Analyzing Public Transportation Offer wrt Mobility Demand

Ala Arman

Distributed Systems and
Internet Tech lab DISIT Lab,
https://www.disit.org
Department of Information
Engineering, DINFO,
University of Florence, Italy,
<name>.<surname>@unifi.it

Pierfrancesco Bellini

Distributed Systems and
Internet Tech lab DISIT Lab,
https://www.disit.org
Department of Information
Engineering, DINFO,
University of Florence, Italy,
<name>.<surname>@unifi.it

Paolo Nesi

Distributed Systems and
Internet Tech lab DISIT Lab,
https://www.disit.org
Department of Information
Engineering, DINFO,
University of Florence, Italy,
<name>.<surname>@unifi.it

Michela Paolucci

Distributed Systems and
Internet Tech lab DISIT Lab,
https://www.disit.org
Department of Information
Engineering, DINFO,
University of Florence, Italy,
<name>.<surname>@unifi.it

ABSTRACT

An important problem in improving mobility services consists in analyzing the transportation offer with respect to the demand of mobility. The purpose is always the assessment of the service for its improvements. This activity can be approached having all the historical data, while in most cases is not realistic due to the expensive process of data collection and lack of details about the movements of travelers at the bus stops in terms of pick-up and drop-off for each bus line. To deal with these issues, in this paper, a model is provided to support mobility analysis in public transport networks. Our model operates first by analyzing the service offer, provided by mobility operators, and the service demands. Then, the model allows to evaluate the number of people who are picked-up and dropped-off at a stop. The performance of the model has been validated by comparing the observed values obtained from a field observation. The research and tool have been developed in the context of MOSAiC research project partially funded by Tuscany Region, with DISIT lab, ALSTOM, Municipia/Engineering, TAGES and CNIT research centers.

CCS CONCEPTS

Simulator / interpreter, Probabilistic algorithms, Model-driven software engineering, Model verification and validation.

KEYWORDS

Public Transport, Offer & Demand algorithms, Origin Destination Matrices, Smart City, What-if analysis.

1 Introduction

A deep understating about how public transportation services are exploited is an additional fundamental step for improving the offered services on the basis of the dynamic demand of mobility. On the other hand, to evaluate the impact of changes in the public

transportation services offered to the commuters, specific analyses and simulations should be performed to save time and costs. For instance, in a real case, the mobility officer of a city may receive complains about recurrent crowding conditions on a bus line and its specific segments, connecting a number of specific bus stops. The corresponding actions to solve the problem could be to perform some on-site analysis and interview and they implement some changes in the public transportation service in agreement with the operator (changes on-time schedule, paths of the bus lines, frequency of the service, addition of bus lines, etc.). The alternatives and combinations can be many, and thus, to avoid proceeding by try and error, a deep analysis by simulation is vitally required to better understand causes of critical issues and impact of changes before performing physical changes.

Therefore, the simulation and analysis of people flows in the city is conquering a growing attention due to a wide spectrum of related applications (e.g., [3–6]). In particular, for predicting human usage of bus lines, several approaches are offered. In most of them, the assumption is that the tracking data of travelers, which can be used to model and predict human mobility, is available. For example, in [5], the tracking data of commuters' area collected by using public transport IC cards. It is assumed that the drop-off probability of passengers at a bus stop follows a uniform or a standard normal distribution (that is quite unrealistic). In other cases, the stops are labeled as "small", "medium", and "large" in terms of volume of passenger exchanged. Then, the drop-off probability at a stop is obtained following the labeling strategy, based on the intuition that larger stops may attract more passengers, and as a result, the drop-off probability at such stops may be higher than others. In [3], a methodology for evaluating the quality of stop boarding and alighting has been presented. A part of the research focuses on estimating the alighting stop of a stage in a multiple-stage trip when the associated boarding stop is available. A different approach has been proposed in [6] to estimate real-time passenger flow for urban bus transit systems. In that case, the number of people, with smart card and on-board tickets, who are picked-up at a station is estimated, considering two consecutive taping records. Finally, considering a bus trip, after a real-time estimation of the number of passengers on the bus, the number of on-board passengers on the remainder of the trip stations is estimated, using a proposed Kalman filter. In addition, in order to contextualize the proposed work with respect to the state of the art, a number of tools for simulating

people flows have been reviewed including, MatSim [7], SUMO [8], and TRANSIMS [9], just to mention a few. Most of them do not address the estimation of pick-ups and drop-offs at stops as well as the analysis of commuters' behavior. The other limitation of these tools is the limited capability in considering contextual data regarding the city structure, and thus, the motivations to get in/out of the bus. Thus, most of the above-mentioned solutions assumed: *(1)* the possibility of tracking passengers (e.g., using public transport IC cards, mobile device tracking data). This fact, however, it is not viable real scenarios. Please note that taping is not mandatory for regular commuters and city common users in most of the modern solutions, and thus relevant errors are produced. The counting of passengers on board may be available and may be performed on busses while the counting of people at the bus stops (with details on drop-off and get-on for bus-line) is typically very expensive and not easy to be performed since they could be waiting for several bus lines, and most of the bus-stops present multiple bus lines; *(2)* to work on single-stage trips where commuters need to take only one bus to reach their destinations.

In this paper, focusing on the bus as the public transport mode, a model and simulator for the analysis of the offer of public transportation services with respect to the demand of mobility is proposed (called ODA, Offer vs Demand Analyzer). In other words, the proposed model aims at *(i)* producing viable and consistent results without the need of detailed data on the bus lines; *(ii)* addressing multi-stage trips, and thus, is in some measure a multimodal simulator and analysis tool for matching demand mobility vs. offer of transportation. The proposed model and tool have been developed in the context of research and development project called MOSAIC founded by Tuscany Region (Italy) with relevant international partners including, ALSTOM (the coordinator), DISIT Lab of UNIFI (us) (https://www.disit.org), Municipia/Engineering, TAGES, CNIT national research center. The model and tool have been built exploiting Km4City knowledge model (https://www.km4city.org), and validated by using the data and services provided by Snap4City (https://www.snap4city.org). The input and contextual data are those covering the Tuscany region, and in particular, the Florence City Metropolitan area which is the capital of the region with about 1.5 Million inhabitants.

This work is organized as follows. Section 2 presents the requirements and data sources. Section 3 describes the architecture. Section 4 provides our solution for service offer and demand analysis. Section 5 provides the evaluation results and validation of the proposed model. Finally, conclusions are drawn in Section 6.

2 Requirements and Data Source Analysis

In this section, the main requirements of a simulation tool in the context of matching demand and offer of mobility are discussed. Among the main requirements, the analysis of the data sources assumes a strong relevance because the tool has to be flexible enough to cope with different kind of data sources. Also, the tool must be **flexible enough** to model the demand starting from (and taking into account) a range of different data that may correlate to; In fact, in urban areas, daily commuters follow different purposes, considering their activities (e.g., work, study). Different points/regions of interest are then needed to be carefully investigated to evaluate the extent of the desirability of commuters for traveling to, considering the contemplated time slot. Therefore, a broad domain of data including places which provide services (work and study, in our scenario) and household data (e.g., residential buildings) must be considered upfront. Moreover, different data regarding daily trips is needed to be assessed, namely, the outbound and inbound trips with different purposes, and inter-area trips. In addition, other static and dynamic data, including the geolocation of the area, daily bus trip schedule (e.g., stop names, stop geometry, arrival times), is vital for efficient offer-demand analysis. Also, in urbanized cities, especially in metropolitan areas with several bus lines and mobility operators, a considerable ratio of daily trips are multi-stage ones, where commuters are needed to change more than a bus to reach their destinations. Therefore, when providing a model to consider the match from service offer and demand, it is also essential to consider such trips to increase the precision of the model. Moreover, the analysis tool has to be **fast enough** to allow performing a WHAT-IF analysis by carrying out a large number of simulations and choose them to assess on the basis of some Key Performance Indicators, KPI, (e.g., the maximum number of people of the bus, number of people moved from the area, the maximum number of people at the bus stop). It is noted that analyzing different scenarios with different input parameters (e.g., area, date, day time, time slot size), which requires digesting a large amount of data, can be notably a time-consuming process.

The **Data Sources** related to the Demand and to Offer for activating the computational model are strongly related to the above-mentioned critical aspects and needs for the city. In particular, the **Demand of Mobility** can be computed by taking into account and analyzing data obtained by:

(i) a census in which also needs of mobility are requested (e.g., city and satellite cities that relate with the commuters) and their needs of transportation to go at work or school, etc. From this kind of data, it is relevant the percentage of commuters (students or workers in our scenario) with respect to the citizens that may need to move for other reasons, maybe in a less systematic and recurrent manner;

(ii) counting the number of vehicles that enter or exit from the city over time (see section 1 for the data regarding Florence). They represent a demand of mobility, expressed by the vehicles entering the city with at least one person each. The counting is typically performed on the city border and may be used to understand how many cars are exchanged with the city from/to each specific external satellite villages. In some cases, the plate number recognition is also used to track vehicles into the city and create origin-destination matrices (ODMs).

(iii) city structure and services. They can be reasons for attracting people (e.g., point of interests (POIs), touristic attractions, schools, industries, banks, entertainment area, parking, shopping centers, stadiums, multi-hub bus-stops) or for trip production (e.g., house and civic numbers, hotels, stadiums) depending on a considered time slot; For example, residential area produce people in the morning and attract them in the late afternoon;

(iv) people flows in the city (e.g., using Wi-Fi network cellular data, mobile App data, collection of PAXCounter data). All of them may produce ODMs and most of them are distinguished with difficulty due to different flow categories (e.g., citizens, tourists, commuters, students);

(v) deployed transport services (e.g., buses, trains). For example, counting people on busses, at the bus stops, on multimodal hubs, exiting from the railways over time, just to mention a few. In some cases, the corresponding data can be provided by transportation companies such as for railways; Recently some of the Bus transportation operators are instrumenting busses with tools for counting people into the bus;

(vi) observation of people flows accessing the city in given time slots. For example, in cities in which the present of Tourists is very relevant (e.g., Venezia, Roma, Firenze) which cannot be neglected with respect to the citizens, commuters, and students. In this case, data that come from the cellular networks [10] and Wi-Fi [11], considering, respectively, access points and regions in the city, can help in analyzing different aspects (e.g., how many of them are daily present in the city, how long they stay, where they go/come in long term/distance) by investigating people's flow.

Also, the **Offer of Public Transportation** can be obtained by taking into account public transportation services offered to the commuters in terms of network of public services for moving in the city, which may be obtained from: (a) General Transit Feed Specification (GTFS) files (e.g., mobility operators, stops, trips, routes) of the several public transportation services in the area including, city bus, intercity bus, tram, ferry, and regional railways; including paths and time schedule of all planned trips; (b) eventual real-time position of the public transportation means which is also more difficult to obtain; (c) alternative transportation means (e.g., bike sharing, car sharing, scooter).

3   ODA System Architecture

In this section, the general architecture of the ODA Model is reported (see Figure 1). The main components of the architecture are data sources, algorithms to transform some data in OD Matrices when possible, the simulator with its algorithms, the integration with Snap4City/Km4City tools via Smart City API, and some visualization tools for presenting the results [12].
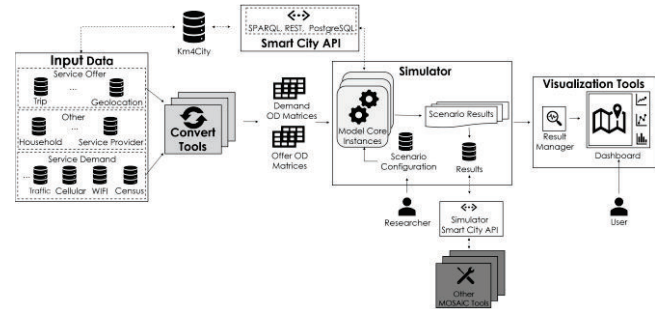

Figure 1: The model architecture

The main **input data** categories include:

- *Service Demand Data* describing people flows in the area and can be gathered from different resources (e.g., WIFI networks, cellular networks, traffic, census). Such data usually is produced by different operators (e.g., mobility, telecommunication).

- *Service Offer Data* describing potential people flows in the studied area. Such data can be gathered from different resources including, trip schedule (e.g., stop list, arrival time list, GTFS), stop information (e.g., name, geolocation), route information, just to mention a few. In this work, thanks to various data supported by the Km4City knowledge model, we adopt it as the source for gathering service offer data.

- *Aggregation and Production Motivations for People flows* describing points and areas where people may start their trip from or may end their trip there. For example, residential buildings, touristic areas, aggregation points (e.g., offices, shopping areas/centers, universities, schools, factories, cinemas, swimming pools), just to mention a few.

All this kind of data may be converted by specific algorithms in ODMs by **Conversion Tools**. ODMs describe the number of people who could/would or are moved from an area to another, the specific meaning of the ODM depends on the Data Source, but structurally are substantially similar. ODMs can be composed of a combination of different data resources, each with a (possibly) different share to obtain the ODMs.

The Simulator performs the demand-vs-offer analysis as described in the next section. Please note that analyzing different scenarios with different input data and configuration scenarios parameters (e.g., studied area, date, day time) is a time-consuming process, considering the input data with additional configuration scenarios. For example, with the current Km4City configuration, it takes around 15 minutes to analyze a single bus trip which passes through 25 stops in the central part of the Florence metropolitan area. Therefore, considering more than 82,000 daily bus trips, 1850 bus stops, 45,200 residential buildings, and 32,000 service providers, it takes two to three days to thoroughly analyze service offer and demand in the Florence metropolitan area, using the proposed model. To avoid such a situation, we adopt a fast-

computational strategy allowing to perform a large number of simulation scenarios, storing the results, and using them for visualization and further analysis. Also, other developed MOSAiC tools can access the results, using the simulator smart city API.

Visualization tools are responsible for presenting customized simulation results in a smart city dashboard (as in Snap4City https://www.snap4city.org). The *Result manager* performs the analysis of the results to be visualized in the dashboard based on the criteria that are selected by the user.

Table 1: Notations

Notation	Description
$O_{t \in T}$	Offer ODMs
$D_{t \in T}$	Demand ODMs
L	Set of localities in region R
$L' \subset L$	Set of localities in area A
$Out_{t \in T}$ and $In_{t \in T}$	Number of outbound and inbound individual trips, respectively, from and to area A
$s_{beg}, s_{tra},$ and s_{fin}	Beginning, transfer, and final stops in a commuter trip, respectively
$c(r, s)$	Radius with stop s in the center and radius r
$P_{t \in T}(s)_{beg}$	Probability that s is the beginning stop in a commuter trip
$P_{t \in T}(s)_{tra}$	Probability that s is a transfer stop in a commuter trip
$P_{t \in T}(s)_{fin}$	Probability that s is the final stop in a commuter trip
$P_{t \in T}(s)_{tra \cup beg}$	Probability that s is the beginning or transfer stop in a commuter trip
$P_{t \in T}(s)_{tra \cup fin}$	Probability that s is the final or transfer stop of in a commuter trip
$W_{t \in T}$	Weight vector associated with time interval H_t
$pck_{t \in T}(s, k)$ and $drp_{t \in T}(s, k)$	Number of pick-ups and drop-offs at stop s in k^{th} interval of H_t, respectively

4 ODA Model Simulator Model and Algorithms

In this section, the structure, the math and the algorithms of the simulator are presented. The main issues addressed are:

- Notation and ODM of the Offer
- The demand of mobility and the ODMs
- BUS Stops assessment
- Analysis Method

4.1 Notation and Offer ODMs

Considering the time set $T = \{mng, aft\}$, which includes two main time slots of a typical working day (i.e., the morning (mng) and the afternoon (aft)), investigated in this work, Table 1 shows notations used in our model. The aim of this section is to produce a model of the people flow that can be satisfied by a certain public transportation service (e.g., in terms of the number of passengers moved from any point of the city to others at a different time slot of the day). In fact, the assessment of the *public transportation services offered* by one or more mobility operators may lead to estimate the number of people who can be moved from a locality

to another. This result can be produced in the form of offer ODMs and computed for different time slots (e.g., morning, afternoon). The morning offer ODM $O_{mng}^{|L| \times |L|}$ is defined as the matrix having in the cell $O_{mng}^{i,j}$ ($O_{aft}^{i,j}$), the total number of people who can be moved from locality l_i to locality l_j, in the morning (afternoon). Formally, the cells $O_{mng}^{i,j}$ and $O_{aft}^{i,j}$ in the morning $O_{mng}^{|L| \times |L|}$ and the afternoon $O_{aft}^{|L| \times |L|}$ offer matrices are calculated as (1) and (2).

$$O_{mng}^{i,j} = t_{mng}^{i,j} \cdot c_{tot} \qquad (1)$$

$$O_{aft}^{i,j} = t_{aft}^{i,j} \cdot c_{tot} \qquad (2)$$

where $t_{mng}^{i,j}$ and $t_{aft}^{i,j}$ denote the number of bus trips from locality l_i to locality l_j, in the morning and in the afternoon, respectively. Also, c_{tot} denotes the total capacity of the bus.

4.2 The Demand of Mobility and ODMs

As a primary step for demand analysis, it is necessary to describe motivations for people flow in the selected area, considering different daily trip purposes. Daily individual trips in our scenario can then be purpose-wise categorized into two groups (see Figure 2): *1)* Home-*to-Work (H2W) trips*, which begin from home to work or study places in the morning; and *2) Work-to-Home (W2H) trips,* which begin from work or study places to home in the afternoon. In the following, considering the morning *H2W* and the afternoon *W2H* trips, the morning and the afternoon demand ODMs are evaluated.
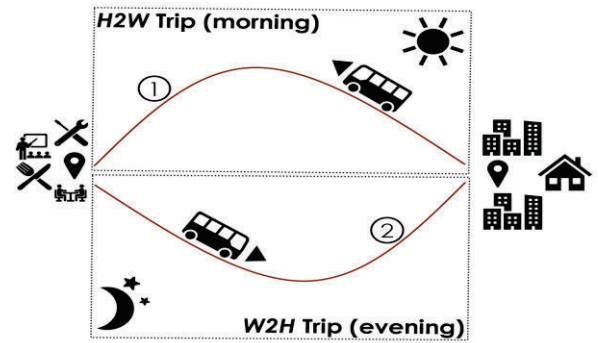

Figure 2: *H2W* and *W2H* trips

For a region, which includes a set L of localities, the morning demand ODM (morning demand matrix, from now on) $D_{mng}^{|L| \times |L|}$ is introduced. It is defined as the matrix which contains in the cell $D_{mng}^{i,j}$, the morning *H2W* outbound trips from locality l_i to locality l_j. Also, the afternoon demand ODM (afternoon demand matrix, from now on) $D_{aft}^{|L| \times |L|}$ is defined as the transpose morning demand matrix $D_{mng}^{|L| \times |L|}$. That is, in the afternoon demand matrix $D_{aft} = D_{mng}^{T}$, which superscript T denotes the transpose operation, the cell $D_{aft}^{i,j}$ describes the number of afternoon *W2H* outbound trips from locality l_i to locality l_j.

According to the census data, commuters who daily travel to work or study places in the morning, usually return home in the afternoon. Then, it can be said that the number of morning $H2W$ outbound trips from locality l_i to locality l_j is equal to the number of afternoon $W2H$ inbound trips to locality l_j from locality l_i (i.e., $OD_{mng}^{i,j} = OD_{aft}^{j,i}$). Figure 3 shows an example of the morning D_{mng} and the afternoon D_{aft} demand matrices for five localities. As illustrated, the number $D_{mng}^{1,3}$ of morning outbound trips from locality l_1 to l_3 is equal to the number $D_{aft}^{3,1}$ of afternoon inbound trips to l_1 from l_3.

For an area A in region R, which includes a set $L' \subset L$ of localities, people flow can be defined as the number of morning outbound and afternoon inbound trips, respectively, from and to area A). As shown in (3), the number Out_{mng} of morning outbound trips from area A, which is equal to the number In_{aft} of afternoon inbound trips to area A, is calculated as the sum of the total number of morning outbound trips from each locality $l \in L'$ to other all localities in set L.

$$Out_{mng} = In_{aft} = \sum_{i=1}^{|L'|} \sum_{l_i \in L'}^{|L|} D_{mng}^{i,j} \qquad (3)$$

Likewise, as shown in (4), the number In_{mng} of morning inbound trips to area A, which is equal to the number Out_{aft} of afternoon outbound trips from area A, is calculated as the sum of the total number of morning inbound trips to each locality $l \in L'$, from other all localities in set L.

$$In_{mng} = Out_{aft} = \sum_{i=1}^{|L'|} \sum_{j=1}^{|L|} D_{aft}^{i,j} \qquad (4)$$

Back to our example, when $L' = \{l_1, l_2, l_3\}$, Figure 3 shows an example of the morning outbound Out_{mng} and inbound In_{mng} trips, from and to area A, respectively. In the next section, our approach for estimating stop popularity, which then will be used for the analysis (see Section 4.4).

Region						
	l_1	l_2	l_3	l_4	l_5	Tot
l_1	14	12	3	13	9	51
l_2	6	11	10	15	31	73
l_3	7	4	13	16	19	59
l_4	0	5	44	32	45	■
l_5	8	18	73	54	33	■
					Out_{mng}	183

Region						
	l_1	l_2	l_3	l_4	l_5	Tot
l_1	14	6	7	0	8	35
l_2	12	11	4	5	8	50
l_3	3	10	13	44	73	143
l_4	13	15	16	32	54	■
l_5	9	31	19	45	33	■
					In_{mng}	228

Figure 3: An example of the morning D_{mng} (left) and the afternoon D_{aft} (right) demand matrices when $L' = \{l_1, l_2, l_3\}$

4.3 Bus Stop Assessment

When it comes to service demand analysis, a vital aspect to be considered is estimating the motivation of commuters for getting-on and -off at a stop. In fact, such motivation indicates the extent of the stop popularity and demand for commuters to be picked-up or dropped-off at the stop. For a proper stop popularity analysis,

daily commuter trips in our scenario transfer-wise are divided into two categories (see Figure 4): *1) single-stage trips,* which the commuter only needs to take one bus trip to reach her destination and *2) multi-stage trips,* which the commuter needs to transfer between a set of buses to reach her destination. Then, for a commuter, a stop in our scenario can be a: *1) beginning stop (s_{beg}),* which he/she gets-on at, to start his/her trip by getting-on a bus; *2) final stop (s_{fin}),* which he/she gets-off at, to finish his/her trip; and *3) transfer stop (s_{tra}),* which he/she gets-off (-on) at in the current stage (in the next stage) of a multi-stage-trip. In the light of these discussions, the following input hypotheses are considered.
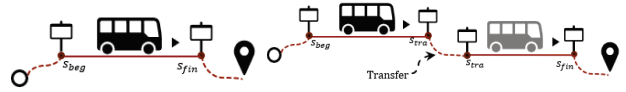

Figure 4: Examples of a beginning s_{beg} and a final s_{fin} stop in a single-stage trip (left) and two transfer stops s_{tra} in a two-stage trip (right)

Stop density. In multi-stage trips, as the number of stops around a stop increases, the probability that the stop is a transfer s_{tra} one is higher because it is more likely to be selected, by a commuter, to transfer to the next trip stage(s). Figure 5, using the ServiceMap interface (http://servicemap.km4city.org/WebAppGrafo), shows the stop density around two sample stops A and B. As one can see, the stop density around stop A is lower than around stop B. As a result, compared to stop A, a higher number of transfers can be done at stop B.

Figure 5: An example of the stop density around two sample stops A and B

Household density. In the morning (afternoon), as the density of residential buildings (which indicates the household density) around a stop increases, the probability that the stop is a beginning s_{beg} (final s_{fin}) one rises because it is more expected that commuters begin (end) their morning $H2W$ (afternoon $W2H$) trips at the stop. Figure 6, using the Overpass Turbo tool (http://overpass-turbo.eu), shows the residential building density around two sample stops A and B. As demonstrated, the residential building density around stop A is lower than around the stop B. Therefore, compared to stop A, it is more probable that stop B is considered as a beginning s_{beg} (final s_{fin}) one in the morning (afternoon).

Figure 6: An example of the residential building density around two sample stops A and B

Service provider density. In the morning (afternoon), the higher the density of service provider places around a stop is, the probability that the stop is a final s_{fin} (beginning s_{beg}) one is higher because more commuters can end (begin) their morning $H2W$ (afternoon $W2H$) trips at the stop. Figure 7 shows the service provider density around two sample stops A and B. It is possible to see that the service provider density around stop A is lower than around stop B. Therefore, compared to stop A, it is more expected that commuters select stop B as their final s_{fin} (beginning s_{beg}) one in the morning (afternoon).

Considering the above-mentioned hypotheses, given a circle $c(s,r)$ with center at stop s and radius r, the following three probabilities are then introduced.

Definition 1. Let N_{stp} denote the total number of stops in the studied area A and $N_{stp}(s)$ denote the number of stops in the circle $c(s,r)$. The probability that the stop s is a transfer one is then defined as (5).

$$P_{t \in T}(s)_{tra} = \frac{N_{stp}(s) - 1}{N_{stp}} \tag{5}$$
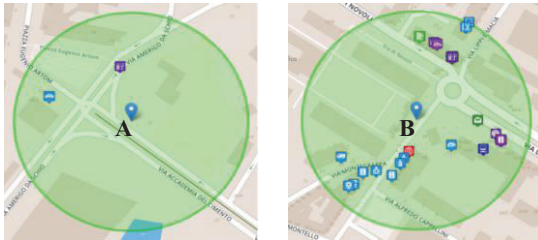

Figure 7: An example of the service provider density around two sample stops A and B

It should be noted that, in a multi-stage trip, the last stop of the current segment of the trip is assumed not to be the same as the first stop of the next trip segment (see Figure 4). This assumption is the rationale behind excluding $N_{stp}(s) - 1$ the stop s from the number $N_{stp}(s)$ of stops in the circle $c(s,r)$, when calculating the probability $P_{t \in T}(s)_{tra}$ that of the stop s is a transfer one.

Definition 2. Let N_{hh} denote the total number of residential buildings in the studied area A and $\overline{N}_{hh}(s)$ denote the average number of residential buildings per stop in the circle $c(s,r)$. The probability that the stop s is a beginning s_{beg} (final s_{fin}) one in the morning (afternoon) is then defined as (6).

$$P_{mng}(s)_{beg} = P_{aft}(s)_{fin} = \frac{\overline{N}_{hh}(s)}{N_{hh}} \tag{6}$$

Definition 3. Let N_{sp} denote the total number of service providers in the studied area A and $\overline{N}_{sp}(s)$ denote the average number of service providers per stop in the circle $c(s,r)$. The probability that the stop is a final s_{fin} (beginning s_{beg}) one in the morning (afternoon) is then defined as (7).

$$P_{mng}(s)_{fin} = P_{aft}(s)_{beg} = \frac{\overline{N}_{sp}(s)}{N_{sp}} \tag{7}$$

Considering that a commuter can select any of stops in the circle $c(s,r)$ as the beginning or final stops of his/her trip. Since we are interested in calculating the probability the stop s is a beginning s_{beg} (or final s_{fin}) one in a trip, the average number of residential buildings $\overline{N}_{hh}(s)$ or service providers $\overline{N}_{sp}(s)$ per stop is calculated. Then, according to the *union probability rule*, the probability that a stop s is a transfer s_{tra} or a beginning s_{beg} (final s_{fin}) one in the morning (afternoon) is defined as (8). Note that being a transfer s_{tra} stop and being a beginning s_{beg} (or final s_{fin}) one are assumed to be independent because their probability do not effect on each other.

$$\begin{aligned} P_{mng}(s)_{tra \cup beg} = P_{aft}(s)_{tra \cup fin} = \ &P_{mng}(s)_{tra} \\ &+ P_{mng}(s)_{beg} \\ &- P_{mng}(s)_{tra} \cdot P_{mng}(s)_{beg} \end{aligned} \tag{8}$$

Likewise, the probability that a stop is a transfer s_{tra} or a final s_{fin} (beginning s_{beg}) one in the morning (afternoon) is defined as (9).

$$\begin{aligned} P_{mng}(s)_{tra \cup fin} = \\ P_{aft}(s)_{tra \cup beg} = \ &P_{mng}(s)_{tra} + P_{mng}(s)_{fin} \\ &- P_{mng}(s)_{tra} \cdot P_{mng}(s)_{fin} \end{aligned} \tag{9}$$

4.4 Analysis Method

Finally, the aim is to assess how the mobility demand is satisfied by the service offer by assessing the number of people who are picked-up and dropped-off at a stop s, in an interval in a morning $H_{mng}[1, ..., p]$ and in an afternoon $H_{aft}[1, ..., q]$ interval vectors. For this, it should be noted that in urban areas, to properly follow their work or study hours (e.g., $9 - to - 17$), commuters usually need to follow a specific time schedule when they travel between home and work or study places. In other words, they usually find some time intervals more desirable for traveling, compared to others. For instance, they typically travel to their work or their study places between $8{:}00$ to $10{:}00$, while they return home from $17{:}00$ to $19{:}00$. A suitable way to express the desirability of time intervals for commuting consists in associating a weight to each time interval, where higher weights, compared to others, model higher

commuting desirability of a time interval. Formally, the commuting desirability of time intervals in the morning H_{mng} and in the afternoon H_{aft} time interval vectors is respectively expressed as a morning $W_{mng}[1, ..., |H_{mng}|]$ and an afternoon $W_{aft}[1, ..., |H_{aft}|]$ weight vector, where $W_{mng}[k]$ ($W_{aft}[k]$) is the weight (relative commuting desirability) of time interval $H_{mng}[k]$ ($H_{aft}[k]$). To enable comparison among the weights, it is assumed the weight vectors to be normalized (i.e., $\sum_{j=1}^{|H_{mng}|} W_{mng}[j] = 1$ and $\sum_{j=1}^{|H_{aft}|} W_{aft}[j] = 1$). Figure 8 illustrates the daily commuting proneness distribution (i.e., the morning W_{mng} and the afternoon W_{aft} weight vectors), used in our experiments, extracted from the DISIT traffic model [13], when the size of time intervals is equal to one hour. For instance, $W_{mng}[07:00\ 07:59:00] = 0.015$ states that time interval $[07:00\ 07:59:00]$ compared to interval $[05:00:00\ 05:59:00]$, has more $(W_{mng}[05:00:00\ 05:59:00] = 0.001)$, and compared to interval $[09:00:00\ 09:59:00]$, has less commuting proneness $(W_{mng}[09:00:00\ 09:59:00] = 0.143)$.
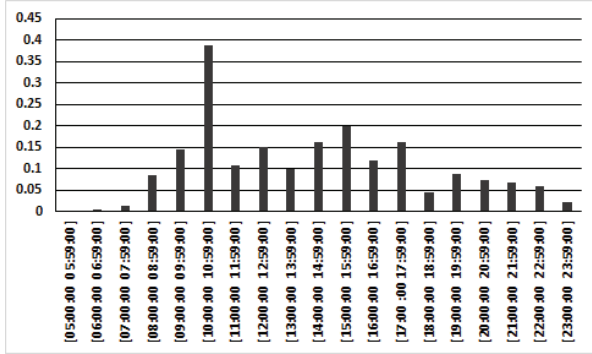

Figure 8: Examples of the morning W_{mng} and the afternoon W_{aft} weight vectors

Then, the number of people who are picked-up $pck_{mng}(s, k)$ (dropped-off $drp_{mng}(s, k)$) at the stop s in the morning time interval $H_{mng}[k] \in H_{mng}$ is equal to the sum of the number of people that begin (end) their trip and those who start their next trip stage (finish their current trip stage) at s which can be calculated as in (10) ((11)).

$$pck_{mng}(s, k) = Out_{mng}.W_{mng}[k].P_{mng}(s)_{tra \cup beg} \qquad (10)$$

$$drp_{mng}(s, k) = In_{mng}.W_{mng}[k].P_{mng}(s)_{tra \cup fin} \qquad (11)$$

Likewise, the number of people who are picked-up $pck_{aft}(s, k)$ and dropped-off $drp_{aft}(s, k)$ at the stop s in the afternoon interval $H_{aft}[k] \in H_{aft}$ can be calculated as in (12) and (13).

$$pck_{aft}(s, k) = Out_{aft}.W_{aft}[k].P_{aft}(s)_{tra \cup beg} \qquad (12)$$

$$drp_{aft}(s, k) = In_{aft}.W_{aft}[k].P_{aft}(s)_{tra \cup fin} \qquad (13)$$

5   Model Testing and Validation

Table 2 presents the model input setup. The number of the morning $\widetilde{Out}_{mng}$ and the afternoon $\widetilde{Out}_{aft}$ outbound trips are respectively obtained from the morning and afternoon demand ODMs, considering 2055 localities in the Tuscany region. The radius r in the circle $c(s, r)$, which is experimentally selected to get the best results, is set to $250\ (m)$. It is worth noting that, to provide more acceptable and realistic results, our model has been tested and validated in the context of typical working days (i.e., neither holidays nor weekend). Therefore, in this case, we focus on the census data since it is the main source of moving people around the city. The census data are publicly available on the Italian National Institute of Statistics (https://www.istat.it) and the Region Tuscany digital portal (http://www.regione.toscana.it). The metropolitan area is served by more than 10 different public transportation operators for more than 1.5 millions of inhabitants/residents, 500,000 daily trips with the purpose of work or study, 14 millions of tourists per year, 300,000 vehicles daily entering and an equal number of those exiting from the city, and around 400,000 inhabitants in the central part of the city in which busses are massively deployed.

Table 2: Simulation input

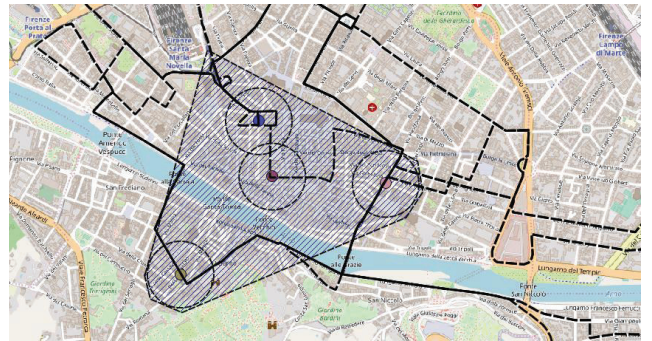| Parameter	Value
Morning interval	$05:00 - 13:59$
Afternoon interval	$14:00 - 23:59$
Time interval size	$60\ (min)$
Morning outbound trips ($\widetilde{Out}_{mng}$)	42393
Afternoon outbound trips ($\widetilde{Out}_{aft}$)	52798
Radius (r) of circle $c(s, r)$ around each stop s	$250\ (m)$

Figure 9: The area, bus stops, and bus lines, considered for testing and validating the model

The model components (implemented in Java) ran on a PC with Intel Xeon E3 − 1235 CPU 3.20 GHz and 32 GB RAM. For model validation, we consider four popular stops in the center area of Florence (see Figure 9) which includes a considerable number of POIs, residential buildings, bus stops, bus lines, and bus trips (i.e., instances of bus lines). To validate the proposed model a field observation was performed in four different time intervals, both in the morning and in the afternoon. Table 3, which shows different criteria to evaluate the complexity of the validation process. As one

can see, the proposed model was suitably analyzed when it was tested and validated, with respect to each considered criterion.

Table 3: Considered criteria for evaluating the complexity of the validation scenario

Criteria	Analyzed	Total
Bus lines	14 (50%)	28
Bus trips	55 (28.4%)	194
Stops	22 (50%)	44
Residential buildings	1560 (64.5%)	2419
Service providers	1371 (62.1%)	2207
Area (km^2)	0.8 (50%)	1.6

According to our model and tool, Figure 10 shows the comparison of the actual number vs the computed number of pick-ups and drop-offs at four selected stops including, Santa Maria Maggiore, Santo Spirito, Verdi, Porta Rossa, in four different time intervals. In those experiments, the model accuracy was evaluated based on R square. The results demonstrated that, considering the R-square values for pick-ups and drop-offs ($> 96\%$ and $> 75\%$, respectively), that the model could provide satisfactory contribution to offer-demand analysis problems in public transport scenarios.
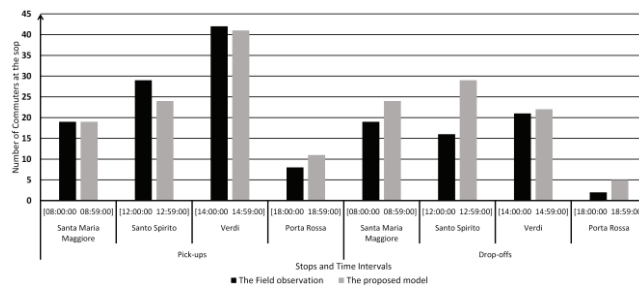

Figure 10: Actual (black bars) vs computed (gray bars) of Pick-ups and drop-offs, respectively, according to the field observation and the proposed model.

6   Conclusions

Providing public transport services with suitable quality is an essential challenge in urban environments. An important step is the evaluation of a mass transport network by comparing the service offer and demand. To address this step, in this work, a model is provided to analyze transport scenarios when daily commuter tracking data is not available. To do so, first, the service offer is evaluated by estimating the number of people who can be moved from a locality to another. Next, service demand is analyzed by evaluating: *1)* the number of people who are daily moved from a locality to another with the purpose of work or study; and *2)* stop popularity. Finally, to compare the service offer and the service demand, the number of people who are picked-ups and dropped-off at stops. The proposed solution can be adopted to analyze the status

of the public transportation services and to detect potential issues (e.g., overloaded bus stops and bus trips) in case of changes (e.g., blocked stops, out of service bus trips) that may emerge. Our work leaves space for future research and developments. In particular, in this work, it is focused on the bus as the mode of transportation. This observation is consistent with our simulation experiment because in the Florence metropolitan area, at the time of writing of this paper, there are only two tram lines and no subway service. Therefore, the bus can be considered as the main mode of public transportation. An interesting alternative can be investigating multi-modal scenarios by considering other public transportation (e.g., tram, subway) or even private (e.g., taxi) modes. Also, considering other metrics for the analysis (e.g., headway, the number of on-board commuters) can be a great source for future work in which it can be used in What-IF analysis.

REFERENCES

[1] M. Drut. 2018. Spatial issues revisited: The role of shared transportation modes. Transp. Policy, vol. 66, pp. 85–95.
[2] S. Jain, P. Aggarwal, P. Kumar, S. Singhal, and P. Sharma. 2014. Identifying public preferences using multi-criteria decision making for assessing the shift of urban commuters from private to public transport: A case study of Delhi," Transp. Res. Part F Traffic Psychol. Behav., vol. 24, pp. 60–70.
[3] M. Arnone, T. Delmastro, G. Giacosa, M. Paoletti, and P. Villata. 2016. The Potential of E-ticketing for Public Transport Planning: The Piedmont Region Case Study. Transp. Res. Procedia, vol. 18, pp. 3–10.
[4] C. Wang and H. Xuan. 2006. A Fair Off-line Electronic Cash Scheme Based on RSA Partially Blind Signature. First International Symposium on Pervasive Computing and Applications, 2006, pp. 508–512.
[5] S. Shang, D. Guo, J. Liu, and K. Liu. 2014. Human Mobility Prediction and Unobstructed Route Planning in Public Transport Networks. IEEE 15th International Conference on Mobile Data Management, 2014, vol. 2, pp. 43–48.
[6] J. Zhang, D. Shen, L. Tu, F. Zhang, C. Xu, Y. Wang, C. Tian. 2017. A Real-Time Passenger Flow Estimation and Prediction Method for Urban Bus Transit Systems. IEEE Trans. Intell. Transp. Syst., vol. 18, pp. 3168–3178, Nov.
[7] A. Horni, K. Nagel, and K. W. Axhausen. 2016. The multi-agent transport simulation MATSim. Ubiquity Press London.
[8] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker-Walz. 2012. Recent Development and Applications of SUMO - Simulation of Urban MObility. Int. J. Adv. Syst. Meas., vol. 3&4, pp. 128–138.
[9] L. Smith, R. Beckman, D. Anson, K. Nagel, M. Williams. 1995. TRANSIMS: Transportation analysis and simulation system. Los Alamos National Laboratory, United States.
[10] K. Sohn and D. Kim. 2008. Dynamic Origin-Destination Flow Estimation Using Cellular Communication System. IEEE Trans. Veh. Technol., vol. 57, pp. 2703–2713.
[11] P. Bellini, D. Cenni, P. Nesi, and I. Paoli. 2017. "Wi-Fi based city users' behaviour analysis for smart city," J. Vis. Lang. Comput., vol. 42, pp. 31–45, 2017.
[12] C. Badii, P. Bellini, P. Nesi, M. Paolucci. 2018. A smart city development kit for designing web and mobile app. IEEE SmartWorld, 28 June 2018. 10.1109/UIC-ATC.2017.8397569.
[13] P. Bellini, S. Bilotta, P. Nesi, M. Paolucci, M. Soderi. 2018. Real-Time Traffic Estimation of Unmonitored Roads. Conference: IEEE-DataCom'2018, Athen. 10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.000-6.