# Vehicular Traffic Simulation in the City of Turin from Raw Data

Marco Rapelli, Claudio Casetti, *Senior Member, IEEE,* and Giandomenico Gagliardi

**Abstract**—The testing of vehicular communication technologies, the study of urban mobility patterns, the evaluation of new traffic policies cannot dispense from vehicle mobility simulation. As is often the case, the larger the dataset, the better. Indeed, in recent years, many projects in the fields of mobility or vehicular communication have sought new traffic simulators with extended areas of investigation, possibly covering a whole city and its suburbs. In this spirit, we have modeled an urban traffic simulation in a 600-Km$^2$ area in and around the Municipality of Turin, leveraging the SUMO tool. This paper aims at reporting in detail the methodology we followed in the creation of this dataset. Our results demonstrate that a complete modeling of such a wide area is possible at the expense of minor simplifications, reaching a very good level of approximation.

**Index Terms**—Urban Mobility, Transportation Modeling, Traffic Simulation, Large-Scale Traffic Simulator.

## 1 Introduction

In August 2019, a report from the Texas A&M Transportation Institute [1] estimated the time spent while driving through heavy traffic by the average American commuter. The result highlighted that 54 extra hours per year were spent in traffic delays. An "extra hour" is considered as the additional time spent traveling at congested speeds rather than free-flow speeds. Though some U.S. cities present more congested traffic conditions than European ones, the American average resembles the majority of big cities in the world. Additionally, traffic jams are expected to increase with the continuous growth of population living in urban areas. Urban planners are therefore tasked with developing sustainable transportation and infrastructure systems for current and future traffic demand.

A tool of choice to achieve this goal is simulation, which, leveraging more powerful computing resources and different sources of information, can paint a detailed picture of increasingly larger regions. A more classic usage for simulated urban models could be the analysis of traffic flows and the identification of residential and industrial districts for a socio-demographic analysis, as performed in [2]. Such type of studies are extremely important for urban planners in order to identify the main arteries between residential districts and work areas and evaluate whether to improve them or create new ones. Furthermore, traffic simulators are the only tools capable of predictive mobility analysis. In such a spirit, a model similar to the one proposed in this paper was already developed by the authors for the feasibility study of the Turin Limited Traffic Zone (Torino ZTL), part of the Torino Centro Aperto project commissioned by the Municipality of Turin in 2018 [3]. The idea of the present research stemmed from the mentioned project, with the new aim of creating an urban-scale scenario.

Besides mobility analysis, large-scale urban models could be effectively used for city-wide pollution studies, by simply modeling the emission class of vehicles in the simulator, as performed in [4]. Possible outcomes of such a simulation-based investigation are the identification of areas of likely congestion where to install air-quality sensors, or a rough quantification of pollution caused by traffic as opposed to other sources, when assessing data from real samples.

Last, but not less important, is the fact that urban traffic models are essential for the testing of vehicular applications. Since road testing of new vehicular technology in support of ADAS (Advanced Driver Assistance Systems) is often too expensive and potentially dangerous, a large portion of the testing process is made via simulated experiments, as in [5]. With this in mind, the choice of the simulated mobility environment is paramount. Due to the lack of existing large-scale traffic simulators, application developers often opt for a small use case map, featuring a couple of intersections at most, where interactions among large number of vehicles, and their repercussions on traffic in a large urban environment, are hardly realistic. Instead, large-scale vehicular testing always leads to more complete, insightful results, as is the case of [6].

It is thus clear why an urban-scale traffic model is necessary. However, a complex, detailed, large-scale model of a city is a challenging project and this could be the reason why examples of urban-scale traffic models in the literature are so few and far between. It is also important to underline that every city has different topological, demographic, public transport and mobility characteristics. Comparing the performance of models of different cities is of little consequence, even for urban areas with a similar extension or

- *M. Rapelli is with the Department of Electronics and Telecommunications Politecnico di Torino, 10129 Torino, Italy. E-mail: marco.rapelli@polito.it*
- *C. Casetti is with the Department of Control and Computer Engineering, Politecnico di Torino, 10129 Torino, Italy. E-mail: claudio.casetti@polito.it*
- *G. Gagliardi is with the ITS Technology & Architecture Department, 5T Srl, 10122 Torino, Italy. E-mail: giandomenico.gagliardi@5T.torino.it*

*Manuscript received MONTH DAY, YEAR; revised MONTH DAY, YEAR.*

population. The city of Turin has distinctive features that are not found in existing large-scale models (such as the ones discussed in Section 2), and which required special care in the modelling effort: as an example, almost all avenues have service roads running alongside, mainly used for parking or by local businesses. The presence of these service roads changes both the way intersections are modelled as well as the distribution of trip starts and ends. For this reason, we maintain that extending the existing literature of urban traffic model to different cities is extremely important.

The present work, which is an extension of our conference paper [7], introduces TuST (Turin SUMO Traffic), a complete large-scale model of the Turin metropolitan area and its neighboring districts over 24 hours, using the SUMO tool. This paper aims at contributing to the discussion on how to create an open-source, large-scale urban simulator, by providing a detailed description of its construction and by demonstrating that creating a model of such a complexity is indeed possible. Aggregated traffic and street sensors data, used respectively for traffic demand construction and system validation, were provided by 5T [8], a local company working on ITS and Info-mobility for its shareholders: Municipality of Turin, Metropolitan City of Turin, and Piemonte Region. Our work improves the existing literature in several ways:

- it describes in detailed steps the construction of a large-scale simulator which inspects a 24-hour scenario;
- as input, it uses real traffic data from a real Origin/Destination (O/D) dataset and real traffic light phases from the local traffic agency;
- it makes publicly available simulated vehicle traces of a complete full-day scenario [9]; in this way, traces can be used as input for more complex mobility models or for building vehicle-to-vehicle (V2V) communication scenarios based on real traffic data;
- it validates simulation results with real traffic data from street sensors provided by the local traffic agency; this represents an objective measure of "realism" and demonstrates the model we developed is a very realistic large-scale simulator of traffic in Turin.

Also, with respect to our previous work [7], the current paper (i) describes a new procedure to incorporate data from traffic-light regulated intersections into the model, (ii) provides a detailed description of the tweaks and refinements of our model that led to stabler results and a more faithful representation of the full-scale daily traffic and (iii) features an entirely new Results Section, accounting for the changes introduced into the model and showing several additional metrics.

The paper is organized as follows: Section 2 is dedicated to related works, where all the most relevant examples of large-scale simulators are discussed. For each of them we compare the complexity based on the extension of the inspected area, number of simulated trips and considered time window. In Section 3, we describe our initial dataset, its characteristics and the area it covers. A complete, detailed description of the modeling methodology is given in Section 4. This section can be used as a step-by-step guide to assemble similar models for other cities. Then, in Section 5, we show a sample collection of results that can be derived from the model. A validation process for our results, using real street sensors data, is also presented. Conclusions and future work appear in Section 6.

## 2 Related Works

In the literature there are just a few examples of traffic simulators that can be considered large-scale models. The TAPASCologne dataset, an open-source project covering the greater urban area of the city of Köln, Germany [10] is an example of a large-scale urban traffic simulator modeling a 400-Km$^2$ area with a total of 1.2 million individual trips in a full day. Thanks to its high level of realism, it is one of the best examples of large-scale simulators today. Another instance of urban mobility model is the traffic simulator of Bologna, Italy [11]. It models a typical weekday traffic pattern (22,000 vehicles in the morning rush-hour), focusing on an area of 28 Km$^2$ around the city center.

Recent literature counts the traffic models of LuST (Luxembourg SUMO Traffic) [12], covering highways and major roads in Luxembourg City and MoST (Monaco SUMO Traffic) [13] focused on the Principality of Monaco and adjacent areas. The LuST scenario covers a 156-Km$^2$ region and simulates morning and afternoon traffic from an original dataset of 14,000 vehicles collected over 16 hours. MoST, instead simulated 35,000 vehicles in the morning rush hour. A much larger model is the ITS Austria West scenario [14] covering a map of around 245,000 nodes and 320,000 edges with a total road length of 27,000 Km. Totalling 1.2 million routes and 1.6 million vehicles over a full day, the ITS Austria West is one of the largest scenarios, although the dataset is not freely available to the research community because some of the information it relies on is proprietary.
Recently, mobility models were proposed by additional works. En Route [15] propose a London macro mobility model. Authors generated an O/D matrix from synthetic data of cameras in London and real taxi traces from other cities. Simulation was performed over two hours with 30,000 vehicle trips modeled. Another example comes from a 800-Km$^2$ simulator of Shanghai [16]. In this work, 13,750 real taxi traces were used in a macro-mobility study to create an Origin/Destination matrix of 17,7 million trips for a multi-days scenario. From this, a SUMO simulator was developed, inspecting vehicles divided between morning and evening peaks: 65,000 on a week day and 60,000 on a weekend day.

The projects mentioned in this section represent all the main existing large-scale traffic models. Our work aims at extending the literature and at proposing a more complex, complete model. Many of the studies previously cited used synthetic traffic data for generating their O/D matrices. Few of them generated a simulator inspecting a full-day scenario. To provide a quantification of the size of our model, 2,200,000 car trips are simulated in the span of 24 hours, making it one of the most extended examples in the literature for a large-scale microscopic traffic simulator, as is also reported in the official SUMO wiki [17]. Compared to the ITS Austria West scenario, our model considers about twice as many vehicles in a much smaller area, thus resulting in a much greater complexity.

# 3 Initial Dataset and Case Study

5T is a private company owned by local public institutions (City of Turin, Metropolitan City of Turin and Piemonte Region) focused on ITS and mobility. It operates the Traffic Operation Centre (TOC), which provides several mobility services, both at urban and regional level, enabled through the integration of ITS (e.g., traffic monitoring, traffic lights control, limited traffic zones control, parking information and others). Regarding traffic monitoring and control, the TOC integrates traffic sensors (approximately 1,700), floating car data, 26 information display panels, 71 traffic cameras, an adaptive traffic control system (managing over 300 urban traffic lights in the city of Turin) and a traffic SuperVision system (SV), devoted to real-time traffic state estimation also in areas not covered by sensors.

The results presented in this paper stemmed from the analysis of traffic flows collected by the 5T fixed sensor network and of the demand data (O/D matrices) provided by the 5T SV system. The last large scale review of O/D matrices was conducted in 2011, using the data derived from the official national Census and from a periodic metropolitan mobility survey. From that date on, matrices were periodically updated with historical data collected by sensors, through the application of a specific algorithm that aims to minimize the difference among estimated and measured traffic flows. Despite this non-uniform composition, O/D matrices are reliable and they have been used over time as private traffic demand data source by local authorities.

The content of the O/D matrices is represented as a database whose records correspond to the traffic flow in terms of number of vehicles per hour (Veh/h) for each Origin/Destination pair. In order to compute it, the SV system applies the principle of traffic macro-simulation to estimate the base traffic states in terms of traffic flows, travel times, speed and vehicle densities. Base traffic states, estimated every hour of the day, are then integrated with traffic data measured in real-time by sensors, taking also into account any event (roadworks, closures, etc.) occurring on the road network. This process is performed every 5 minutes.

Of course, different O/D matrices can be observed in different days of the week, specifically weekdays, pre-public holiday days (including Saturdays) and public holidays (including Sundays). A further subdivision accounts for three main periods of the year: school days, non-school days and "summer closure" days (typically 2 weeks in August). For the scope of this work, we considered a typical school weekday O/D matrix.

Origins and Destinations in the matrix are outlined as Traffic Assignment Zones (TAZs), which represent aggregation of Census Cells (i.e., basic elements of the National Statistics Institute official zoning). TAZs have been defined in Turin in '90s by the Turin Metropolitan Mobility Agency and used by local authorities as reference for their planning activities.

The case study we focused on is a 600-Km$^2$ area in and around the Municipality of Turin, including its suburbs: Moncalieri, Nichelino, Borgaretto, Orbassano, Beinasco, Grugliasco, Collegno, Cascine Vica, Pianezza, Druento, Venaria Reale, Borgaro Torinese, Mappano, Settimo Torinese, San Mauro and Revigliasco for a total of 257 TAZs and an estimated population of more than 1,200,000 residents.

# 4 Methodology

We now summarize the methodology followed in order to construct such a complex model, highlighting the three main steps we followed. Further details are also available in [7]. First of all, we describe the process of creating an accurate road graph of the study area. In the following step, we discuss the procedure to create all vehicles routes starting from the original O/D matrix. Lastly, we review all the additional improvements and tweaks adopted to reach the goal of a complete full-day traffic simulation that adheres as much as possible to the traffic situation reported by city sensors.

## 4.1 Road Graph Creation and Map Editing

The graph representing the road network of the study area was created with the JOSM tool [18], an extensible editor for OpenStreetMap [19] in Java 8+, yielding an OSM file. The OSM file has an eXtensible Markup Language (XML) format and it contains all the information that can be Georeferenced and linked to a bi-dimensional topography map representing highways, roads, railways, bike paths and waterways. Additional elements, such as buildings, parks, rivers and other points of interest can also be imported.

The mobility simulator we used in this project is SUMO (Simulator of Urban Mobility) [20], an open source, highly portable, microscopic and time-continuous road traffic simulation package designed to handle large road networks. In particular, we used version 1.1.0, developed in December 2018. SUMO is a microscopic simulator, thus explicitly modeling each vehicle as it travels on its own route. We used NETCONVERT [21] (one of SUMO's helper tool) to create a network file in SUMO-readable XML format from the OSM file of the area of interest. Since this study is focused on vehicular traffic only, the infrastructure considered in the OSM file was filtered. The resulting network is composed of almost 33,000 nodes, 66,000 edges and more than 6,500 Km of roads (more than 7,700 Km considering multiple-lanes roads).

This initial set-up is however far from accurate, and the resulting network has many errors. Streets are often tagged incorrectly and require manual post-processing, while, in other instances, it is the NETCONVERT conversion process which is faulty. One must never forget, though, that while we aim at creating a faithful representation of the real Turin road map, it nevertheless remains a model and it carries inconsistencies and approximations like all models do. Most corrections in the map were applied either manually or through scripts, acting preferably on the ones that were responsible for high, unrealistic traffic congestion situations.

### *Speed limits*

An essential editing job targeted the nominal speed limits of network edges, which is assigned by SUMO following traffic rules. This clearly differs from the average speed of real vehicles on the corresponding street. On the one hand, a real vehicle often has to contend with asphalt conditions

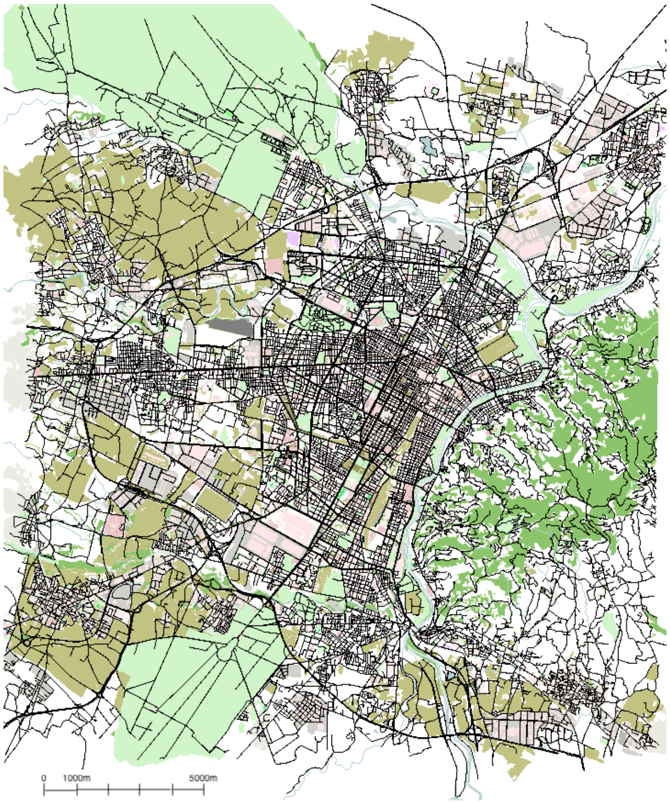


Fig. 1. Map of the Municipality of Turin and suburbs, object of our model

or lane restrictions due to, e.g., double-parked cars. On the other, it is not infrequent that a real vehicle travels faster than the limit, if traffic and road conditions allow it.

We thus set the speed limit of a network edge to the average speed that would be experienced by a vehicle if it were the only one in that street section. This situation is called Free-Flow or Flow-0 traffic condition and it is the input we introduced in order to study how speed decreases in congestion periods. To extract the Flow-0 average speeds we used an API of Google Maps called Distance Matrix [22]. This matrix yields the realistic time in seconds that a vehicle would take to travel from a given Origin to a Destination choosing the depart time and the traffic conditions (optimistic, medium, pessimistic). Since we are interested in an ideal situation with no vehicles at all, we used the optimistic traffic condition option, starting at 4:30am, a very low congestion situation for Turin (and for any city, for that matter).

### *Street width and lanes*

Many roads in Turin lack a lane-dividing line, but are wide enough to allow two or more vehicles to travel in parallel. These roads are incorrectly modeled in SUMO as being single-lane, or, in general, as having fewer-than-expected lanes. As a result, we would often detect congestion in areas where traffic is actually fluid, because vehicles were unnecessarily squeezed into a single lane.

Lacking an online database reporting the exact number of lanes, the full length of some primary roads was checked using Google Maps Street View and the precise number of nominal or theoretical lanes was updated. Specifically, we manually edited more than 100 primary roads.

After the described correction, the network topology we ended up with is shown in Fig. 1.

### *Intersection management*

The management of intersections presented further problems: priority for intersections in SUMO is only right-before-left, while roundabouts are usually regulated by a left-before-right priority. The priority of more than 500 roundabouts was therefore manually fixed.

Another major editing job targeted the phases of traffic lights. NETCONVERT creates surreptitious traffic lights phases in intersections tagged as regulated by a traffic light in the OSM file. These phases are clearly oblivious of the actual traffic flows and, in the specific case of the Turin map, they often resulted in long queues at the intersection. We corrected the phases of more than 900 traffic lanterns using data provided by the 5T SV system, although a new problem presented itself: phases in Turin traffic lights change throughout the day according to a phase-adaptation algorithm called UTOPIA [23] which takes into account live traffic patterns. Importing such a behavior in our model would have increased its complexity manyfold. We thus averaged phase durations during a single day and then manually inserted the averaged values in more than 800 intersections regulated by traffic lights. Fig. 2 shows their distribution over the model map. We privileged the traffic lights that had a larger impact on the fluidity of the traffic.

Additionally, some synchronization was needed in order to have the so-called green-wave effect on a road. SUMO provides a Python script called `tlsCoordinator.py` that manages the traffic light synchronization, given the traffic pattern of a single hour. The `tlsCoordinator.py` script does its job but it nevertheless needs some manual corrections. A compromise was therefore needed to account for the unavoidable discrepancies between our manual interventions and the real traffic light situation in Turin. We noticed that the activation of all traffic-light-regulated junctions was responsible for the increase in congestion. Thus, through an appropriate SUMO option, we enabled traffic lights only

Fig. 2. Traffic lights distribution over the map.

for traffic-saturated edges, while all the intersections with low traffic conditions were treated as priority-based, i.e., without a traffic light. As a result, a mixed scenario is modeled, where traffic lights are activated only in congested intersections, preventing inaccurately-set traffic lights from hindering the flow in low-traffic situations. When a low-traffic situation occurs in a traffic-light-regulated junction, SUMO allows cars to move following the classic priority-based scheme as if all traffic lights were flashing amber.

### 4.2 Traffic Assignment

Creating a realistic map would not serve any purpose unless it were coupled with realistic input traffic. We thus used the 5T O/D matrices to populate the map with traffic that matched the actual one as closely as possible. As a first step, we isolated the traffic flows with Origin and Destination in a TAZ inside the map described in the previous Section, considering only a typical school weekday. The O/D matrix was then translated into SUMO trips, mapping each of them into origin edge, destination edge and suitable depart times (the latter generated randomly within each hour).

Generating an O/D edge pair and a depart time according to the flow is quite a trivial task. Given an O/D pair and a time of departure for each vehicle, the challenge is the choice of a route that leads the system to a stable situation even in congested conditions. Consider how a human driver behaves (especially when equipped with a navigator) when confronted with increasing road congestion: if possible, it will not add one more car to the queue, but will look for alternatives.

In mobility simulation, this task traditionally falls to Traffic Assignment algorithms, of which there are many examples in the literature, starting from some early studies by Wardrop [24] back in 1952, and later by Liu and Fricker [25] in 1996 on a Stochastic User Equilibrium (SUE) with the Logit method. In 1998, Gawron's dissertation [26] defined the SUE-Gawron assignment. Other recent studies, like [27] or [28], compare the performances of different Traffic Assignment algorithms, reaching the conclusion that there is no optimal Traffic Assignment method and its performance highly depends on the use case. For this reason, in [29] different use cases are analyzed and the Traffic Assignment problem of many cities are addressed. Unfortunately, Turin is not among the considered use cases.

After a thorough analysis of alternatives, described in detail in [7], we chose the Incremental Traffic Assignment (ITA) method. ITA is a greedy algorithm in which fractions of traffic volumes are assigned in steps, through a Dijkstra algorithm for shortest path search in a graph with weights. In our case the used weights are the edge travel times. In SUMO, the method to compute a shortest path is called DUAROUTER [30] and, in order to apply ITA to our scenario, a DUAROUTER algorithm is executed for each vehicle at its depart time. This allowed a vehicle to compute its optimal route based on the current traffic conditions, just like a human driver is expected to do.

It must be pointed out that ITA is not optimal if used in a single run. Indeed, in order to achieve a good assignment, it should be used iteratively. There are two methods for such an iterative assignment: the Microscopic Traffic Assignment method and the Macroscopic Traffic Assignment method. In the former, a simulation is run for each iteration and travel times on network edges are measured. Then, travel times obtained in this way are given as input to the ITA algorithm on the next iteration step and the cycle goes on until the maximum number of iterations is reached. Unfortunately, this approach is too time consuming: in order to have an optimum Traffic Assignment, thousands of simulations have to be run.

The Macroscopic Traffic Assignment aims to imitate the Microscopic one, but in a faster way. Indeed, instead of running a simulation for each step, it leverages mathematical resistive functions that approximate the travel time increase when traffic flows surge. In SUMO this method is implemented by MAROUTER [32], a tool that uses a hard-coded capacity-constraint function based on speed limits, lane numbers and edge priorities to compute travel times and flows from traffic density. The Macroscopic Traffic Assignment method is, of course, less precise than the Microscopic one: due to approximations of mathematical functions, travel times are computed instead of measured. Nevertheless, its outputs have proven to be consistent, as described at length in [7], and this is our solution of choice. Specifically, we let the MAROUTER tool go through 50 iterations, after which no further significant improvements were observed.

### 4.3 Improvements and Assumptions

Despite all the corrections on the model topology previously described, our simulator is still not capable of handling the nominal full-day traffic pattern. Only a small percentage of a full-day traffic can be successfully introduced, at this stage. Therefore, in order to increase the traffic demand, some improvements have been adopted.

We noticed how the route creation, during the traffic assignment phase, tends to favor the usage of main roads with multiple lanes and, conversely, it assigns few or even zero vehicles to side roads and little residential streets. This behavior can be considered quite realistic, since most drivers are likely to choose for their trips fast, multiple-lanes roads instead of single-lane streets in residential areas. However, this is not always the case for the start and the end of a car trip. Indeed, it is reasonable that drivers start and/or end their trip near a parking area or close to home, which usually are located in residential zones. Additionally, it is a typical feature of the city of Turin to pair its main avenues with service roads ("controviali" in Italian). They are single-lane streets running alongside both directions of an avenue, and are mainly used for parking or by local businesses to avoid hindering faster traffic on the main avenue. Even if service roads have a slow average velocity and are not considered main roads, it is quite reasonable that a car trip starts and ends there.

Thus, we used a Python script to extend the first and last parts of car trips. To achieve this goal, origins and destinations inside TAZs are no longer chosen randomly among all the road edges of the zone, but only among residential streets and service roads inside the area. A shortest path procedure with DUAROUTER is then performed to connect the new trip origin with the old one and the same operation


Fig. 3. Snapshots from the final TuST dataset. Red vehicles are slow or stopped, blue ones are moving fast. Morning peak of 8:00am (left) and afternoon peak of 17:00pm (right).

is repeated for the destination. The new route obtained in this way is finally substituted to the corresponding old path in the route file.
Since such operation is quite time consuming, it is not reasonable to perform this process on all the 2,200,000 car trips. Consequently, only a portion of all trips was edited in this way, in order to achieve a moderate vehicles scattering across all the map. After this improvement, we also noticed an increase in the traffic our model is capable to address, since now the route assignment spreads more traffic among some secondary and tertiary streets, which were unused before. We will get back to this topic when discussing results, in the next section.

The modification that has the greatest impact on the outcome was the use of the SUMO option for mesoscopic simulations, called MESO [33]. Using SUMO with the MESO option enabled has vehicles on every edge grouped in traffic queues, where a vehicle at the top of the queue can “jump” to the next queue in a neighboring edge. Every queue is treated separately by the system, resulting in a much simpler management of the entire simulation. Indeed, computing vehicle movements with queues, makes simulations run up to 100 times faster than the microscopic model of SUMO without the MESO option. Moreover, due to the jump movements of vehicles from one queue to an other, it is more tolerant of network modeling errors than SUMO without the MESO option. For this reason, it enormously increases the percentage of traffic demand our model is capable of accepting as input.
Despite all the described benefits, the MESO option carries some limitations. Its queue-based model is optimal for managing edges of very large-scale systems, but it loses all the microscopic mobility at intersections. Furthermore, the mesoscopic model is edge-based, so no lane-based output is possible. All the information related to single lanes of a street are aggregated on an edge-based output. Anyway, as detailed later, output traces can still be considered microscopic and can be used as an input for a microscopic model, too.

TABLE 1
TuST Scenario in Numbers

| | |
|---|---|
| Area | 602.61 $Km^2$ |
| Traffic Assignment Zones | 257 |
| Total nodes | 32,936 |
| Total edges | 66,296 |
| Total junctions | 22,209 |
| Traffic lights | 856 |
| Roundabouts | 501 |
| Total length edges | 6,570.28 Km |
| Total length lanes | 7,723.40 Km |
| Vehicle length | 4.3 m |
| Total vehicle trips | 2,202,814 |
| Ended vehicle trips | 2,197,672 |
| Final waiting vehicles | 0 |
| Final running vehicles | 5142 |
| Teleports | 334 |
| Collisions | 0 |
| Halting | 19 |

Launching the SUMO command, we are eventually able to run a full-day traffic pattern simulation. As a result, the road traffic at morning and afternoon peaks, in Fig. 3, looks like a very reasonable picture of the actual traffic pattern of Turin (as will be quantitatively assessed later). In the figure, the orbital motorway around the city is outlined in bright blue, corresponding to speeds higher than 90 Km/h. Large portions of urban roads are in violet, indicating fluid traffic conditions. The traffic congestion is denoted by red regions: dark red regions indicate vehicles that move at slower speeds, between 30 and 50 Km/h; bright red areas denote congested traffic situations, where vehicles are stuck in queues.

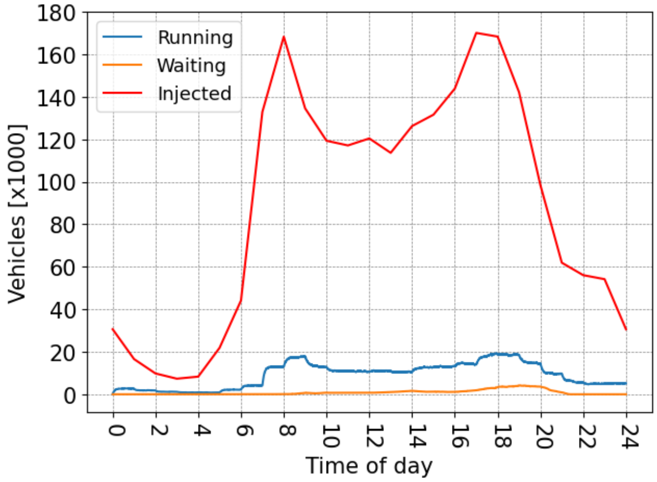


Fig. 4. Stability study result on full-day traffic demand.

Following the described methodology, we ended up with a huge traffic model, covering a vast area in and around the city of Turin. To better understand the magnitude of our model, some characteristics are reported in Table 1.

## 5 Results

This section showcases sample results collected through the TuST simulator. The results we present aim at demonstrating the goodness of the model: a mobility analysis of the inspected area is not the purpose of this paper. Since many types of measurements can be performed with TuST, we decided to categorize the results based on the map element they focus on. The global output covers the entire map and the simulation as a whole, providing useful indication to assess whether the simulation is meaningful or not. We then shift our attention to traffic on single street portions. These kind of measurements are very important for the validation process with real traffic data from street sensors. Finally, the inspection of vehicle routes is discussed. They constitute the vehicle traces we also provide in our GitHub repository [9], and they can be used as a starting point for any kind of mobility analysis or testing of vehicular application.

### 5.1 Global output

In order to inspect any metric from simulations, some important files are output by TuST. The *Summary* file contains, for each time step, all the aggregated data regarding vehicles in the simulation, i.e., the total number of vehicles running, the amount of vehicles that ended their journey and the number of cars waiting to enter the model. The waiting vehicles are those that are scheduled to enter at a given time, but, due to congestion in the origin area, do not find room in the map. Those vehicles have to wait for the congestion to alleviate in that area, and will be inserted later than it was planned. Together with the number of active cars in the model, for each time step, it constitutes the main index of congestion in the model.

From the *Summary* output, we can extract global information of the whole simulated environment. In particular, if the number of running cars, as well as of the waiting ones, increases exponentially, it is a clear index of an unstable simulation, which means that the traffic demand given in input has not been sized correctly. A stability study is thus required.

The stability study plot representing a stable simulation is shown in Fig. 4, where the injected vehicles per hour, according to the O/D matrix, are shown in red; the blue line depicts the number of running vehicles and the orange line represents the waiting ones. In the plot a full-day traffic pattern is presented, showing that a complete simulation, with all trips according to the O/D matrix, can be successfully achieved. It is possible to observe how the blue line of running vehicles resembles the red one of input traffic demand throughout the day. The waiting vehicles peak around 7pm, highlighting the evening rush hour. Nevertheless, the traffic congestion is solved in the next hour. Performing a full-day simulation, with no time limit, the remaining cars end their trips in less than 28 minutes and 57 seconds, highlighting that no persistent congestion affects the simulation.

### 5.2 Road-level traffic inspection

If we want to inspect an edge-based output, i.e., the traffic on a specific road, different data have to be extracted from the simulation. To this end, on every edge of our map, a detector object has been created. Detectors are in charge of counting the number of vehicles passing on a road, thus creating a *Detectors Output* file for edge-based output with a given sample frequency.

From a *Detectors Output* file we can study how vehicles distribute on the map, as shown in Fig. 5. Edges highlighted in red are the streets travelled on by more than two vehicles per minute; in yellow, we can observe edges with more

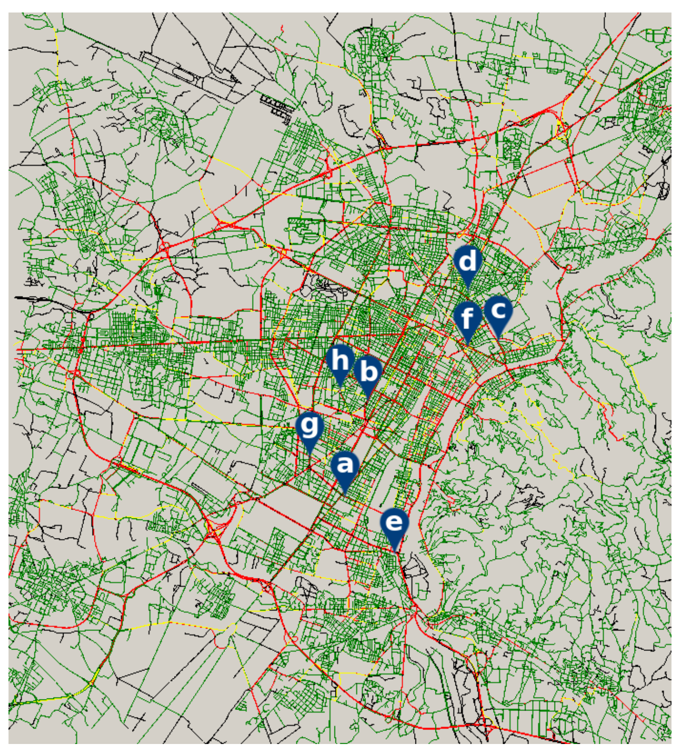


Fig. 5. Vehicles distribution over the map. Blue icons represent the location of street sensors used for model validation.

than one vehicle per minute, while less than one vehicle per minute travels on green streets. The edges colored in black are unused streets, with no traffic throughout the simulation. As we can see, red roads are the orbital motorway or primary roads, built for handling a large portion of traffic demand. On the contrary, green and black edges are residential and neighborhood streets on which few cars travel even in real urban environments. Blue icons represent the exact location of street sensors used for model validation, as described below.

As previously mentioned, we extended the very first and the very last parts of car trips in order to make our model more realistic and to obtain a better vehicle scatter on the map. Thanks to this correction up to 94% of all edges of the map is used, as shown by the tail of the distribution in Fig. 6. Using the path-extension methodology on *all* edges, we could expect to further increase the percentage of edges travelled on by at least a vehicle. However, the procedure would be too time consuming, compared to its added benefits, which would a have mere statistical relevance and very limited impact of the performance.

To validate the edge-based output, data from 5T street sensor on many primary roads were used. A street sensor counts the rate of vehicles transiting on the portion of road under which it is placed. 5T installed hundreds of sensors under Turin's road surfaces. However, over the years, many of them became inactive or were damaged by roadworks. Nowadays, there are still more than 300 active sensors placed under the surface of many roads. Setting our detectors sample frequency to 5 minutes, to match that of 5T sensors, we were able to compare data from simulations with real data of traffic, validating in this way our model.
In Fig. 7 we can see some examples of sensors data compared with data from simulation. The exact locations of the streets for which the comparisons are run are pinpointed in Fig. 5 by blue icons. As we can observe, there is little discrepancy between sensed data, represented by orange dots, and measured data, shown in blue. Interestingly, there is no bias in the comparison, with sensed data either overestimated or underestimated by some percentage, depending on the street.

Finally, a quantitative validation was performed. Indeed, it is reasonable that simulation data represent a more realistic picture of primary roads than of service and residential streets. For this reason, an affinity index was measured for sensors placed under the surface of main streets only. The affinity index was computed as the difference between simulated and sensed data. Fig 8 shows how 80% of the considered sensors have more than 75% affinity with simulation data, and that only 5% of sensors have less than 50% affinity. All in all, this comparison proves that the simulator represents with a very good approximation the actual traffic in Turin on all primary roads of our map.

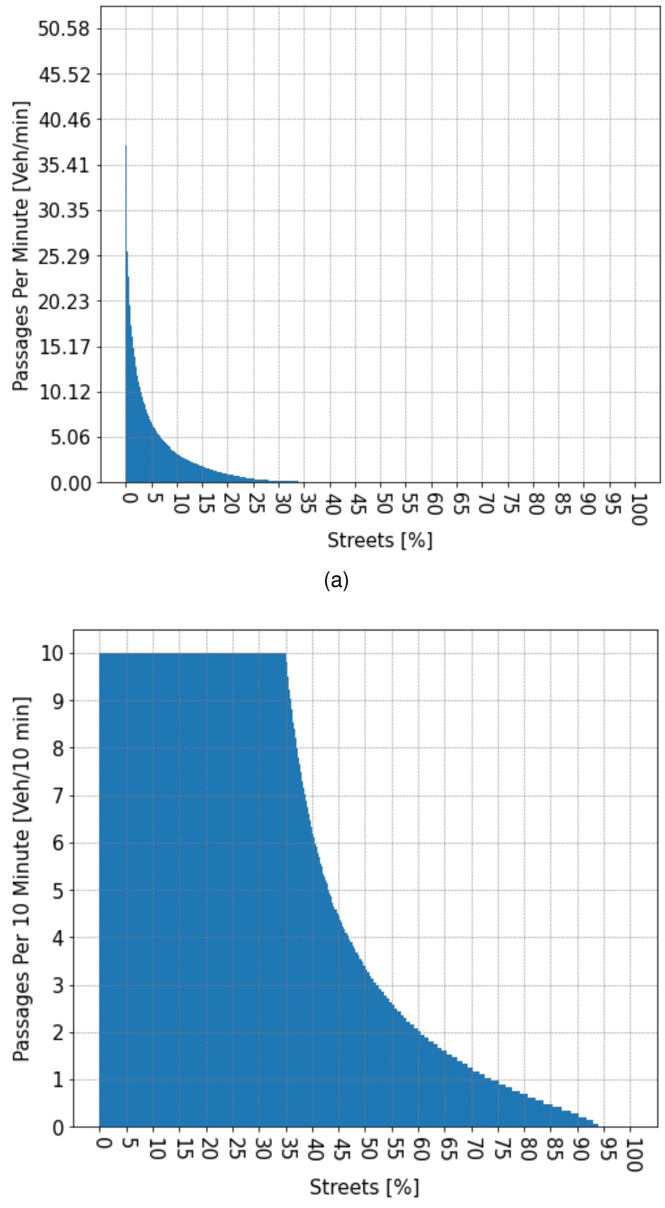


Fig. 6. Distribution of vehicle densities on map edges: (a) total distribution; (b) distribution zoomed over the tail.

## 5.3 Vehicle route inspection

The third type of output we can extract from simulations is the *Vehicle Route Output* file. It constitutes a very important type of output, containing all vehicle traces. For each trace, alongside the complete route flow, some important characteristics of the vehicle path are measured, i.e., exit times of every edge of the route, total length of the complete path and the time at which the considered vehicle reaches its final destination.

We can thus compute some very significant metrics regarding vehicle paths. We can see an example in Fig. 9, where the average route length is shown in blue and the average travel time of a vehicle is plotted in red. Morning and afternoon peaks are recognizable in both plots. In those periods of the day we can notice longer car trips on average, due to commuters' trips, which can be reasonably considered longer than others. Similarly, we can notice higher travel times on average, due to both the greater length of trips and the increasing congestion. The night period is, instead, quite unpredictable. We could classify the trips in this period as night-life trips, resulting in various average lengths. However, regardless of route length, drivers en-

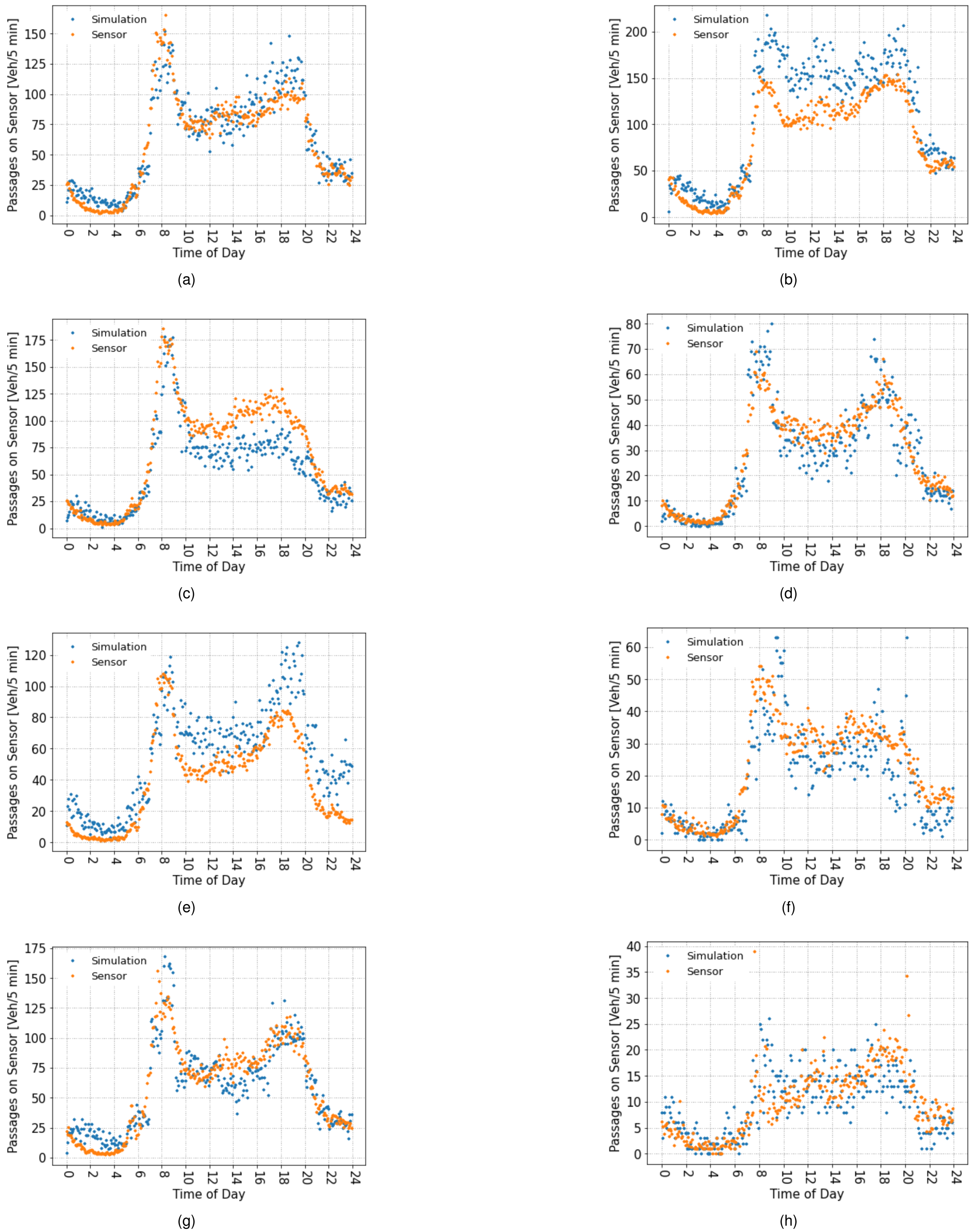


Fig. 7. Validation of simulation measurements (blue dots) with data from 5T street sensors (orange dots). Refer to icons in Fig. 5 for the exact location. (a) Corso Eusebio Giambone. (b) Corso Mediterraneo. (c) Corso Novara. (d) Corso Palermo. (e) Corso Piero Maroncelli. (f) Corso Regina Margherita. (g) Corso Siracusa. (h) Via Monginevro.

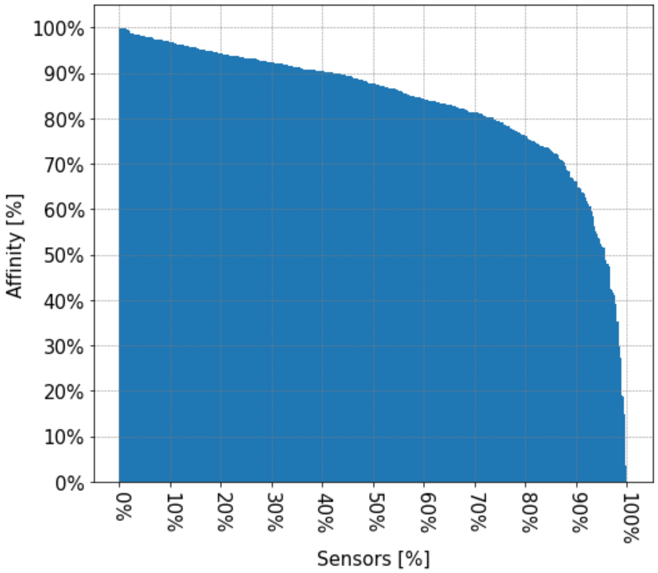


Fig. 8. Quantitative validation analysis on main street sensors.

counter very low traffic conditions at this time, enjoying faster journeys. Overall, we observed that the majority of car trips lasts less than 10 minutes and, in particular, 95% of them were less than 15 minutes and 36 seconds.

Besides this simple route analysis, many other significant results can be obtained from vehicle-based outputs, such as the analysis of traffic congestion areas, or a deep analysis of zones with high pollutant levels. However, vehicle routes constitute a very important output by themselves, since they can be used as realistic car traces for other simulations.
In such a spirit, full vehicle traces and other simulation output have been made available in open-source fashion for everyone. They can be found on the GitHub main page dedicated to TuST [9], which is also linked by the SUMO wiki page related to large-scale scenarios [17].

As stated before, using SUMO with the MESO option enabled for mesoscopic modeling, we lost all the microscopic mobility in our system and we can consider the vehicle routes as mesoscopic. In spite of that, the *Vehicle Route Output* file contains all route edges which could be given

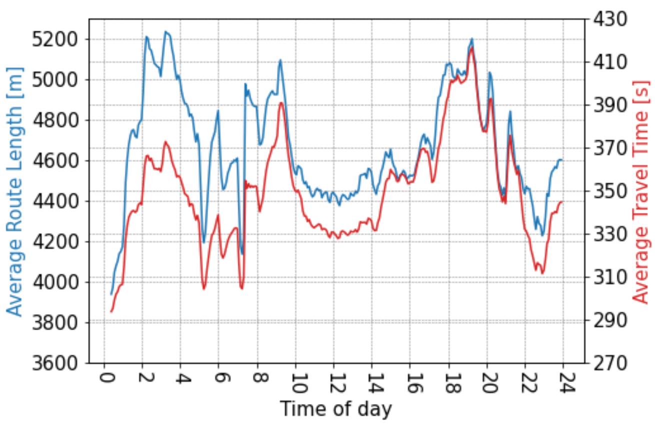


Fig. 9. Mobility analysis: average route length vs. average travel time.

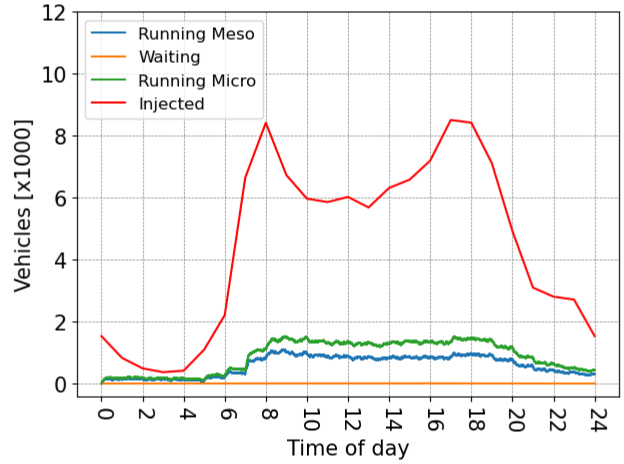


Fig. 10. Comparison between mesoscopic and non mesoscopic simulations.

in input to a microscopic model, too. In order to prove that, we compared a microscopic simulation with 5% of injected traffic (i.e., the greatest traffic demand we are capable of giving as input to a microscopic simulation in order to obtain a stable result) with its correspondent macroscopic one. As we can observe from Fig. 10, there is a slight difference between the running vehicles in a mesoscopic simulation, shown in blue, and the ones of a microscopic simulation, in green. Like previous plots, injected traffic is plotted in red and waiting vehicles are orange. In order to compare those models, a measure of "discrepancy" was defined. We compared the transit rate of vehicles on edges in each simulation and we measured an average discrepancy of 9.18% between the two models.
We can therefore claim that also traces carrying greater traffic can be fed to microscopic models. Even if using the complete dataset on a 600-Km$^2$ map will surely yield an unstable result in a microscopic model, a more relaxed map could be built from only few TAZs. In such a model, from the complete *Vehicle Route Output* file, only flows that travel on the considered TAZs can be extracted and a corresponding O/D matrix can be computed, creating a realistic portion of Turin's traffic with origin, destination and transient flows for the considered areas. Building a microscopic simulator in this way, could lead to a scenario on a district scale for V2V and V2I (vehicle-to-infrastructure) communication scenarios with real traffic traces, yielding results of unprecedented detail.

## 6 Conclusions and Future Work

Continuously increasing traffic congestion in urban areas results in wasting time and money for thousands of drivers stuck in queues and in the worsening of already critical pollution conditions, calling for hard future urban planning choices. More and more research relies on large-scale simulation tools to conduct urban analysis on mobility and emissions as well as evaluation of ITS applications for vehicular technologies and Smart Cities. However, the literature concerning models that can investigate wide-ranging scenarios, up to entire cities, is surprisingly scarce.

In this work we presented a contribution in the form of a large-scale urban mobility simulation model and dataset, describing how to build a complex model, showing which kind of data are necessary for its realization and highlighting all the critical aspects in assembling the system. We focused on a 600-Km$^2$ area including the municipality of Turin and its surrounding districts, starting from O/D matrices extracted from the local ITS authority data and TAZs defined by Turin Metropolitan Mobility Agency and used for every transportation/economical/statistical analysis.

We compared mesoscopic and microscopic simulations methods, showing that both methods are valid. The choice of what should be used depends on whether the models have to address large mobility analysis or to study micro-mobility aspects. Results demonstrate how such a large model is capable to absorb a full-day traffic demand in input at a cost of minor simplifications.

Validations using traffic sensors of the local Info-mobility Agency were successfully performed, resulting in acceptably little discrepancy levels on many sensors of the primary roads we inspected.

As future work, the model could be further improved. An automated deep learning method for image processing could be used in order to correct the remaining map errors on secondary and residential streets. This would lead to a model with higher realism also on not-crowded roads. Envisioned uses of our model could lead to critical emissions studies. One can also test large-scale data dissemination policies or evaluate how V2V and V2I applications perform in a scenario as wide as an entire city, using real traffic flows.

## ACKNOWLEDGMENTS

The present research was conducted in the Future *Urban Legacy* Lab (FU*L*L), an interdisciplinary and interdepartmental laboratory of Politecnico di Torino.

## REFERENCES

[1] D. Schrank, B. Eisele and T. Lomax, "2019 urban mobility report," Texas A&M Transp. Inst., College Station, TX, USA, Aug. 2019.

[2] S. Wang, S. Djahel and J. McManis, "An adaptive and vanets-based next road re-routing system for unexpected urban traffic congestion avoidance," *IEEE Veh. Netw. Conf. (VNC)*, pp. 196–203, December 2015.

[3] *Torino Centro Aperto*, Muoversi a Torino (MATO) and Città di Torino (Municipality of Turin), [Online], Available: https://www.muoversiatorino.it/torinocentroaperto/

[4] A. T. Akabane, R. L. Gomes, R. W. Pazzi, E. R. Madeira and L. A. Villas, "Apolo: A mobility pattern analysis approach to improve urban mobility," *IEEE Glob. Comm. Conf. (GLOBECOM)*, pp. 1–6, December 2017.

[5] G. Manzo, M. A. Marsan and G Rizzo, "Performance modeling of vehicular floating content in urban settings," *IEEE 29th Int. Teletraff. Congr. (ITC 29)*, pp. 99–107, September, 2017.

[6] A. Memedi, C. Sommer and F Dressler, "On the need for coordinated access control for vehicular visible light communication," *IEEE 14th Conf. Wir. On-demand Net. Sys. and Serv. (WONS)*, pp. 121–124, February 2018.

[7] M. Rapelli, C. Casetti and G. Gagliardi, "TuST: from raw data to vehicular traffic simulation in Turin," *IEEE/ACM 23rd Int. Symp. Distrib. Simul. and Real Time Appl. (DS-RT)*, pp. 1–8, 2019.

[8] 5T Srl, "5T Torino," 5t.torino.it, http://www.5t.torino.it

[9] M. Rapelli, *Turin SUMO Traffic Scenario (TuST)*, [Online], Available: https://github.com/marcorapelli/TuSTScenario

[10] S. Uppoor, O. Trullols-Cruces, M. Fiore, and J. M. Barcelo-Ordinas, "Generation and analysis of a large-scale urban vehicular mobility dataset," *IEEE Trans. Mobile Comput.*, vol. 13.5, pp. 1061–1075, 2013.

[11] V. Caiati, L. Bedogni, L. Bononi, F. Ferrero, M. Fiore and A. Vesco, "Estimating urban mobility with open data: A case study in Bologna," *IEEE Int. Smart Cities Conf. (ISC2)*, pp. 1–8, September 2016.

[12] L. Codecá, R. Frank, S. Faye and T. Engel, "Luxembourg sumo traffic (lust) scenario: Traffic demand evaluation," *IEEE Intell. Transp. Syst. Mag.*, vol. 9.2, pp. 52–63, 2017.

[13] L. Codecá, J. Erdmann and J. Härri, "A SUMO-based parking management framework for large-scale smart cities simulations," *IEEE Veh. Netw. Conf. (VNC)*, pp. 1–8, December 2018.

[14] K. H. Kastner and P. Pau, "Experiences with SUMO in a real-life traffic monitoring system," *SUMO 2015–Intermodal Simul. for Intermodal Transport*, vol. 1, 2015.

[15] R. Ketabi, B. Alipour and A. Helmy, "'En route: Towards vehicular mobility scenario generation at scale," *2017 IEEE Conf. Computer Comm. Workshops (INFOCOM WKSHPS)*, pp. 839–844, 2017.

[16] F. Xia, A. Rahim, X. Kong , M. Wang, Y. Cai, and J. Wang, "Modeling and Analysis of Large-Scale Urban Mobility for Green Transportation," *IEEE Trans. Ind. Inform.*, vol. 14.4, pp. 1469–1481, 2017.

[17] *Turin SUMO Traffic (TuST)*, Eclipse Foundation, [Online], Available: https://sumo.dlr.de/docs/Data/Scenarios.html

[18] *JOSM*, [Online], Available: https://josm.openstreetmap.de

[19] *OpenStreetMap*, OpenStreetMap Community, [Online], Available: https://www.openstreetmap.org

[20] *SUMO - Simulation of Urban MObility*, (versio 1.1.0), Eclipse Foundation, [Online], Available: http://sumo.sourceforge.net/userdoc

[21] *NETCONVERT*, (versio 1.1.0), Eclipse Foundation, [Online], Available: http://sumo.sourceforge.net/userdoc/NETCONVERT.html

[22] *Distance Matrix API*, Google Maps, [Online], Available: https://developers.google.com/maps/documentation/distance-matrix

[23] MIZAR Automazione Spa, *UTOPIA – Urban Traffic Control System Architecture*, Verona, 2012.

[24] J. G. Wardrop, "Wardrop of Some Theoretical Aspects of Road Traffic Research-Road Paper," *Road Engineering Division*, 1952.

[25] S. Liu, and J. D. Fricker, "Estimation of a trip table and the Θ parameter in a stochastic network," *Transp. Research, Part A: Policy and Practice*, vol. 30.4, pp. 287–305, 1996.

[26] C. Gawron, "Simulation-based traffic assignment – computing user equilibria in large street networks," Ph.D. Dissertation, University of Köln, Germany, 1998.

[27] T. V. Mathew and K. V. Krishna Rao, "Introduction to Transportation Engineering," *Traffic Data Collection*, NPTEL, 2007.

[28] M. Behrisch, D. Krajzewicz and Y. P. Wang, "Comparing performance and quality of traffic assignment techniques for microscopic road traffic simulations," *Proceedings of DTA2008*, 2008.

[29] *Transportation Networks*, [Online], Available: https://github.com/bstabler/TransportationNetworks

[30] *DUAROUTER*, (versio 1.1.0), Eclipse Foundation, [Online], Available: https://sumo.dlr.de/docs/DUAROUTER.html

[31] *DUAITERATE*, (versio 1.1.0), Eclipse Foundation, [Online], Available: https://sumo.dlr.de/docs/Demand/Dynamic_User_Assignment.html

[32] *MAROUTER*, (versio 1.1.0), Eclipse Foundation, [Online], Available: https://sumo.dlr.de/docs/MAROUTER.html

[33] *Simulation/Meso*, (versio 1.1.0), Eclipse Foundation, [Online], Available: http://sumo.sourceforge.net/userdoc/Simulation/Meso.html

**Marco Rapelli** received his B.Sc. in Telecommunications Engineering (2015) and his M.Sc. in Computer And Communication Networks Engineering (2017) both at Politecnico di Torino. He is part of FULL (Future Urban Legacy Lab), an inter-disciplinary center of Politecnico di Torino, where, in November 2018, he started his Ph.D. In October 2020 he joined the CCS Lab as a visiting Ph.D. at TU Berlin. His main research interests focus on mobility studies and large-scale urban traffic simulators.

**Claudio Casetti** is a full professor with Politecnico di Torino, Turin, Italy. His research interests include vehicular networks, 5G networks, IoT. He is a senior editor of IEEE Vehicular Technology Magazine. He is the chair of the Turin Urban Digital Mobility working group within the Smart Roads project fostered by the City of Turin.

**Giandomenico Gagliardi** is a Transport Engineer in deal with Intelligent Transport Systems since 2008. As Solution Designer of the 5T's "Traffic" Business Unit, he is the company's key person for the Traffic Supervision System, a software platform based on DTA and devoted to real-time traffic monitoring and forecast on a 35,000-Km road network. He is also in charge of the Floating Car Data integration system and of the DATEXII node, the Regional Access Point in the National framework of real-time traffic information exchange. He is member of the Italian DATEXII WG and is involved in several Italian and international projects related to ITS Integration, Smart Road and Cooperative ITS, Transport Planning, Info-mobility.