Regression based estimation of traffic flows using multi-source data

Francesco Bianchi[1[0009−0009−0380−1093]], Giovanni Felici[1,2[0000−0003−0544−5407]], and Giuseppe Stecca[1,2[0000−0001−5876−4538]]

[1] Istituto di Analisi dei Sistemi ed Informatica “Antonio Ruberti" CNR-IASI, Rome, Italy, francesco.bianchi@iasi.cnr.it
[2] CNR-CISC - Interdepartmental Centre for City Science, Rome, Italy

Abstract. The successful exploitation of emergent digital technologies and advanced location and sensor-based data analytics can effectively support urban mobility. In this context, the digital twin of mobility framework considers the real-time availability of data from diverse sources. The integration of these sources enables the simulation of mobility behavior and facilitates decision support at the urban scale.

In this work we propose a procedure to integrate multi-source data to road network arcs with the aim to calibrate traffic models.

Data are gathered from different types of road sensors but also from cellular network. Presence data are allocated to the road arcs. Regression models are run on the collected dataset and results allow to infer relations among static presence data and traffic flow patterns. A test case on the Italian city of Matera is presented.

Keywords: regression analytics · digital twin of mobility · multi-source data.

1 Introduction and Literature Review

Urban flow analysis based on sensor data has provided a huge push on developing methods to analyze and synthesize the way people move on urban areas and how they use mobility services.

The use of digital technology in smart mobility is reported in several papers, such as in [1]. The recent advances of the research in Digital Twins (DT) consider the application of DT to urban mobility. In this case advanced algorithms are required in order to use sensor data to calibrate traffic simulation model [3].

Several sources of data can be integrated in order to improve the underlying digital twin models. In this work we deal also with occupancy data. [4] describes different technologies and different applications of using data for occupancy profiles. The possibility to integrate these data to forecast mobility behavior of people but also occupancy has been deepened in [2].

2 Methodology

We present an approach to integrate several sources of data in order to calibrate mobility macrosimulation data for the development of a Mobility Digital Twin.

The methodology is based on the imputation of the data to the road graph that is successively used for traffic simulation. The graph considers both real data and simulation data; values assigned to arcs can be compared in order to calibrate the model, but also to infer mobility behavior. Specifically, we use flow sensors to monitor traffic generated by vehicles and pedestrians across various sections of the road network. Additionally, we leverage data from a telecom company that can detect the presence of mobile devices within a grid of 150x150 meters categorized by user profile, with high frequency (i.e., every 15 minutes). Occupancy data cover all the urban area but is less precise than sensor data. Our methodology considers the proportional imputation of presence to road arcs. The allocation is done considering tiles crossed by each arc; then, for each tile, the total length of arcs contained in that tile is cumulated. Finally, for each arc and for each crossed tile, the imputation percentage of occupancy is computed. The procedure can be adjusted considering the class of the arcs (e.g., single lane, high speed, and the like).

After allocating on the road graph all the multi-sensor data, regression models are used to relate sensor and occupancy data. Features considered are: time span, day of the week, arc id, data allocated, type of arc, urban zone type, and tile density, as expressed by the number of Points of Interest.

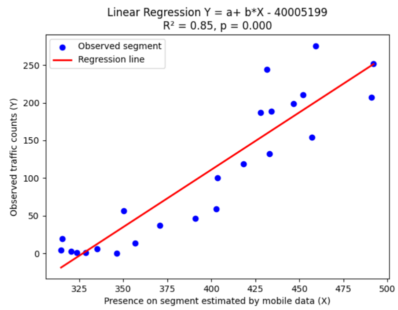
3 Results and conclusion

We tested the approach on the road network of the city of Matera. Data from TIM cellular network and road sensor data coming from standard counters based in infrared and cameras, and traffic-restricted zone surveillance cameras and combined. Regression models are used on a subset of location of the network are used to extend traffic flow calibration and refinements to the rest of the network with a good degree of precision.

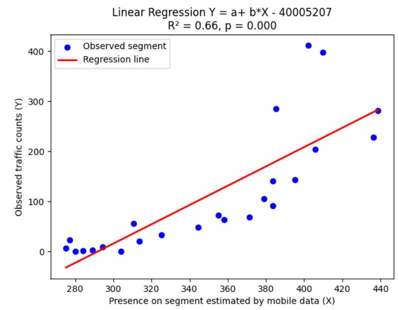
Current tests comprise the projection of presence data from tiles to arcs of the network, based on the actual proportion of the arc covered by one or more tiles.

Multiple regression results with occupancy and hour of the day as independent variables show promising R^2 coefficient, often above 0.80. Figure 1 shows simple regression lines for a subset of points of the network where both sensor data and mobile presences are present (the latter are projected onto arcs as described above). Two locations for each type of sensor (infrared, radar, LTZ radar) are presented.

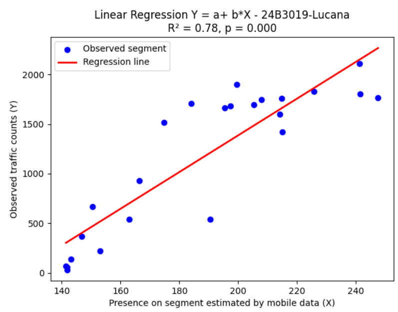
Future works will consider comparison among different models, and integration of other territorial information, as type of buildings, and use type.



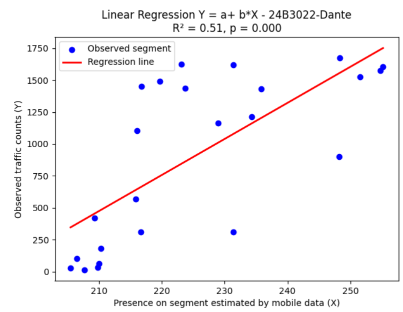
(a) Simple regression for Infrared pedestrian - Sensor 40005199



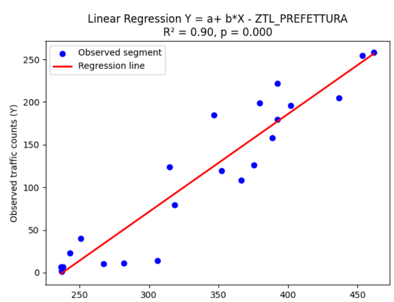
(b) Simple regression for Infrared pedestrian - Sensor 40005207



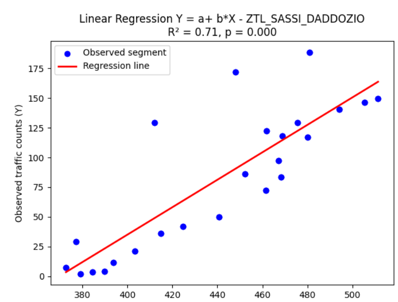
(c) Simple regression for vehicular radar - Via Lucana



(d) Simple regression for vehicular radar - Via Dante



(e) Simple regression for LTZ radar - Prefettura



(f) Simple regression for LTZ radar - Sassi D'Addozio

Fig. 1: Regression between sensors data and Mobile occupancy from tiles.

Acknowledgments. This work is partially supported by the project "House of emerging technologies of Matera" (CTEMT) funded by the Ministry of Economic Development of Italy, CUP I14E20000020001, and and CNR DIT.AD106.097 project UISH - Urban Intelligence Science Hub, CUP B51B1200430001.

Disclosure of Interests. No competing interests.

References

1. L. Butler, T. Yigitcanlar and A. Paz Smart Urban Mobility Innovations: A Comprehensive Review and Evaluation *IEEE Access*, vol. 8, pp. 196034-196049, 2020.
2. R. De Benedictis, A. Cesta, R. Pellegrini, M. Diez, D. M. Pinto, P. Ventura, G. Stecca, G. Felici, A. Scalas, M. Mortara, *et al.* Digital twins for intelligent cities: the case study of Matera. *Journal of Reliable Intelligent Environments*, vol. 11, no. 1, p. 6, 2025.
3. AL. Pala, G. Felici, V. Torrisi, A. Salvatore, P. Leonardi, G. Stecca, M. Ignaccolo, and G. Inturri. "Digital Twin for mobility: simulation integrated approach for demand matrices calibration with empirical data." *Transportation Procedia Research*, to appear, 2024.
4. S. Nejadshamsi, U. Eicker, C. Wang, and J. Bentahar. "Data sources and approaches for building occupancy profiles at the urban scale—A review." *Building and Environment*, vol. 238, p. 110375, 2023, Elsevier