Estimating Urban Mobility with Open Data: A Case Study in Bologna

Valeria Caiati*, Luca Bedogni†, Luciano Bononi†, Francesco Ferrero*, Marco Fiore‡, Andrea Vesco*

*Istituto Superiore Mario Boella, Italy
e-mail: name.surname@ismb.it
†Università di Bologna, Italy
e-mail: name.surname@unibo.it
‡ CNR-IEIIT, Italy
e-mail: marco.fiore@ieiit.cnr.it

***Abstract*—Real-world data are key to the implementation and validation of urban transport models, and their availability and accuracy can dramatically affect the reliability of the resulting estimates. This paper discusses the potential of open data as a mean to gain insights in urban mobility, so as to supplement traditional methodologies that are often complex and expensive. We propose a methodology –fully based on publicly accessible data– for the development of Origin-Destination Matrices (ODMs). The methodology uses as input *(i)* a baseline morning-peak-hour ODM and *(ii)* road traffic count data. We test our proposed approach in a real-world case study, *i.e.*, the city of Bologna, Italy. We also employ open geospatial data, from socio-economic sources, to validate the ODMs, and find that these reproduce the typical urban travel demand profile in the target city during a whole day. Our results demonstrate the suitability of the methodology for transport policy design and urban planning. We also discuss current difficulties of gathering open data and the lessons learned when attempting to leverage such data.**

***Index Terms*—Open Data, Origin Destination Matrix, Urban Mobility, Transportation Modelling, Road Traffic Simulation.**

I. INTRODUCTION

As urban population expands, travel demand increases and city streets become increasingly congested. These trends have a significant effect not only on the urban mobility system, but also on environmental quality, economic development, and everyday life of cities. Hence, city planners need to develop wise and comprehensive urban transport policies and create efficient transportation services that address the ongoing social and demographic change, with minimal environmental impact and at reasonable costs.

Urban mobility models can allow urban planners and other urban decision makers to better manage all these challenges and make informed decisions about transportation investments and policies. Urban mobility models produce in fact a multitude of information, both qualitative and quantitative, about transportation infrastructures and services performances, traffic data and travel demand. These information are essentials to predict the impact of alternative policies and choose the most efficient ones for implementation. However, modelling real world urban transport scenarios needs input data about the real world transport system, and the quality and reliability of these data are crucial for a dependable performance assessment. For example, one of the key aspects in the transportation modelling is the estimation of travel pattern within an urban area, which means estimating realistic data about where, why, when and how people travel during a given time period.

Origin-Destination Matrices (ODMs) are used to describe the travel patterns showing the number of trips between different zones of an urban area. Several methods have been explored for the estimation of ODMs, either conventional (*e.g.*, based on house hold survey, or on data from traffic count detectors) or technological (*e.g.*, using GPS and Bluetooth technologies or mobile phones data) [1]. Each method is characterized by different pro's and con's, concerning the time required to obtain the ODM, the technical and economical feasibility of the process, and the extension and accuracy of the estimated ODM. Anyhow, finding new and less expensive technique to collect data for estimating reliable ODMs remains a major issue in transport planning.

Moreover, knowing traffic pattern between various zones of a city is significant not only for urban mobility planning purposes, but also to understand demographic, social and economic dynamics. Urban mobility is a complex dynamic system deeply interrelated with other systems made by infrastructural, economic and social networks. In order to gain isights into the relationships that govern such systems requires a variety of urban datasets (*e.g.*, transportation, population, employment, buildings, economic activities, education opportunities, etc.). Having access in an open fashion to a huge volume of all these data can be very helpful for understanding the correlation between travel demand and demographics, socio-economic and land use characteristics of a city. In recent years, the surge in open data platforms has created new opportunities for the scientific and technical communities in that sense: they can now take advantage of a vast amount of data to develop urban mobility models useful for different decision makers. Unfortunately, publicly available datasets do not always feature a sufficient quality level and may be complex to cleanse and manipulate. Moreover, the full value of open data has yet to be explored among the urban science, as opposed to other sciences, such as the biological science, for which the use of open data is already a usual practice [2].

This study contributes to the discussion about the powerful opportunities offered by open data in the smart city context. It finds its inspiration in the interest of exploring how open data could play a key role in having an insight of urban mobility, when they are used for transportation modeling purpose. In particular, the contribution of the present paper lies in proposing a general approach for estimating and validating a daily travel demand –in the form of 24 hourly ODMs– using open source data. For this purpose, we present a case study based on the City of Bologna, Italy. With its large amount of freely available datasets, Bologna provides an ideal scenario for our performance evaluation. The developed ODMs are available at *http://www.cs.unibo.it/projects/bolognaringway/* as open data.

II. Transportation Modelling: A Primer

Transportation models are both analysis and forecasting tools, which provide a structured framework for the representation of multiple and interrelated aspects of transportation systems. For what concerns private vehicle transports, which are the object of our study, the essential components of the related transportation model are as follows.

- Road network data, or rather the supply data of the transport network described in a network model consisting of basic objects. On a road network, these objects mainly consist in (*i*) links representing streets segments, (*ii*) nodes symbolizing street intersections, (*iii*) turns specifying which movements are permitted at a street intersections and (*iv*) zones which are areas containing a number of inhabitants, work places, schools, shopping center and other characteristics.
- Travel demand data, typically described using ODMs containing the number of trips between the origin and destination zones of the area under investigation, modeled as a network. These data also capture the temporal variability of the travel demand.
- Macroscopic traffic flows, which consist in information about traffic volumes of the different network objects, vehicle speeds and vehicle density. They are the results of the application of traffic assignment procedures, which allocate current or anticipated travel demand to existing or planned transport supply and return the paths between the origin and destination zone.
- Validation tests that compare the results with traffic measurement data. These tests are a crucial element, since, like all models, a transportation model represents an abstracted simplification of the real world.

III. Urban Case Study Scenario and Input Datasets

We focus on a case study representing a wide portion of the city of Bologna, a mid-sized city of around 380,000 inhabitants located in North-Central Italy. According to recent Smart City rankings [3], Bologna is one of the Italian cities with the largest number of datasets available on open data portals. The geographical area under analysis covers a total of around 28 km^2, encompassing the city downtown, the surrounding ringway, and part of the periphery. In the following, we review the open-source datasets that refer to this region and that have been employed in our study.

Fig. 1. Road network in Bologna, from OSM and VISUM.

A. Road network model

Road network information for the selected area was collected from OpenStreetMap (OSM) [4], a well-known crowdsourcing project aiming at developing a free high-detail world map. Being up-to-date and providing global coverage, the geographical data from OSM constitute an important –and undoubtedly very popular– input for modelling urban transport networks. These data are in form of digital street maps composed by different geometric elements (*e.g.* links, nodes) with meaningful attributes that determine the main characteristic of road segments (*e.g.* number, direction, capacity and free flow speed of lanes, permitted transport systems) and road intersections (*e.g.* geometry, control type including presence of traffic light, rights-of-way, allowed movements). They thus offer more than simple geometric information.

However, accuracy is a major problem with OSM data. As far as the data layers relevant to our specific case are concerned, OSM road network data models the real-world transportation network as a graph, composed by directed links that map to road segments, and nodes that map to intersections. Clearly, we would like such representation of the topology of the transport network to be as close as possible to the actual structure it mimics, both topologically and geographically; otherwise, there is a risk to compromise the accuracy of results, with cascading impacts on calculations and analysis [5]. Checking the consistency of the network model is thus essential, and so are the identification and fixing of all critical errors.

In order to verify the correctness of the OSM road network data, we loaded it into PTV VISUM [6], the leading software program for computer-aided transport planning. The PTV VISUM representation provides an accurate abstraction of the real transport system, including roads, intersections, roundabouts and traffic lights. The initial OSM data represents

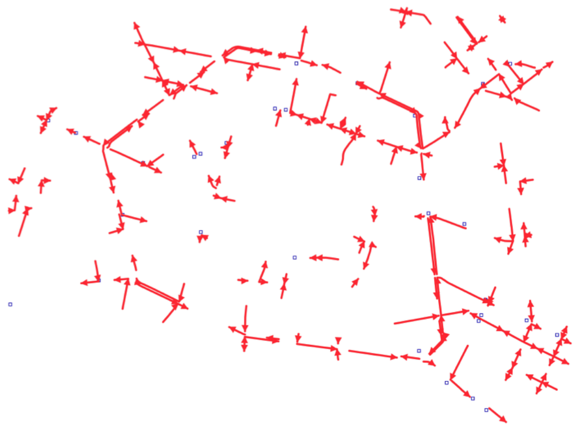

Fig. 2. Location and direction of traffic counters in Bologna.

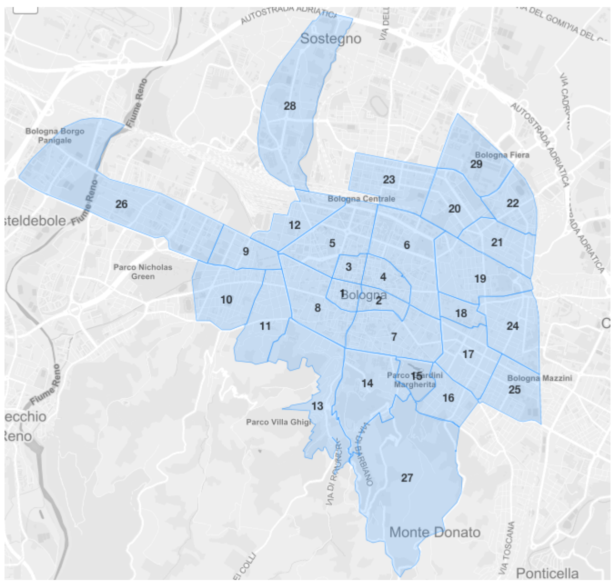

Fig. 3. City statistical areas in Bologna.

the basis of our urban scenario consisted of 2,909 nodes, 7,589 links and 23,552 turns. A consistency network check in PTV VISUM allowed us to detect the presence of isolated nodes, nodes with more than one straight turn per incoming link, links with no connection to the network and closed links, which have a speed limit of 0 km/h. Other recurring critical errors are the incorrect direction and number of road lanes. We proceeded to fixing all inconsistencies, until the final road network consisted of 1,351 nodes, 3,936 links and 12,368 turns. The layout of the road network is in Fig. 1. Different road segment colors indicate different types of roads characterized by diverse capacity.

B. Road traffic count data

Fig. 2 shows the geographical position on the Bologna road network of inductive loops installed by the local municipality to monitor the traffic dynamics. The area under evaluation has 187 detectors that record traffic flow data in terms of number of vehicles passing over each detection site at every 5-minute interval. In our case, they do not provide information about the vehicle speed. The Municipality of Bologna provided us with datasets containing inductive loops data, aggregated into hourly time periods and referring to 24 hours of a typical Wednesday. The road traffic count data give a good coverage of the main city roads, including the fast-transit ringway, the main entry/exit roads to/from downtown and other several roads crossing the city center. Most of these roads are also multilane roads with bi-directional traffic. Having traffic count data for both road directions is essential for the estimation of inbound versus outbound flows, and consequently for mining the target 24 hourly ODMs.

C. Partitioning of the study area in statistical areas

The definition of an ODM refers to some tessellation of the geographical area into traffic zones: indeed, traffic zones are the (aggregated) geographic locations of trip origins and destinations, and thus connect the transport supply (network model with nodes, links, etc.) and the travel demand (in form of Origin-Destination Matrix, which contain the number of trips of all OD pairs of the model). Traffic zones are represented as polygons, denoting the boundaries of the spatial extension of each zone, with centroids connected with the road network within the zone through access and egress routes.

Fig. 3 shows the 31 traffic zones identified in our Bologna case study. They correspond to the city statistical areas, *i.e.* territorial units used by public administrations in Italy for common statistical purposes. They are characterized by homogeneous demographics and socio-economic characteristics, and are the result of the aggregation of census sections that belong to the same neighborhood and account for physical boundaries (*e.g.* railways or urban slip roads). We remark that these are the same criteria commonly adopted in the definition of traffic zones [7]. The map of statistical areas of the city of Bologna is publicly available via the Open Data Portal of the Bologna municipality [8] as a shape file format.

D. Travel demand model

The baseline travel demand we consider is structured as an ODM containing the number of trips from an origin traffic zone to a destination traffic zone, for all traffic zone pairs within the study area. It is representative of vehicular flows during the morning traffic peak hour, between 8 am and 9 am, in a typical working day.

We used the Bologna Ringway dataset as initial ODM [5]. It is a publicly available dataset describing the individual trips, during the morning rush hour, of more than 22,000 vehicles. These cars start their trips at 93 different links and end them at 81 links in an area that covers the center and outskirts of Bologna. For the purpose of our study, we converted the initial point-to-point ODM, which contains the number of trips between two different points belonging to different links, into a zone-to-zone ODM, where each element of the matrix represents the number of trips between two zones within the reference area. As the region under study consists of 31 zones, the final ODM is composed of 961 origin-destination pairs.

IV. Generation Of ODMs

We can find in the literature a variety of studies using traffic counts from loop detectors for correcting a shadow ODM and generating a new up-to-date ODM [9], [10], [11], [12]. Among different data sources, traffic data provided by traffic counts are commonly regarded as those that are the most readily available, frequently updated, and qualitatively accurate: they are thus often preferred to alternative data sources, especially from the perspective of cost effectiveness [13]. In the following, we present a methodology for mining 24 hourly ODMs of a common workday using the 24-hour traffic counts data within the study area and the morning-peak-hour ODM from 8 am to 9 am as a shadow demand matrix in order to capture the pattern of travel demand.

The first step consists in assigning the shadow ODM to our road network model to calculate links traffic volumes. VISUM provides several assignment procedures for the private transport, and for the purpose of our analysis we selected the equilibrium assignment procedure. It distributes travel demand according to Wardrop's user equilibrium principle: every road user selects his route in such a way that the travel time on all alternative routes is the same, and that switching to a different route would increase personal travel time [14]. The equilibrium assignment problem is solved as an optimization problem with a convex objective function and linear secondary conditions. The procedure consists in a series of iterations and terminates only when the equilibrium condition is reached, *i.e.*, when all routes of any OD pair are in the balanced state. The resulting traffic volume assigned to each link of the network depends then on the private transport user behavior: drivers choose their routes –among a choice set that includes all the possible routes connecting the origin and destination zones of the road network– according to the perceived travel time of the corresponding links.

The route distribution over the network at the end of the first simulation time period (8 am - 9 am) is shown in Fig. 4. Red lines on the networks represent the traffic volume in terms of number of vehicles circulating on each link. The thicker the red line the higher the traffic volume on the road segment. This clearly reveals that ringway road segments support most of the travel demand during the morning peak hour; a similar consideration holds for the main entry roads connecting the peripheral regions of the city with downtown. Overall, this reflects the typical morning traffic, which consists of commuting drives from residential areas in the city peripheries towards the city centre, where offices, schools, universities and commercial activities are located [5].

Once the traffic volumes have been calculated through the assignment procedure, it becomes possible to apply the correction technique to update the shadow ODM with traffic counts values measured between 8 am and 9 pm. The datasets presented in Sec. II were used as an input to the matrix correction and generation performed with a fuzzy set based

Fig. 4. Distribution of routes computed by VISUM from the shadow ODM.

matrix correction [15]. This procedure is also known as TFlowFuzzy (TFF) tool within VISUM [6]. TFF allows the correction of an ODM previously assigned to the modeled road network, so that the traffic volumes resulting from the assignment procedure actually match the real traffic flows observed by the traffic counters loaded into the network model as link attributes. However, the traffic count is not a certain value, since the measured value represents only a snapshot situation that could be affected by sampling errors, and thus by statistical uncertainty. Using the Rosinowski approach [15], traffic counts have been modelled as imprecise values based on Fuzzy sets theory, which means that the counted values oscillate within a bandwidth. By adding a fuzzy value for each individual traffic count value, TFF calculates a new ODM, which is the most likely demand matrix representing the count values within the bandwidth. We set the value of the fuzzy parameter to 20%, as indicated in [16]: traffic count data becomes then a fuzzy set, and the correction algorithm runs until all modelled link flows were within $\pm 20\%$ of the measured link flows. This is equivalent to assuming that the number of trips originating in one zone to fluctuate by about 20% from day to day.

Following the TFF procedure, the new generated one-hour ODM has been assigned to the road network. We tested the goodness of fit of our model measuring the coefficient of determination R^2 between the traffic volumes calculated in the assignment and the observed values (*i.e.*, traffic counts), which results equal to 0.87. This reveals an improvement of our model accuracy since R^2 measured before the application of TFF procedure was equal to 0.69.

The others 23 hourly ODMs are obtained by iterating the same TFF and traffic assignment procedure. Let $M(t) = \left[m_{ij}^t\right]$ be a 1-hour ODM at time $t = 0, .., 23$, where m_{ij}^t is the number of trips between the ODM elements i and j in the hourly travel time band t, with $i = 1, .., 31$ and $j = 1, .., 31$. $M(t)$ represents what we are calling shadow ODM. Using it as

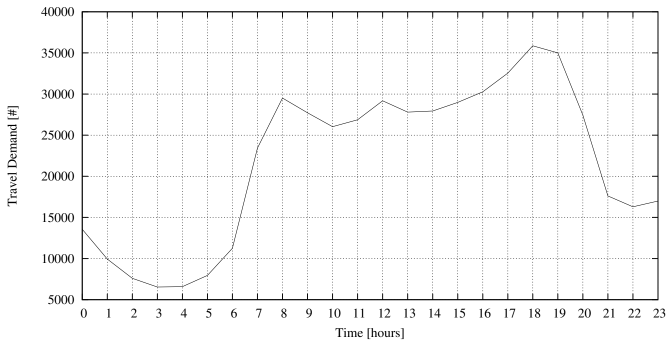

Fig. 5. Daily time series of the travel demand obtained from the iterative ODM calculation process.

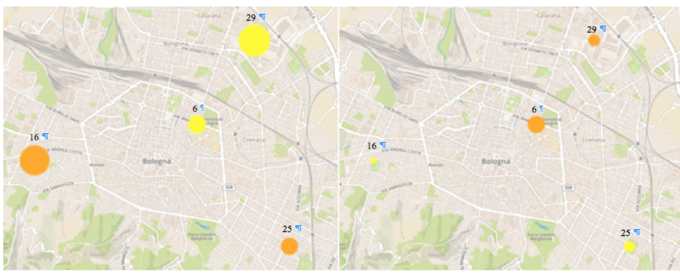

Fig. 6. Generator and attractor zones during morning (left) and afternoon (right) traffic peak hours.

an input for the TFF procedure, we then generated $M(t+1)$ as $M(t+1) = f(m_{ij}^t, c^{t+1})$, where c^t is the traffic counts vector at time t. We repeated the TFF procedure 23 times using at the first iteration the corrected ODM at the morning-peak-hour (8 am - 9 am) as a shadow matrix.

At the end of the iterative process, we obtained the travel demand profile traversing the study area of Bologna during an entire day, as shown in Fig. 5. There is only a small amount of trips during the night. In the morning hours the number of trips rises until there is a morning peak between 8 am and 9 am time period. Then, the traffic flow decreases a little bit and remains at a certain level with a small peak between 12 am and 1 pm. In the afternoon the number of circulating vehicles rises until it reaches a pm peak between 6 pm and 7 pm. This represents a credible profile of the daily road traffic observed in many cities, as illustrated in some urban mobility reports published by urban authorities or mobility agencies [17].

Furthermore, we note that the final mobility model resulting from the assignment of the 24 hourly ODMs well captures the traffic flow inversion during a 24-hour day. Fig. 6 shows a representative example of how traffic direction varies at different hours of the day. To make the figure clearer, only four zones are shown. The selected zones are situated in different parts of the study area. More specifically zones 16 and 25 are basically residential areas, while several workplaces are located in zones 6, which includes university buildings, and 29, which corresponds to the fairgrounds of Bologna. The circle dimension is proportional to the module of the difference between the generated trips and the attracted trips for each zone. Orange (yellow) circle corresponds to zone generating (attracting) more trips in a given time interval. The left and right plots refer to the morning traffic peak hours (7 am - 9 am) and the afternoon traffic peak hours (5 pm - 7 pm), respectively. By comparing these two figures, we clearly observe a colour inversion for the four selected zones, demonstrating a reverse traffic flow during the day.

V. Validation Of ODMs With Open Data

Our validation procedure leverages socio-economic and land use data from open data portals to verify the travel demand contained in the ODMs. The 24 hourly ODMs are the results of a modelling technique based on the application of the TFlowFuzzy algorithm, making validation a much-needed step. Previous proposed validation techniques are based on tract-tract worker flow census data (*e.g.*, from Census Transportation Planning Package) [18], commuting matrices from National Statistical Institutes [19], travel survey results [20], [21]. Following this trend, we explore the use of socio-economic data from different open data sources.

Primary sources of data are the Italian National Institute of Statistics (ISTAT) [22], and the Municipality of Bologna with its Open Data Portal [8]. Both authorities provide demographics and socio-economic data on statistical areas basis. Useful information concerning the economic activities are made available on another free accessible website [23] that provides a detailed list of workplaces organized by geographical location according to ATECO classification [24]. However, since this kind of websites are more likely to be used for marketing purposes, the format in which the data are available to the public results fairly complex to manage and analyse.

Assuming that trips are categorized by purpose (*i.e.*, the activity undertaken at a destination location), we divided trips datasets into two broad categories for validation purposes: home-based trips and work-based trips. Home-based trips datasets contain the data used to validate the total number of trips leaving or returning to homes in each zone (trips originating or ending at home). Work-based trips datasets contain the data used to validate the total number of trips going to or leaving the work in each zone (trips ending or originating at work). Consequently, the hourly travel time bands chosen for the validation procedure refer to morning and afternoon peak hours for home-work travels (presuming morning peak hours of home-work travel from 7 to 9 am and afternoon peak hours of work-home travel from 5 to 7 pm). The statistical areas were also used as units of analysis since travel demand and socio-economic characteristics were aggregated to the statistical area level.

The basic principle of the validation procedure is simple: the more business-related the activities (and thus the jobs) available in a certain zone, the more that zone could be a potential travel destination for nearby or distant citizens during the morning peak hours and a travel origin during the afternoon peak hours; likewise, the more housing units (and thus the residents, or more specifically the worker residents) present in a certain zone, the more that zone could be a potential travel origin for worker citizens during the morning peak hours and a destination during the afternoon peak hours.

Since the object of our study is the private vehicle transport, other indicators may be used to validate the obtained ODMs. In fact, other demographic and socio-economic factors that have been found to impact travel behavior include age, household size and income (*e.g.*, higher household size and higher incomes are associated with high travel demand) [25]. Moreover, the presence of traffic restriction measures, such as pedestrian areas and Limited Traffic Zones (ZTL) which close or limit parts of the city road network to motor vehicle traffic, affects the automobile accessibility for each traffic zone.

In order to validate home-based and work-based trips using the below-mentioned demographics and socio-economic data, a correlation analysis is performed. The Pearson correlation coefficient (R) is used to assess whether there is a linear relationship between two selected variables (travel and socio-economic data)[1].

A. Home-based trips validation

We assumed that the total number of trips originating or ending at home in each zone depends on the number of residents, number of workers residents, number of housing units, residents age, household size, and resident income.

The data warehouse of the population and housing census published by ISTAT [22], free and openly accessible, contains the information disaggregated to the census section level on (*i*) the resident population by age and by current working status, (*ii*) on housing units occupied by residents and (*iii*) on size of private households. All these data are available in form of CSV file and are organized in tables showing the quantitative value of each indicator referred for each census section. For the purpose of our analysis, we aggregated the collected data by statistical areas by means of QGIS [26], a free and open source software application for spatial analysis. Data on resident income by statistical area are available on Bologna Open Data Portal [8], in form of CSV file. The reason of introducing also this data in the validation of home-base trip was based on the assumption that members of high income households are more likely to extra expenses which may induce to do more trips.

Validation tests are performed in order to compare through the R coefficient the six collected demographic and socio-economic datasets and the home-base trips resulting from our urban mobility model in the four peak-hours for home-work travel. Table I shows the validation results of home-based trips. We observe a good correlation between all the selected demographic and socio-economic data and traffic model results. All the correlations are strong, with a maximum mean correlation coefficient R=0.75 between the density per square km of households with more than 3 components and the number of home-based trips in the four selected hours. This matches with the assumption that households of large size may produce more trips compared to small one. We remark a minimum value of the mean correlation coefficient R=0.69 between the density per square area of housing units and the number of home-based trips. In fact, although it is credible that a zone with a highest amount of housing units may produce a highest amount of trips, housing units density could not represent the best indicator to evaluate the trip production rate since it doesnt take account of the effective quantity of potential travelling people living in it.

[1]The absolute value of the correlation coefficient is in the range 0 to 1, where a value of 0 indicates that there is no relationship, whereas a value of 1 indicates that there is a perfect correlation.

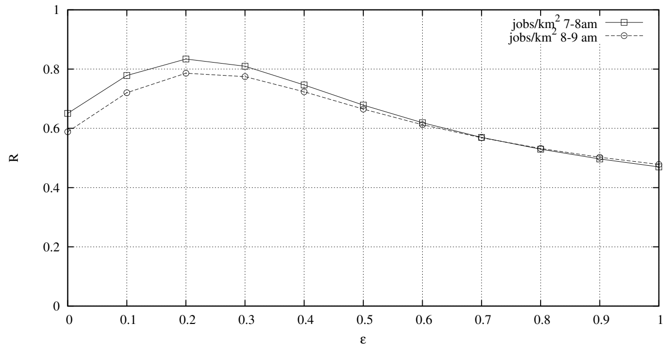

Fig. 7. Effect of the reduction coefficient ϵ on the correlation coefficient for inbound trips.

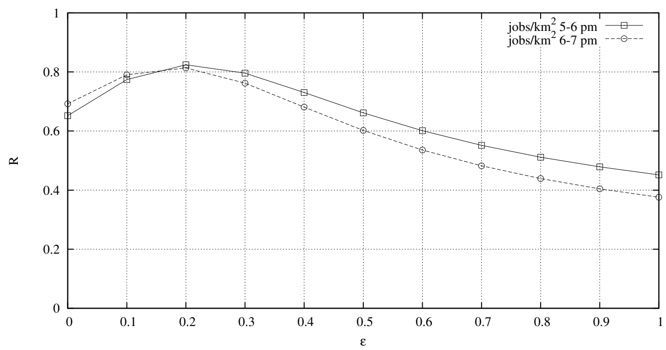

Fig. 8. Effect of the reduction coefficient ϵ on the correlation coefficient for outbound trips.

B. Work-based trips validation

The total number of trips destined for or originating at work in a zone may depend on the number of non-residential buildings and jobs (employees) in workplaces (office, retail, manufacturing, education, health, etc.).

The number of non-residential buildings per statistical areas was retrieved by the data warehouse of the population and housing census published by ISTAT. For what concerns the number of workplaces we used as primary data source the Elenco Aziende Italiane website [23], which references the address and the ATECO classification for each workplaces within the city of Bologna [24]. After getting this information in form of tables in CSV format, it was possible to map the spatial position of each workplace within the statistical areas by means of QGis. However, it is worth noting that the workplaces have different sizes (*e.g.*, ranging from a home-office to a large building or industrial plant) and consequently hold different number of employees. Thus, a more significant role in attracting and generating work-based trips is certainly played

TABLE I
CORRELATION COEFFICIENT (R) VALUE FOR HOME-BASED TRIP VALIDATION

Home-based trips	Residents nr/km^2	Workers Residents nr/km^2	Housing units nr/km^2	Residents with age 30-64 nr/km^2	Households nr/km^2	Income €/km^2
Outbound trips 7-8 am	0.73	0.70	0.68	0.71	0.74	0.61
Outbound trips 8-9 am	0.75	0.72	0.69	0.73	0.78	0.69
Inbound trips 5-6 pm	0.73	0.72	0.70	0.72	0.73	0.74
Inbound trips 6-7 pm	0.74	0.73	0.70	0.68	0.76	0.76

TABLE II
CORRELATION COEFFICIENT (R) VALUE FOR WORK-BASED TRIP VALIDATION

Work-based trips	Non residential buildings nr/km^2	Jobs nr/km^2
Inbound trips 7-8 am	0.70	0.83
Inbound trips 8-9 am	0.72	0.79
Outbound trips 5-6 pm	0.65	0.82
Outbound trips 6-7 pm	0.69	0.81

but the number of jobs in a certain area. It was indirectly obtained matching the datasets of the number of workplaces by categories previously collected and the information related to the mean number of employees for each ATECO category, provided by ISTAT.

After performing the validation test, we noted that the correlation between jobs density per square km and work-based trips was initially weak, with R ranging from 0.38 to 0.45 at the four selected time periods. A reduction coefficient $\epsilon \in [0,1]$ applied to the jobs density data in the first eight zones of the studied area increases the correlation coefficient value. As shown in Fig. 7 and Fig. 8, R value was maximized for ϵ=0.2 for both inbound and outbound work-based trips. We remarked that the first eight zones correspond to the downtown, which is most affected by car traffic restriction, such as the presence of ZTL and pedestrian areas. Shapefiles on pedestrian and ZTL coverage area available on Bologna Open Data Portal [8] were helpful to test that such zones have this particular characterization compared to the remaining zones. More precisely, they reveal that the mean area covered by both pedestrian areas and ZTL on the first eight zones was about 80%. No pedestrian areas and ZTL covered the other 23 zones. The results of the second validation tests of work-based trips are shown in Table II.

As expected, we remark a very strong correlation between jobs density per square km and work-based trips resulting from the urban mobility model, with a correlation coefficient R ranging from a minimum of 0.79 to a maximum of 0.83. This result confirms the belief that employment opportunity attracts more work trips. Still, we observe that the correlation is good enough also for the density per square km of non-residential buildings, with a mean correlation coefficient equal to 0.69.

All in all, the results confirm the quality of the generated urban mobility model applied to a wide portion of the City of Bologna. Specifically, the daily travel demand expressed in terms of 24-hourly ODMs generated using traffic counts data and applying TFF technique well reflect the demographic and socio-characteristics of the study area.

VI. CONCLUSION

With the expansion of urban population and consequent increase of daily traffic the development of mobility models useful to urban decision makers and planners is becoming crucial. These models play a key role in helping urban decision makers and planners to understand urban dynamics and to assess and predict the impacts of new transport policies, infrastructures and services. However, building large-scale truthful models of urban mobility is not a simple task. They require a significant amount of input data, which need to be as accurate as possible. Even though open data platforms are experiencing a rapid expansion over the last years, a very limited number of urban mobility dataset that describes real world travel demand and road traffic data over an urban area is publicy available today. Furthermore, a high quality of open data is not always guaranteed. In this work we presented a contribution to the use of open data for building and validating realistic urban mobility model. We described the generation process of 24 hourly ODMs portraying the daily car travel demand in an area of about 28 km^2 around the City of Bologna, Italy, and showed how the availability of open dataset combined with other real datasets allows the feasibility of this process. Having access in an open fashion to a huge amount of datasets concerning the demographics and socio-economics characteristics of an urban area, is helpful for understanding the correlation between urban mobility system and other urban systems made by infrastructural, economic and social networks. Thus, we emphasized the value of a wide variety of open data, by demonstrating how they could be used also in the validation phase necessary for completing the entire modelling process. In future works, we intend to continue the open data collection process, extending it also to other data sources, such as location-based social networks (*e.g.* Twitter). We will also investigate how this urban mobility model based on 24-hours ODMs can be used as a decision support tool, whose application could be enhanced by the development of a web-based mapping tool (*i*) based on KPIs useful to predict the demand for innovative transport services (*e.g.* car sharing, demand responsive transport) and (*ii*) capable to perform specific what-if analyses regarding changes to the socio-economic and land use data correlated with travel demand changes.

ACKNOWLEDGMENT

The research leading to these results was also supported by the People Programme (Marie Curie Actions) of the European Unions Seventh Framework Programme (FP7/2007-2013) under REA grant agreement n.630211 ReFleX.

REFERENCES

[1] T. Saini, S. Srikanth, and A. Sinha, "Are Ubiquitous Technologies the future vehicle for Transportation Planning: An analysis on various methods for Origin Destination," *International Journal of Ad Hoc, Sensor & Ubiquitous Computing*, vol. 4, no. 6, pp. 17–25, 2013.

[2] Jappinena S. and Toivonena T. and Salonena M., "Modelling the Potential Effect of Shared Bicycles on Public Transport Travel Times in Greater Helsinki: An Open Data Approach," *Applied Geography*, vol. 43, pp. 13–24, 2013.

[3] Ernst & Young, "Italia Smart: rapporto Smart City Index," 2016.

[4] OpenStreetMap. [Online]. Available: http://www.openstreetmap.org

[5] L. Bedogni, M. Gramaglia, A. Vesco, M. Fiore, J. Harri, and F. Ferrero, "The Bologna Ringway Dataset: Improving Road Network Conversion in SUMO and Validating Urban Mobility via Navigation Services," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 12, pp. 5464–5476, Dec 2015.

[6] PTV VISUM. [Online]. Available: http://vision-traffic.ptvgroup.com/en-us/products/ptv-visum/

[7] K. Baass, "Design of Zonal Systems for Aggregate Transportation Planning Models," *Transportation Research Record*, no. 807, pp. 1–6, 1981.

[8] Open Data Bologna. [Online]. Available: http://dati.comune.bologna.it/

[9] M. G. Bell, "The Estimation of an Origin-Destination Matrix from Traffic Counts," *Transportation Science*, vol. 17, no. 2, pp. 198–217, 1983.

[10] Y. Shafahi and R. Faturechi, "A New Fuzzy Approach to Estimate the OD Matrix from Link Volumes," *Transportation Planning and Technology*, vol. 32, no. 6, pp. 499–526, 2009.

[11] E. Cascetta, A. Papola, V. Marzano, F. Simonelli, and I. Vitiello, "Quasi-Dynamic Estimation of OD Flows from Traffic Counts: Formulation, Statistical Validation and Performance Analysis on Real Data," *Transportation Research Part B: Methodological*, vol. 55, pp. 171 – 187, 2013.

[12] A. Talebian and Y. Shafahi, "The Treatment of Uncertainty in The Dynamic Origindestination Estimation Problem Using a Fuzzy Approach," *Transportation Planning and Technology*, vol. 38, no. 7, pp. 795–815, 2015.

[13] J. Xie, Chiand Duthie, "An Excess-Demand Dynamic Traffic Assignment Approach for Inferring Origin-Destination Trip Matrices," *Networks and Spatial Economics*, vol. 15, no. 4, pp. 947–979, 2015.

[14] J. G. Wardrop, "Road Paper. Some Theoretical Aspects of Road Traffic Research," *Proceedings of the Institution of Civil Engineers*, vol. 1, no. 3, pp. 325–362, 1952.

[15] J. Rosinowski, "Entwicklung und Implementierung eines PNV-Matrixkorrekturverfahrens mit Hilfe von Methoden der Theorie unscharfer Mengen," Master's thesis, University of Karlsruhe, 1994.

[16] L. Novako, L. Simunovic, and D. Krasi, "Estimation of Origin-Destination Trip Matrices for Small Cities," *Traffic and Transportation*, vol. 26, no. 5, pp. 419–428, 2014.

[17] Agenzia Mobilita Metropolitana Torino, "IMQ 2010: Indagine sulla Mobilit delle Persone e sulla Qualit dei Trasporti. Area Metropolitana e Provincia di Torino," 2010.

[18] F. Calabrese, G. DiLorenzo, L. Liu, and C. Ratti, "Estimating Origin-Destination Flows Using Mobile Phone Location Data," *IEEE Pervasive Computing*, vol. 10, no. 4, pp. 36–44, 2011.

[19] V. Frias-Martinez, C. Soguero, and E. Frias-Martinez, "Estimation of Urban Commuting Patterns Using Cellphone Network Data," in *Proceedings of the ACM International Workshop on Urban Computing*. New York, NY, USA: ACM, 2012, pp. 9–16.

[20] Y. Zhang, "Travel Demand Modeling Based on Cellular Probe Data," Ph.D. dissertation, University of Wisconsin, 2012.

[21] S. Bekhor, Y. Cohen, and C. Solomon, "Evaluating Long-distance Travel Patterns in Israel by Tracking Cellular Phone Positions," *Journal of Advanced Transportation*, vol. 47, no. 4, pp. 435–446, 2013.

[22] ISTAT. [Online]. Available: http://www.istat.it/it/archivio/open+data

[23] Elenco Aziende Italiane. [Online]. Available: http://www.elencoaziendeitaliane.it

[24] ATECO. [Online]. Available: http://en.istat.it/strumenti/definizioni/ateco/

[25] D. A. Badoe and E. J. Miller, "TransportationLand-Use Interaction: Empirical Findings in North America, and Their Implications for Modeling," *Transportation Research Part D: Transport and Environment*, vol. 5, no. 4, pp. 235–263, 2000.

[26] QGIS. [Online]. Available: http://www.qgis.org/en/site/