Using Google Transit Feed Specification in Travel Modeling

BY
Sasanka Gandavarapu
Wilbur Smith Associates
1301 Gervais Street, Suite 1600
Columbia , SC, 29201
Phone: 803-251-2974
Email: sgandavarapu@wilbursmith.com

Submitted for presentation at 4th Transportation Research Board Conference on Innovations in Travel Modeling, 2012
Word count: 2150 words + 1 Table (250 words) + 2 Figures (500 Words) = 2900 total approx.

I. Background

Modeling transit systems is an integral part of travel demand modeling. To accomplish this task, modeling agencies go through the tedious process of coding the entire transit network and update it periodically to reflect the changes in service characteristics. To this end, agencies spend significant time and resources. Recently, many transit agencies are releasing data in the specifications laid out by Google. These specifications known as Google Transit Feed Specification (GTFS) serve as a platform for standardizing the way transit agencies publish their service characteristics. This GTFS feed is a valuable source of free data for MPO's that could be used in a variety of ways during the modeling process. The main focus of this paper is to illustrate the main applications of GTFS feed for various modeling purposes.

II. Objectives

This paper attempts to showcase the usefulness of the GTFS feed to the transportation modeling community. Specifically, three ideas for using GTFS in model estimation and application are discussed in this paper. Firstly, a methodology to compute a measure of transit accessibility using Google transit data is described. Secondly, application of these accessibility measures in model estimation procedures for auto ownership and trip attraction models is discussed. Thirdly, the resulting implication of this methodology towards implementing a feedback loop in travel demand models is illustrated. The remaining paper is organized as follows: Section III briefly describes the data available through GTFS. Section IV provides a brief description of earlier studies that used GTFS in travel modeling. Section V describes the framework to use GTFS in computing transit accessibility measures. Further, Sections VI and VII discuss the implications for model estimation and application respectively. Finally, Section VIII provides concluding remarks of the study.

III. What is GTFS?

GTFS defines a common format for public transportation schedules associated geographic information. Transit agencies release their service characteristics in a set of text (.txt) files. Each of the files contains information about agency, stops, routes, schedules, calendar, fares and frequencies. These are the files that help provide transit direction to users on the Google maps website. Currently, many big transit agencies have made their GTFS publicly available such as TriMet (Portland), BART (San Francisco), DART (Dallas) and so on. Transit agencies provide different services during peak and off peak and often the schedule for peak services account for congestion.

IV. Earlier Applications of GTFS

GTFS was first released in the year 2005 beginning with Portland's TriMet data. As more and more transit agencies published the data in GTFS format, transportation modelers started to see the value of this data source and used it for various purposes. The main motivation for this study is the implementation of transit extraction method employed in California Statewide Model (Circella et al, 2011). This study models in-vehicle and out-of-vehicle transit travel times based on transit level of service and highway travel times. Also, Brookings Institute conducted a nationwide study to examine transit accessibility to jobs using GTFS data (Adie Tomer et al, 2010). Puchalsky and Scherr, 2011 demonstrated how GTFS and Open street Map were used to build a transit network as an input for

Delaware valley region commissions' model. In these studies, GTFS has been used to substitute the necessity of explicitly modeling transit in statewide models and in computing aggregate accessibility measures, the data source can be tapped in for usage in demand models, more directly. Next section outlines a methodology to extract transit accessibility measures for direct usage in commonly used demand model components.

V. Transit Accessibility Measures

Figure 1 shows the proposed methodology for computing accessibility measures starting from the GTFS feed. There are a few software applications available in the market today that import GTFS feed and allow the user to build multimodal paths on the transportation network. For example, GraphServer which is a Linux based open source software integrates the highway network (from Open Street Map) and the transit network (from GTFS) thereby allowing the users to build multimodal paths from user-supplied origins to destinations. Using this software, a shortest path tree can be constructed from a specific location by providing its latitude and longitude. This feature can be used to compute a shortest path tree using transit service from a TAZ centroid. In typical travel models, a measure of 30 minutes or 45 minutes travel time is used to examine the proximity of population and employment to each of the TAZ centroids. Overlaying such a shortest path tree of 30 or 45 minutes with population and employment data can be used to compute accessibility measures for employment and households for each TAZ. For example, the transit accessibility to employment would be the percentage of regional employment within this 30 or 45 minute shortest path tree polygon. Additionally, using land use data at parcel level would provide a finer resolution of the population and employment characteristics.

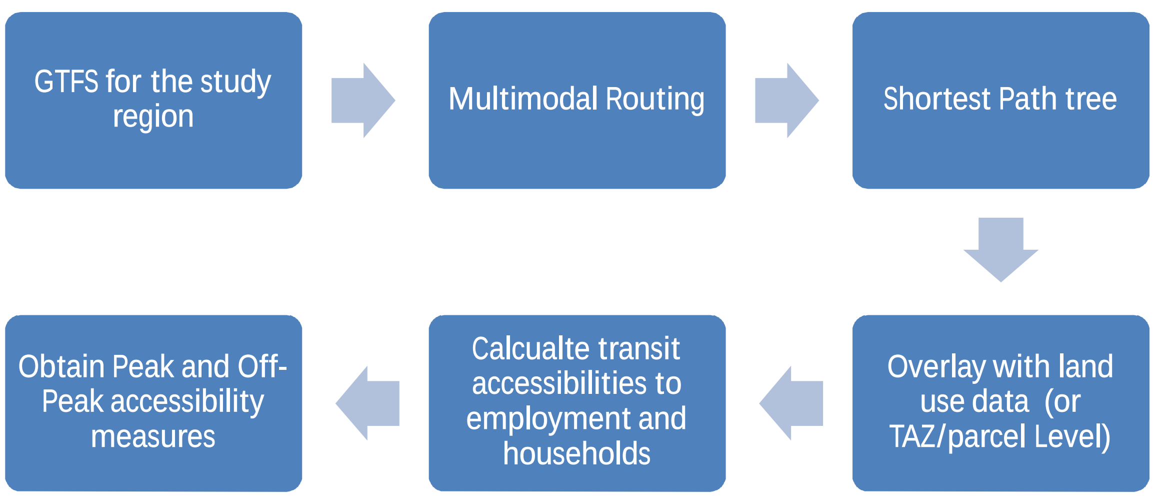

Figure 1 Using GTFS for accessibility measures

There are two advantages of computing accessibility measures this way. Firstly, accessibility can be computed for peak and off peak separately as the shortest path trees can be generated for different times of day. The accessibility in congested conditions is generally an output from the demand models but by using this methodology, the base year transit accessibility during peak conditions can be computed outside the model. This has implications during the phases of both, model estimation and application, since peak time transit accessibility would be crucial for commute while off peak measures potentially influence travel behavior for other trip purposes. These implications are discussed in detail in the following two sections. Secondly, any policy decisions relating to transit investments could be quickly tested by editing the GTFS feed to obtain the accessibility measures for peak and off peak conditions.

VI. Implications for Model Estimation

Obtaining transit accessibility measures for peak and off peak conditions in base year has significant implications during the estimation of various demand models. In this section, potential improvements to two such models namely, the auto ownership model and the trip attraction models are discussed.

Auto ownership models help in forecasting the automobile ownership levels of household and are play a key role in evaluating trips generated by each household. Most trip generation models segment household based on auto availability and hence it is essential to obtain the distribution of households by automobile ownership for reach TAZ. As the vast literature on the subject indicates, a household's decision of owning an automobile (or more) could depend on a variety of factors such as household size, income and number of workers, neighborhood characteristics (such as density, diversity and design) and accessibility characteristics (particularly relating to transit). Commonly used model specifications are multinomial logit model or an ordered response logit model with the alternatives being the number of vehicles owned by the household. A list of all the commonly used variables in a typical auto ownership model is shown in Table 1. Many auto ownership models demonstrated the significance of transit accessibility in determining auto ownership levels. Using GTFS to obtain peak and off peak accessibility measures gives the analyst an option to test the fit of the model based on these variables. It can be hypothesized that transit access while commuting to work would play a key role in determining auto ownership levels and hence the specification with peak period accessibility measure would yield a better model fit. The analyst also has the option of building two separate models, one with peak and the other with off peak accessibility and use the appropriate model in the application stage. These implications in model application are further discussed in the next section.

Table 1 Typical explanatory variables in an auto ownership model

Typical explanatory variables in an auto ownership model	
Household characteristics	Household income, Household size, Number of workers, Household type (Single family vs. multi family)
Density variables	Employment density, Household density
Accessibility characteristics	Percentage of employment accessible by transit within 30 minutes from the TAZ centroid Percentage of households accessible by transit within 30 minutes from the TAZ centroid

Another potential application of accessibility computed from GTFS is during the estimation of trip generation models. Specifically, consider a multinomial logit model being employed for a home based work trip attraction model. Typically, in such a model, the alternatives for each zone would be the market segments of people that get attracted to that particular zone for a work trip. For example, consider a specification that segments the commuters into seven market segments based on their household sizes and income: Household Size 1 and Income quintile 1; Household Size 1 and Income quintile 2; Household Size 1 and Income quintiles 3,4 and 5; Household Size 2 and Income quintile 1; Household Size 2 and Income quintile 2; Household Size 2 and Income quintiles 3,4 and 5; Household Size 3 and Income quintiles 3,4 and 5.

Such a logit model would yield the split of HBW trips attracted to each zone for each of the market segments above. Independent variables that would be considered for such a model would be – employment by sector and transit accessibility measures. Given that this is an attraction model for HBW trips that mostly take place during the peak periods, this would serve as a good example where an analyst could use peak transit accessibility skims to reflect the congestion present in the highway network during commuting.

VII. Implications for Model Application

A common feature of many travel demand models is the ability to feed the congested travel times back into the model to account for the prior knowledge of travelers with regard to the transportation system. This feedback loop is often implemented after traffic assignment step and the congested times would be used as impendence in trip distribution step and the iterative process is continued till convergence of link flows is reached in successive iterations. Accessibility measures in auto ownership models and trip generation would entail the feedback process to include these steps in the iterative procedure as well.

It should be noted that many models use uncongested time to calculate accessibilities that feed into the auto ownership and trip generation models simple because of unavailability of congested times at the model estimation stage. In such a case, there is no need to implement a feedback loop on these two models. However, that would be a limitation to the methodology since travelers could very well base their automobile buying decisions on the transit accessibility to their jobs, which are trips made during peak hour. This limitation can be addressed by using congested times from GTFS that would help the analyst to estimate accessibility measures for peak periods and employ a feedback loop as shown in Figure 2. The boxes colored in yellow would be executed in the first iteration while the green boxes are executed part of the feedback loop. The scope of this illustrative figure is limited in terms of showing how congested accessibility measures can be used during a demand model application.

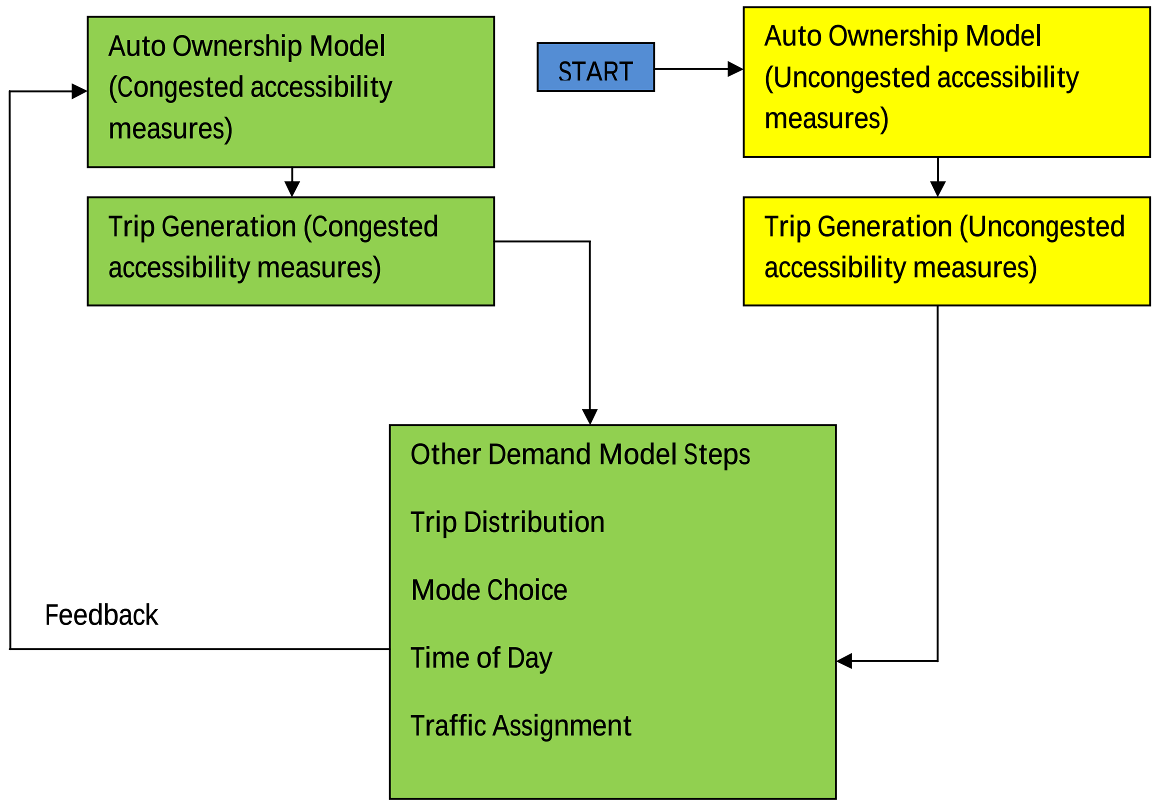

Figure 2 Feedback loop

VIII. Concluding Remarks

This paper explored the option of using publicly available Google transit data for transportation modeling applications. Given that the data is available at no additional cost to the modeling agencies, GTFS could be a valuable source of transit data. This paper demonstrated three possible GTFS uses in model application process. The transit accessibility measures for peak and off peak can be computed using the GTFS data and further this can be employed in auto ownership and trip generation model estimation. A framework for implementing a feedback loop for these newly estimated models is also shown.

There are a lot of challenges in practical implementation of GTFS. Firstly, there is considerable amount of work involved in converting the GTFS data into the standard modeling software. Given so many advantages of the GTFS data, commonly used modeling software need to quickly embrace this developments and build a bridge between GTFS and their platforms. Secondly, as GTFS data is constantly updated by the transit agencies, modelers need to keep track of the versions compatible to the modeling years.

References

Circella, G., J. D. Hunt, K. J. Stefan, A. T. Brownlee and M. McCoy, 2011. Simplified Model of Local Transit Services, TRB 90th Annual Meeting, January 23-27, 2011.

Adie Tomer, Elizabeth Kneebone, Robert Puentes, Alan Berube, 2011. Missed Opportunity: Transit and Jobs in Metropolitan America Cities, The Brookings Institution study.

Christopher Puchalasky and Wolfgang Sherr, 2011. DVRPC's New Regional Network Model - Using Google Transit Feeds and OpenStreetMap in Transportation Planning, Presented at TMIP webinar.