



HAL
open science

Refining SUMO Simulation Strategies for Realistic Traffic Patterns: Insights from Field Experience

Jean Tshibangu-Muabila, Sebti Mouelhi, Patrick Leserf, Amar
Ramdane-Cherif

► **To cite this version:**

Jean Tshibangu-Muabila, Sebti Mouelhi, Patrick Leserf, Amar Ramdane-Cherif. Refining SUMO Simulation Strategies for Realistic Traffic Patterns: Insights from Field Experience. 2023 7th International Conference on System Reliability and Safety (ICSRS), Nov 2023, Bologna, Italy. pp.237-246, 10.1109/ICSRS59833.2023.10381116 . hal-04408938

HAL Id: hal-04408938

<https://hal.science/hal-04408938v1>

Submitted on 27 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Refining SUMO Simulation Strategies for Realistic Traffic Patterns: Insights from Field Experience

Jean Tshibangu Muabila
ESTACA'Lab, LISV, Université Paris-Saclay
Montigny-le-Bretonneux, Vélisy, France
jean.tshibangu-muabila@estaca.fr

Sebti Mouelhi and Patrick Leserf
ESTACA'Lab, ESTACA, Groupe ISAE
Montigny-le-Bretonneux, Laval, France
first.last@estaca.fr

Amar Ramdan-Cherif
LISV, Université Paris-Saclay
Vélisy, France
rca@lisv.uvsq.fr

Abstract—Urbanization and the surge in vehicle numbers have posed significant challenges, including tackling traffic congestion, reducing ecological footprint, and enhancing safety. In the context of developing efficient and sustainable transportation technologies, this research focuses on gaining insights into traffic patterns, human behavior, and their impact on urban livability. Drawing on field experience, this study investigates the intricate traffic conditions in the regional shopping center of Saint-Quentin-en-Yvelines using advanced SUMO simulations. By refining traffic models and fine-tuning their parameters, the simulation strategy is gradually adjusted to accurately simulate the traffic dynamics. Generated data are then analyzed to ascertain the usefulness of the whole approach in capturing realistic traffic patterns. The outcomes pave the way for designing smart vehicular applications to improve traffic flow in comparable urban contexts.


Index Terms—Traffic patterns, Simulation strategies, Model refinement, Data visualization and analysis, SUMO package.

I. INTRODUCTION

The exponential urbanization and the escalating growth of vehicles have given rise to a convoluted lattice of interlinked challenges, including effective traffic congestion management, accident prevention, and promoting environmental stewardship and sustainability. Effectively tackling these matters requires a focused examination of traffic patterns in specific congested urban regions, which is a crucial aspect of contemporary traffic engineering. The research work presented in this paper aims to acquire a holistic comprehension of these challenges and offer valuable insights into traffic patterns, human behavior (drivers, riders and pedestrians), and their ramifications on urban living standards. Additionally, this study operates within the context of smart cities, prioritizing the development of powerful and sustainable transportation technologies across diverse domains such as energy supply systems, communication networks, and data analytics. By embracing a systemic and interdisciplinary approach, it becomes feasible to effectively investigate traffic in densely urbanized regions. This endeavor sets the stage for breakthroughs in future smart mobility.

This work is part of a two-stage PhD project. In the first stage, the primary focus is on the development of a simulation-based analytical approach to tackle the difficulties associated with managing traffic jams in highly populated commercial areas. The findings showcased in this paper represent the results achieved during this phase. The objective of the second

stage is to create a novel predictive simulation-based solution by incorporating vehicular communication at both network and application layers into the existing model. This involves the development of IoV (Internet of Vehicles) [1], [2] application utilities aiming at enhancing traffic conditions in the studied area, while considering the requirements identified in the first stage. Furthermore, particular emphasis will be placed on maximizing network performances by smartly distributing relevant connectivity metrics, thereby ensuring real-time cooperation throughout the IoV interconnected systems.

The simulation model discussed in this paper was designed using the SUMO  (Simulation of Urban Mobility) [3], [4] toolkit. It is a powerful open-source traffic simulation software widely used for modeling and analyzing urban transportation systems. It offers a comprehensive set of features and tools to simulate vehicle movements, traffic flow, and road networks. SUMO enables researchers, engineers, and planners to study traffic patterns, evaluate different scenarios, and appraise the impact of various strategies on traffic management efficiency. With its extensive user community and continuous development, the tool has become a popular choice for transportation systems design in both academia and industry.

The developed model is meticulously crafted, drawing upon extensive field experience observations conducted within the regional shopping center of Saint-Quentin-en-Yvelines (SQY) agglomeration in Paris metropolitan area, during peak hours, particularly on Saturday afternoons. By capturing the intricate observation details of traffic patterns and dynamics, the model provides profound insights into the challenges faced within the congested environment of the considered shopping center. It faithfully replicates approximate real-world conditions of traffic congestion and bottlenecks encountered on major roads and highways, ensuring an accurate portrayal of traffic conditions within the studied region.

Through manipulation of key parameters in the simulation model, the traffic flows were deliberately designed through an iterative refinement procedure. Vehicle and pedestrian arrivals and fluctuations were characterized using Poisson distributions that underwent incremental adjustments. These distributions were applied under microscopic car following, lane changing, and junction traffic flow-specific schemes [4], [5]. Simulation realism and effectiveness, stemming from this process, were evaluated by tuning diverse parameters of these schemes.

Upon performing simulations using SUMO, we collected a significant amount of data related to traffic behavior and patterns. This data serves as a valuable resource for further analysis and evaluation of our simulation strategy. Through the graphical representation of generated data, we have confirmed several of our apprehensions regarding various aspects such as traffic densities and congestion levels across multiple zones within the studied area. By performing data analysis, we are able to assess how different parameters and adjustments affect the overall traffic performance, thus identifying potential areas for further improvement. Ultimately, this data analysis not only validates our methodology, but also offers valuable insights that shape the trajectory of the second stage of the project. As previously mentioned, our objective in this subsequent phase is to develop predictive simulation-based solutions by integrating vehicular communications into the existing traffic model.

The structure of this paper is as follows. Section II provides an overview of related works, offering insights into the existing literature. Our research direction is outlined in Section III. In Section IV, we present the foundational knowledge in traffic engineering that underpins our contributions. The simulation methodology and case study are both detailed and discussed in Section V. Section VI delves into data visualization and analysis. The manuscript concludes in Section VII by summarizing the key findings and charting future research paths.

II. RELATED WORKS

In their work [6], the authors propose a framework aimed at calibrating microscopic traffic simulations in order to faithfully reproduce real-world traffic flow conditions. The framework leverages a dataset gathered from the Ohio State University campus bus service, encompassing GPS (Global Positioning System) sensor data, to construct microscopic simulation models and capture traffic conditions using SUMO. To incorporate variability, their approach introduces probabilistic traffic flow generation. Besides, a communication interface is provided to establish real-time hardware-in-the-loop simulations, enabling virtual testing of advanced driver-assistance systems.

In [7], the authors employed a novel simulation approach to emulate traffic in Jiangnan Zone, located in Wuhan, China. Instead of traditional data collection methods, they used geographic population distribution data from public sources. Simulations were carried out using the SUMO tool `activitygen` allowing to produce realistic traffic flow. The accuracy of the simulated traffic was evaluated face to real-time road conditions and historical traffic predictions. The findings highlight the valuable insights provided by the generated traffic data. By employing refined model classifications, the simulated traffic closely resembles real-world conditions.

The authors of [8] tackle the pressing problem of road traffic in major cities, which is exacerbated by the growing global population and the persistent need for transportation. They perform a microscopic traffic SUMO simulation of Coimbra city in Portugal. Through various experimental scenarios involving adjustments in simulated vehicle numbers and other variables,

the findings validate the simulator's ability to accurately reproduce traffic conditions in Coimbra. The integration of origin-destination matrices and routing algorithms based on real data further enhances the simulator's potential for future research pertaining to the case study.

In [9], the authors introduce an object-level mobility traffic perception module to Eclipse MOSAIC framework [10]. This module uses advanced spatial indexing methods for efficient access to traffic entities, particularly moving objects, within a specified field of view. The computational performance of the indexing techniques and their integration with SUMO are evaluated using TraCI [11]. Using this module, Eclipse MOSAIC enables the evaluation of distributed applications that analyze and share information about the close-by surroundings of each object to enhance traffic state estimations.

The methodology presented in [12] describes the development process of a comprehensive traffic simulation covering an expansive area of approximately 400-km² surrounding Turin, Italy. Despite minor simplifications, the initial findings suggest that it is possible to successfully model the traffic in such a large-scale urban region. In [13], the authors introduce a traffic scenario that meets essential criteria in terms of size, realism, and duration, with the objective of establishing a standardized foundation for assessments. They use data from Luxembourg city, which exhibits a typical topology observed in European medium-sized cities, and incorporates genuine traffic demand and mobility patterns. The paper outlines the employed method for constructing LuST (Luxembourg SUMO Traffic) scenario, providing a compendious overview of its key features, along with evaluations and validation results.

The authors of [14] propose a solution to reduce traffic congestion at intersections through temporary control strategies such as optimizing traffic signal timing, implementing two-stage left turn boxes, and assigning traffic directions. Based on a simulation-based optimization model, they estimate vehicle travel time, jam lengths, and key indicators about intersection performance. Analysis of peak hour traffic data from a main intersection in Can Tho city, Vietnam, reveals the necessity for temporary control strategies to address recurrent congestion and mitigate traffic-related air pollution exposure. Their model serves as a valuable tool for evaluating and exploring various traffic scenarios prior to real-world implementation.

III. RESEARCH METHODOLOGY

As previously mentioned in the introduction, the contributions presented in this paper are an integral part of a doctoral research project. The overarching objective is to establish a tooling methodology that effectively tackles the challenges related to traffic flow optimization and network connectivity within IoV environments. In order to lay the foundation for the entire project, the first stage places emphasis on the examination of practical and pertinent urban traffic scenarios involving SUMO simulations. Specifically, it targets the issues associated with congestion and traffic jams in crowded commercial zones. The outcomes and findings detailed in this paper directly stem from this particular phase of the project.

The objective of the second stage is to develop original predictive solutions by integrating networking layers into the existing simulation SUMO model using the multi-domain/scale Eclipse MOSAIC framework [10], and network simulation packages like OMNet++ [15]. This involves the development of IoV applications aiming at enhancing traffic flow in the studied traffic areas, while considering the requirements identified in the first stage of the project. Furthermore, particular emphasis will be placed on network optimization through intelligent distribution of pertinent connectivity metrics, thereby guaranteeing improved communication during simulations.

This second stage is twofold. Firstly, it entails generating simulation datasets that capture the connectivity metrics of the communicating entities (such as vehicles and roadside equipment) over representative time periods while making them in relation with traffic data. Secondly, we can harness the power of machine learning techniques on these datasets in order to forecast network performance and assess the quality of service QoS for each scenario over extended time periods. This process continues to be iterated until configurations align with the desired system and network requirements.

IV. TRAFFIC ENGINEERING

Traffic flow is influenced by various factors such as drivers, traffic regulation devices, and road geometry and signs. These factors, along with their stochastic and unpredictable nature poses challenges in providing theoretical descriptions of traffic flow. As a consequence, simulations and quantitative analysis techniques are commonly employed to study urban road traffic and design more efficient transportation networks.

The study of traffic flow relies on different levels of modeling: *macroscopic*, *microscopic*, and *mesoscopic*. The macroscopic level focuses on overall vehicle flow, including volume, average velocity, density, road capacity, air quality, and travel time. The microscopic level examines individual vehicles, considering their positions, speeds, spacings, accelerations, lane changes, overtaking, pedestrian and cyclist characteristics, and driver choices. The mesoscopic level serves as an intermediate between the microscopic and macroscopic levels.

Traffic volume can be measured and studied daily or hourly. Daily volume represents the average amount of traffic at a specific location each day, while hourly volume captures the traffic during a specific hour. Since traffic flow varies throughout the day, daily volumes are inadequate for analyzing urban traffic. Instead, focusing on hourly volume, especially during peak hours, is essential for understanding and addressing urban traffic issues. Peak hours reflect the heaviest traffic period and offer valuable insights for policymakers to enhance efficiency, road safety, and sustainability in urban transportation systems.

Fundamental traffic flow diagrams [16], [17] (see Figure 1) depict relationships between vehicle density, speed, and flow rate, providing valuable insights for the traffic dynamics and infrastructure planning. These diagrams rely on fundamental traffic flow models defined by Equations (1) and (2), outlining

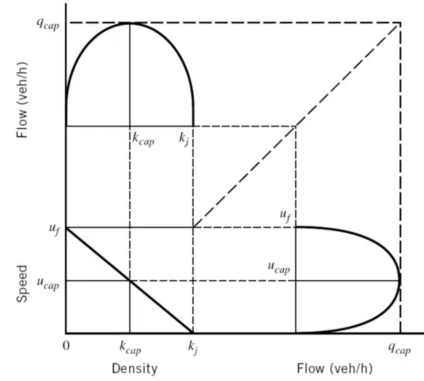


Fig. 1. Fundamental diagrams

the relationship between flow, space-mean speed, and density as per the Greenshields model.

$$q = u \times k \quad (1)$$

where q is the *flow* (in veh/h), u is the *space-mean speed* (in mi/h or km/h), and k is the *density* (in veh/mi or veh/km). The space-mean speed refers to the average speed of vehicles along a designated road section, distinct from the time-mean speed, which represents the mean speed measured at a specific point over a time period:

$$u = u_f \left(1 - \frac{k}{k_j}\right) \quad (2)$$

where u_f is the *free-flow speed* (in mi/h or km/h) and k_j is the *jam density* (in mi/h or km/h) when vehicles are stopped on the road close to each other.

Traffic flow distributions in urban areas measure the vehicle arrivals at a specific position. Vehicle arrivals are not uniformly distributed, and statistical laws, such as the Poisson distribution (3), estimate the probability of observing exactly n vehicle arrivals within a specified time interval t . It is important for urban transport planning, real-time traffic management and simulation, congestion prediction, and infrastructure design.

$$P(n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad (3)$$

The (negative) exponential distribution is derived from the assumption of Poisson vehicle arrivals. It describes the probability of having headways (time intervals) between vehicle arrivals longer than a specific duration t . The distribution is defined by the following equation.

$$P(t) = e^{-\lambda t} \quad (4)$$

Here, λ signifies the *average rate of vehicle arrivals* in veh/s, the same parameter employed in Equation (3).

Given that our simulation model is built using SUMO, the tool incorporates all the theoretical aspects discussed earlier. This includes the effective capture of traffic flow distributions such as Poisson and negative exponential distributions, along with the incorporation of the Greenshields model.

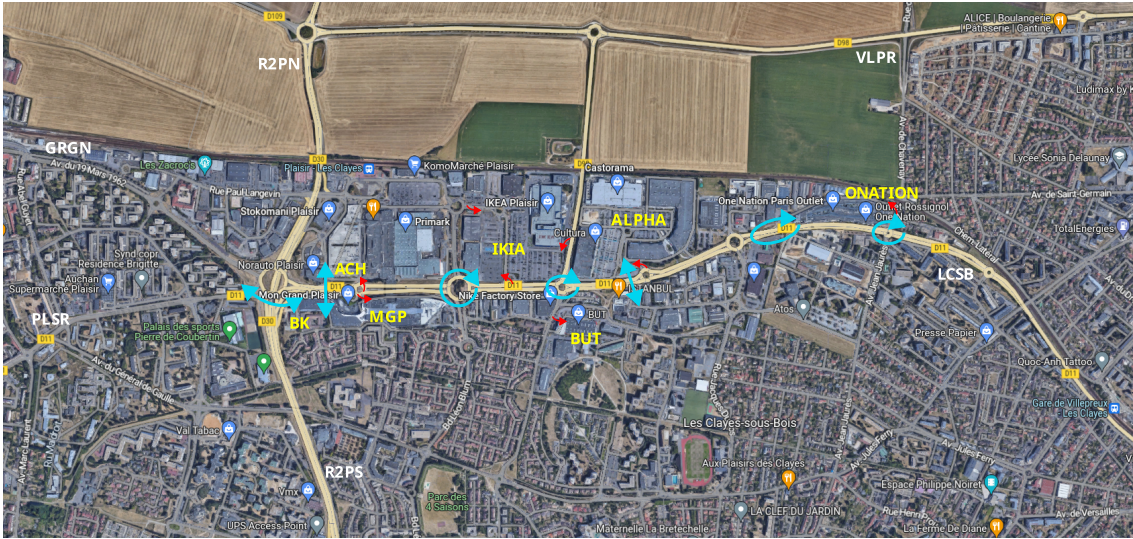


Fig. 2. The regional shopping center of SQY agglomeration. People gathering points and stores (yellow): ACH (Auchan), MGP (Mon Grand Plaisir), IKIA, BUT, ALPHA (Alpha Park), ONATION (One Nation). Arrival (leaving) directions towards (from) the shopping center (white): R2PS (Route des Deux Plateaux, South), R2PN (Route des Deux Plateaux, North), PLSR (Plaisir), GRGN (Gare de Plaisir Grignon, Plaisir downtown), VLPR (Villepreux), LCSB (Les Clayes-sous-Bois). Red arrows represent the entrances to the parking lots. Cyan arrows indicate the presence of significant pedestrian traffic.

This juncture provides an apt occasion to reinforce SUMO's selection. The toolbox is widely respected and used for traffic simulation and modeling. However, toolkit choice hinges on study requirements and goals. While there are alternative commercial tools available like Vissim¹, Aimsun², Pramaics³ and TransModeler⁴, each with its own strengths and weaknesses, SUMO stands out for several reasons. Firstly, SUMO is an open-source software, allowing for free access, customization, and collaboration. Secondly, SUMO offers a modular and extensible framework, enabling the integration of custom models and scenarios to address specific research questions. Thirdly, SUMO benefits from an active community of users, providing extensive documentation and support. Lastly, SUMO seamlessly integrates with other tools and frameworks, facilitating comprehensive analyses of complex transportation systems by incorporating, among others, traffic signal controllers, routing engines, and communication simulators.

V. SIMULATION PROCESS

The simulation process encompasses several essential steps aimed at accurately modeling and analyzing traffic behavior (refer to [18] and [19]). To begin, a relevant traffic scenario is carefully chosen – in this case, the study focuses on the peak hours of Saturday afternoons at the regional shopping center situated in the Saint-Quentin-en-Yvelines (SQY) agglomeration within Greater Paris. Next, the road network is imported from OpenStreetMap (OSM) and meticulously tailored using *netedit* [20] (a tool of SUMO), guaranteeing an accurate portrayal of the shopping center's layout and its vicinity.

Subsequently, both vehicle and pedestrian flows are defined and tailored. Finally, forming an essential part of the simulation setup, an iterative refinement process is applied, driven by repeated simulations, to fine-tune the average vehicle rate parameters (λ in Equations (3) and (4)) across the various flows. This iterative approach culminates in the generation of traffic behavior that closely resembles real-world conditions during the designated peak hours. Together, these steps contribute to a comprehensive and lifelike analysis of the traffic dynamics in the selected regional shopping center.

Figure 2 illustrates the geographic extent of the regional shopping center in SQY (as sourced from Google Maps). The figure's legend provides comprehensive information about the primary stores, arrival (leaving) directions towards (from) the shopping center, the entrances to parking lots, and the areas marked by dense pedestrian footfall. The roads experiencing the highest congestion levels are D11 and D13 (D representing "Départementale", which refers to departmental roads). The study is particularly directed towards examining the traffic of both vehicles and pedestrians along these roads.

A. Road network

The road network generated by *osmWebWizard* underwent several editing steps using *netedit*. The aim was to declutter the map by eliminating non-relevant parking, cycle and service lanes. To guarantee the correctness of highway lane counts, a deep verification process was conducted, correcting any errors found in the imported map. Special attention was given to resolving ambiguities in junction turnarounds (see the example shown in Figure 3) to enhance the road network clarity and precision. Moreover, adjustments were made to improve road geometries with sharp turns, acute radii, and abnormal angles, resulting in smoother and more realistic road layouts.

¹<https://www.ptvgroup.com/fr/solutions/produits/ptv-vissim/>

²<https://www.aimsun.com/aimsun-next-overview/>

³<https://www.paramics.co.uk/en/>

⁴<https://www.caliper.com/transmodeler/default.htm>

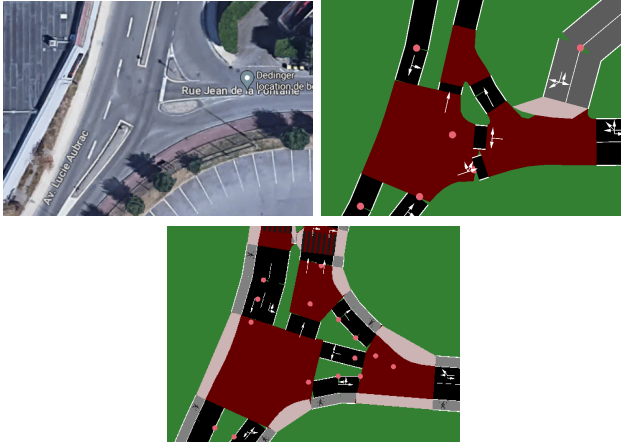


Fig. 3. North junction of Avenue Lucie Aubrac: Google Maps (top left); OSM imported ambiguous junction (top right); Corrected junction (bottom).

In response to abnormal speed reductions observed during initial simulations, a comprehensive analysis was conducted. Factors such as turning radius and angle were carefully examined to pinpoint areas of concern. By analyzing these factors, necessary adjustments were made to improve the traffic flow. In order to mitigate potential conflicts between left-turning lanes at junctions, we resolved the issue by enlarging the radius of critical junctions based on real map data. This modification afforded vehicles greater space to navigate left turns, resulting in a reduced risk of abnormal congestion.

By enacting these adjustments, our objective was to rectify the problematic sections in the road network and improve the simulation accuracy. These interventions enhance the fidelity of portraying vehicle movement in the affected areas, capturing the intricacies and dynamics of real-world traffic scenarios.

B. Vehicle types

Various approaches are available to define vehicular demand in SUMO using existing input data. `netedit` can be employed to visually design traffic demand, generating the corresponding XML definitions. Alternatively, demand files can be textually created and edited, as practiced in this work. The structure of the XML demand specifications used for our simulation model will be outlined later in Section V-C.

A single demand specification comprises two main components: i) one vehicle type `vtype` (several ones could define distinct physical properties and behavior patterns), and ii) the definition of a `vehicle` or repeated vehicles (`flow`) including the depiction of connected `edges` in a `route` entry representing the followed road. It is important to note that specifying a vehicle type is optional, with a default type applied if omitted. Route information can be directly included as origin (`from`) and destination (`to`) edges within a `trip` (single vehicle) or a `flow`, without the need for a complete embedded `route` entry. In these situations, the simulation employs fastest-path routing based on real-time traffic conditions. Note that the possibility to define intermediate edges is available through `via` attribute.

The `vTypeDistribution` entry allows the combination of multiple vehicle types (attribute `vTypes`). As previously stated, these different vehicle types (`vType` entries) represent various demand categories, each characterized by unique physical attributes and traffic behavior. This feature permits designers to assign weights to these types and combine them to influence simulation outcomes. This is particularly useful for replicating real-world traffic, where certain vehicle types may dominate specific scenarios. Urban areas might see more passenger cars with varying driving styles depending on the traffic conditions, while industrial zones might experience consistent traffic flow primarily involving trucks and freight vehicles. The subsequent text outlines the descriptions of the distinct vehicle types we established. These vehicle types are meant to be consolidated into `vTypeDistribution` entries and assigned to the model's demand specifications.

Two global vehicle types `t_pg` and `t_tg` were defined to represent two distinct demand categories: trucks and passenger cars, respectively. These categories are characterized by these attributes: `accel` (acceleration in m/s^2), `decel` (deceleration in m/s^2), emergency deceleration `emergencyDecel`, `length` (the vehicle's length in m), `maxSpeed` (maximum velocity in m/s), `minGap` (the distance to the rear of the leading vehicle in m), `sigma` (driver imperfection ranging in $[0,1]$ with a value of 0 denoting perfect driving), and the individual `speedFactor` which is the vehicles expected multiplier for lane speed limits and the desired maximal speed (defined as truncated normal distribution). Detailed attribute specifications for each vehicle type can be found in [5], [21].

```
<vType id="t_cfm_k" carFollowModel="Krauss"/>
<vType id="t_cfm_wm" carFollowModel="Wiedemann"
  estimation="0.5" security="0.5" />
```

Introducing our second set of vehicle types, we adopted the Krauss car-following model [22], SUMO's default microscopic model, for simulations. This choice is based on its simplicity, utilizing few parameters for computational efficiency and ease of use. Derived empirically, the model captures common driver behavior when following the preceding vehicle. Its versatility spans urban and highway settings, balancing simplicity with realism. Additionally, the model's parameters can be calibrated using real-world data to reproduce the desired traffic features and the common driver actions. The attributes of the previously introduced `t_pg` and `t_tg` are exploited by this model.

We also explored the Wiedemann model [23] with nearly the same attributes, except for two scaled factors `security` and `estimation` ranging in $[0,1]$. The first sets the drivers' desire for security, and the second represents drivers' ability to accurately estimate the situation by anticipating the leading vehicle behavior. This model accurately simulates human driving dynamics by considering factors like anticipation, reaction time, and comfort. Despite expectations with moderate values, it didn't outperform the Krauss model. The simulation uncovered unexpected traffic congestion in areas where a smooth traffic flow was expected. Further research is planned to explore these car-following model aspects more deeply.


```
<vType id="t_lcm_e_m" laneChangeModel="LC2013"
lcStrategic="0.5" lcCooperative="0.0"
lcSpeedGain="0.5" lcKeepRight="0.25"
lcOvertakeRight="0.0" lcOpposite="0.25" />
```

The vehicle type provided above employs the default lane-changing model used in SUMO [24], exhibiting a *moderate* lane-changing behavior. The attribute `lcStrategic` indicates the inclination for strategic lane changes, with higher values resulting in earlier changes. The attribute `lcCooperative` signifies the readiness for engaging in cooperative lane-changing, where decreased values correspond to diminished cooperation. A null value was assigned to it as this lane-changing mode is presumed to be unfavored in the studied area. `lcSpeedGain` is the eagerness towards lane-changing to increase speed, with higher values leading to more lane-changing. `lcKeepRight` is the willingness to comply with the “keep right” rule, resulting in earlier lane-changing with higher values. The `lcOpposite` attribute indicates the willingness for overtaking through the counterflow lane. Higher values lead to more lane-changing.

A second vehicle type, `t_lcm_e_kr`, shares the default lane-changing properties with the moderate type but features different values. It strictly adheres to “keep right” rule, avoiding lane changes whenever possible. The attributes `lcStrategic`, `lcCooperative`, and `lcSpeedGain` are all set to 0, indicating no strategic lane changes, cooperation, or speed gain. Conversely, `lcKeepRight` is set to 1, prioritizing keeping the right lane. `lcOpposite` retains the same value as in `t_lcm_e_m`, indicating a slight tendency for overtaking through counterflow lanes. This will ensure that, especially in congested conditions on fast roads, drivers promptly follow the practice of keeping right, as elaborated further in Section V-C.

In contrast to the previous two types, the third, `t_lcm_e_a`, embodies an assertive lane-changing behavior while utilizing the same default lane-changing model. It exhibits a heightened inclination for strategic lane changes (`lcStrategic` set to 0.75) and a stronger preference for more lane changes to gain speed (`lcSpeedGain` also set to 0.75). However, both attributes `lcCooperative` and `lcKeepRight` are retained at 0, reflecting limited cooperation and a disregard for keeping right. As with the previous types, `lcOpposite` maintains a value of 0.25.

It’s noteworthy that in all the three types, `lcOvertakeRight` (probability of violating rules for overtaking on the right) is consistently set to 0. This decision is motivated by our focus on safe and compliant driving behaviors within the simulations, aligning with real-world practices ideally adhering to rules.

```
<vType id="t_jm_m" impatience="0.25" jmCrossingGap="3"
jmIgnoreKeepClearTime="300" />
```

Concluding our definitions, the type `t_jm_m` defines moderate driver behaviors at junctions. The `impatience` attribute is set to 0.25, reflecting slight considerate driving with respect to higher-priority vehicles [5]. `jmCrossingGap`, equal to 3m, indicates the minimum gap to pedestrians when they approach the vehicle’s trajectory. The last attribute means the cumulative waiting time beyond which a vehicle enters a junction, despite the possibility of generating traffic jam.

In contrast, the `t_jm_a` type characterizes aggressive junction behavior. The `impatience` attribute is set at 1, reflecting a more assertive approach when interacting with higher-priority vehicles. The value of `jmCrossingGap` is set to 1m, denoting a minimal gap as pedestrians approach, leading to more daring driving choices. `jmIgnoreKeepClearTime` is configured for 60s, indicating a threshold at which the vehicle proceeds into an intersection without considering the risk of congestion.

```
<vTypeDistribution id="t_p_jm_m_lcm_e_kr"
vTypes="t_pg t_cfm_k t_jm_m t_lcm_e_kr"/>
<vTypeDistribution id="t_p_jm_m_lcm_e_m"
vTypes="t_pg t_cfm_k t_jm_m t_lcm_e_m"/>
<vTypeDistribution id="t_p_jm_a_lcm_e_kr"
vTypes="t_pg t_cfm_k t_jm_a t_lcm_e_kr"/>
<vTypeDistribution id="t_p_jm_a_lcm_e_a"
vTypes="t_pg t_cfm_k t_jm_a t_lcm_e_a"/>
<vTypeDistribution id="t_t_jm_a_lcm_e_a"
vTypes="t_tg t_cfm_k t_jm_a t_lcm_e_a"/>
```

As stated before, our goal was to allocate unique driving behaviors to the various defined vehicle flow entries in our simulation model. We have created various `vTypeDistribution` entries by combining the established vehicle types, and carefully designed to encapsulate the diverse driving characteristics within the simulated environment:

- The `t_p_jm_m_lcm_e_kr` entry involves moderate lane-changing and cooperative junction behavior for passenger cars, emphasizing right-lane adherence;
- The `t_p_jm_m_lcm_e_m` entry maintains a similar profile while allowing moderate lane-changing behavior;
- The `t_p_jm_a_lcm_e_kr` entry showcases assertive junction behavior while prioritizing right-lane adherence and cooperative lane-changing;
- The `t_p_jm_a_lcm_e_a` entry combines aggressive junction behavior and lane-changing;
- The `t_t_jm_a_lcm_e_a` entry expands assertive behavior to trucks while retaining same junction behavior and lane changes as the previous entry.

C. Demand specifications

Vehicle flows were defined, configured and XML-generated over specific routes chosen from the “*flow (embedded route)*” mode in the “Demand / Vehicle mode” menu of `netedit`. It’s noteworthy that we intentionally avoided using the alternative mode “*flow (from-to junctions)*”. This choice was influenced by the observation that vehicles operating under this mode tend to compute alternate routes to bypass congestion and expedite their arrival at destination junctions. Resultant vehicle trajectories from such flow types would not faithfully depict actual traffic behaviors.

By initially selecting embedded routes, we aimed to capture a more realistic depiction of traffic behavior. Drivers tend to stick to well-known and frequently used routes, often relying on GPS navigation apps on their mobile devices. Deviations from these established routes are typically minimal. That is why *embedded routes* flow type align better with real driving practices, enhancing the precision and realism of our portrayal of traffic dynamics.

The XML code below represents the demand element automatically generated by `netedit` for one of the flows linking R2PS and ACH (see Figure 2). The average arrival rate λ of the flow, specified as a Poisson distribution (see Section IV), can be expressed using the attribute `period="exp(λ)"` [25] where λ is the average arrival rate in veh/s. `begin="0.00"` and `end="3600.00"` indicate that the flow is designed to span at least one hour (3600s), specifically at the peak hour. To capture the peak traffic hour and gain valuable insights on traffic issues of the studied area, we specifically analyze the hourly traffic between 2:30 PM and 3:30 PM on Saturdays (see Section IV). This timeframe, drawn from our field experience and Google statistics, offers essential help to emulate the traffic patterns during peak hours.

```

<!-- Manually defined -->
<flow id="f_r2ps_to_ach" type="t_p_jm_a_lcm_e_kr"
  begin="0.00" end="3600.00" color="white"
  departLane="random" departSpeed="max"
  period="exp(0.034)">
  <route edges="814259719 689277284#0 689277288#0
    689277288#1 689277288#2 881567591#0
    881567591#1 881567591#2 542209564#3
    542209564#4 542209564#5 542209564#6
    542209564#7 760744245"/>
</flow>
<!-- Randomly generated -->
<trip id="veh677" type="t_p_jm_a_lcm_e_a"
  depart="1947.11" departLane="best"
  departSpeed="max"
  from="-88439574#1" to="44419464#0"/>

```

The application of the `t_p_jm_a_lcm_e_kr` distribution to the given flow `f_r2ps_to_ach` captures its specific driving attributes. This distribution is characterized by assertive junction behavior, prioritization of the right lane, and cooperative lane-changing. The complete list of flows, along with their corresponding type distributions and attributes, is available for download. The link to access this detailed configurations will be provided in a later section of the paper.

In the randomly generated passenger car trips, the default type has been replaced with `t_p_jm_a_lcm_e_a`. This modification is particularly significant as it seeks to embody their assertive junction behavior and lane changes. The usage of random traffic with this specific assertive behavior in this scenario mirrors the traffic patterns of the local residents, who possess an intimate familiarity with the area and knowledge of the surroundings, especially during peak hours. Their driving behavior tends to be assertive, reflecting their comfort and circulation expertise.

```

<vType id="t_d_a" speedFactor="norm(1,0.1)" impatience="1" />
<vType id="t_d_m" speedFactor="norm(1,0.1)" impatience="0.5" />

<personFlow id="p_ach_to_mgp_1" type="t_d_a" color="red"
  begin="0.00" end="3600.00" period="exp(0.03)">
  <personTrip from="E5" to="1084069664#3" lines="ANY"/>
  <personTrip from="1084069664#3" to="E5" lines="ANY"/>
</personFlow>

<personFlow id="pf_bus_to_onation" type="t_d_m" color="red"
  begin="0.00" end="3600.00" period="exp(0.02)">
  <personTrip from="E53" to="E52" lines="ANY"/>
  <personTrip to="E53" from="E52" lines="ANY"/>
</personFlow>

```

The preceding XML entries represent the structure of pedestrian flows. Our selection of the default pedestrian behavior model, namely *striping*, facilitates pedestrian interaction and collision avoidance. Alternatively, the *non interacting* model is an alternative choice where pedestrians don't interact with vehicles or other pedestrians, optimizing execution speed. This model might be advantageous in scenarios where pedestrian dynamics play a less pivotal role in the simulation [26].

Attribute `impatience` determines the extent to which pedestrians are willing to cross the street at an unprioritized crossing, even if it requires vehicles to brake. It takes values from 0 to 1, with 0 representing patient pedestrians and 1 representing impatient pedestrians. In our simulations, most of the pedestrian flows are defined as impatient, except for a few specific ones (particularly the flows near ONATION) which are set to have a moderate level of impatience (`impatience="0.5"`).

D. Gradual fine-tuning of the flows' average arrival rate

As discussed in Section III, our main objective is not to carry out an exhaustive study of vehicular traffic in the specific area under investigation. Rather, our focus is on establishing a method for developing intelligent IoV applications, aiming at enhancing the flow of vehicles in densely crowded areas. While conducting a real-world case study on vehicular traffic holds significance, our main interest lies in building realistic simulation models within SUMO. These models will serve as a robust foundation for creating the envisioned IoV application, following a systematic scientific approach. Delving into an in-depth analysis of challenges related to traffic takes a secondary position in terms of importance. Instead, our methodology is driven by the aim to design effective IoV applications focused on improving the smoothness of traffic, ultimately aiming to implement such applications in large-scale real-world projects.

The absence of actual data regarding average vehicle flows in the examined area introduces an additional challenge when investigating traffic dynamics within this case study. However, this limitation offers an important opportunity to harness simulation and modeling tools like SUMO, enabling virtual mirroring of the real-world traffic. Through meticulous calibration of simulated traffics with genuine parameters and validating them against existing data, we obtain crucial insights on traffic patterns, useful to assess the effectiveness of IoV applications. This approach enables the formulation of innovative strategic solutions that can be further honed and executed in real-world contexts, ultimately contributing to the development of more efficient and sustainable transportation systems.

To address the absence of actual data concerning the average vehicle rates λ (refer to Section III) of the different considered flows within the studied area, our proposed approach involves an iterative calibration procedure. Here, we make adjustments to the λ parameter's value of each of the traffic flows defined in the simulation model. Through a comparison of the simulated traffic patterns with any available data (such as Google Maps travel time statistics or practical experiences), we can gradually fine-tune the λ values until they closely fit to real-world traffic.

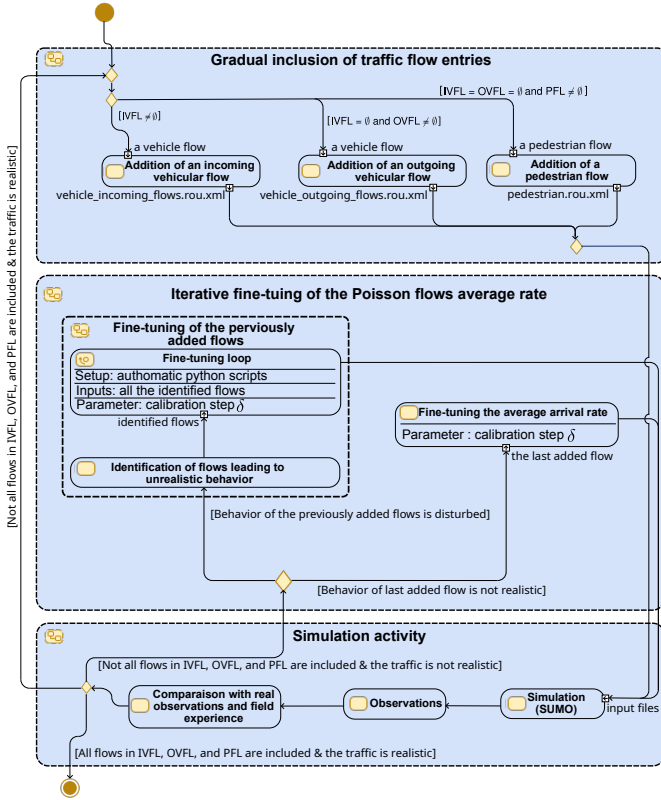


Fig. 4. Refinement of average arrival rates λ . IVFL (incoming vehicle flows list), OVFL (outgoing vehicle flows list), PFL (pedestrian flows list).

By methodically modifying λ parameters, we can simulate multiple scenarios, evaluate different modeling strategies, and gauge their impact on traffic realism. For the sake of clarity, we have introduced a SysML [27] activity diagram (see Figure 4) to illustrate the steps of the procedure, outlined as follows:

- 1) The process starts by gradually last adding flow entries to the model in a specific sequence: beginning with incoming vehicular flows (IVFL), then the outgoing flows (OVFL), and finally pedestrian flow entries (PFL).
- 2) Subsequently, simulations under SUMO are performed, accompanied by observations and comparisons against available real-world data and field experiences.
- 3) If not all the flows are included and traffic is unrealistic, two scenarios arise: (a) if the most recently added flow appears unrealistic, its λ parameter value is adjusted (by adding or subtracting a step δ), and revert to step (2); (b) if the previously added flows are disrupted, an automated Python script is used to fine-tune the affected flows in a similar way as (a), and the process returns to step (2).
- 4) If not all flows are included, but the traffic patterns are realistic, the process loops back to step (1).

This procedure iterates until all flows are fully added and assessed. Following these steps guarantees a comprehensive and systematic enhancement of the traffic simulations process. The SUMO model input files are accessible for download⁵.

⁵https://github.com/Jean-Tshibangu-jtm/sumo_ccr_sqy

Our study employs a carefully conducted data analysis, serving as the final step to derive insights from the simulation data we have gathered. This analysis involves pattern recognition and correlation examinations to systematically uncover trends and relationships within the generated dataset. This approach validates hypotheses and draws informed conclusions from the simulation results. This analysis taps into our simulation data's potential, contributing to our study's objectives.

While acknowledging the significance of data analysis, it is crucial to grasp that this phase occupies a subordinate position compared to the intricate simulation process, which constitutes the cornerstone of our work. The demanding nature of refining the simulation process absorbed the most significant part of our time, owing to its inherent complexity. Ensuing data analysis, while pivotal, primarily serves to corroborate the observations gleaned from the simulation.

Our data metrics undergo meticulous selection, guided by a comprehensive criteria. The foremost among these criteria include *relevance*, which ensures that each metric effectively evaluates the specific phenomenon or objective at hand. In our context, this translates to accurately assessing congestion levels within the studied area. *Sensitivity to traffic variations* stands as the second criterion, spotlighting a metric's capacity to promptly capture data shifts. This criterion holds particular significance within our framework, as it empowers the detection and quantification of congestion fluctuations. Progressing to the third criterion, *granularity*, we examine a metric's ability to deliver finely nuanced measurements. This attribute supports precise analysis by furnishing information at a level of detail that aligns with our needs.

Three types of detectors were utilized to generate data: *lane area detectors*, *induction loop detectors*, and *edge data*, each contributing unique information essential for studying traffic patterns, congestion, and the effectiveness of the simulation strategies [28]. Lane area detectors replicate real-world traffic measurement devices placed in specific areas or lanes of a road segment to provide data about vehicle movement. These detectors are valuable for studying lane-specific traffic patterns, densities, and lane-changing behavior, allowing the simulation of various traffic scenarios. Induction loop detectors represent real-world traffic sensors to monitor vehicular movements on the road network, offering essential traffic data for analyzing behavior, congestion, vehicles speed, travel times, and more. Edge dataset [29] refers to macroscopic information collected from specific road segments (edges), including various metrics like vehicle counts, density, mean speed, and occupancy rates.

Figure 5 showcases the seamless integration of the mentioned detectors into our simulation model. It is important to mention that although edge data entries are not explicitly shown in the figure, they are textually included as additional in `osm.sumocfg` (SUMO configuration file). The most critical congested paths include R2PS to RA2 via RA1, PLSR to RA1, RA3 to/from RA4, CCB to RA4, and RA5 to RA4. The plots presented below are restricted to these paths.

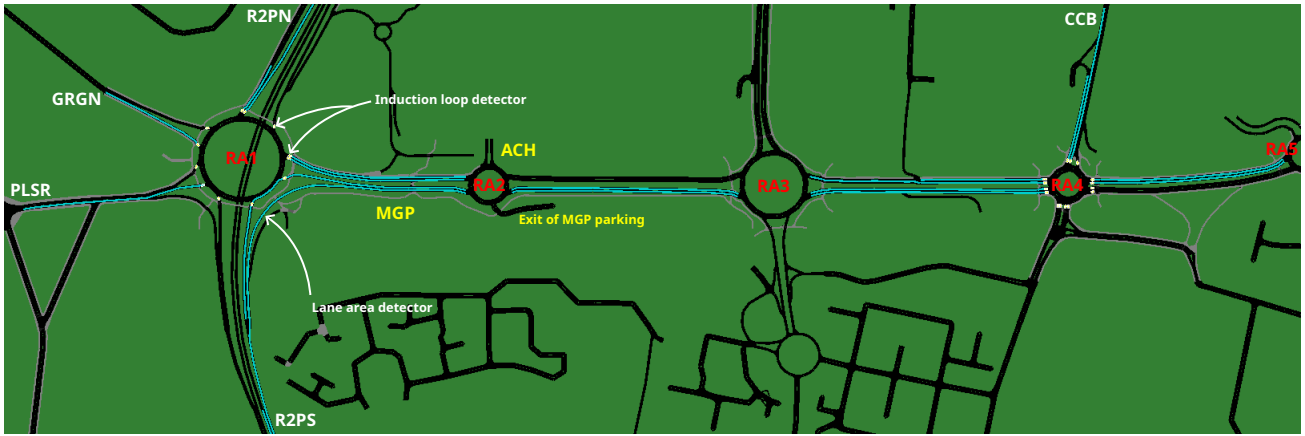


Fig. 5. Detectors added to our simulation framework. CCB stands for “Chemin de la Croix Blanche” and RA x represents Roundabout n° x (refer to Figure 2 for the rest of acronyms).

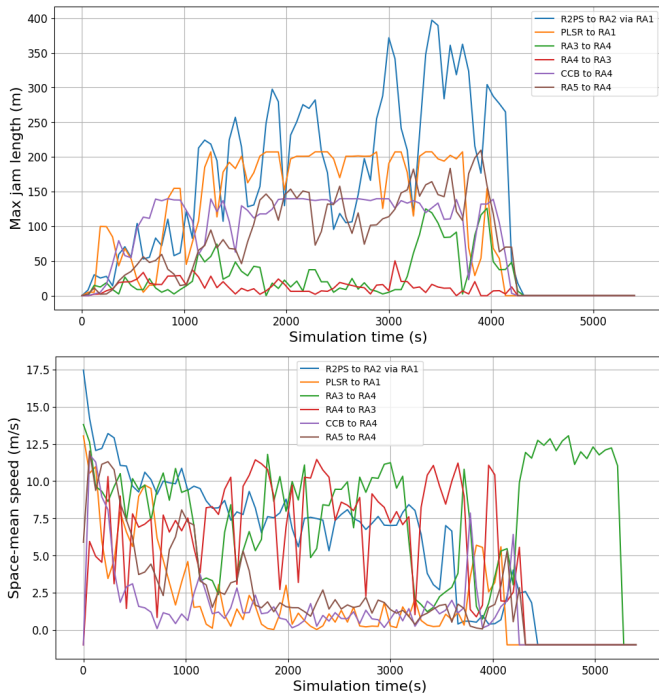


Fig. 6. Maximum jamming lengths (top), and space-mean speeds experienced (below) along the identified routes (bottom).

The data depicted in Figure 6 (top) illustrates The length of the longest observed traffic jam during the simulation time, including the gaps between blocked vehicles. This initial plot reveals that the lengthiest congestion occurs along the routes R2PS to RA2 through RA1 and PLSR to RA1, which aligns with our observations.

The plot in Figure 6 (below) shows the evolution of space-mean speeds (see Section IV) experienced along the identified routes. Only the speed recorded on route RA3 to RA4 remains non-null at the end of the simulation. This phenomenon can be attributed to the release of vehicles obstructed by traffic congestion at the MGP parking exit around RA2. As all flows

are exhausted, these vehicles are eventually liberated, and their presence becomes detectable only by the RA3 to RA4 detector. We can also note that the average speed drops significantly as detected on the routes R2PS to RA2 through RA1 and PLSR to RA1, which is consistent with the high jam lengths observed on these routes.

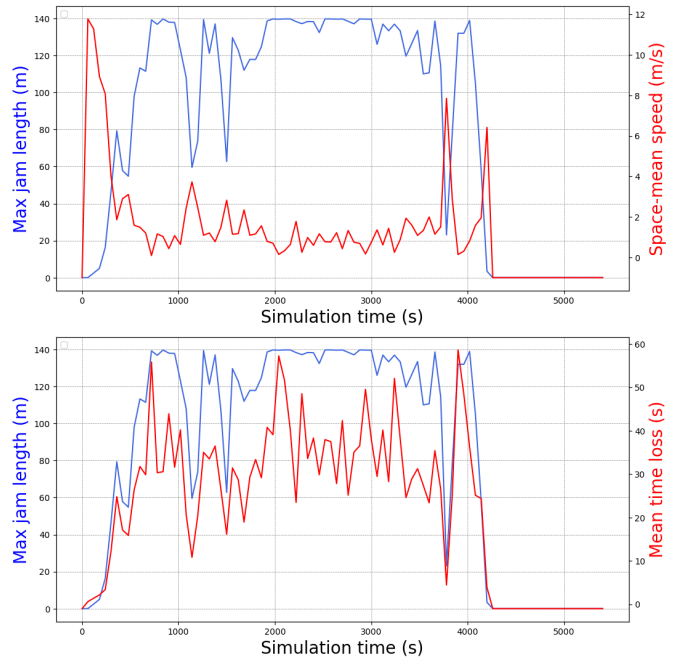


Fig. 7. CCB to RA4: maximum jamming lengths vs. space-mean speed (top), and vs. mean time loss (bottom).

The data `meanTimeLoss` is the average time loss per vehicle in the corresponding interval. The total time loss is computed by multiplying this value with `nVehSeen`, which represents the count of vehicles that were present at the detector during the simulation timeline. The correlations between the maximum jam length observed on the CCB to RA4 route and both the space-mean speed and time loss are evident (see Figure 7).

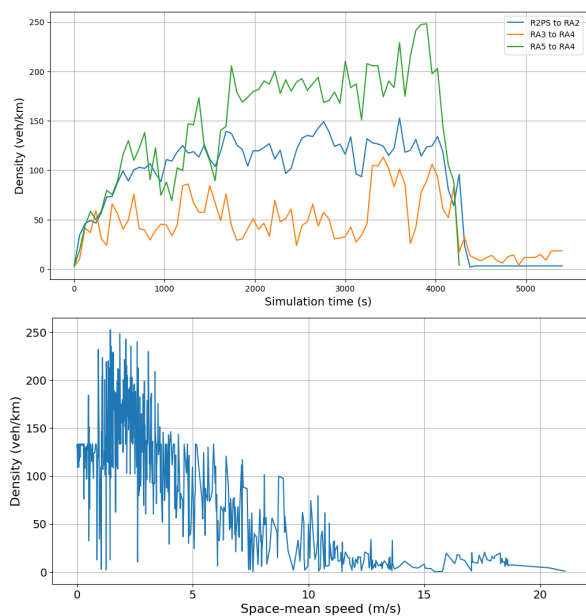


Fig. 8. Density (refer to Section IV) recorded respectively on paths R2PS to RA2, RA3 to RA4, and RA5 to RA4 (top). Fundamental diagram of density vs. space-mean speed for the path R2PS to RA2 (bottom).

In Figure 8 (top) we can see the evolution of densities for the routes R2PS to RA2 through RA1, RA3 to RA4, and RA5 to RA4. Density signifies the filling capacity of the road rather than the length of congestion. This is why the density of the edge RA5 to RA4 is higher than that of edge R2PS to RA2. In the middle of the simulation, the edge RA5 to RA4 remains consistently filled, while the edge R2PS to RA2 becomes less congested, causing lower density (refer to the video recording in Github). Figure 8 (bottom) shows an empirical plot of the fundamental diagram (see Figure 1) relating density to space mean-speed of the route R2PS to RA2 via RA1.

VII. CONCLUSIONS AND PERSPECTIVES

Our research makes substantial strides in addressing the complexities of studying urban traffic congestion, particularly in densely populated commercial zones. Our meticulous development and refinement of a simulation-based approach have yielded valuable insights into the intricate interplay of traffic patterns and behaviors. This study's significance is heightened as it resonates within the realms of smart cities, vehicular networks, and data analytics. By delving into these challenges, our work lays a robust foundation for implementing transformative solutions through IoV applications and machine learning techniques, promising enhanced smarter urban mobility.

REFERENCES

[1] F. Yang, S. Wang, J. Li, Z. Liu, and Q. Sun, "An overview of Internet of Vehicles," *China communications*, vol. 11, no. 10, pp. 1–15, 2014.
 [2] B. Ji, X. Zhang, S. Mumtaz, C. Han, C. Li, H. Wen, and D. Wang, "Survey on the Internet of Vehicles: network architectures and applications," *IEEE Commun. Stand. Mag.*, vol. 4, no. 1, pp. 34–41, 2020.

[3] D. Krajzewicz, G. Hertkorn, C. Rössel, and P. Wagner, "SUMO (Simulation of Urban MObility): an open-source traffic simulation," in *Proceedings of the 4th Middle East Symposium on Simulation and Modelling (MESM'02)*, 2002, pp. 183–187.
 [4] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, "Microscopic traffic simulation using SUMO," in *Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC'18)*. IEEE, 2018, pp. 2575–2582.
 [5] SUMO user documentation, "Definition of Vehicles, Vehicle Types, and Routes," 2023, Documentation, URL: https://sumo.dlr.de/docs/Definition_of_Vehicles%2C_Vehicle_Types%2C_and_Routes.html.
 [6] V. Kalra, P. J. Tulpule, and J. A. Isaman, "Multi-modal traffic simulation calibration and integration with real-time hardware in loop simulator," in *SUMO Conference Proceedings*, vol. 3, 2022, pp. 105–119.
 [7] X. Ma, X. Hu, T. Weber, and D. Schramm, "Evaluation of accuracy of traffic flow generation in SUMO," *Applied Sciences*, vol. 11, no. 6, p. 2584, 2021.
 [8] J. C. Dias, P. H. Abreu, D. C. Silva, G. Fernandes, P. Machado, and A. Leitão, "Preparing data for urban traffic simulation using SUMO," in *SUMO Conference Proceedings*, 2013, pp. 25–34.
 [9] R. Protzmann, K. Schrab, M. Schweppenhäuser, and I. Radusch, "Implementation of a Perception Module for Smart Mobility Applications in Eclipse MOSAIC," in *SUMO Conference Proceedings*, vol. 3, 2022, pp. 199–214.
 [10] The Eclipse Foundation, "Eclipse MOSAIC documentation," 2020, documentation, URL: <https://www.eclipse.org/mosaic/docs/>.
 [11] SUMO user documentation, "TraCI," 2023, Documentation, URL: <https://sumo.dlr.de/docs/TraCI.html>.
 [12] M. Rapelli, C. Casetti, and G. Gagliardi, "TuST: From raw data to vehicular traffic simulation in Turin," in *2019 IEEE/ACM 23rd International Symposium on Distributed Simulation and Real Time Applications (DS-RT'19)*. IEEE, 2019, pp. 1–8.
 [13] L. Codecá, R. Frank, S. Faye, and T. Engel, "Luxembourg sumo traffic (lust) scenario: Traffic demand evaluation," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 2, pp. 52–63, 2017.
 [14] P. T. K. Phung, N. T. Thi, and V. T. K. Cuc, "A study on urban traffic congestion using simulation approach," in *2020 5th International Conference on Green Technology and Sustainable Development (GTSD'20)*. IEEE, 2020, pp. 555–561.
 [15] A. Varga, "OMNeT++: Discrete event simulation system," in *Proceedings of the European Simulation and Modelling Multiconference (ESM'01)*, 2001, pp. 1–7.
 [16] C. F. Daganzo, *Fundamentals of Transportation and Traffic Operations*. Emerald Group Publishing Limited, 1997.
 [17] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759–770, 2008.
 [18] D. A. Guastella and G. Bontempi, "Traffic Modeling with SUMO: a Tutorial," *arXiv preprint, 2304.05982*, 2023.
 [19] M. L. Clemente, "Building a real-world traffic micro-simulation scenario from scratch with SUMO," in *SUMO Conference Proceedings*, vol. 3, 2022, pp. 215–230.
 [20] SUMO user documentation, "Netedit," 2023, Documentation, URL: <https://sumo.dlr.de/docs/Netedit/index.html>.
 [21] —, "Vehicle Type Parameter Defaults," 2023, Documentation, URL: https://sumo.dlr.de/docs/Vehicle_Type_Parameter_Defaults.html.
 [22] S. Krauß, "Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics," DLR, Tech. Rep., 1998, 98-08.
 [23] R. Wiedemann, "Simulation des straßenverkehrsflusses," Thesis, Schriftenreihe des Instituts für Verkehrswesen, Karlsruhe, 1974.
 [24] J. Erdmann, "SUMO's lane-changing model," in *SUMO Conference Proceedings*. Springer, 2015, pp. 105–123.
 [25] SUMO user documentation, "Randomness," 2023, documentation, URL: <https://sumo.dlr.de/docs/Simulation/Randomness.html>.
 [26] —, "Pedestrian Models," 2023, documentation, URL: https://sumo.dlr.de/docs/Simulation/Pedestrians.html#pedestrian_models.
 [27] Object Management Group (OMG), "OMG Systems Modeling Language SysML," 2017.
 [28] SUMO user documentation, "Additional elements," 2023, documentation, URL: <https://sumo.dlr.de/docs/Netedit/elements.html>.
 [29] —, "Lane/Edge-based Traffic Measures," 2023, documentation, URL: https://sumo.dlr.de/docs/Simulation/Output/Lane-_or_Edge-based_Traffic_Measures.html.