentage of pairs of vehicles in a noisy situation above the level of 20% (we name this kind of check Check1($\bar{T}^*_1$)). A similar situation could occur also with the percentage of pairs of vehicles in a non-congested state, which could be less than 80% (we name this kind of check Check2($\bar{T}^*_1$)).
Unfortunately, a reason for the checks to fail could be that of having chosen a too large duration $S$ for the state of congestion of interest. This would mean that for many pairs of subsequent cars the following holds: the congested (or non-congested) state a first vehicle incurs in does not last in time, as a second vehicle does not find the same state any longer. However, this could be a problem simply concerned with the duration of the $S$ we have chosen, while a smaller value for $S$ could exist, in principle, for which both the subsequent cars incur in the same state of congestion. The idea is hence that of looking for such value, by reducing $S$ until a situation is captured where both the subsequent vehicles of the pair experience a similar state of congestion (or no congestion). Reminding the reference value of 27.5 minutes as discussed in Section I, our algorithm starts its search from $S = 27.5 \times 3 = 82.5$ minutes and completes when it reaches the final value of 2 minutes. Obviously, if the search completes without the possibility of identifying any congestion or non-congestion states, whatever is the value of the chosen $S$, this means that for that given road segment it is not possible to distinguish any congestion state of interest. In such particular case our algorithm completes by returning an adequate alert message.

To conclude, we sketch in Table I below the main phases of the algorithm we have so far discussed. It shows an algorithm which, after some iterations on $S$ values, finds the congestion thresholds $\bar{T}^*_1$ and $\bar{T}^*_2$ of a given road segment (provided that both these thresholds exist).

### Table I

**Congestion Threshold Detection Algorithm**

| input: A list of traversal times. |
| output: $S, \bar{T}^*_1$ and $\bar{T}^*_2$. |

$S \leftarrow 82.5$ minutes;

$(\bar{T}^*_1, \bar{T}^*_2) \leftarrow T^*$ s.t. $\text{Max}(T^*_1, T^*_2)$;

**while** $\neg \text{Check1}(T^*_1) \land \neg \text{Check2}(T^*_2) \land S > 2$ **do**

$S \leftarrow S - 1$ minutes;

$(\bar{T}^*_1, \bar{T}^*_2) \leftarrow \text{Max}(T^*_1, T^*_2)$;

**end**

C. Representing Congestion with Graphs

We now proceed showing how the results of our algorithm can be graphically shown. To do so, we define what a congestion graph is.
Simply, a congestion graph of a road segment $R$ is a scatter graph of points $(x = T_i, y = T_j)$, where the $x$ and $y$ axis values of each point represent the traversal time of a pair of subsequent vehicles. In particular, for each pair of vehicles the $x$ value of a point on the graph is equal to the traversal time of the first vehicle of the pair that entered $R$, whereas the $y$ value equals the traversal time of the second vehicle that entered $R$ within $S$ seconds later. Each congestion graph can be indexed with a value of $S$, $S$ being the maximum difference in time between the moment when the first vehicle of a pair entered $R$ and the moment when the second vehicle of the same pair entered $R$, for any given pair of vehicles on that graph.

As significant examples, consider the leftmost, rightmost and bottom graphs in Figure 1 representing three different congestion graphs for three different $S$ values. The leftmost graph has been filled with data coming from vehicles running on a road segment which either suffers from congestion or not, depending on the specific moment of the day (in particular that road was under observation for seven hours). Within any given $S$ there was a traversal of only two cars. As expected, running our algorithm on the data shown on this graph returns the values of $S$, $\bar{T}_1^*$, $\bar{T}_2^*$, $N$, $M$, $H$ and $K$ respectively as follows: $\bar{T}_1^* = 93$ seconds, $\bar{T}_2^* = 89$ seconds, $N = 92\%$, $M = 8\%$, $H = 84\%$ and $K = 16\%$. Indeed, the $\bar{T}_1^*$ and $\bar{T}_2^*$ thresholds are plotted on the graph by means of two different couples of intersecting lines.

An interesting, even if expected, phenomenon, shown by this example, is that if $\bar{T}_1^*$ and $\bar{T}_2^*$ exist, as returned by our algorithm, they take very close values (often almost coincident). This is a natural consequence of the physical reality where in practice only one congestion threshold exists, above which we have congestion and below which we do not. Further, in all the experiments we have carried out (and discussed in the following Section IV) the value $\delta = [(\bar{T}_1^* - \bar{T}_2^*)/\bar{T}_1^*] \times 100\%$ was always smaller than $3\%$. For this reason, with the aim of simplifying this matter, in the remainder of the paper we will exploit the value $\bar{T}_1^*$ as a unique representative of the congestion threshold of a given road.

The congestion graph at the right of Figure 1 represents instead a clear situation where no reasonable values for $\bar{T}_1^*$ and $\bar{T}_2^*$ can be found. This is not surprising as the data for the traversal times plotted in this graph were sampled with a value of $S = 3$ hours for a road where congestion states did last always less than 30 minutes. Clearly, if we take such a huge value for $S$ in this situation, we are making the mistake of establishing a correlation between two cars which traversed the same road segment in very different moments, subject to very different states of November 16, 2010 DRAFT
Fig. 1. Left: Congestion graph with Equations 1 and 2 satisfied and $S = 6$ minutes. Right: Congestion graph with $S = 3$ hours, Equations 1 and 2 not satisfied. Bottom: Congestion graph with $S = 5$ seconds, no interesting situation.

congestion.

As a final and interesting example, consider the case of the bottom graph in Figure 1. There was sampled data with an extremely small value for $S$ ($S = 5$ seconds). If we tried to apply the conditions expressed in Equations 1 and 2 to this graph we would find values for $\bar{T}_1^*$ and $\bar{T}_2^*$ which could apparently look like reasonable. Again, this would be a mistake as the problem is that there is no interest in identifying congestion when its duration is 5 seconds. Indeed, there is
no reasonable congestion state that lasts so shortly. This is why our algorithm has a lower limit on $S = 2$ minutes.

Obviously, with the cases of the rightmost and the bottom graphs of Figure 1 we wanted to represent the abnormal situations we would get if we used wrong values for $S$. Specifically, if we chose a too large value of $S$ we would have obtained sparse points in the graph, accounting for a situation where no relationship between vehicle pairs exist. Instead, if $S$ was chosen too low, we would have obtained points concentrated along the x-y bisector, accounting for a situation where vehicles are too close to each other to be useful for any kind of decision.

IV. EXPERIMENTAL ASSESSMENT

We carried out a set of nine different experiments to verify the validity of our congestion detection and forecasting algorithm discussed in the previous Section. Each of these experiments was conducted by managing the vehicular data sampled and transmitted by a set of cars driven over a real section of road. With the term section we both indicate a single road segment and two or more adjacent segments separated by one or more intersections. Eight out of nine sections of roads taken into consideration were in Los Angeles, CA, while one was in Pisa, Italy. All the information concerning these roads is listed in Table II, where the name of the road, the section of the road under analysis, its length, its traversal time under free flow conditions, its entire traffic light cycle time and, finally, the green time duration of the cycle are provided.

<table>
<thead>
<tr>
<th>Road</th>
<th>Section, Direction</th>
<th>Road section length [m]</th>
<th>$T_{FFTT}$ [s]</th>
<th>CT [s]</th>
<th>GT [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Via B. Croce</td>
<td>P.zza Guerrazzi-Via Queirolo, left</td>
<td>380</td>
<td>34</td>
<td>85</td>
<td>55</td>
</tr>
<tr>
<td>2 S. Monica Blvd</td>
<td>Veteran-Sepulveda, left</td>
<td>380</td>
<td>61</td>
<td>120</td>
<td>15</td>
</tr>
<tr>
<td>3 S. Monica Blvd</td>
<td>Wilshire-Roxbury, straight</td>
<td>280</td>
<td>17</td>
<td>90</td>
<td>54</td>
</tr>
<tr>
<td>4 S. Monica Blvd</td>
<td>Wilshire-Bedford, right</td>
<td>390</td>
<td>30</td>
<td>90</td>
<td>54</td>
</tr>
<tr>
<td>5 Lincoln Blvd</td>
<td>Fiji-Venice, back</td>
<td>2300</td>
<td>205</td>
<td>120</td>
<td>60</td>
</tr>
<tr>
<td>6 Wilshire Blvd</td>
<td>Midvale-Westwood, right</td>
<td>130</td>
<td>7</td>
<td>150</td>
<td>80</td>
</tr>
<tr>
<td>7 S. Monica Blvd</td>
<td>Roxbury-Bedford, right</td>
<td>100</td>
<td>7</td>
<td>90</td>
<td>54</td>
</tr>
<tr>
<td>8 Wilshire Blvd</td>
<td>Veteran-Westwood, right</td>
<td>340</td>
<td>33</td>
<td>150</td>
<td>80</td>
</tr>
<tr>
<td>9 S. Monica Blvd</td>
<td>Westwood-Sepulveda, right</td>
<td>680</td>
<td>75</td>
<td>120</td>
<td>50</td>
</tr>
</tbody>
</table>
Of a certain importance is also the consideration that all the examined roads in Los Angeles are wide and provided with an advanced ATMS infrastructure, while the road in Pisa is narrow and crowded with both pedestrian and vehicular traffic. The motivation behind the choice of these two cities is that they represent two very different traffic situations, both from a traffic management and driving style standpoint.

To collect data, each vehicle was equipped with an onboard system comprised of a laptop with a GPS and an EVDO interface used to store a digital map of the area under analysis. As discussed in the previous Section, upon traversal of a given road section $R$ a car transmitted its traversal time to a centralized entity. As soon as a sufficient amount of data required to distinguish congestion from non congestion on $R$ was available (typically after a dozen hours during the daytime), our system computed an estimate of $T^*$. 

Road section traversal times were sampled performing loops on tracks over those roads. Tracks were in general comprised of a first part of the road section chosen as it presented a high varying traffic pattern plus a second part with little or no traffic. The rationale underlying this choice was to be able to perform subsequent observations of a given road section as close in time so as to exploit the same set of cars. In Pisa, for example, we chose a track that included Via Bonaini, Via B. Croce, P.zza Guerrazzi and Via Gian Battista Queirolo, as Via B. Croce could become very crowded, while the other road sections in the track rarely experienced intense vehicular flows (the track is highlighted in Figure 2).

![Experiment site map in Pisa.](image-url)
In the following subsection we are going to present the results we have obtained from our experiments.

A. Results

Our results are shown in Table III where for each road are respectively listed the number of loops on tracks, the congestion threshold $\bar{T}^*$ computed using our mechanism, the duration of the congestion $S$, our measure $N$ of how many pairs of subsequent vehicles suffer of a stable congestion situation and, the measure $H$ of how many pairs of subsequent vehicles experience no congestion. Further, to verify the validity of the results we have obtained, we have computed for each examined road section the value $\hat{T} = T_{FFT} + (CT - GT)$ based on the methodology proposed in [28]. $\hat{T}$ amounts to the time a car would spend traversing that road, in the case it does not incur in a traffic congestion situation, but that of waiting for the traffic light situated at the end of the intersection to become green. Cars are assumed to stop in queue for an entire red light time (this resembles a typical situation where drivers enjoy an experience of a very moderate congestion).

<table>
<thead>
<tr>
<th>Road</th>
<th>Section, Direction</th>
<th># of loops</th>
<th>$\bar{T}^*$ [s]</th>
<th>$S$ [s]</th>
<th>N</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Via B. Croce</td>
<td>P.zza Guerrazzi-Via Queirolo, left</td>
<td>111</td>
<td>93</td>
<td>362</td>
<td>92%</td>
<td>84%</td>
</tr>
<tr>
<td>S. Monica Blvd</td>
<td>Veteran-Sepulveda, left</td>
<td>134</td>
<td>175</td>
<td>608</td>
<td>80%</td>
<td>87%</td>
</tr>
<tr>
<td>S. Monica Blvd</td>
<td>Wilshire-Roxbury, straight</td>
<td>77</td>
<td>62</td>
<td>987</td>
<td>94%</td>
<td>99%</td>
</tr>
<tr>
<td>S. Monica Blvd</td>
<td>Wilshire-Bedford, right</td>
<td>77</td>
<td>82</td>
<td>987</td>
<td>92%</td>
<td>99%</td>
</tr>
<tr>
<td>Lincoln Blvd</td>
<td>Fiji-Venice, back</td>
<td>30</td>
<td>354</td>
<td>900</td>
<td>100%</td>
<td>97%</td>
</tr>
<tr>
<td>Wilshire Blvd</td>
<td>Midvale-Westwood, right</td>
<td>71</td>
<td>36</td>
<td>454</td>
<td>39%</td>
<td>98%</td>
</tr>
<tr>
<td>S. Monica Blvd</td>
<td>Roxbury-Bedford, right</td>
<td>77</td>
<td>42</td>
<td>987</td>
<td>46%</td>
<td>83%</td>
</tr>
<tr>
<td>Wilshire Blvd</td>
<td>Veteran-Westwood, right</td>
<td>71</td>
<td>74</td>
<td>454</td>
<td>37%</td>
<td>100%</td>
</tr>
<tr>
<td>S. Monica Blvd</td>
<td>Westwood-Sepulveda, right</td>
<td>67</td>
<td>121</td>
<td>493</td>
<td>90%</td>
<td>54%</td>
</tr>
</tbody>
</table>

Values for $\hat{T}$ have been inserted in Table IV, as they represent a figure of merit against which our obtained $\bar{T}^*$ results must be contrasted.

Examining the results in Table IV, we can draw the following considerations. First, roads from 1 to 5 were all roads where situations of high traffic congestion alternated with situations...
of no congestion. This is revealed by the values of $N$ and $H$ that were both above the threshold of 80%. The validity of these results is also witnessed by the value of the congestion threshold $\hat{T}^*$ we derived, which was always larger than the value of $\hat{T}$, thus confirming that our algorithm was able to find a congestion threshold value above which cars really incurred in congestion. Indeed, the alternating between states of high congestion and no congestion revealed by our algorithm matches with the empirical knowledge of the traffic situation over those roads.

### TABLE IV
**Comparison: $\hat{T}$ and $\hat{T}^*$**

<table>
<thead>
<tr>
<th>Road</th>
<th>Section, Direction</th>
<th>$\hat{T}$ [s]</th>
<th>$\hat{T}^*$ [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Via B. Croce</td>
<td>P.zza Guerrazzi-Via Queirolo, left</td>
<td>64</td>
<td>93</td>
</tr>
<tr>
<td>S. Monica Blvd</td>
<td>Veteran-Sepulveda, left</td>
<td>166</td>
<td>175</td>
</tr>
<tr>
<td>S. Monica Blvd</td>
<td>Wilshire-Roxbury, straight</td>
<td>53</td>
<td>62</td>
</tr>
<tr>
<td>S. Monica Blvd</td>
<td>Wilshire-Bedford, right</td>
<td>63</td>
<td>82</td>
</tr>
<tr>
<td>Lincoln Blvd</td>
<td>Fiji-Venice, back</td>
<td>265</td>
<td>354</td>
</tr>
<tr>
<td>Wilshire Blvd</td>
<td>Midvale-Westwood, right</td>
<td>77</td>
<td>36</td>
</tr>
<tr>
<td>S. Monica Blvd</td>
<td>Roxbury-Bedford, right</td>
<td>43</td>
<td>42</td>
</tr>
<tr>
<td>Wilshire Blvd</td>
<td>Veteran-Westwood, right</td>
<td>103</td>
<td>74</td>
</tr>
<tr>
<td>S. Monica Blvd</td>
<td>Westwood-Sepulveda, right</td>
<td>145</td>
<td>121</td>
</tr>
</tbody>
</table>

Of particular interest are also roads from 6 to 8. We deliberately chose these road sections as they are well known to be roads which almost never experience states of congestion. Our results confirm this fact in two different ways. First, for each of these roads the value $\hat{T}$ is always larger than that provided by our algorithm, thus confirming that cars over these roads enjoy a smooth drive, rarely incurring in a red light, due to the existence of a green wave. Second, the very small values we obtained for $N$ (corresponding to failures of the Check1 procedure in our algorithm) further corroborate the fact that almost no congestion is visible over those roads.

Finally, a very specific case is given by road # 9. This road section, in fact, would seem to reveal a stable high congestion state as resulting from a high value of $N$ and a small value of $H$. However, the obtained $\hat{T}^*$ value is smaller than the value of $\hat{T}$, thus providing a contrasting argument against our initial statement. The motivation for this paradox is as follows. Indeed, this road section often experiences severe congestion on S. Monica Blvd.. Nonetheless, the fleet of cars we used for our experimentation was instructed to take a right turn on Sepulved Blvd. to
maintain the route on the predetermined tracks. Hence, as this traffic light permits to turn right on red, only very seldom our cars incurred in the delay given by a red light time. This motivates the low value of $\bar{T}^*$, especially in comparison with that of $\hat{T}$. To overcome this, we carried out a couple of additional experiments with cars going straight at that intersection. As expected in this case, the value of $\bar{T}^*$ always surpassed that of $\hat{T}$.

Moreover, in addition of what just discussed, we regarded as important to further extend our experimentation to include a comparison of our results with those that can be obtained exploiting the scheme proposed in [23], where samples coming from probe vehicles are contrasted with a set of static thresholds to determine the existence of congestion. While authors of [23] defined four different thresholds to distinguish among five different levels of congestion, we are simply interested in differentiating between only two situations: the presence of congestion (including medium, congested and very congested states in [23]) or the absence of congestion (including smooth and very smooth states in [23]). Hence, as the authors distinguish between these two situations with a 25 km/h threshold on roads where the enforced speed limit is 40 km/h, we adapted such value to the roads we drove on, maintaining the same threshold value for the experiment carried out on the narrow road in Pisa, while we increased to 28 km/h the congestion threshold for the experiments conducted on the wide Los Angeles streets. At this point we converted these speed thresholds into traversal time thresholds $\bar{T}^*$ and inserted those values in Table V. Table V also reports the percentages of traversals classified as congested ($C_{\bar{T}^*}$) and those classified as non-congested ($NC_{\bar{T}^*}$), computed based on the scheme of [23].

Examining $C_{\bar{T}^*}$ and $NC_{\bar{T}^*}$, it results evident how the method described in [23] has the tendency to overestimate the situations of congestion, while states of no congestion are underestimated. This bad attitude is confirmed not only by contrasting the values reported in Table III with those obtained with our method in Table V, but also by resorting to the empirical knowledge of the traffic situation over those roads.

As a comparison with [23] cannot be considered sufficient to validate our method because of the tendency of [23] to overestimate congestion, we carried out an additional and final comparison. In particular, we contrasted the values of $\bar{T}^*$ with those provided by the HCM method for each of the road sections under analysis. As reported in Section II, the HCM method exploits the value $T_{HCM}$ that should be interpreted as the average traversal time cars experience when driving on a given road, as a function of the intensity of traffic and of the peak capacity of
TABLE V
ROAD DATA: NUMBER OF LOOPS, CONGESTION THRESHOLD $\bar{T}$, $C_{\bar{T}}$ AND $NC_{\bar{T}}$.

<table>
<thead>
<tr>
<th>Road</th>
<th>Section, Direction</th>
<th>$\bar{T}$ [s]</th>
<th>$C_{\bar{T}}$</th>
<th>$NC_{\bar{T}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Via B. Croce</td>
<td>P.za Guerrazzi-Via Queirolo, left</td>
<td>55</td>
<td>76%</td>
<td>24%</td>
</tr>
<tr>
<td>2 S. Monica Blvd</td>
<td>Veteran-Sepulveda, left</td>
<td>49</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>3 S. Monica Blvd</td>
<td>Wilshire-Roxbury, straight</td>
<td>36</td>
<td>70%</td>
<td>30%</td>
</tr>
<tr>
<td>4 S. Monica Blvd</td>
<td>Wilshire-Bedford, right</td>
<td>50</td>
<td>75%</td>
<td>25%</td>
</tr>
<tr>
<td>5 Lincoln Blvd</td>
<td>Fiji-Venice, back</td>
<td>296</td>
<td>47%</td>
<td>53%</td>
</tr>
<tr>
<td>6 Wilshire Blvd</td>
<td>Midvale-Westwood, right</td>
<td>17</td>
<td>61%</td>
<td>39%</td>
</tr>
<tr>
<td>7 S. Monica Blvd</td>
<td>Roxbury-Bedford, right</td>
<td>13</td>
<td>71%</td>
<td>29%</td>
</tr>
<tr>
<td>8 Wilshire Blvd</td>
<td>Veteran-Westwood, right</td>
<td>44</td>
<td>77%</td>
<td>26%</td>
</tr>
<tr>
<td>9 S. Monica Blvd</td>
<td>Westwood-Sepulveda, right</td>
<td>87</td>
<td>84%</td>
<td>16%</td>
</tr>
</tbody>
</table>

that road [30]. We provide in Table VI a comparison between $\bar{T}^*$ and $\bar{T}_{HCM}$. As it can be seen from the rightmost column of Table VI, the two parameters give almost converging results. In particular, for roads # 5, 6, 8 and 9 the matching between $\bar{T}^*$ and HCM is nearly perfect. This is due to the fact that the Los Angeles Transportation Authority has provided very accurate and updated estimates for the peak capacities of those road sections to be used in the HCM formula [10]. Instead, the value of the peak capacities for roads # 1, 2, 3, 4 and 7 to be used in the HCM method were less precise as simply drawn from the HCM general manual. This motivates why the difference between $\bar{T}^*$ and the $\bar{T}_{HCM}$ in percentage may exceed 20% in these particular cases.

B. A Study of Two Specific Cases

To provide further evidence of the validity of our approach, we here discuss in more detail a set of specific cases drawn from the two tracks shown in Figures 2 (traversing Via B. Croce, Pisa, in counter-clockwise direction) and 3 (traversing N. S. Monica Blvd., Los Angeles, in clockwise direction), respectively. In particular, we studied the cases and contrasted the results coming from a comparison between road # 1 and road # 4, as well as those coming from a comparison between road # 3 and # 7. The motivation for choosing these cases is as follows.

Roads #1 and #4 were chosen as they exhibited similarities (e.g., length, traffic light phase and speed limit), but differed in terms of the number of traffic lights (two on the S. Monica
Blvd. and one at the end of the via B. Croce), lanes reserved for through traffic (two for S. Monica Blvd. and one for via B. Croce) and driving discipline (although we cannot provide any supporting data, our drivers reported that traffic discipline was more strictly adhered to in Los Angeles). The purpose of this first comparison is to argue how these characteristics influenced the value of $S$ on both roads.

The case of roads # 3 and # 7 was taken into consideration as these two sections are different but consecutive parts of the same street. Nonetheless, as discussed previously, our algorithm was able to distinguish congestion from non-congestion on road # 3, while it was unable to perform as well on road # 7. Our aim, hence, is to better clarify such situation, showing how the results given by our algorithm should be interpreted, avoiding a misuse that may lead to a contradictory and inefficient selection of travel routes.

**Road # 1 vs. Road # 4:** What of interesting emerges from this comparison is that these two roads, namely # 1 and # 4, have very different values of $S$ in spite of a series of similarities, both in terms of road characteristics (see Table II) and in terms of the results provided by our algorithm (see Table III).

A clear explanation of this phenomenon can be given by observing the congestion graph for these two road sections (Figures 4 and 5, respectively). What emerges is the following. The points in the graph of Figure 4 are more clustered and concentrated almost exclusively in the two regions of congestion (top-right area) and of no congestion (bottom-left area). Hence, as in

---

### TABLE VI

**Comparison:** $T_{HCM}$ and percentage difference.

<table>
<thead>
<tr>
<th>Road</th>
<th>Section, Direction</th>
<th>$T_{HCM}$ [s]</th>
<th>$(T_{HCM} - T^*) \times 100$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Via B. Croce</td>
<td>P.zza Guerrazzi-Via Queirolo, left</td>
<td>78</td>
<td>-16%</td>
</tr>
<tr>
<td>2 S. Monica Blvd</td>
<td>Veteran-Sepulveda, left</td>
<td>147.5</td>
<td>-16%</td>
</tr>
<tr>
<td>3 S. Monica Blvd</td>
<td>Wilshire-Roxbury, straight</td>
<td>54</td>
<td>-13%</td>
</tr>
<tr>
<td>4 S. Monica Blvd</td>
<td>Wilshire-Bedford, right</td>
<td>67</td>
<td>-18%</td>
</tr>
<tr>
<td>5 Lincoln Blvd</td>
<td>Fiji-Venice, back</td>
<td>341.4</td>
<td>-4%</td>
</tr>
<tr>
<td>6 Wilshire Blvd</td>
<td>Midvale-Westwood, right</td>
<td>35</td>
<td>-3%</td>
</tr>
<tr>
<td>7 S. Monica Blvd</td>
<td>Roxbury-Bedford, right</td>
<td>33</td>
<td>-21%</td>
</tr>
<tr>
<td>8 Wilshire Blvd</td>
<td>Veteran-Westwood, right</td>
<td>75</td>
<td>+1%</td>
</tr>
<tr>
<td>9 S. Monica Blvd</td>
<td>Westwood-Sepulveda, right</td>
<td>122</td>
<td>+1%</td>
</tr>
</tbody>
</table>
this case it is easier to intercept both congested and non-congested states, also, the horizon of predictability (the $S$ value) grows larger. This does not apply, instead, to the graph of Figure 5 where the points are in some sense more scattered on the left semi-plane. Here, the 16% of points lies in the top-left area (noisy situation). As a results, the value of $S$ concerning the length of forecasting drops lower.

Obviously, there are real facts behind the explanation we just provided. The fact essentially is related to the number of lanes per considered road. Normally, roads with higher capacities (two lanes or more) or even smaller capacity fluctuations (no presence of obstructing vehicles) experience a faster transition from a congested to a non-congested state. This is exactly what happened to the two-lane road # 4, as confirmed by Figure 4, while the relatively high percentage of scattered points in the left semi-plane of Figure 5 reveals that road # 1 is a one-lane street which easily transitions from a non-congested to a congested state.

Road # 3 vs. Road # 7: What could seem a paradox here is that even though # 3 and # 7
are sections of roads belonging to the same street, the former alternates states of no congestion to states of congestion, while the latter almost always experiences a non-congested situation. These situations are highlighted by the corresponding graphs of Figure 6 (road # 3) and Figure 7 (road # 7). Indeed, the graph of Figure 6 has almost no scattering, while the graph of Figure 7 presents a high percentage of scattered points, especially in the right semi-plane devoted to represent congestion. Again, this graphical pattern is easy to explain based on what happens in reality. Indeed, road # 7 has the road section between Roxbury Dr. and Bedford Dr. of very short length. Further, the traffic light at Bedford Dr. is coordinated with all the downstream traffic lights, thus easing the outflow of the vehicles that are stuck in queue. For these two reasons, this street experiences a fortunate situation where the shift from a congested to a non-congested state is made easier.

To conclude, the cases we here discussed confirm that our algorithm is really precise in identifying even those situations where abnormal facts may occur.
C. Dealing With Non-Recurrent Traffic Congestion Causes

Our mechanism has been designed with the aim of detecting congestion both in the situation when it is determined by recurrent traffic patterns and when its cause is due to non-recurrent (or abnormal) events. In fact, the suitability of our mechanism in recognizing congestion in all type of circumstances derives from its ability to follow the evolution in time of the congestion threshold $\bar{T}^*$ of a given road segment as any of its physical properties change (e.g., diminished capacity due to maintenance work) or as it experiences non-recurrent traffic patterns (e.g., accidents).

To better explain how our mechanism can deal with non-recurrent congestion events, take the following example where a two-lane road, due to scheduled maintenance work, is reduced to a single lane road starting from 1pm. In such scenario, clearly, vehicles running through that road between 8am and 1pm enjoy a non-congested situation. After 1pm, since the capacity of the road is halved, vehicles will take a longer time to traverse it, thus experiencing congestion. If a sufficient number of vehicles traversed that road during the morning, as well as after 1pm, our mechanism would have been able to determine an adequate congestion threshold $\tilde{T}^*$ (of value, say, 100 seconds) that allows one to distinguish states of no congestion from states of congestion. We can also assume that maintenance work lasts for a few days. Suppose now that
the day after, at a given time, an accident occurs, blocking the flow of cars for a very long time (i.e., a large number of cars incur in a state of severe congestion). Under these circumstances, our scheme would have identified a much larger congestion threshold, say 200 seconds (provided that a sufficient number of cars has incurred in this abnormal event). Up to this point we have described how our algorithm works in its present form. Obviously, it is not difficult to devise an extension of our model that is able to distinguish different causes of congestion (recurrent, non-recurrent), simply by observing how the congestion threshold fluctuates, provided that a sufficient number of cars experience that specific traffic situation.

As a real example, take into consideration what we observed during one of our experiments in Via Benedetto Croce in Pisa. During a situation where a moderate level of congestion was experienced (traversal times fluctuated around 100 seconds), a pedestrian suddenly fell on the sidewalk and, very rapidly, a few minutes later an ambulance arrived, at first stopping and then slowing the flow of vehicles. The problem was fixed taking no more than 15 minutes, thus only a limited amount of vehicles experienced this abnormal event, with traversal times reaching approximately 220 seconds. As the duration of this event was too short, it did not significantly influence the magnitude of the value of the congestion threshold \( \bar{T}^* \). Needless to say, a longer duration of this abnormal event would have had a more serious impact on the value of the congestion threshold, thus allowing one to distinguish it as a new and more severe cause of congestion, with respect to the previous situation (moderate congestion).

In summary, while other mechanisms exist which are designed to detect non-recurrent congestion states [32], [33], these typically address the problem of identifying the cause of the events that are at the basis of congestion (e.g., accidents). Instead, our mechanism, as the abovementioned examples confirm, is concerned with the issue of revealing the severity and duration of a congestion state. Nonetheless, our mechanism can be easily adapted to distinguish recurrent congestion from non-recurrent congestion, thus resulting also a valuable tool for detecting accident events, for example.

V. CONCLUSION

We presented a simple and efficient general-purpose vehicular congestion detection and short-term forecasting algorithm. Our algorithm has been checked on a real test-bed driving over 450 miles throughout Pisa and Los Angeles.
We proposed a new definition of congestion, where a road is congested only when the likelihood of finding it in the same congested state is high in the near future. This makes our algorithm easy to implement and effective in providing significant results.

Given its characteristics, we believe this algorithm is well suited for modern advanced traveler information systems that aim at guiding drivers around critical traffic states.

REFERENCES


