Comparing the Daily Schedules in the NHTS from 2009 and 2017

Greg Macfarlane PhD PE and Josie Kressner PhD, Transport Foundry

8 August 2018

1 Abstract

Travel modelers commonly use passive transportation-related data, such as cellular or GPS origin-destination matrices, to calibrate or validate their planning models. The passive data are generally used, in this case, to expand and adjust behavioral models estimated from a small-sample local household survey. Recently, there has been interest in deriving synthetic records of individual travel directly from the passive data, with the support of other datasets. Kressner (2017) previously developed a method to build synthetic daily activity and travel patterns for a complete population. In this method, simulated individuals used the 2009 NHTS as a basis for tour patterns while passive origin-destination matrices spatially locate the simulated tours within a specific region.

In this work, we update our tour patterns to the 2017 NHTS and comment on observed differences in the resulting simulated travel patterns. Specifically, we consider the temporal distribution of weekday trips. We also examine the consequences of these differences in the synthesized daily schedules and in their traffic assignment. The results show that the NHTS has a similarly high proportion of mid-day, off-peak travel in both 2009 and 2017, and that this proportion is higher than other comparable local household surveys. The results also show that the change from the 2009 to the 2017 dataset does not substantively affect the simulated demand or assignment in a data-driven travel model using NHTS data.

2 Background

In a conventional travel modeling paradigm, planners estimate econometric models of travel behavior — mode choice, destination choice, etc. — and expand the survey to match the total population distribution. Under this paradigm, a local dataset on which to estimate the models is important to ensure the models capture region-specific behavior. Passively collected travel data such as cellular phone traces and highway counts are used to validate and calibrate the models, but are not useful in estimation because they do not contain information on the behavior and choices of individuals.

The general National Household Travel Survey (NHTS)(U.S. Department of Transportation 2017) is not typically suitable for a conventional travel modeling exercise, because the local sampling rate is too small and the geographic resolution is too large for model estimation or calibration (with some modifications, state or MPO-specific NHTS add-ons may be used).

In the past several years, there have been efforts in the travel demand modeling community to explore the use of so-called "data-driven" travel demand models. In contrast to conventional models, a data-driven model uses the large-scale passive data explicitly in the model process. Kressner (2017) previously described a methodology for such a model; in this methodology, synthetic individuals generate tours by referencing the NHTS for individuals living in similarly-sized cities and then draw locations for their tour activities from origin-destination datasets specific to the model region. The resulting synthetic tours can then be assigned to highway and transit networks to understand route and mode choice. In this research we use an open-source transport network simulator, MATSim.

The most recent two iterations of the NHTS were in 2009 and 2017. Though the methodologies between successive survey collections did change somewhat, enough similarities exist that a comparison of travel trends between is useful and important. In this research we consider the diurnal distribution of tour start times and trips-in-motion reported in the NHTS and other related household surveys, and we compare the results to the data-driven model developed by Transport Foundry.

3 Methodology

We developed a data-driven model for Asheville, North Carolina, which is a medium-sized city in western North Carolina. The core-based statistical area (CBSA) in Asheville has a population of approximately 400,000 people. We used the public data files for the 2009 and 2017 NHTS for respondents living in a CBSA with a population between 250,000 and 500,000 people. In both years, there are more than 20,000 respondents in cities of this size.

We compare the distribution of tour types and trips-in-motion in the data-driven simulation against the public NHTS schedules data from which the model is derived.

4 Results

4.1 Tour Start Distribution

For the distribution of tour types, we calculate the probability of archetype tours beginning in each 90 minute period for workers only, accounting for person weights in the NHTS data. Non-workers are excluded in this plot for simplicity. An archetype tour is a means of tour classification: a "simple work" tour archetype includes a single work activity, a "multi-part work" tour includes more than one work activity, and a "composite" tour includes both work and non-work activities.

The distribution of probabilities for each archetype of tours by time of day are given in Figure 1. A distribution of the difference between the distributions by purpose and period is in 2. From this plot, it appears that workers in the 2017 survey are somewhat less likely to make tours that involve a work activity, and somewhat more respondents take multi-part work tours. There is also less of a peak in the late afternoon in 2017. Overall however, the distributions are comparable.

That is, the largest increase in probabilities are for multi-part and simple work tours, and the decrease has come primarily from simple work and composite from-work tours. This may indicate a true trend in travel behavior, or a discrepany in how respondents classified the purpose of their activities, or simply a random variation. That said, there are over 20,000 respondents to the NHTS in cities of this size in both survey years, so a random variation is unlikely.

With this information as a background, we now turn to understanding how this change in tour information affects the simulated trips and tours in the data-driven model. To do this, we consider the temporal distribution of trips-in-motion.

4.2 Trips-in-Motion

For a distribution of trips-in-motion, we calculate the number of trips moving during each 15 minute interval throughout a 24-hour weekday period. To be counted in a 15 minute interval, a trip either needs to be currently underway or completed entirely within the next 15 minutes. We generate separate distributions for home-based work (HBW) trips and all trips.

We also obtained comparable distributions from two other data sources: a local household travel survey for the Hickory, North Carolina region and a 2009 NHTS add-on for the Blacksburg, Virginia region. Though these regions are both smaller than Asheville in terms of population, they share common elements in culture, geography, and infrastructure. Figure 3 shows the distribution of these trips-in-motion in each dataset. All

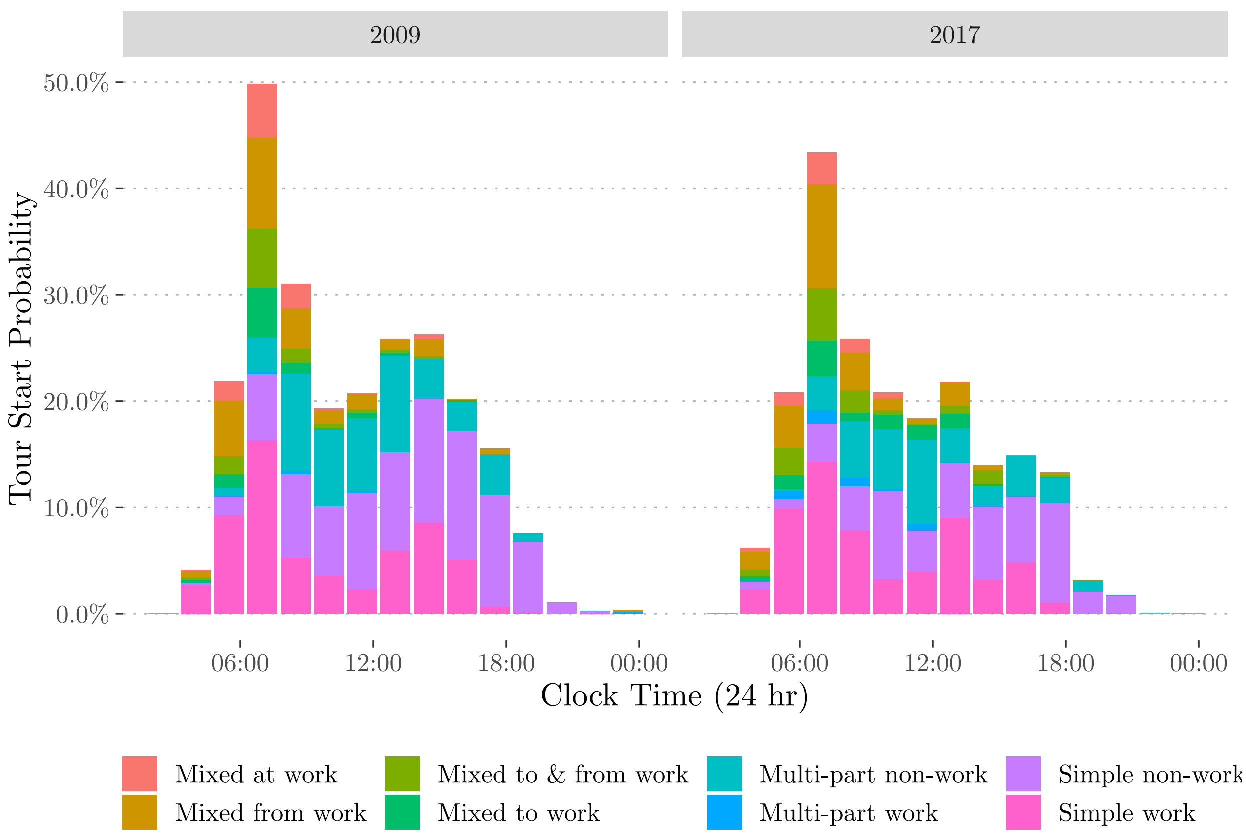

Figure 1: Probability of an archetype tour starting for workers by time of day.

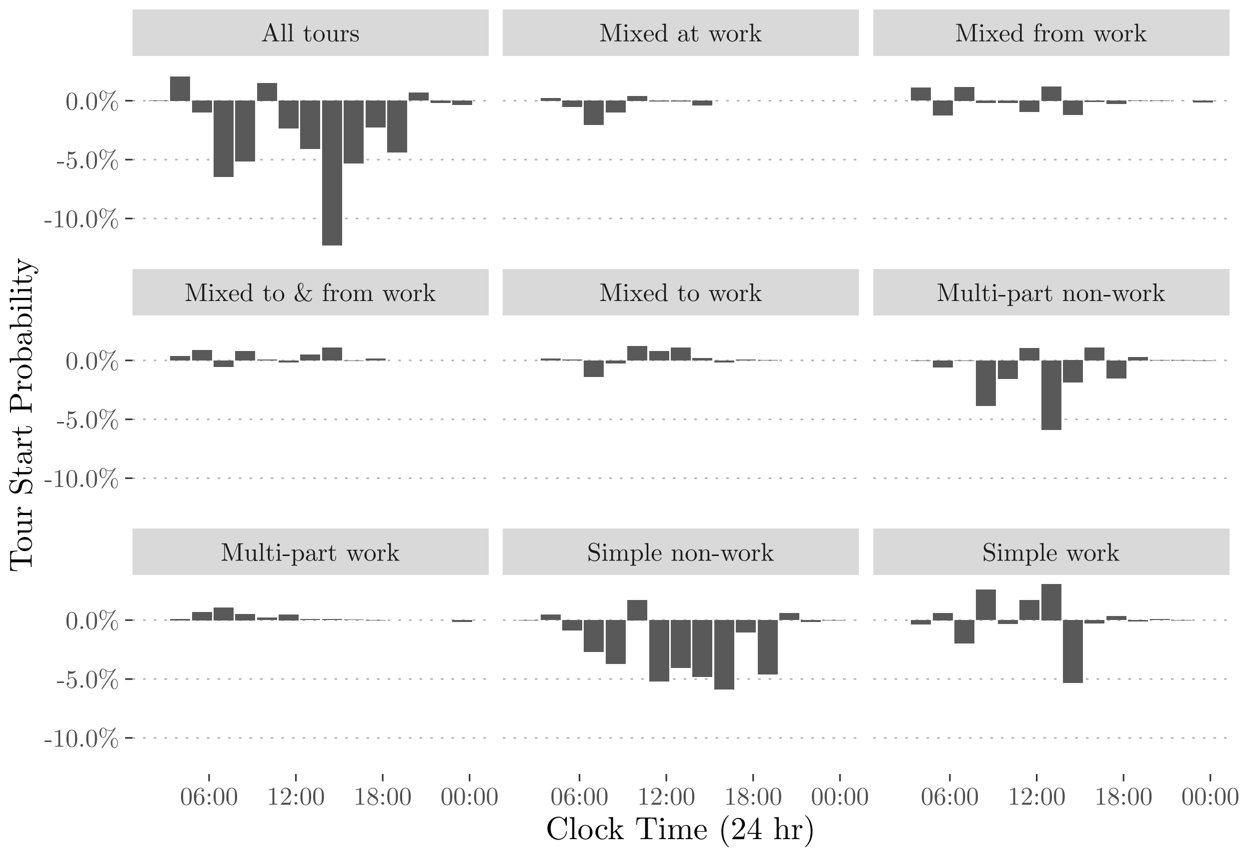

Figure 2: Difference in archetype tour start probability between 2017 and 2009 NHTS for workers by time of day.

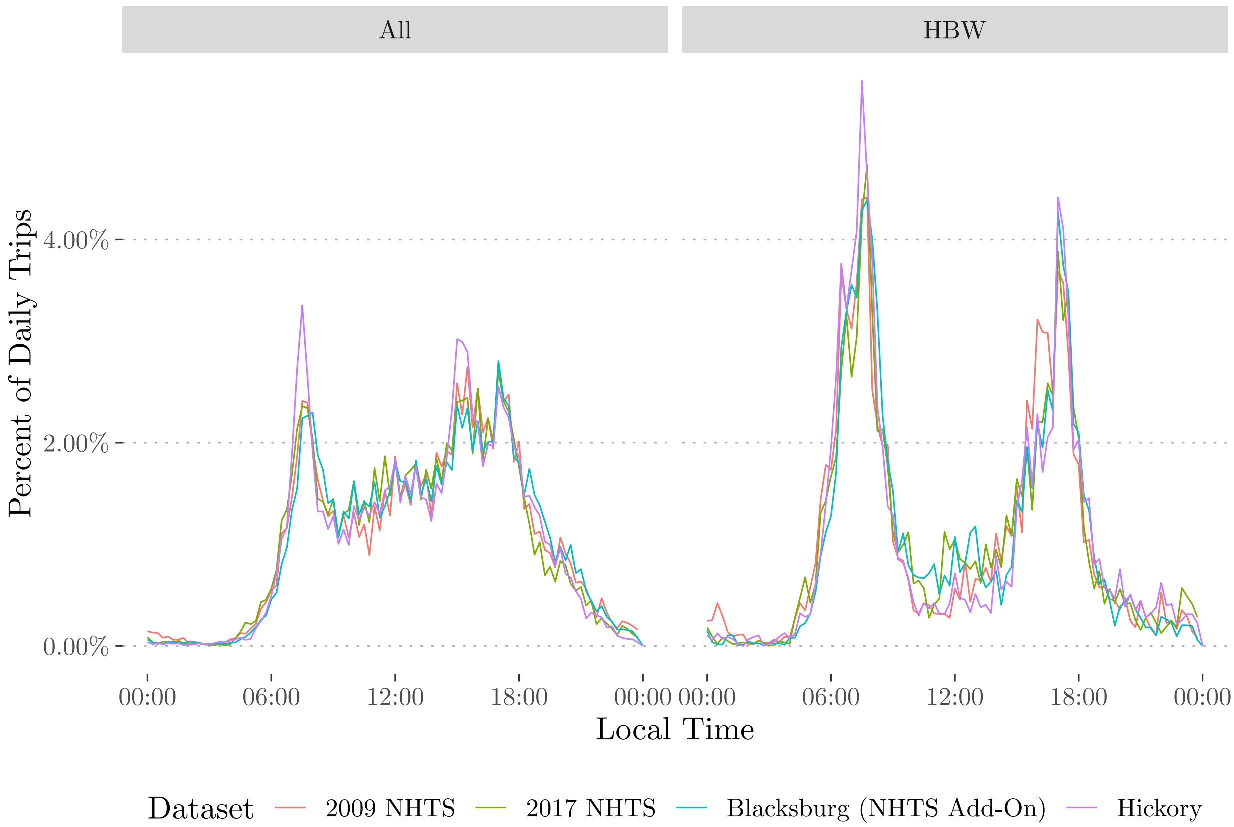

Figure 3: Distribution of trips-in-motion for NHTS for medium-sized cities and comparison datasets.

the data show a pattern typical of daily travel: there are peaks in the AM and PM representing the beginning and end of the work day. The peaks are somewhat more pronounced in the Hickory survey, but overall the four datasets are remarkably similar.

The data-driven model generates synthetic travel demand in two stages. First, the model generates an initial demand considering the tour patterns and origin-destination data. This initial demand is then fed into MATSim, which adjusts the activity start and end times, the travel routes, and the travel mode in an attempt to account for congestion (considering all other people's travel) and to iteratively improve the utility of each person's day. Travel time and cost is minimized, and time spent doing activities is maximized.

We ran the data-driven model to generate travel demand for a synthetic population in Asheville, North Carolina using the 2009 and 2017 NHTS as a tour archetype database and subsequently ran the routing and replanning simulation in MATSim. Figure 4 shows the distribution of trips-in-motion in 15 minute bins.

The demand simulation initially generates a higher share of mid-day trips than the NHTS source data, but the replanning simulation seems to re-peak the distribution, particularly for work trips. Both the demand and replanning simulations seem to have a longer tail in the evening, with more trips still on the road than the NHTS data suggests.

It is worth noting that the NHTS data are considerably rougher than either simulated outcome. This is both a function of the limited sample size of the NHTS (20,000 respondents versus 400,000 simulated agents) as well as the tendency for self-reporting diary respondents to round trip departures to nearest half-hours.

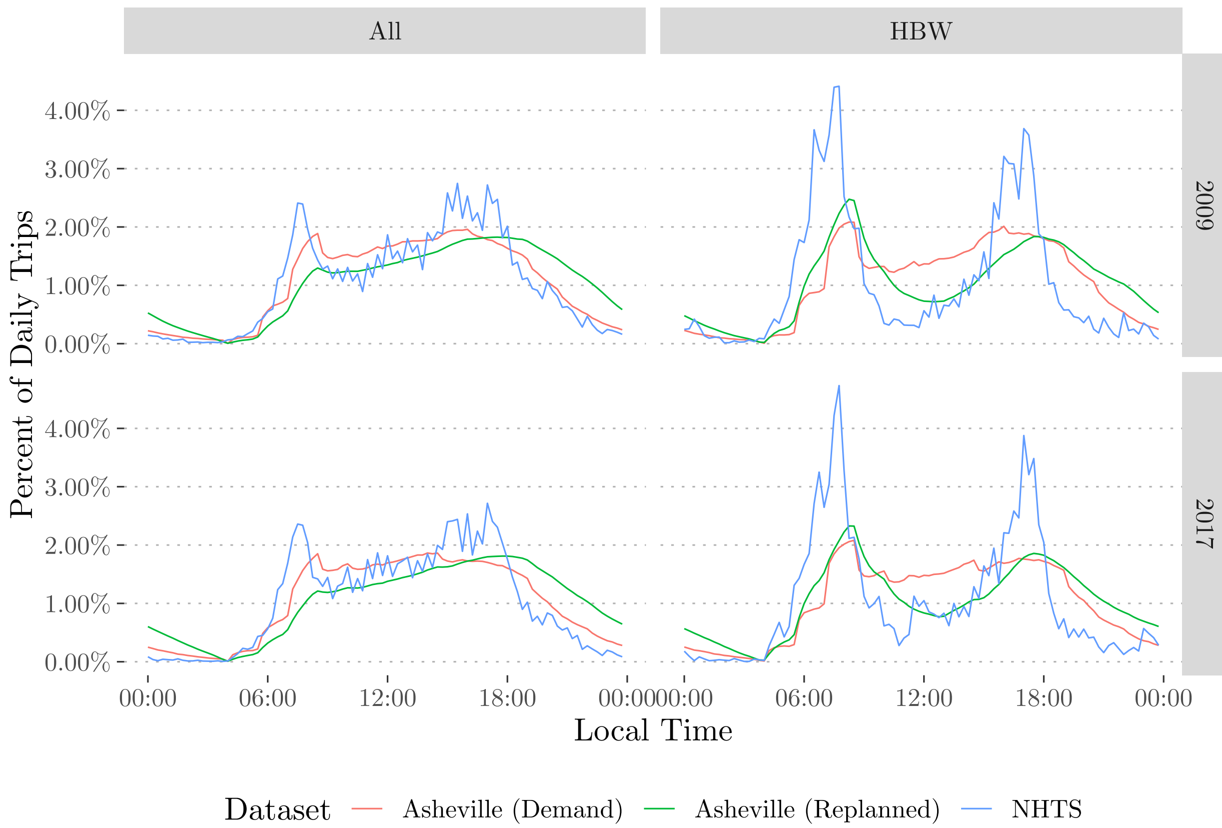

Figure 4: Trips-in-motion for NHTS and simulated Asheville by time of day.

5 Conclusion

The change from the 2009 to the 2017 dataset does not substantively affect the temporal distribution of tour start times by tour type, or the distribution of vehicle trips-in-motion for medium-sized cities. Applying a data-driven generation process can smooth out the demand distribution and lead to more natural distributions of travel throughout the day for traffic assignment, but additional research is needed to match peaking behavior with more complete fidelity.

5.0.1 Acknowledgements

We are grateful to Kyle Ward of WSP who provided trips-in-motion data for Hickory and Blacksburg.

References

Kressner, Josephine. 2017. "Synthetic Household Travel Data Using Consumer and Mobile Phone Data." *NCHRP IDEA Program*. Transportation Research Board. http://www.trb.org/Main/Blurbs/176216.aspx.

U.S. Department of Transportation. 2017. "National Household Travel Survey." https://nhts.ornl.gov.