

Regression Analysis to Create New Truck Trip Generation Equations for Medium Sized Communities

Mehrnaz Doustmohammadi¹, Michael Anderson^{1*}, Ehsan Doustmohammadi²

¹Department of Civil Engineering, The University of Alabama in Huntsville, Huntsville, USA ²Jacksonville, USA

Email: *andersmd@uah.edu

How to cite this paper: Doustmohammadi, M., Anderson, M., & Doustmohammadi, E. (2019). Regression Analysis to Create New Truck Trip Generation Equations for Medium Sized Communities. *Current Urban Studies, 7*, 480-491. https://doi.org/10.4236/cus.2019.73024

Received: August 24, 2019 Accepted: September 23, 2019 Published: September 26, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

http://creativecommons.org/licenses/by/4.0/

Abstract

This paper uses data from a trucking origin/destination study conducted with global positioning system (GPS) technology to develop a truck trip generation model for medium sized urban communities-in this study taken to be communities between 200,000 and 1,000,000 people. The difficulty with developing truck trip generation equations centers on the limitation of data. For passenger transportation, data are collected from household surveys. For truck transportation, if available, data are typically collected from a small collection of shippers/businesses within the urban area and extrapolated to cover the entire study area. Because of the data limitations, truck transportation is typically indirectly modeled or as an after-thought. Increasing truck volumes, coupled with cost saving strategies such as just-in-time delivery systems, require that transportation policymakers analyze infrastructure needs and make investment decisions that explicitly include truck volumes as a component. This paper contains a case study using a medium sized urban area and a GPS collected set of truck origins and destinations to develop a truck specific trip generation equation using standard employment data. The paper presents the models developed and validates the models to the case study community. The paper concludes that the trip generation equations developed can be incorporated into medium sized community travel models to provide a framework for truck planning that can be used to improve resource allocation decisions.

Keywords

Truck Trip Generation Models, Travel Models, Employment Data

1. Introduction

The efficient and effective movement of freight is a critical component in the

transformation and growth of the economy. Transportation infrastructure improvements can assist in moving freight and growing the economic vitality of a community. To evaluate the best investment decisions, travel demand models, which are representations of the existing transportation infrastructure, are used to evaluate "what if" scenarios to future socio-economic growth and infrastructure decisions.

This paper examines the potential to use a database of freight origin/destination locations in a medium sized community to develop a calibrated truck trip generation equation. Medium sized communities are taken to be between 200,000 and 1,000,000 people for the focus of this work. The relationship between truck stops and employment was developed and then tested to determine the validity of the equations for use in practice and the level of accuracy associated with the use in the community travel demand model. The paper concludes that using a 20 sector model for truck trip generation can provide the level of detail necessary to incorporate truck transportation needs into the travel demand models, improve the results, and potentially lead to improved investment decisions for the community.

2. Literature Review of Freight Transportation

Truck demand modeling issues are prevalent as the development of methods for freight forecasting have lagged behind that of passenger transportation forecasting in both theoretical and simulation modeling analyses (Samimi et al., 2010; Alho, 2011; Fischer et al., 2005; Jong et al., 2004; Tavasszy & Jong, n.d.; Hunt & Stefan 2007; Jansuwan et al., 2014; Yang et al., 2010; North, 2009; Wheeler & Figliozzi, 2011; Chow et al., 2010; Cohen et al., 2008; Holguin-Veras et al., 2011). Additionally, data collection and availability are proving to be a large obstacle in the development and improvement of truck demand models (Bastida & Holguin-Veras, 2009; Samimi et al., 2010; Hunt & Stefan, 2007; Ruan et al., 2011; Greaves & Figliozzi, 2008; Roorda et al., 2010). Fortunately, some research studies have tested the use of GPS in collecting better data for modeling and found it to increase the quality of the models (Kuppam et al., 2014; Wheeler & Figliozzi, 2011; Greaes & Figliozzi, 2008; Doustmohammadi & Sisiopiku, 2016; Doustmohammadi et al., 2016b). Despite the issues being faced, researchers have been working on several variations of truck demand models. Activity-based, or tour-based, modeling approaches have been shown to provide the most accurate movement predictions (Samimi et al., 2010; Figliozzi, 2007; Kuppam et al., 2014; Gliebe et al., 2007; Hunt & Stefan, 2007; Ruan et al., 2011; Kim et al., 2011; Bradley et al., 2010; Doustmohammadi & Sisiopiku, 2016; Doustmohammadi et al., 2016b). Additionally, the ability to combine modeling techniques, primarily combing tour-based modeling with other types of modeling can meet specific needs (Alho, 2011; Fischer et al., 2005; Jansuwan et al., 2014; Holguín-Veras & Thorson, 2000; Holguín-Veras & Patil, 2008; Daly, 1982; Doustmohammadi et al., 2016a; Doustmohammadi et al., 2016b; Holguín-Veras et al., 2008; Holguín-Veras et al., 2013; Jong &

Ben-Aliva, 2007).

Many medium sized communities are still using traditional four step modeling techniques. The initial step in the model, trip generation, is the key area where advances in truck modeling can be made. The Quick Response Freight Manual (QRFM) and its updated version QRFM II were prepared to provide tools to assist transportation professionals in truck modeling (Cambridge, 1996; Cambridge, 2007). The models in QRFM have been shown to be useful in forecasting truck in a medium sized community (Anderson et al., 2013; Cambridge, 2007). However, more work can be performed and this paper seeks to develop new truck trip generation equations based on GPS collected truck origin and destination locations and zonal employment in the area.

3. Data and Methodology

The two main datasets used to create the truck trip generation equations were the GPS collected truck origins/destinations and employment data from the U.S. Census Department Longitudinal Employer-Household Dynamics (LEHD).

The GPS collected truck origin-destinations were collected from a sample of vehicles for four days during 2011. The data collected were purchased from the American Transportation Research Institute (ATRI) and contained over 4 million observations. When the data were analyzed, it was determined that there were over 119,000 locations, where the truck remained stationary for a sufficient time to allow for loading/unloading of freight to avoid identification of traffic congestion and red light stops to be considered in the database.

The employment data were collected for all Birmingham, AL area census blocks. The LEHD data, for 2010, included employment data divided into 2-digit NAICS code, see Table 1.

The employment data were combined with the GPS data for truck origin/destinations in a geographic information system (GIS) to determine the number of trucks that stopped in each census block and distribution in employment for the block. The combined data were then utilized to develop equations that related the number of trucks having an origin/destination in the census block with the number of and type of employment in the block.

The truck trip generation equations were developed using standard statistical methodologies for relating independent variables, employment by NAICS sector, with the dependent variable, the number of truck trips into and out of the location. Two different modeling techniques were applied to the data: Stepwise Linear Regression and Bayesian Linear Regression. The techniques represent common statistical regression methodologies for model development. For modeling purposes, a random sample of 70 percent of the blocks were used for model development and 30 percent of the blocks were used for validation of the model.

4. Model Results and Validation

The models developed from the different regression techniques were analyzed

| Table 1. NAICS data categories. |
|--|
| NAICS sector 11—Agriculture, Forestry, Fishing, Hunting |
| NAICS sector 21—Mining Quarrying, and Oil and Gas Extraction |
| NAICS sector 22—Utilities |
| NAICS sector 23—Construction |
| NAICS sector 31-33—Manufacturing |
| NAICS sector 42—Wholesale Trades |
| NAICS sector 44-45—Retail Trade |
| NAICS sector 48-49—Transportation and Warehousing |
| NAICS sector 51—Information |
| NAICS sector 52—Finance and Insurance |
| NAICS sector 53—Real Estate and Rental and Leasing |
| NAICS sector 54—Professional, Scientific and Technical Services |
| NAICS sector 55—Management of Companies and Enterprises |
| NAICS sector 56—Administrative and Support and Waste Management and Remediation Services |
| NAICS sector 61—Educational Services |
| NAICS sector 62—Health Care and Social Assistance |
| NAICS sector 71—Arts, Entertainment and Recreation |
| NAICS sector 72—Accommodations and Food Services |
| NAICS sector 81—Other Services (except Public Administration) |
| NAICS sector 92—Public Administration |

based on accuracy. For the Stepwise Linear Regression and Bayesian Linear Regression models, **Table 2** shows the equations obtained and the R-squared value.

For the model validation, the 30 percent of the data that were withheld from the data used to develop the models were tested to examine the model accuracy. The validation was performed through a plot of the results and calculating the Percent Root Mean Square Error (%RMSE). The %RMSE error was calculated as (Etz, 2018):

Percent RMSE =
$$\sum ((\text{Obi} - \text{Exi})2/n)^{1/2} *100/(\text{Average Traffic Count})$$
 (1)

The validation scatter plots for the 20 sector Linear model and 20 sector Bayesian model are shown in **Figure 1**. The model accuracy, as determined by the %RMSEs, was calculated as 71.2 for the Linear model and 69.9 for the Bayesian model.

5. Accuracy of the Trip Generation Equations

To test the accuracy of the two 20 sector trip generation equations developed in this study, a comparison of the number of trucks observed on the roadways in Birmingham and the number of truck assigned to the Birmingham travel demand

Table 2. Model resluts and R-squared value.

Stepwise Linear Regression

Trucks = 54.297 + 0.184 * sector 22 employment + 0.263 + sector 23 employment + 0.182 * sector 33-33 employment + 0.551 * sector 51 employment + 0.146 * sector 52 employment + 0.361 * sector 54 employment + 0.333 * sector 55 employment + 0.225 * sector 56 employment + 0.190 * sector 61 employment + 0.173 * sector 62 employment + 0.316 * sector 72 employment + 0.311 * sector 92 employment R-squared = 0.673)

Bayesian Linear Regression

Trucks = 52.91 + 0.821 * sector 11 employment + 0.095 * sector 21 employment + 0.181 * sector 22 employment + 0.235 + sector 23 employment + 0.176 * sector 33-33 employment + 0.085 * sector 42 employment + 0.057 * sector 44-45 employment + 0.007 sector 48-49 employment + 0.54 * sector 51 employment + 0.144 * sector 52 employment - 0.037 * sector 53 employment + 0.352 * sector 54 employment + 0.325 * sector 55 employment + 0.201 * sector 56 employment + 0.185 * sector 61 employment + 0.175 * sector 62 employment + 0.093 * sector 71 employment + 0.246 * sector 72 employment + 0.079 * sector 81 employment + 0.311 * sector 92 employment

R-squared = 0.676)





Figure 1. Validation plot for 20 sector models.

model network was performed. This step was accomplished as the accuracy of the model must be a function of the application of the model to forecast truck trips in the community. The travel demand model for Birmingham was obtained and the truck trip generation equations were applied using the employment data from the LEHD for blocks, aggregated as necessary to model traffic analysis zones. For external stations, the number of trucks entering and leaving the study area were determined from traffic counts and assigned to each location in the model. The distribution of pass-through trucks was taken from the travel demand model methodology for dividing external-internal trips and external-external trips. For all zones, intrazonal truck trips were not allowed.

The truck trips were distributed to the network using a gravity model and friction factors designed for non-home-based-trips. Finally, an assignment was performed using an equilibrium assignment technique. However, as this study is solely focused on truck trips, only truck trips were assigned to the network. The results of the assignment were compared to the actual truck counts and the results are shown in **Figure 2**, the X-axis represents truck counts and the Y-axis represents the model assigned truck counts. The %RMSEs for the two methodologies were calculated as 47.262 for the Linear model and 47.147 for the Bayesian model.

From the %RMSE values and model validation plots, there is not a clear difference between the two models. This is to be expected as the parameters in the models were very similar and the key differences in the number of sectors used in the analysis.

6. Transferability of the Trip Generation Equations

To test the transferability of the two 20 sector truck trip generation equations, the travel demand models for Huntsville, AL and Montgomery, AL were used. The LEHD data for these two locations were collected and the two 20 sector truck trip generation equations were used to determine the number of truck productions and attractions for each zone. The external station truck numbers were obtained from traffic counts and a similar adjustment to remove pass-through was performed.

The assignment of trucks using the methodology and the actual number of truck on the roadway was compared. The results of the comparison for Huntsville are shown in **Figure 3** and **Figure 4** and the results of the comparison for Montgomery are shown in **Figure 5** and **Figure 6**. The %RMSEs for Huntsville were calculated as 80.21 for the Linear model and 87.06 for the Bayesian model and for Montgomery were calculated as 84.35 for the Linear model and 78.41 for the Bayesian model.

Examining the difference between the two communities used for model transferability, the key difference is the distribution of employment within the two cities. **Figure 7** shows the employment in the urbanized areas for the two cities using the 20 sectors. As can be seen, Huntsville has a very high employment





Figure 2. Model validation plot for 20 sector models.







Figure 4. Transferability plot for Huntsville using the Bayesian model.



Figure 5. Transferability plot for Montgomery using the linear model.







Figure 7. Difference in Employment between Huntsville and Montgomery.

in Sector 54—Professional, Scientific and Technical Services and Sector 62—Health Care and Social Assistance which account for 19 percent and 16 percent of the total employment, respectively. Montgomery has a very high employment in Sector 92—Public Administration—accounting for 16 percent of the total employment.

The additional sectors included in the Bayesian model versus the Linear model might have benefitted Montgomery as several of the sectors where there was higher employment are not included in the Linear model. All of the sectors related to trade, transportation and warehousing, which accounts for 18 percent of the total employment in Montgomery, are only included in the Bayesian model. Additionally, other sectors that are higher in employment for Montgomery are not included in the Linear model, reducing the number of truck trips and leading to under assignment for trucks in the community and lowering the %RMSE.

7. Conclusion

The results of the study show that the combination of truck GPS data and LEHD employment data can be used to create new truck trip generation equations for medium sized communities. The 20 sector models both validated well statistically and in practice for Birmingham, with almost similar results. The final decision on the model selected relies with the community performing the modeling, however, the recommendation is that the 20 sector Bayesian model has the advantage of including all sectors of employment in the model that make the model more applicable in more communities as a high number of employment in the sectors not included in the Linear model would reduce the effectiveness.

Overall, the contribution of this paper is a truck trip generation model that can be used in medium sized communities (populations between 200,000 - 1,000,000) and the presentation of the means to develop a community specific truck trip generation model. While some researchers will claim that the results in this paper are not as accurate as necessary, the presentation and potential use

of the models presented in this work will be an improvement over doing mothing and/or waiting until perfect truck data is available to have more accurate models. The choice of the models presented in this paper or using the methodology followed in this work to develop community specific models, if the GPS truck data is available, is left to the agency. Including truck trips in the model is vital as the correct forecasting of truck in a community is vital for the economic growth of the community.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Alho, A. (2011). *The Adequacy of Freight Modeling Methods to Study Policy/Regulations' Implementation Impacts in a City Logistics Context.* M.S. Thesis, Lisbon: Technical University of Lisbon.
- Anderson, M. D., Dondapati, M. C., & Harris, G. A. (2013). Effectively Using the QRFM to Model Truck Trips in Medium-Sized Urban Communities. *Journal of Transportation Technologies, 3*, 185-189. <u>https://doi.org/10.4236/jtts.2013.33018</u>
- Bastida, C., & Holguin-Veras, J. (2009). Freight Generation Models: Comparative Analysis of Regression Models and Multiple Classification Analysis. *Transportation Research Record: Journal of the Transportation Research Board, 2097*, 51-61. https://doi.org/10.3141/2097-07
- Bradley, M., Bowman, J. L., & Griesenbeck, B. (2010). SACSIM: An Applied Activity-Based Model System with Fine-Level Spatial and Temporal Resolution. *Journal of Choice Modeling*, 3, 5-31. <u>https://doi.org/10.1016/S1755-5345(13)70027-7</u>
- Cambridge Systematics, Inc. (1996). *Quick Response Freight Manual*. Federal Highway Administration. DTFH61-93-C-00075, DTFH61-93-C-00216.
- Cambridge Systematics, Inc. (2007). *Quick Response Freight Manual II*. Federal Highway Administration. Publication No. FHWA-HOP-08-010.
- Chow, J. Y. J., Yang, C. H., & Regan, A. C. (2010). State-of-the-Art of Freight Forecast Modeling: Lessons Learned and the Road Ahead. *The Journal of Transportation*, *37*, 1011-1030. https://doi.org/10.1007/s11116-010-9281-1
- Cohen, H., Horowitz, A., & Pendyala, R. (2008). NCHRP Report 606: Forecasting Statewide Freight Toolkit (pp. 5-15). Washington DC: Transportation Research Board of the National Academies, Report by Cambridge Systematics, Inc., & Global Insight (Formerly Reebie Associates). <u>https://doi.org/10.1016/0191-2615(82)90037-6</u>
- Daly, A. (1982). Estimating Choice Models Containing Attraction Variables. Transportation Research Part B: Methodological, 16, 5-15. <u>https://doi.org/10.17226/14133</u>
- Doustmohammadi, E., & Sisiopiku, V. (2016). Comparison of Freight Demand Forecasting Models. *International Journal of Engineering Science Invention*, *5*, 45-51.
- Doustmohammadi, E., Sisiopiku, V., & Sullivan, A. (2016b). Modeling Freight Truck Trips in Birmingham Using Tour-Based Approach. *Journal of Transportation Technology*, 6, 436-448. <u>https://doi.org/10.4236/jtts.2016.65035</u>
- Doustmohammadi, E., Sisiopiku, V., Erson, M., Doustmohammadi, M., & Sullivan, A. (2016a). Comparison of Freight Demand Forecasting Models. *International Journal of*

Traffic and Transportation Engineering, 5, 19-26.

- Etz, A. (2018). Introduction to the Concept of Likelihood and Its Applications. *Advances in Methods and Practices in Psychological Science*, *1*, 60-69. http://journals.sagepub.com/doi/pdf/10.1177/2515245917744314
- Figliozzi, M. A. (2007). Analysis of the Efficiency of Urban Commercial Tour: Data Collection, Methodology, and Policy Implications. *The Journal of Transportation Research Board*, 41, 1014-1032. <u>https://doi.org/10.1016/j.trb.2007.04.006</u>
- Fischer, M. J., Outwater, M. L., Cheng, L. L., Ahanotu, D. N., & Calix, R. (2005). An Innovative Framework for Modeling Freight Transportation in Los Angeles County. Los Angeles, CA: Los Angeles County Metropolitan Transportation Authority by Cambridge Systematics, Inc. <u>https://doi.org/10.1177/0361198105190600113</u>
- Gliebe, J., Cohen, O., & Hunt, J. D. (2007). Dynamic Choice Model for Urban Commercial Activity Patterns of Vehicles and People. *Transportation Research Record: Journal* of the Transportation Research Board, 2003, 17-26. <u>https://doi.org/10.3141/2003-03</u>
- Greaves, S. P., & Figliozzi, M. A. (2008). Collecting Commercial Vehicle Tour Data with Passive Global Positioning System Technology: Issues and Potential Applications. *Transportation Research Record, 2049,* 158-166.
- Holguín-Veras, J., & Patil, G. R. (2008). Integrated Origin-Destination Synthesis Model for Freight with Commodity-Based and Empty Trip Models. *Transportation Research Record: Journal of the Transportation Research Board, 2008*, 60-66. https://doi.org/10.3141/2008-08
- Holguín-Veras, J., & Thorson, E. (2000). Trip Length Distributions in Commodity-Based and Trip-Based Freight Demand Modeling. *Transportation Research Record: Journal* of the Transportation Research Board, 1707, 37-48. <u>https://doi.org/10.3141/1707-05</u>
- Holguin-Veras, J., Jaller, M., Destro, L., Ban, X., Lawson, C., & Levinson, H. (2011).
 Freight Generation, Freight Trip Generation, and Perils of Using Constant Trip Rates.
 Transportation Research Record: Journal of the Transportation Research Board, 2224, 68-81. https://doi.org/10.3141/2224-09
- Holguin-Veras, J., Sanchez-Diaz, I., Lawson, C., Jaller, M., Campbell, S., Levinson, H., & Shin, H.-S. (2013). Transferability of Freight Trip Generation Models. *Transportation Research Record: Journal of the Transportation Research Board, 2379*, 1-8. https://doi.org/10.3141/2379-01
- Hunt, J. D., & Stefan, K. J. (2007). Tour-Based Microsimulation of Urban Commercial Movements in Alberta, Canada. *Transportation Research Part B*, 41, 981-1013. <u>https://doi.org/10.1016/j.trb.2007.04.009</u>
- Jansuwan, S., Ryu, S., Chen, A., & Heaslip, K. (2014). *A Two-Stage Approach for Estimating a Statewide Truck Trip Table* (pp. 5-12). Report, Logan, UT: The Utah Transportation Center, Utah State University.
- Jong, G., & Ben-Akiva, M. (2007). A Micro-Simulation Model of Shipment Size and Transport Chain Choice. *Transportation Research Part B*, *41*, 950-965. https://doi.org/10.1016/j.trb.2007.05.002
- Jong, G., Gunn, H., & Walker, W. (2004). National and International Freight Models: An Overview and Ideas for Further Development. *Transport Reviews, 24*, 103-124. https://doi.org/10.1080/0144164032000080494
- Kuppam, A., Lemp, J., Beagan, D., Livshits, V., Nippani, S., & Vallabhaneni, L. (2014).
 Development of a Tour-Based Truck Travel Demand Model Using Truck GPS Data. In Proceedings, 93rd Transportation Research Board Annual Meeting (No. 14-4293).
 Washington DC: National Research Council.

- North Carolina Department of Transportation (2009). North Carolina Truck Network Model Development Research.
- Roorda, M., Hain, M., Amirjamshidi, G., Cavalcante, R., Abdulhai, B., & Woudsma, C. (2010). Exclusive Truck Facilities in Toronto, Ontario, Canada: Analysis of Truck and Automobile Demand. *Transportation Research Record: Journal of the Transportation Research Board, 2168*, 114-128. <u>https://doi.org/10.3141/2168-14</u>
- Ruan, M., Lin, J., & Kawamura, K. (2011). Modeling Commercial Vehicle Daily Tour Chaining. *Transportation Research Part E: Logistics and Transportation Review, 48,* 1169-1184.
- Samimi, A., Mohammadian, A., & Kawamura, K. (2010). Behavioral Paradigms for Modeling Freight Travel Decision-Making. In 12th International Conference on Travel Behavior Research.
- Tavasszy, L., & Jong, G. (n.d.). *Modelling Freight Transport* (pp. 166-191). Amsterdam: Elsevier.
- Wheeler, N., & Figliozzi, M. (2011). Multi-Criteria Freeway Performance Measures for Trucking in Congested Corridors. *Transportation Research Record: Journal of the Transportation Research Board*, 2224, 82-93. https://doi.org/10.3141/2224-10
- Yang, C., Regan, A. C., & Son, Y. T. (2010). Another View of Freight Forecasting Modeling Trends. *KSCE Journal of Civil Engineering*, 14, 237-242. https://doi.org/10.1007/s12205-010-0237-y