Methods for Validation and Expansion of Passively Collected Origin-Destination Data

EXECUTIVE SUMMARY

Passive data from mobile electronic and in-vehicle devices present an opportunity for accurate origin-destination data at unprecedented levels of geographic resolution, but only if they are properly expanded. Failure to correctly account for the under- or over-representation of certain groups of travelers, areas of origins and destinations, and types of trips can lead to faulty analyses and false conclusions.

- Passive data can present a complete, high resolution picture of origin-destination travel patterns that surveys cannot.
- However, passive data are not representative, but demonstrated to be systematically biased both with regard to the demographics of travelers and with regard to the length of trips represented.
- Data validation and expansion must therefore be the first step in any analysis using passive origin-destination data.
- The most commonly used methods cannot correct for systematic trip length / duration bias, but several methods can.
- Since all currently mature methods have important pros and cons, robust expansion schemes combine an ensemble of methods.

2 SYSTEMATIC BIAS IN PASSIVE DATA

Although passive data provides large sample data, often including millions of trips, it is still only a sample, and because it is not a controlled random sample, it is not representative of all travelers or trips. Commercially available datasets include only travelers with certain devices, carriers, and/or apps installed. Moreover, short-distance trips or short-duration activities are often under-represented in the data because they require more frequent observations of position which are not always available due to several factors including battery management, device, and app usage.

3 EXPANSION METHODS FOR PASSIVE ORIGIN-DESTINATION DATA

The nine methods now in use for expanding passively collected OD data can be categorized in terms of what control data they expand the passive data to match. There are now three sources of control data for expanding passive data: demographic data generally from the census, traffic counts on the roadway network, and disaggregate trace data from smartphone surveys. Methods based on demographic data such as market penetration-based factors and trip-generation based scaling are important for addressing demographic bias in some datasets, but cannot address trip length bias. Some methods for expanding to traffic counts can correct for trip length bias. Iterative screenline fitting using matrix partitioning does not rely on a network assignment model but is generally limited by several factors and therefore often must be supplemented by other methods. ODME is powerful but must be used with careful constraints, and even then provides little insight into the nature of the expansion. Parametric scaling provides a transparent expansion method, but is difficult to implement. New trace auditing methods using disaggregate smartphone survey data are perhaps the most promising, but require a smartphone survey dataset. While all of the methods have some usefulness, not all methods are equally robust or appropriate for certain datasets or analyses. Multiple complementary expansion methods are therefore often used together as an ensemble.