

This is the peer reviewed version of the following article:

Traffic analysis in a smart city / Bachechi, C.; Po, L.. - (2019), pp. 275-282. (19th IEEE/WIC/ACM International Conference on Web Intelligence Workshop, WI 2019 Thessaloniki, Greece 14 October 2019 through 17 October 2019) [10.1145/3358695.3361842].

Association for Computing Machinery, Inc
Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

05/05/2026 15:39

(Article begins on next page)

Traffic Analysis in a Smart City

Chiara Bachechi*

Laura Po*

chiara.bachechi@unimore.it

laura.po@unimore.it

'Enzo Ferrari' Engineering Department
Modena, Italy

ABSTRACT

Urbanization is accelerating at a high pace. This places new and critical issues on the transition towards smarter, efficient, livable as well as economically, socially and environmentally sustainable cities.

Urban Mobility is one of the toughest challenges. In many cities, existing mobility systems are already inadequate, yet urbanization and increasing populations will increase mobility demand still further.

Understanding traffic flows within an urban environment, studying similarities (or dissimilarity) among weekdays, finding the peaks within a day are the first steps towards understanding urban mobility.

Following the implementation of a micro-simulation model in the city of Modena based on actual data from traffic sensors, a huge amount of information that describes daily traffic flows within the city were available.

This paper reports an in-depth investigation of traffic flows in order to discover trends. Traffic analyzes to compare working days, weekends and to identify significant deviations are performed. Moreover, traffic flows estimations were studied during special days such as weather alert days or holidays to discover particular tendencies.

This preliminary study allowed to identify the main critical points in the mobility of the city.

CCS CONCEPTS

• **Computing methodologies** → **Model development and analysis**; **Modeling methodologies**; • **Human-centered computing** → *Visualization*.

KEYWORDS

Traffic analysis, Time series, Data Modelling, Real Time Data, Big data analytics, Traffic modelling, Smart mobility, Smart city

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WI '19 Companion, October 14–17, 2019, Thessaloniki, Greece

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6988-6/19/10...\$15.00

<https://doi.org/10.1145/3358695.3361842>

ACM Reference Format:

Chiara Bachechi and Laura Po. 2019. Traffic Analysis in a Smart City. In *WI '19 Companion: IEEE/WIC/ACM International Conference on Web Intelligence (WI '19 Companion)*, October 14–17, 2019, Thessaloniki, Greece. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3358695.3361842>

1 INTRODUCTION

Traffic influences the life of every citizen in several aspects: the time required to move from home to work, the quality of the air he/she breathes, the stress generated from being stuck in a traffic jam and the lack of sleep and exercise as a consequence of time wasted caught in traffic congestion. Since individual drivers cannot see the traffic system as a whole, to raise citizens awareness in their mobility choices and the consequent impact on the environment and to boost sustainable mobility, a representation of the whole traffic system of their city is needed. Traffic simulations models are mathematical tools that help the city planners to manage and analyze urban mobility by producing an overview of the traffic situation. This representation would also help Public Administrations to identify critical points and to respond to new mobility demands that might emerge thanks to a comprehensive traffic analysis.

In this work, we focus on the analysis of the outputs of a micro-simulation model implemented in the city of Modena [1]. We started running traffic flows simulations since October 2018, and we are still running them every day. In this paper, we focus our analysis on daily traffic flows simulations of November 2018. This trend analysis has provided insight to better understand the mobility issues related to traffic in the urban area of Modena.

To evaluate the results obtained from the real traffic simulations in Modena and to compare outputs of different simulations, it is necessary to employ the right metrics. We take advantage of the MULTITUDE¹ guideline [7] that uses a time series based approach to evaluate the overall performance of the simulation model. The rest of the paper is organized as follows. Firstly, we introduce related work that inspires us 2; secondly, in Section 3, we describe the simulation model employed, the input required and the output obtained; thirdly, in Section 4, the metrics used to represent the output of the traffic model are described; then in Section 5 the procedure employed to represent the distance between two simulation is described. Then in Sections 6, results of the implemented model are reported, in particular, the traffic flow on an average weekday is calculated and compared to weekend and special days. Then, in Section 8, we listed the critical

¹www.multitude-project.eu

points detected by our traffic analysis. Conclusions and future research direction are summarized in Section 9.

2 RELATED WORK

In transportation engineering, traffic flow analysis is the study of interactions between travelers (that might be pedestrians, cyclists, drivers, and their vehicles) and infrastructure (i.e. highways, signage, and traffic control devices), with the aim of understanding how people and means of transport are moving within a city or an area, that aims at detecting traffic congestion problems and optimizing the transport network with more efficient movement of people and vehicles.

Some approaches focus on traffic trajectories analysis [11], while others analyzed the speed profiles on the road network in [10]. These approaches could not be used in city of Modena, where we do not have enough GPS data to trace trajectories and the speed measurements collected by the traffic sensors are not meaningful, since sensors are placed near traffic light intersections. Therefore, we implemented a micro-simulation traffic model based on sensor data in [1] and, in this paper, we focus on the analysis of traffic flows profiles. A spatio-temporal data analysis on traffic data is crucial, in [5] particular attention is given to the data profiling method which is based on the periodicity and similarity characteristics of traffic flow during a week or a year. Also in this paper, we focus on studying the periodicity and trends.

3 THE MODEL

A micro-simulation traffic model was implemented for the city of Modena based on actual data from traffic sensors, the model is described in [1]. The model employs an open source micro-simulator: SUMO (Simulation for Urban Mobility). Traffic is simulated in the urban area, starting from the data collected every minute from sensors in the street.

The model queries the database where sensor measurements and sensor locations are collected and creates input files for SUMO. When input files are generated, the simulation is started on a remote HPC resource. The outputs produced are uploaded in the database, creating an historical resource of traffic flows, ready for being analyzed. The model interactions with the database and the collection of sensors data are detailed described in [9].

Since the model is dynamic, the produced output contains a detailed evolution of the traffic situation over time. The output is represented by a collection of time series of traffic flows values, one for each sensor location and each road lane in the map. The model is able, starting from traffic punctual information on 300 sensors locations, to provide an overview of traffic intensity on more than 800 km of roads. The generated output includes information about vehicle count, lane density and average speed for every road lane in the map and every minute of the simulation. Moreover, the simulated values of vehicle counts and average speeds are available as punctual information on every sensor location.

Therefore, the information about the number of vehicles passing and their average speed is given on the sensors locations and cumulative for every lane section. The output files, provided in XML format, are converted and elaborated to extract further knowledge on the traffic situation during the simulation. A mean daily trend for weekly days was estimated, then an evaluation of the similarity between simulations has been provided. To identify the most congested roadways for every simulation the value of traffic density in every lane has been taken into account. A map was produced for every simulation time step, where more populated roads were colored in red.

The workflow of the model is depicted in Figure 1. The data of the traffic sensors provide real traffic flows in the specific position where the sensors are located.

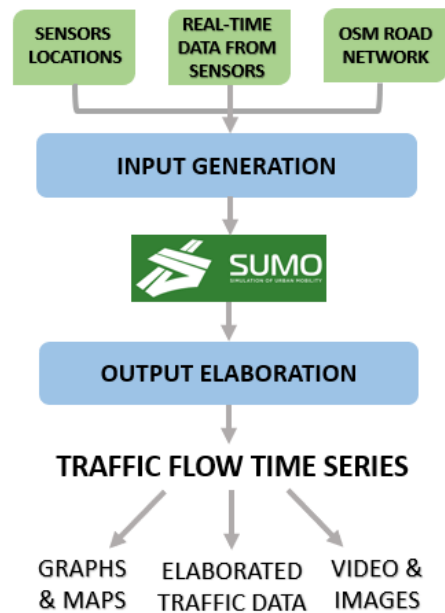


Figure 1: Overview of the model structure.

4 REPRESENTATION AND METRICS

The estimation of traffic flow is based on the counting of vehicles that cross a given location during some time interval. Therefore urban traffic flows are spatio-temporal data series. SUMO output files are composed by series of values, one for every time interval and every lane or edge in the map. The time interval is an input parameter that can be modified. Time is discrete and a value is normally collected for every step. Therefore outputs can be represented as time series. A data series is a sequence of values ordered along some dimension. In the case of time series, this dimension is time. Therefore, the analysis of this kind of data has to consider the evolution of values across the time dimension. Time series analysis concerns in analyzing the evolution of values across the time dimension [4]. This allows to identify some trends that characterize data. There are several possible data series representations and measures of distance used to perform similarity matching among time series.

Evaluating the similarity of two different time series describing the same phenomena is very hard, due to the big dimension of each time series (i.e. in a daily simulation, 1440 values are produced for every lane and for every sensor). Therefore effective representation techniques have to be adopted. These techniques can significantly reduce processing time. A description of different representation methods of a time series is given in [6], this paper introduces and compare theoretically and empirically different metrics and demonstrate that *Piecewise Aggregate Approximation* is faster to compute and supports more sophisticated querying techniques. For these reasons, the *Piecewise Aggregate Approximation* (PAA) is the representation method we chosen to calculate the distance among two time series. It is better described in [13].

PAA approximates a time series X of length n in a vector $\vec{X} = (\vec{x}_1, \dots, \vec{x}_M)$ of any arbitrary length $M \leq n$ where each of x_i is calculated as follows:

$$\vec{x}_i = M/n \sum_{j=n/M(i-1)+1}^{(n/M)i} \vec{x}_j$$

PAA represents a time series with a linear combination of box basis functions where each box has the same length. PAA use segments to represent data series, so it is extremely fast to calculate and supports non-euclidean measures.

Once represented time series through PAA, a metric is needed to evaluate the distance between two different . *DynTimeWarping* (DTW) is the metric we adopted. DTW allows sequences to be stretched along the time axis, in [2], a more detailed description of this metric is given. DTW is computed by dynamic programming: given two different sequences, scores between every two points in the sequence are calculated in a matrix and then, starting from the bottom-left cell, all paths can be selected to reach the top-right cell. The chosen path pass through the minimum possible value of cells surrounding the current cell. Hence, DTW allows to finding the optimum warping path with the smallest matching score. This path has to satisfy the following conditions:

- **Boundary:** the path has to include the left-bottom cell and the right-top cell;
- **Continuity:** all cell belonging to the path has to be continuous to each other;
- **Monotonicity:** the values of the cell in the path have to be ordered, the value of a subsequent cell in the path has to be equal or greater than the value of the previous cells.

DTW can be used to find corresponding regions between the two time series DTW can tolerate noise, time shifts and even scaling in the y-axis. The main disadvantage of dynamic time warping is computational complexity. For this reason, in our work, we use FastDTW, a less complex version of DTW where not all matrix values are filled. In [12], Salvador and Chan describe FastDTW as a multilevel approach, because the length of the path grows linearly with the length of the series (n). This solution speeds up Dynamic Time Warping and reduces its complexity from an $o(n^2)$ to an $o(n)$ algorithm. To

determine if two simulations are similar to each other, FastDTW distances have been calculated between their time series. These distances have been evaluated in all the points of the map where sensors are located. This choice was taken because in these points the reliability of the obtained vehicles counted is higher and a comparison with real measurements is possible. All these values of distance are then divided for the dimension of the matrix, to make the distance value less dependant on the duration of the simulation. Then, dividing for the total number of points, a mean value is calculated. The metric adopted employs FastDTW to calculate the distance between the time series of two different simulations in every point. Then, it calculates the mean distance between all the values obtained in the simulation observed points.

To consider also the spatial dimension of the output the model generates also several images. These images are maps with the road network of the city where streets with a different number of vehicles are drawn in different colors. Moreover, to consider both the spatial and temporal dimensions of the output, the traffic flow of the city can be visualized in a video where different colors (green, yellow, red) describe the intensity of the traffic flow on the road ².

5 TRAFFIC SITUATION EVALUATION

To communicate the traffic model results and to identify the most congested area of the city, additional elaborations are needed.

Through images and videos, a clear idea of the traffic situation in the city of Modena will be depicted. MULTITUDE studies pointed at investigating the uncertainty enclosed in the estimation process, which could affect the reliability of the next analysis. As described in [7], in the field of micro-traffic models calibration, studies showed that the use of error measures, as well as statistical GOF (Goodness Of Fit) functions, is not the correct way to calibrate the model. This is caused by the integral nature of the traditional objective functions that locally cumulates the errors employed to calibrate the model and do not take into account the dynamic of the observations. This observation lead us to employ an analysis based on the comparison of time series. To represent the simulation with a unique curve, considering only the time dimension, the space dimension is aggregated, averaging the different values in all the observed points. The obtained time series are, then, compared with the DTW based metric described in section 4 to obtain a unique value representing the distance between them.

An analysis of the trend of the produced time series will be conducted to obtain an average mean trend. The average trend, then, will be compared with every daily trend to identify days in which the temporal evolution of the number of vehicles was different than usual.

²An example of video is available at <http://trafair.eu/sumo-simulation/>.

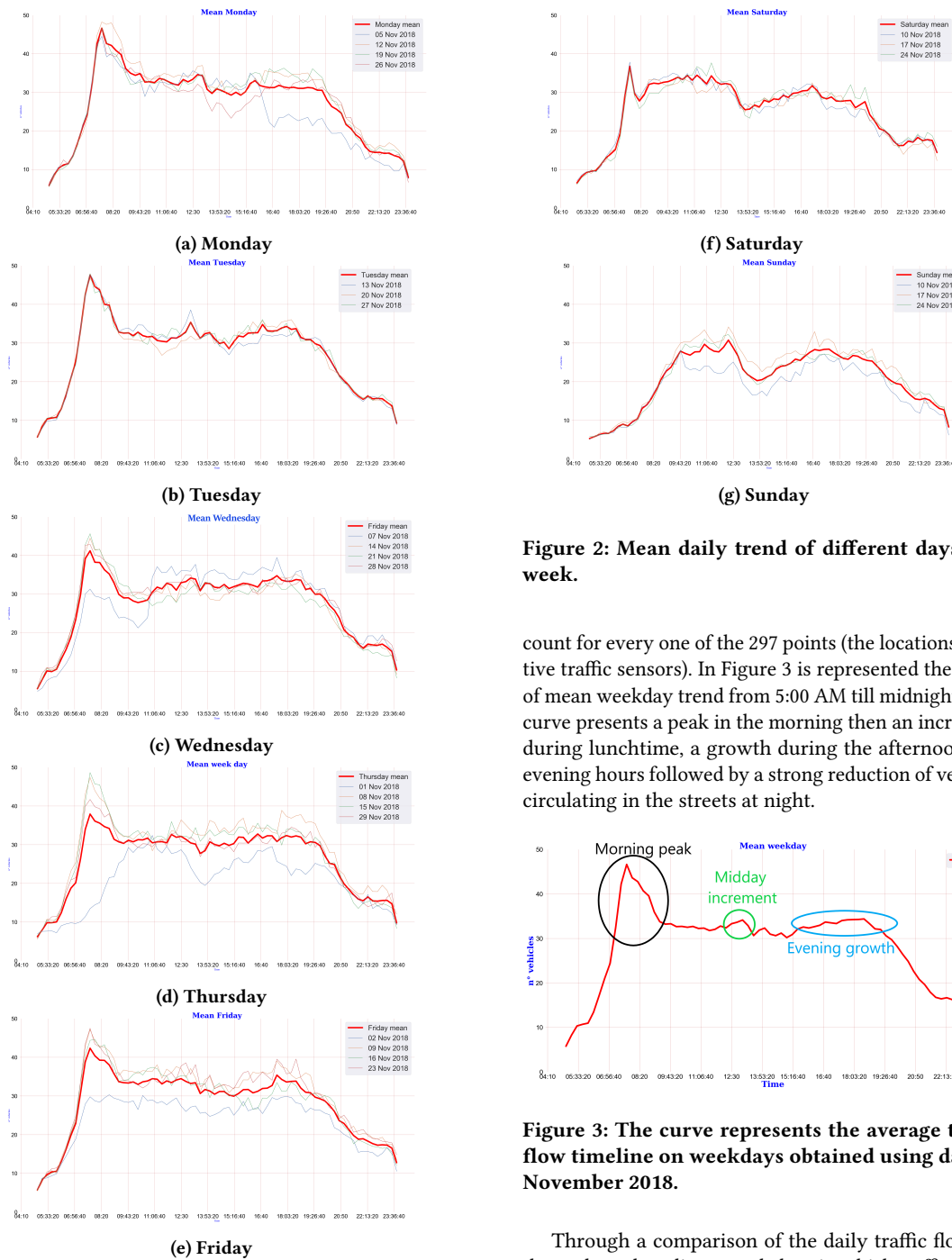


Figure 2: Mean daily trend of different days in a week.

count for every one of the 297 points (the locations of active traffic sensors). In Figure 3 is represented the curve of mean weekday trend from 5:00 AM till midnight. This curve presents a peak in the morning then an increment during lunchtime, a growth during the afternoon and evening hours followed by a strong reduction of vehicles circulating in the streets at night.

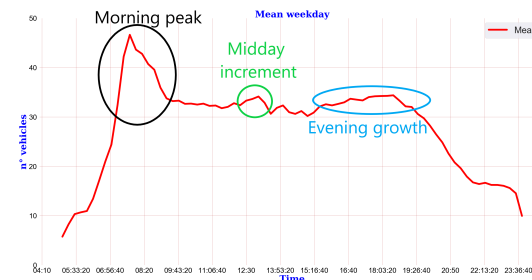


Figure 3: The curve represents the average traffic flow timeline on weekdays obtained using data of November 2018.

6 EVALUATING AN AVERAGE WEEKDAY

Through the collection of numerous simulation outputs, a medium trend for every day of the week was evaluated. Then, these medium trends were compared with daily simulations to identify anomalies and detect the location of the city where traffic was atypical.

Several outputs obtained simulating November weekly days were considered to evaluate a mean value of vehicle

Through a comparison of the daily traffic flows to the evaluated medium trend, days in which traffic conditions are different from the habitual can be identified. In Figure 4, a table displays values of DTW based distances between the given day simulation and the average trend. As can be observed, only Wednesday 28th November is similar to the medium trend. The other days (Tuesday 6th in particular) have a high distance, therefore a sensibly different trend. An analysis of these days is provided in the subsequent section.

In Figure 2, the medium trend of every singular weekday has been estimated to investigate if there was a difference in the distinct days of the week. As expected, working days curves, from Monday to Friday, appear more similar to each other than Saturday and Sunday

curves. From Monday to Friday there is a high peak in the morning between seven and half-past nine. A similar peak is present also in the Saturday trend but it has lower vehicles count and it is centred between seven and eight in the morning. A good interpretation could be that in Modena most schools are open on Saturday and lessons start at eight o'clock in the morning. On Sunday morning there is no peak since in Italy people usually do not work and schools are closed.

Some November days were not used to evaluate the average curve. Thus their simulation output was compared to the evaluated mean weekday to understand if their trends were regular or their daily traffic situations were anomalous.

	Distance from mean weekday	Total n° vehicles	UPS
Monday 5th Nov	0,497	892853	5238,43
Tuesday 6th Nov	0,808	644686	4926,99
Wednesday 7th Nov	0,453	977280	6498,31
Friday 2nd Nov	0,448	1075136	5161,06
Thursday 1th Nov	0,450	789396	5994,70
Wednesday 28th Nov	0,369	1172618	5663,28

Figure 4: DTW based distances calculated between the simulation mean time series and the calculated average weekday, the total number of vehicles circulating and UPS. Green color indicates that the DTW values are low and red that these values are high.

6.1 Identification of anomalous daily trends

Studying the result obtained as outputs from our model some days appears distant from the evaluated trend. An anomalous day can be identified by a DTW based metric distance from the medium trend above 0.4. This value was defined through an empirical analysis of the obtained time series and their distance from the average trend. The simulation can also be plotted in the same graph with the average trend to visualize the main difference between them. Even if a daily simulation can be represented with a unique distance value, it is also possible to compare the medium evaluated trend with everyday simulation in all the observed points singularly. This can help to identify which are the zones of the city that show an unusual traffic situation.

6.2 Detecting location affected by Traffic Anomalies

The traffic anomaly is considered to occur in sub-areas or punctually in a city when the values of the corresponding traffic flows indicators deviate significantly from the expected values. Traffic anomalies often happen in urban traffic networks and can be caused by traffic accidents, traffic congestion, and large gatherings and events, such as construction, occur. The detection of traffic anomalies is an important aspect for traffic management.

In Figure 4, Tuesday 6th November has a total number of vehicles fewer than other days. The UPS (Updates Per Seconds), which must be interpreted as the average

number of vehicles simulated per second of computation time, is very low in this Tuesday. Figure 6 shows that even if the number of vehicles counted is lower, the trend and the shape of the curve is similar to the average Tuesday: the daily peak is still in the morning and in the evening the number of vehicles grows as usual. Instead, the midday peak is not around lunchtime as usual, but at ten a.m.. That anomaly was studied comparing this Tuesday to the mean weekly day. A function has been created to identify where traffic anomalies were located which returns the points in which DTW is higher between the two simulations.

Points in which the adopted distance metric was higher than 2 in every comparison have been identified. In Figure 5, the position of these points is showed on OSM map. As it can be seen, their location is distributed in several areas of the city on the most populated roads. Therefore, the origin of the anomaly is not confined to a specific area of the city.

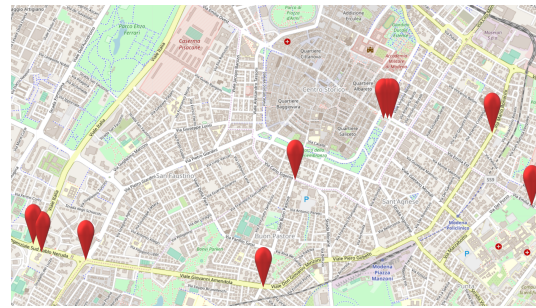


Figure 5: OSM map [8] where, in red color, are highlighted the points in which the number of vehicles counted was lower on Tuesday 6th November than on the other weekly days.

In order to investigate the cause of the anomaly detected on Tuesday 6th, we examined also the adjacent days. While Wednesday 7th does not have a significant deviation from the average Wednesday, Monday 5th has a regular evolution until half past four p.m., then there is a consistent reduction of vehicles circulating as can be observed in Figure 7. We examine, also weather conditions in these days, and realized that Monday 5th and Tuesday 6th were rainy days and Emilia-Romagna civil protection reported a yellow weather alert. In particular, the Modena area was in alert due to river flooding.

In Figure 4, the particularity of Monday 5th and Tuesday 6th November is confirmed by their high DTW based distance. Their distance values from the average weekday are higher also than all the other distances, among weekdays and weekend days. The total number of vehicles circulating in the whole simulation is lower than expected in both days.

The 1st of November, in Italy, is a festivity, called All Saints' Day, and most of the people do not go to work. Traditionally, Italians observe All Saints' Day by visiting cemeteries in order to lay flowers and candles on the graves of their deceased loved ones. We examine the traffic flows in this day, paying special attention to

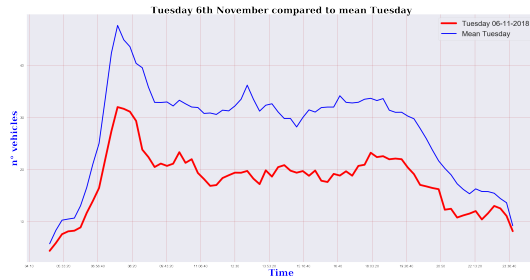


Figure 6: Comparison between Tuesday 6th November and the average Tuesday evaluated.

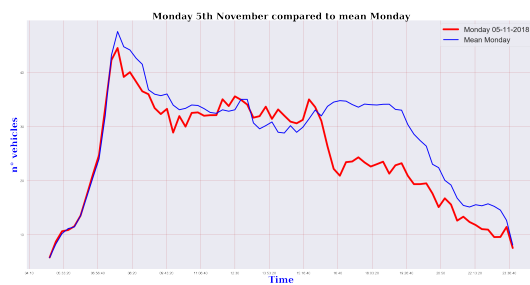


Figure 7: Comparison between Monday 5th November and the average Monday evaluated.

one of the most affected locations around the cemetery. As it is shown in Figure 4, the 1st November appears different from the mean weekday evaluated. In Figure 9, the average weekday calculated in two observed points is compared with the 1st November simulation in the same points. Considering the position of the cemetery, the sensor located near it shows higher values than the average weekday. A sensor located in a secondary road to access the city instead shows a lower vehicles flow, as expected in a not working day.

These examples underline that the traffic flows generated by SUMO can be employed to study how the traffic changes during days and which are the points of the city that are more influenced by some events.

7 SIMULATION STATISTICS

In Figure 4 some statistics of daily simulations are presented. Through a comparison of values in different days of the week, Monday and Friday appear the days with the highest number of routes. Instead, Sunday was not as much populated by vehicles. Tuesday 06/11/2018 has a very low number of vehicles compared to other days. This is coherent with its daily trend that showed lower values than other weekdays. UPS stands for Unit Per Seconds, thus gives us an idea of how many vehicles were on average every time step in the simulation. On Wednesday 7 November 2018, even if the total number of vehicles is lower than weekly days, UPS value is the highest. Therefore these vehicles were more distributed during the day. In fact, on the other weekday, there was a high peak in the morning and then the number of vehicles decreases. Contrariwise on Wednesday (see Figure

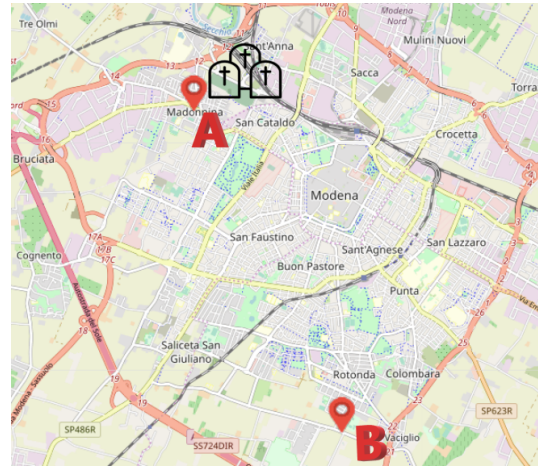
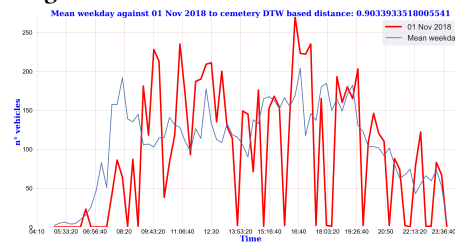
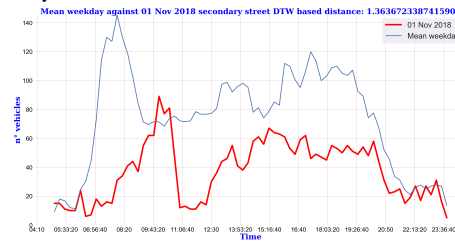


Figure 8: Location of sensors and cemetery



(a) Trend of vehicles circulating near the cemetery



(b) Trend of vehicles circulating far away the cemetery

Figure 9: Time series of the average week day calculated in two different positions of the map compared with the time series of the 1st November simulation in the same positions.

10), there was no evident peak but the number of vehicles remains high for all afternoon and evening.

8 UNCOVERING MOBILITY CRITICAL ISSUES

Observing the average curves obtained for every weekday, the daily peak is always between 7 and 9 a.m., more precisely the highest value is around 8 a.m..

Referring to the peak hours, some areas of the city have been analyzed summing the total number of vehicles counted by sensors located inside them.

Four areas have been individuated:

- An area with 3 schools of different grades;
- A residential area;

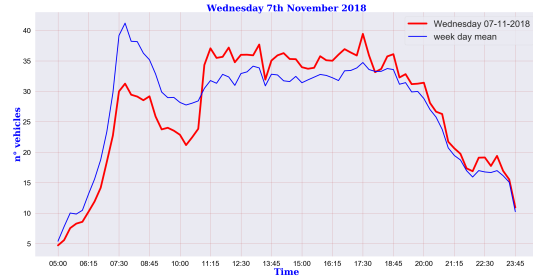


Figure 10: Comparison between Wednesday 7th November and the average Wednesday evaluated.

- An industrial area;
- The main hospital area (which is also a university area).

In Figure 11 the total numbers of vehicles counted in the 2 peak hours have been compared through a histogram. graph refers to a November Monday (19th November 2018). The school area seems to be the most critical, contrarily the industrial area is the one with fewer vehicles.

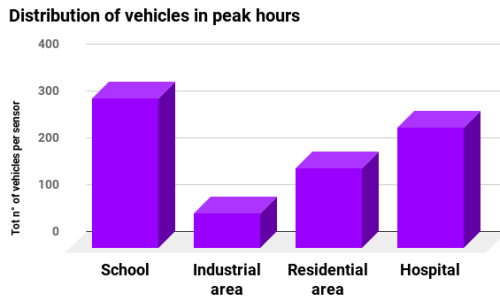


Figure 11: Histogram: comparison between vehicles circulating in peak hours in different areas of the city.

Saturday and Sunday have a different peak, Saturday has a lower and shorter peak around 8 o'clock, then Sunday has its peak at lunchtime around midday. The peak on Saturday morning is related to school, this can be demonstrated comparing a November Saturday (in which school were open) and Saturday 8th June, the first day of summer break in Italy. In Figure 12, a comparison between a November school day Saturday and Saturday 8th June shows that the latter has not the morning peak but has a really higher number of vehicles circulating during the day, due to people moving from the city to go on vacation.

To compare different lanes and identify which are the most congested roadways in the city, the lane density value was taken into account. Lane density is a measure of the number of vehicles in a lane every kilometre. This measure was chosen since even different edges, with more than one lane, have the same traffic situation if their lane density is equal. The flow was not a good measure to compare traffic situations in different lanes because the number of vehicles counted in an hour

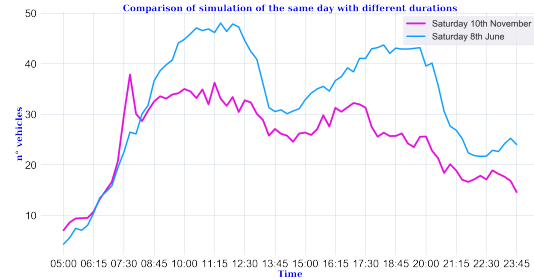


Figure 12: Comparison between a November Saturday (when schools were open) and the first Saturday of summer vacation.

depends on the maximum speed allowed in that lane. Studies described in [3] demonstrate that an increase in the speed limit leads to an increase in the saturation flow rate. Saturation flow is a very important road traffic performance measure indicating the maximum rate of flows of traffic. Saturation flow describes the number of passenger car units (pcu) in a dense flow of traffic for a specific intersection lane group. If vehicles are moving at the speed of 30 km/h in a lane where the maximum allowed speed is 130 km/h, this means that there is a queue. If the same speed is reached by vehicles in residential areas (where limits cannot overcome the 50 km/h) then there is no congestion. Flow value is calculated multiplying density and speed. In Figure 13 the relation between flow and density is illustrated. When density grows over a critical value, the flow starts to decrease and vehicles are congested. Therefore there is a cut off value of lane density which can be interpreted as a signal of congestion. Traffic density is a good indicator of traffic estate.

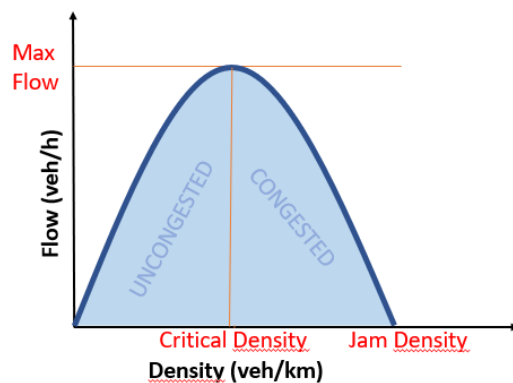


Figure 13: Relation between flow and density. The maximum flow is reached when the density is at its critical value.

Since the number of vehicles that could be over a one-kilometre section of road, to avoid traffic congestion, is similar in different types of road. Considering this lane-based value, since in our simulation all the vehicles were of the same type and thus had the same length (4,3 m), and taking into account the security distance between every two vehicles, it was evaluated that in one

