Estimating travel demand based on OpenStreetMap in the context of urban digital twins

Lotte Notelaers*[1], Chris M.J. Tampère[2], Joachim Landtmeters

[1]PhD candidate, KU Leuven Institute for Mobility (CIB), KU Leuven, Belgium

[2]Professor, KU Leuven Institute for Mobility (CIB), KU Leuven, Belgium

SHORT SUMMARY

A digital twin of a smart city captures the dynamics of people interacting within the urban environment composed of several interconnected systems. Urban mobility, which is simulated by a traffic model within an urban digital twin, is becoming an increasingly complex component within this ecosystem. Demand modeling is an important part of the set-up of a traffic model for a city. Nevertheless, most cities do not have the data, budget, or experience to estimate the demand in their region. Therefore, this study develops a demand generation method that enables to easily estimate travel demand for any region. It is a trip-based modeling approach based on the freely-available land-use data of OpenStreetMap (OSM). The model is applied to a case study of Antwerp as part of the Flemish DUET pilot. It is shown a crude estimate of travel demand can be generated from widely available open OSM data.

Keywords: Demand modeling, OpenStreetMap (OSM), Multi-modal transportation, Big data analytics, Digital Twins

1. INTRODUCTION

According to IBM, a digital twin is “a virtual representation of an object or system that spans its lifecycle, is updated from real-time data, and uses simulation, machine learning and reasoning to help decision-making” (IBM, n.d.). Considering cities as systems, this technology is applied to develop urban digital twins to model the dynamics of these complex socio-technical ecosystems.

Within a city’s ecosystem, mobility and transportation form one of the main domains, along with buildings, spatial planning, water and energy management, etc. The field of traffic modeling yields many algorithms and techniques to model the interactions between people, transport services and the available infrastructure.

An essential component of a traffic model within a city`s digital twin is the projection of travel demand in the region. As most local authorities do not have such an estimate, nor the data or experience to generate one, this study proposes a method that allows for easily estimating travel demand for any region given values for model parameters are available, estimated or calibrated based on a comparable region.

As part of the DUET - Digital Urban European Twins – project (Tampère, 2021), this study proposes a demand generation method that combines trip-based modelling (Ortúzar and Willumsen, 2011) with the use of disaggregated data from OSM. This short paper provides proof of this concept by building a travel demand model for a particular Flemish region based on data extracted from OSM. The performance is evaluated, and the transferability of model parameters to a different but comparable region is analyzed to validate the concept.

2. METHODOLOGY

The methodology is divided into five parts (Figure 1).

Figure 1: Methodology

The **first step** in the demand generation process is to delineate the study area and divide it into TAZs for which an OD-matrix will be generated. For the time being, no external demand (trips going into or out of the study area from or to the surrounding cities) has been taken into account.

The **second step** includes the extraction of relevant OSM data and the removal of inconsistencies in the data, such as, overlapping and contained building and land-use polygons. Further, the geometries are classified according to their attribute values. For example, for the building attribute a categorization into accommodation, commercial, civic/amenity, etc. exist. Lastly, because the extracted geometries may have non-null values for multiple attributes and hence, be classified in various categories, a priority ranking is specified to reduce the multi-dimensionality to only one dimension.

The **next step** uses this pre-processed data to create separate unidimensional layers for the residential and activity type. These two layers can be summarized by two questions: Where do people reside? Where do people do things?

Residential layer
The process to attain a layer with residences of different types starts by selecting geometries with the right attribute values such as geometries being labelled as accommodation buildings. However, the labeling of objects is sometimes incomplete or incoherent, making some extra filtering methods necessary. Six types of houses are distinguished, three of them follow from available attribute information (large residential, student residential and single residential). The other types classify the undefined building features in accordance with some additional applied rules.

Activity layer
The approach creating the activity layer starts from the POI dataset which disregards land-use polygons and objects already included in the residential layer. For classifying the remaining POIs, the NACE classification system is used. Going over existing attribute-value pairs, every pair was mapped manually to an activity class. Features with an attribute-value priority not belonging to any activity class are dropped from the activity layer.

The **fourth step** is trip generation. The created residential and activity layers give insights into possible origin and destination locations of trips. The number of POIs of different types can serve as the starting point to estimate travel between zones. In this paper, multiple linear regression is used for estimating production and attraction separately.

The **final step** towards an OD-matrix is the distribution step. The goal of the distribution model is to match trip makers' origins and destinations into actual trips. For this purpose, a standard gravity model is used.

3. RESULTS AND DISCUSSION

The trip generation and trip distribution models are calibrated for the case study of Antwerp that considers the morning peak car traffic. For this purpose, available demand data from the Flemish department of Mobiliteit en Openbare Werken (MOW) is considered as ground-truth. The true production and attraction per zone were calculated by, respectively, summing the values of each row (origin) and column (destination) in the extracted OD-matrix.

Modeling the **zonal production** is based on the residential layer. This is because residences can be expected as important origins of morning peak trips. A significant regression equation for production was found with an R^2 of 0.754.

For modeling the **zonal attraction**, the activity layer is used. Following the same analogy as used for zonal production, activities can be considered the main destinations of morning peak trips. A significant regression equation for attraction with an R^2 of 0.640 was obtained.

For modeling the **trip distribution**, a combined power-exponential function is calibrated and used by the gravity model as a deterrence function:

$$F(c_{ij}) = \alpha \cdot c_{ij}^{-\beta} \cdot e^{-\gamma c_{ij}} \tag{1}$$

In this study, the travel impedance c_{ij} is calculated as the shortest path free flow travel time between the centroids of the origin and destination zone. The extracted OD-matrix from the MOW data is used for calibration of α, β, and γ. Based on the average number of trips made in each travel time range, the trip cost distribution is derived (Figure 2).

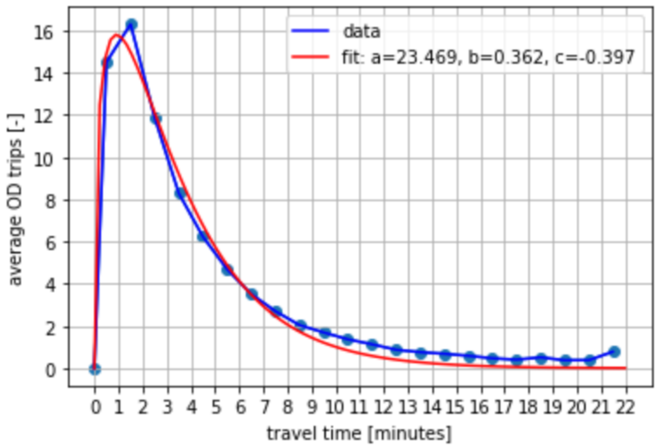


Figure 2: Calibrated deterrence function

Next, the **performance** of the trip generation and distribution models is evaluated. The estimated and predicted zonal production and attraction are visualized in Figure 3. The MAPE metric shows errors ranging from 10% up to 30% for production and from 10% up to 40% for attraction (excluding the outliers[1]).

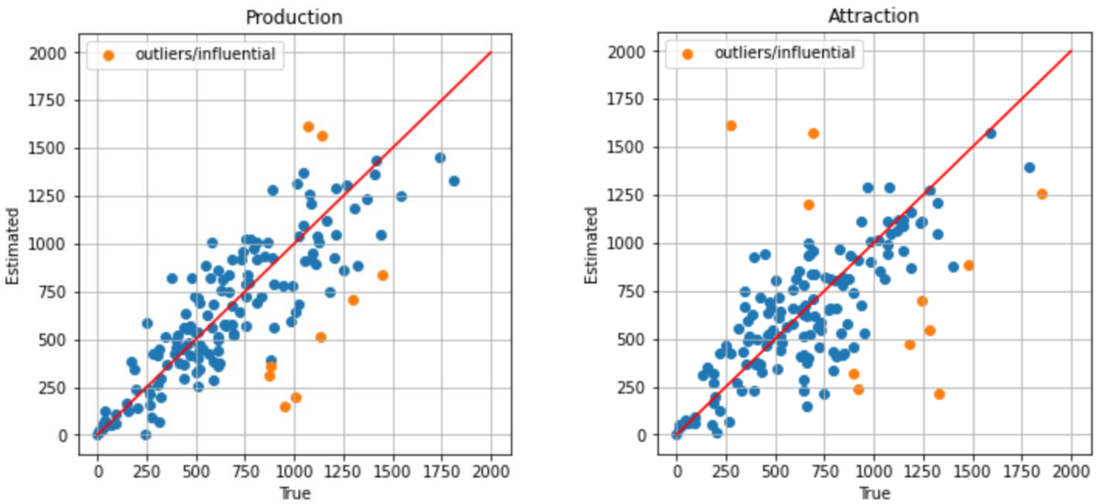


Figure 3: True and estimated (a) production and (b) attraction values

To evaluate the performance of the trip distribution model itself (and not the performance of the combined trip generation and distribution model which is discussed further in the abstract), first, the true production and attraction values derived from the MOW OD-matrix are used as inputs for the model. Overall, the R^2 value between the estimated and true OD-values is equal to 0.765. The prediction error slightly increases for OD-pairs with more OD-trips (Figure 4).

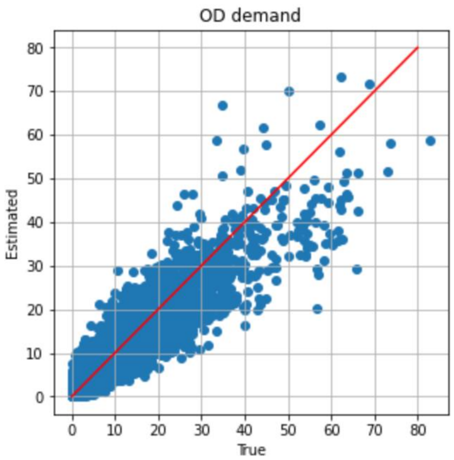
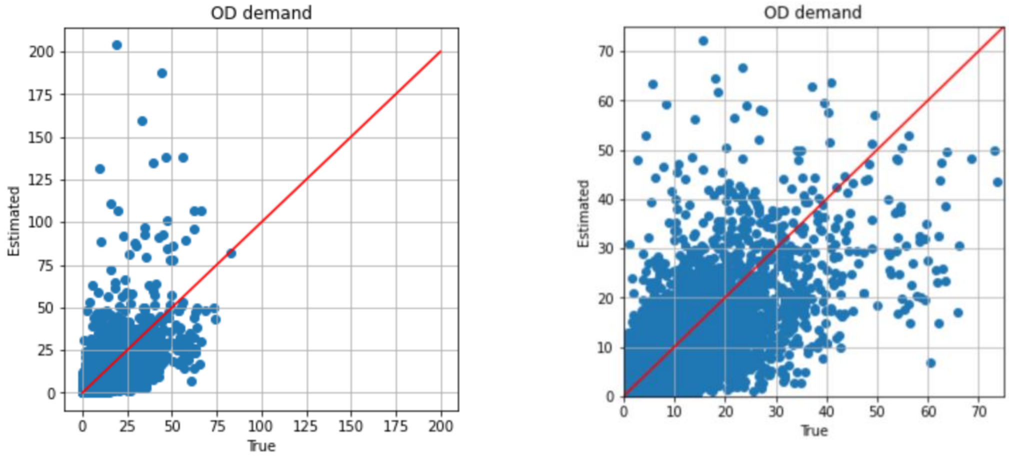


Figure 4: True and estimated trips per OD-pair

Both the trip generation and distribution models separately proved to be imperfect. Sequentially performing both steps (as depicted in Figure 1), these errors will be accumulated. As error-containing zonal estimates for production and attraction are used as inputs, the distribution model distributes these errors further over the OD cells. The resulting R^2 value is -0.087 (Figure 5). Overall, the number of trips is underestimated. This is the result of total production and attraction being underestimated by the trip generation models. The MAPE ranges between 40 and 55%. This may at first seem a disappointingly large error. However, considering the very limited POI

[1] Based on statistical analyses, influential observations were dropped

information used as input to the procedure, it is remarkable that so much structure of aggregated and disaggregated OD-demand can be captured.



(a) Full result

(b) Zoom excluding estimates above 75 trips

Figure 5: Trips per OD-pair (combined generation and distribution)

Validation on Ghent

The model is validated on the city of Ghent. The R^2 values between the estimated and true production, attraction, OD-values starting from the true production and attraction, and the OD values obtained after combined generation and distribution are equal to 0.667, 0.247, 0.682, and -0.822, respectively. This validation assesses the portability of the calibrated parameters to estimate production and attraction or generate an OD-matrix for other comparable cities in Flanders. From these results one can conclude that the performance is only slightly lower than for Antwerp.

4. CONCLUSIONS

Within this study, a proof of concept showing that a crude estimate of travel demand can be generated from widely available open OSM data is given. This substitutes expensive and time-consuming travel surveys and reduces the dependency on inflexible governmental data to estimate travel demand.

ACKNOWLEDGMENT

This research is developed as part of the Flemish DUET pilot which is funded from the EU Horizon 2020 research and innovation program under grant agreement No.870697. Moreover, MOW is acknowledged for providing travel demand data of Flanders.

REFERENCES

IBM. (n.d.). *What is a digital twin?* Retrieved January 13, 2022, from https://www.ibm.com/topics/what-is-a-digital-twin

Ortúzar J. de D., Willumsen L.G. 2011. *Modelling Transport*, 4th edition. ed. Wiley, Chichester, West Sussex, United Kingdom.

Tampère, C. (2021). *DUET Deliverable D3.4: Smart City domains, models and interaction frameworks v2*. Retrieved from www.digitalurbantwins.eu