Enriched travel demand estimation by including zonal and traveler characteristics and their relationships

Van Der Vliet, J. P.; Pel, A. J.; Van Lint, J. W.C.

DOI
10.1109/MTITS.2017.8005696

Publication date
2017

Document Version
Accepted author manuscript

Published in
5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2017 - Proceedings

Citation (APA)
Van Der Vliet, J. P., Pel, A. J., & Van Lint, J. W. C. (2017). Enriched travel demand estimation by including zonal and traveler characteristics and their relationships. In *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2017 - Proceedings* (pp. 351-355). Article 8005696 Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/MTITS.2017.8005696

Enriched travel demand estimation by including zonal and traveler characteristics and their relationships

J.P. van der Vliet*†, A.J. Pel*, J.W.C. van Lint*
*Dept. of Transport and Planning
Delft University of Technology
Delft, The Netherlands
email: j.p.vandervliet@tudelft.nl
†Studio Bereikbaar
Rotterdam, The Netherlands

Abstract—**Current procedure in travel demand estimation models is to separately deal with attraction, production and trip distribution, where the latter typically assumes inverse distance proportionality. We show that this procedure leads to errors in the demand estimation, particularly when dealing with very specific zones and heterogeneous travel behavior. We argue that this traditional procedure is rooted in traditional ways of data collection, while new (big) data sources allow direct observation of travel demand patterns. Using such data, we propose an enriched travel demand estimation method in which zonal and traveler characteristics and their relationships are consistently carried over from the empirical data into the demand model. This can improve both the validity and richness of demand estimations.**

Keywords—travel demand; demand estimation; big data

I. ASSUMPTIONS IN CURRENT DEMAND ESTIMATION MODELS

Transport models are an important tool in the development and evaluation of policy and planning measures. A central component within these models is the origin-destination matrix (OD-matrix) that describes the travel demand pattern: the number of trips and their origins, destinations and modes. Considering currently used demand estimation models [1, 2], the production and attraction of zones (i.e. trip generation) are estimated separately from the trip distribution. Although this separation may appear practical, as it allows separate data collection and model estimation, it also inherits a number of errors.

A consequence of this separation is that any unit of attraction or production is equal to any other. Considering, for example, a general gravity model:

$$T_{ij} = \alpha O_i D_j f(c_{ij}) \quad (1)$$

Thus, the number of trips T from zone i to zone j is proportional to the attraction D of zone j and the production O of other surrounding zones. This proportionality is equal for all travelers. Given this assumption of homogeneity, travel resistances are the only governing variable for destination choice. That is, the number of trips from i to j is inversely proportional to the travel resistance $f(c_{ij})$, which is typically ultimately a function of distance.

These assumptions can be poorly warranted by data, but are instead the result of limitations in the data the models are based on. In past decades, the primary source of data has been travel diaries, which are necessarily limited in size and scope [3]. Recently, new data sources have become available. Smartphone applications and cellular network data, for instance, allow the collection of travel itineraries of many participants over extended periods of time at reasonable costs. This new information allows scrutiny of the mentioned assumptions, on a larger scale than was previously possible [4]. In this paper, we scrutinize the issue of distance proportionality, particularly with respect to very specific zones and heterogeneous travel behavior.

As we will show, the heterogeneity in travel behavior causes very different patterns than those estimated by models that operate by the assumption of independent trip generation and trip distribution. This problem manifests itself in two ways: Firstly, differences in the characteristics of the destination zones (e.g. local versus regional relevance, activity and socio-cultural characteristics) cannot be captured, as nearby destination zones with similar attraction will attract trips from the same origin zones in the same proportions, when using only distance proportionality. Secondly, differences in traveler characteristics that affect the destination choice are not properly modeled under distance proportionality.

The problem is illustrated using three case studies. In the first study, we consider two nearby zones that are very similar in terms of model parameters (size and type of activities), while in reality, there are marked differences in zonal characteristics. Here we inspect the face-validity of the model results. In the second study, we consider a zone dominated by a hospital and associated with relatively uniform local attractiveness. The modeled travel patterns are compared to the results of a survey conducted among hospital staff, patients and visitors. Finally, the third study concerns a comparison of modeled and observed origins of traffic towards a medium-sized city. Each of these cases illustrates part of the problem of disregarding zonal and traveler characteristics, which occurs when separating trip generation and trip distribution.

In the ensuing, in section II we introduce the case studies in further detail, and present the study results in section III. Then in section IV we propose a new approach to include zonal and traveler characteristics in demand estimation. And finally, we draw conclusions on the identified problem and the new approach in section V.

II. METHOD

This paper recounts an empirical investigation of demand estimation model results. Firstly, we investigate their face-validity in the Hilversum case. Secondly, the results are compared visually to empirical data in both the Jeroen Bosch Hospital and Den Bosch cases.

Model-estimated OD-matrices used in the case studies are derived from the VENOM model system [5]. This is an adaptation of the NRM model system, developed and maintained by the Dutch Ministry of Infrastructure and the Environment [6]. It encompasses modules for trip generation, distribution, modal split and traffic assignment on a macroscopic level, for the whole of the Netherlands. It is used for mobility prognoses on both the national and regional level. The model system is considered to be state-of-practice [7]. In the trip distribution step, the system uses a nested logit model for combined estimation of destination and mode choice, instead of a gravity model. The assumption of distance proportionality still applies to this model type as the trip generation and trip distribution steps are separate and the utility function of the model is based on travel time and cost, both of which are functions of distance. All matrices used describe passenger car trips during the morning peak period, for all trip purposes.

A. Hilversum

This first case study considers the face-validity of model results when dealing with specific zones: Two zones in the city of Hilversum, The Netherlands, are compared that are similar in terms of model parameters: they are nearby and connected to the highway network through the same secondary roads. Additionally, they both contain commercial and/or industrial activity. The real-world characteristics of the two zones are very different: the first zone is the complex that houses the television and radio studios of the Dutch national public broadcasters (i.e. the zone is of national relevance). The second zone comprises general commercial and industrial activities, of local relevance. The model-estimated origins of traffic to the zones in the morning peak period are compared visually in subsection III-A. The 2020 Global Economy scenario matrix was used for both visualizations.

B. Jeroen Bosch Hospital

The second case study compares a modeled and empirical trip distribution for one zone with local relevance. More specifically, it is a visual comparison of observed and model-estimated origins of morning peak period traffic to a single zone containing a hospital. The base-year matrix (2010) is compared to the results of a survey conducted in 2012 among hospital staff, patients and visitors [8]. In the survey, respondents were asked to state the place of origin of their trip to the hospital. The survey results are used here as a proxy for the trips to the zone the hospital is in, in Den Bosch, The Netherlands.

C. Den Bosch

The third case study is also a comparison of an empirical and modeled trip distribution, here pertaining to a medium-sized city with regional significance. The ability of the model to capture the regional relevance of the area and heterogeneity in travel behavior is considered. We visually compare the origins of model-estimated morning peak car traffic to the urban area of the city of Den Bosch, The Netherlands, to observations. The observations in question are cell-phone network detections, collected and processed by Mezuro [9]. The matrix visualized is the weekday morning peak period average of observed car trips in September 2016. The model-estimated matrix visualized is the 2020 Global Economy scenario matrix.

III. RESULTS

In the following figures, we visualize the model results and empirical data, indicating origins of traffic (in blue) and the destination zones (in orange). The shade of the blue color indicates the number of trips originating from that zone, corrected for the size of the origin zone. The darker the color, the more trips originate per unit of area. A proportional color scale with the same number of bins (9) is used in each image. The visualizations in all case studies use the same color scheme.

A. Hilversum case study

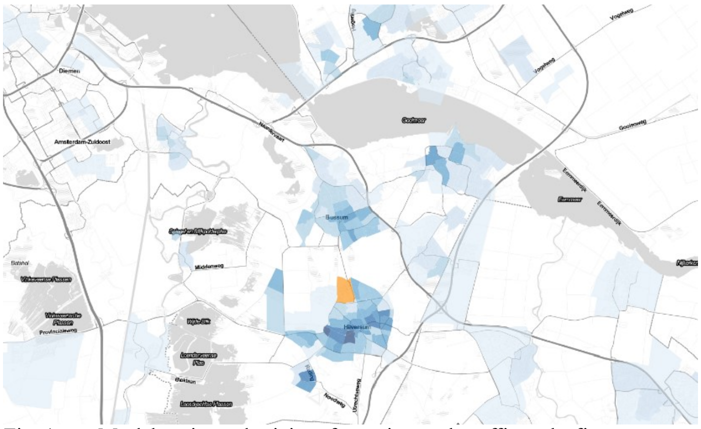

Fig. 1. Model-estimated origins of morning peak traffic to the first zone in Hilversum.

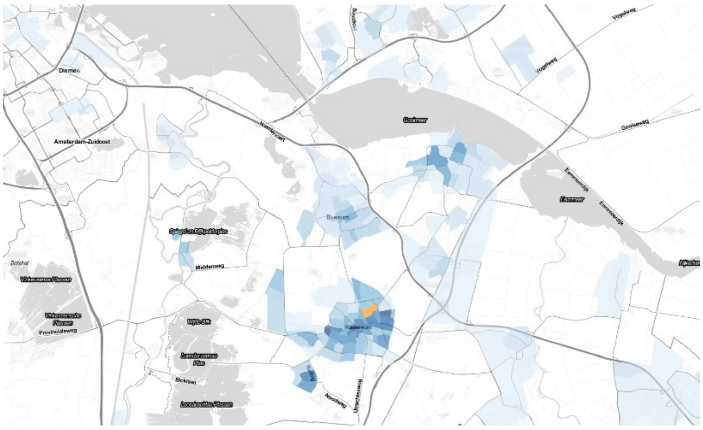

Fig. 2. Model-estimated origins of morning peak traffic to the second zone in Hilversum.

Figures 1 and 2 show the model-estimated origins of morning peak traffic to two destination zones in close proximity. One can observe that both images are very nearly identical. This confirms that the zones are indeed similar in terms of model parameters. Both images show the same nearby population centers, as would be expected given the assumption of distance proportionality. Thus, the model is not able to account for any differences in characteristics of the two zones, as expected. While the second zone has a local relevance, the first has national relevance, which the model estimates do not show. A zone with a national relevance will attract trips from more distant origins. Thus, for such a zone, one would expect a lower proportion of traffic to originate from nearby population centers, and more from large population centers further away.

B. Jeroen Bosch Hospital case study

The figures 3 and 4 contain visualizations of origins of traffic to the Jeroen Bosch Hospital zone as estimated and as observed respectively.

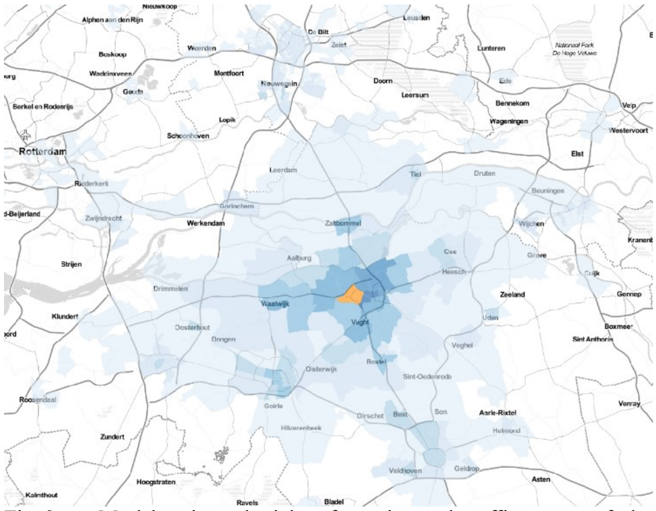

Fig. 3. Model-estimated origins of morning peak traffic to zone of the Jeroen Bosch hospital.

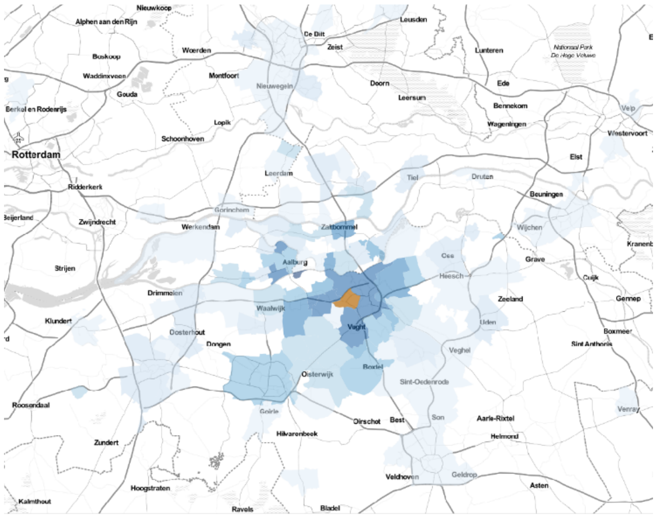

Fig. 4. Actual origins of morning peak traffic to the Jeroen Bosch hospital.

The images are again very similar. The extents of the attractiveness and its distribution exhibit the same general pattern. The only noticeable difference is that the model gives a more uniform distribution of traffic origins than the survey responses. The hospital and its surrounding zone (generic housing, industrial and commercial activity) have a local and uniform attractiveness. The model appears to perform well in these conditions.

C. Den Bosch case study

Figure 5 is a visualization of the origins of traffic to the Den Bosch urban area as modeled. Figure 6 shows the observations from cellular network detections. The data is visibly courser; this is due to the required aggregation in the processing of cellular phone detections to prevent identification of individuals.

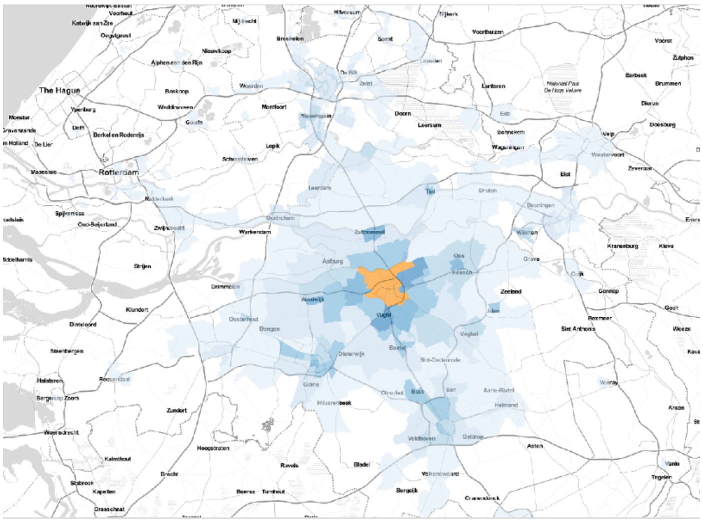

Fig. 5. Model-estimated origins of morning peak traffic to the city of Den Bosch.

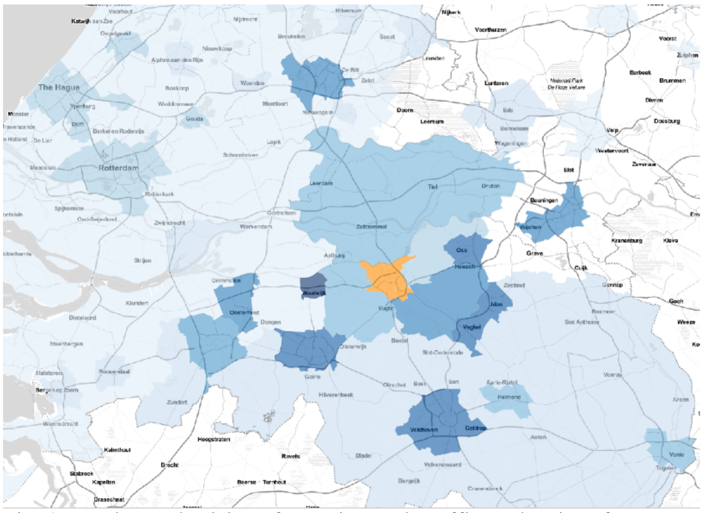

Fig. 6. Observed origins of morning peak traffic to the city of Den Bosch.

The city of Den Bosch (152,472 inhabitants as of January 2017 [10]) has much more than just local attractiveness, as is visible in figure 6. Regional urban centers show clearly, especially in the province of Noord-Brabant, of which Den Bosch is the capital. The model in figure 5 shows mostly local origins, just as in figures 1, 2 and 3. Even though the model does show traffic originating in several urban areas in the region, the proportions are very different: Observations show that a much larger proportion of trips originate from urban areas, and from further away. Apparently, the Den Bosch urban area is more attractive to those living in other urban areas than those living in more rural areas. The model is not able to account for the regional relevance of the Den Bosch urban area, nor the predominantly urban origin of trips.

D. Summary of results

In the Hilversum case, we expect differences in pattern between the two zones, as one has national relevance, while the other has only local relevance. The model does not capture these differences and shows uniform, local relevance for both zones. The face-validity of model estimations can be considered poor in this case. Next, we compare model estimations to observations from a zone with local relevance. The model performs well for that zone. Lastly, in the Den Bosch case study, we look at a city with regional relevance and show that the model is unable to capture both the regional relevance, and the heterogeneity in behavior (i.e. predominantly urban trip origins).

These results show that models operating by the assumption of independent trip generation and trip distribution can perform well in the case of local, uniform relevance. However, when a zone has a wider relevance, or in the case of behavioral heterogeneity, errors occur as the model cannot take either of this into account.

IV. Discussion: Towards a New Approach

To counter the problem demonstrated, we propose an alternative approach to demand estimation that explicitly uses and conserves observed travel behavior patterns, and their relationship to local/regional characteristics. In essence, the estimation of production, attraction and distribution are integrated. Figure 7 contains a sketched overview.

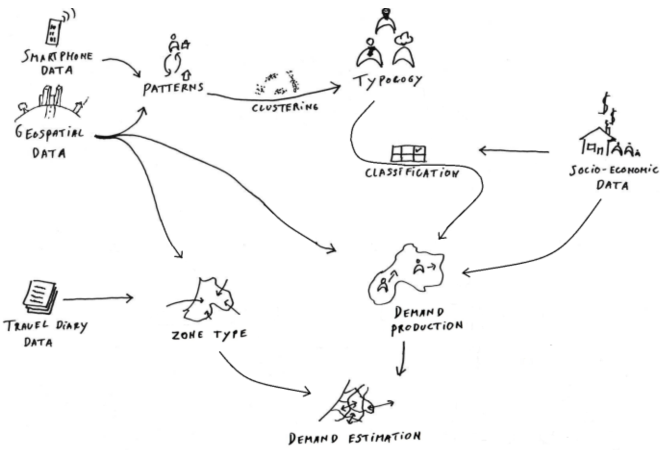

Fig. 7. Overview of proposed approach.

Starting point is a large database of recorded travel itineraries, i.e. multiple trips per person spanning an extended period, on the order of several weeks. This data is collected through smartphone apps. Typical travel patterns are then identified using machine-learning clustering algorithms [11-13]. Using classic travel diary survey data -- with a stratified sample -- and a classification algorithm, the travel patterns are connected to socio-economic variables.

The (future) socio-economic and spatial characteristics of a zone ("zone type") can now provide information on the travel behavior that the zone produces, through the typical travel patterns associated with those characteristics [14, 15]. Additionally, the travel patterns contain information on the destination zone type, modality, trip length constraints and departure time. Conversely, the zone type and patterns provide information on the attraction of that zone. Together, this results in significantly reduced difference in dimensionality between the input data and resulting OD matrix.

Most importantly, this approach enables predicted flows to be traced back to the type of behavior that they originate from, and accounts for the relationship between travel behavior patterns, and local/regional characteristics. This method has two further advantages: Firstly, it produces richer demand predictions, that support the development and evaluation of policy targeted at specific types of travel behavior and/or zones. Secondly, the face validity is enhanced by explicitly including zonal and traveler characteristics and their relationships. The issues originating from separating the trip generation and trip distribution steps demonstrated in this paper are thus sidestepped. Possible drawbacks are that this approach is significantly more complex than current methods, and that it relies on the availability of large numbers of recorded travel itineraries.

V. CONCLUSIONS

Current models based on the assumption of distance proportionality are not able to account for the relationship between travel behavior and local/regional characteristics. The case studies in this paper show that this leads to errors in estimations. Our approach proposed here will allow the use of information on travel behavior, and its relationship to the built environment, contained in new (big) data sources to improve both the validity and richness of demand estimations. In the near future, we will further develop this approach into a viable method for estimating travel demand.

Acknowledgment

Both the results from the Jeroen Bosch hospital survey and the Mezuro observations dataset were made available for this research by Studio Bereikbaar (www.studiobereikbaar.nl).

References

[1] M. G. McNally, "The Four-Step Model," in *Handbook of Transport Modelling*, ed, 2008, pp. 35-53.

[2] J. D. Ortúzar and L. G. Willumsen, *Modelling Transport*. Chichester: Wiley, 1990.

[3] P. R. Stopher and S. P. Greaves, "Household travel surveys: Where are we going?," *Transportation Research Part A: Policy and Practice,* vol. 41, pp. 367-381, 6// 2007.

[4] P. K. Sikdar and B. G. Hutchinson, "Empirical studies of work trip distribution models," *Transportation Research Part A: General,* vol. 15, pp. 233-243, 1981/05/01/ 1981.

[5] S. Kieft and T. van der Linden, "VENOM Handboek," Metropool Regio Amsterdam, Amsterdam, 2017.

[6] C. Bannink, R. Clement, F. Hofman, M. Meijer, M.-J. Olde Kalter, H. van der Rest, *et al.*, "NRM Handboek," Adviesdienst Verkeer en Vervoer, Rotterdam, 2006.

[7] L. Tavasszy, M. Snelder, M. Duijnisveld, R. Haaijer, H. Meurs, R. van Nes, *et al.*, "Audit LMS en NRM Syntheserapport," TNO, Delft, 2012.

[8] R. van den Hamsvoort and B. Veroude, "JBZ onderweg 2013," DTV Consultants, Breda, 2013.

[9] W. Steenhakkers, "Method for collecting data of users of active mobile telephones," 2014.

[10] C. B. S. Netherlands. (2017). *Bevolkingsontwikkeling; regio per maand*. Available: http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=37230NED&D1=17-18&D2=101-650&D3=l&LA=EN&HDR=T&STB=G1,G2&VW=T

[11] A. Orriols-Puig. (2009, October). *Introduction to Machine Learning*. Available: http://www.slideshare.net/aorriols/presentations

[12] Q. Cao, B. Bouqata, P. D. Mackenzie, D. Messier, and J. J. Salvo, "A grid-based clustering method for mining frequent trips from large-scale, event-based telematics datasets," in *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, 2009, pp. 2996-3001.

[13] F. H. O. Cordón, F. Hoffmann, L. Magdalena, *Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases*. Singapore: World Scientific Pub Co Inc, 2001.

[14] H. J. P. Timmermans and P. van der Waerden, "Space-time consumption on the move," Technische Universiteit Eindhoven, Den Haag, 1998.

[15] S. Zhou, L. Yang, and L. Deng, "The spatial-temporal pattern of people's daily activities and transportation demand analysis-a case study of Guangzhou, China," in *2010 International Conference on Management and Service Science, MASS 2010*, 2010.