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

Vince Bernardin, PhD RSG
Hadi Sadrsadat, PhD RSG

RSG
the science of insight

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.

FIGURE 2. Systematic Age Bias in Columbus, OH

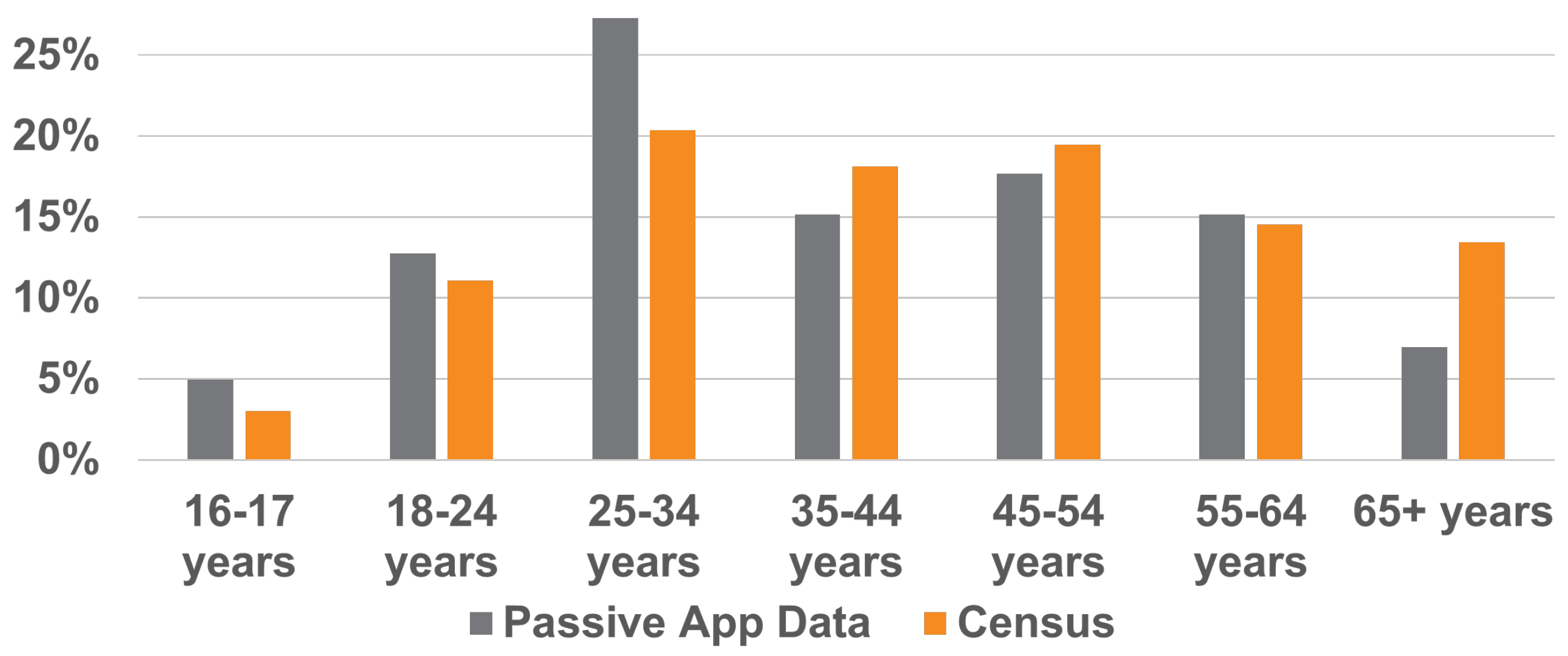
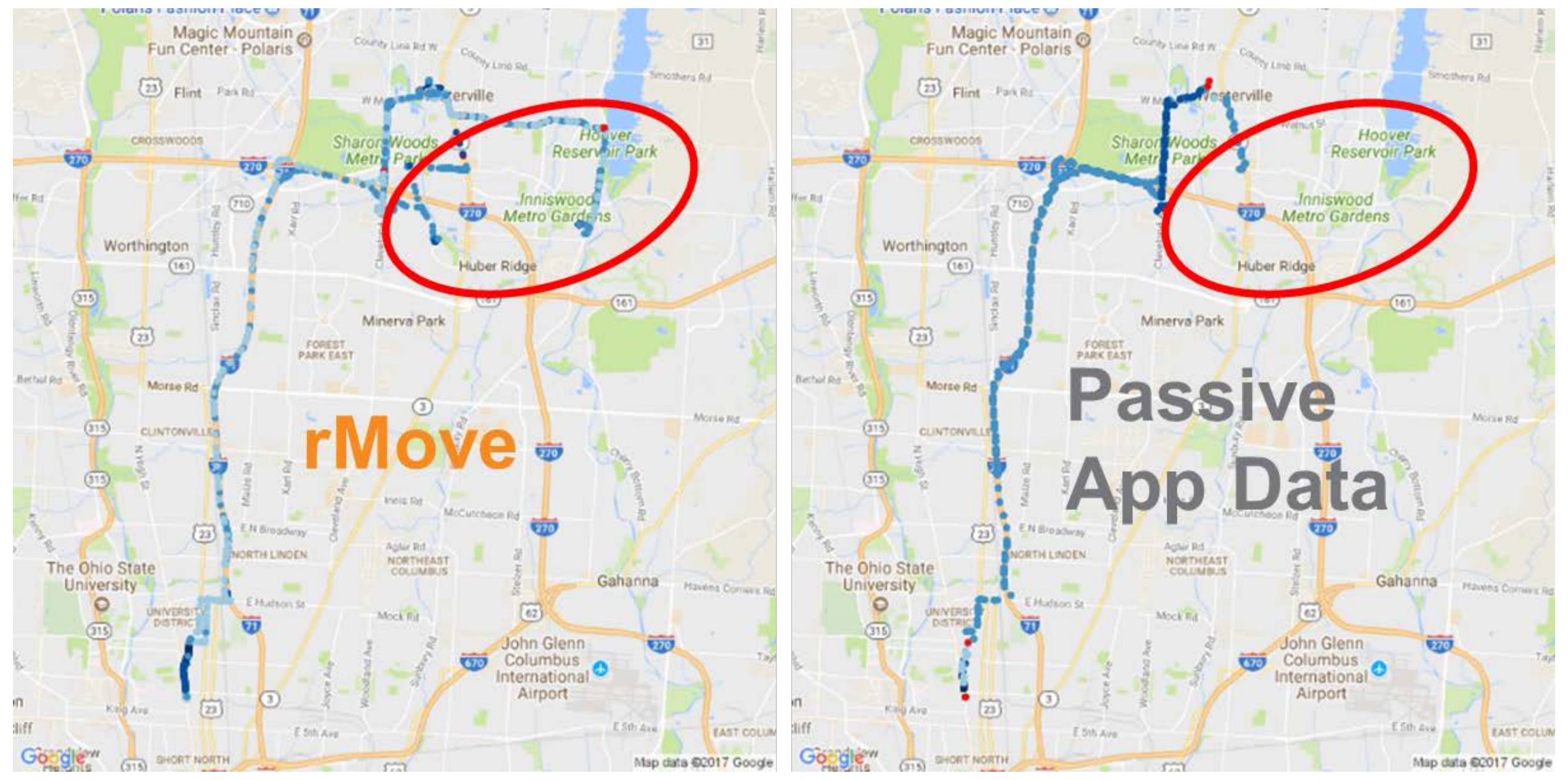


FIGURE 3. Trace Audit showing Missing Short Trips



1 TYPES OF PASSIVE ORIGIN-DESTINATION DATA

There are currently three main types or general sources of passive OD data: cellular signaling data between towers and cellular devices, location-based services (LBS) data from smartphone apps, and global positioning data, largely from in-vehicle devices. Each type of data and particular dataset has its own characteristics including limitations. Moreover, these characteristics are not necessarily constant between regions or over time.

FIGURE 1. Comparison of Types of Passive Origin-Destination Data

	CELLULAR	LBS	GPS	
Description				
Universe	All Travel	All Travel	Trucks	Private Autos
Precision and Coverage				
Locational Precision	> 100 m often ~ 200 - 2000 m	10-100 m often ~ 50 m	1 - 10 m	1 - 10 m
Sample Penetration	6-10%	5-8%	9-12%	~0.5%
Data Collection Time Period	Typically 1 month	One or more months	1 month - 2 years depending on provider & pricing	
Segmentation & Applications				
Number of Zones	Limited by pricing and locational precision	Depends on pricing scheme	Relatively unlimited in most pricing schemes	
Select Link / Corridor Analysis	Generally indirect only	Indirect only currently but a subset may support direct in the future	Limited or Unlimted direct depending on provider, or indirect	
Filtering of Intermediate Stops on Long Trips	Premium option	Premium option	Depending on provider may be possible as a post-process	
Residency Information	Premium option	Premium option	Not available due to ID persistence limitations	
Trip Purpose	Premium option for imputed purposes	Premium option for imputed purposes	Not available due to ID persistence limitations	

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.

FIGURE 4. Comparison of Expansion Methods for Passive Origin-Destination Data

		Fix Trip Length Bias	Fix Coverage Problems	Fix Demographic Bias	Independent of Network	Ease of Application	Holdout Count Sample	Transparency
1	Market Penetatraion-based	✗	✓	✓✓	✓	✓	✓	-
2	Trip-Generation-based	✗	✓	✓✓	✓	-	✓	✓
3	Single-factor Scaling	✗	✗	✗	✓	✓	-	✓
4	Frataring	✗	✓	✗	✓	✓	✗	✓
5	Iterative Screenlines	✓	✓	✓	✓	-	✓	✓
6	Direct ODME	✓	✓	✓	✗	✓	-	✗
7	Indirect ODME	✓	✓	✓	✗	✗	-	-
8	Parametric Scaling	✓✓	✓	✓	✗	✗	-	-
9	Disaggregate Trace Auditing	✓✓	✓	✓✓	✓	✗	✓✓	✓

FIGURE 5. Trip-Generation Based Validation of LBS Data for Fort Wayne, IN

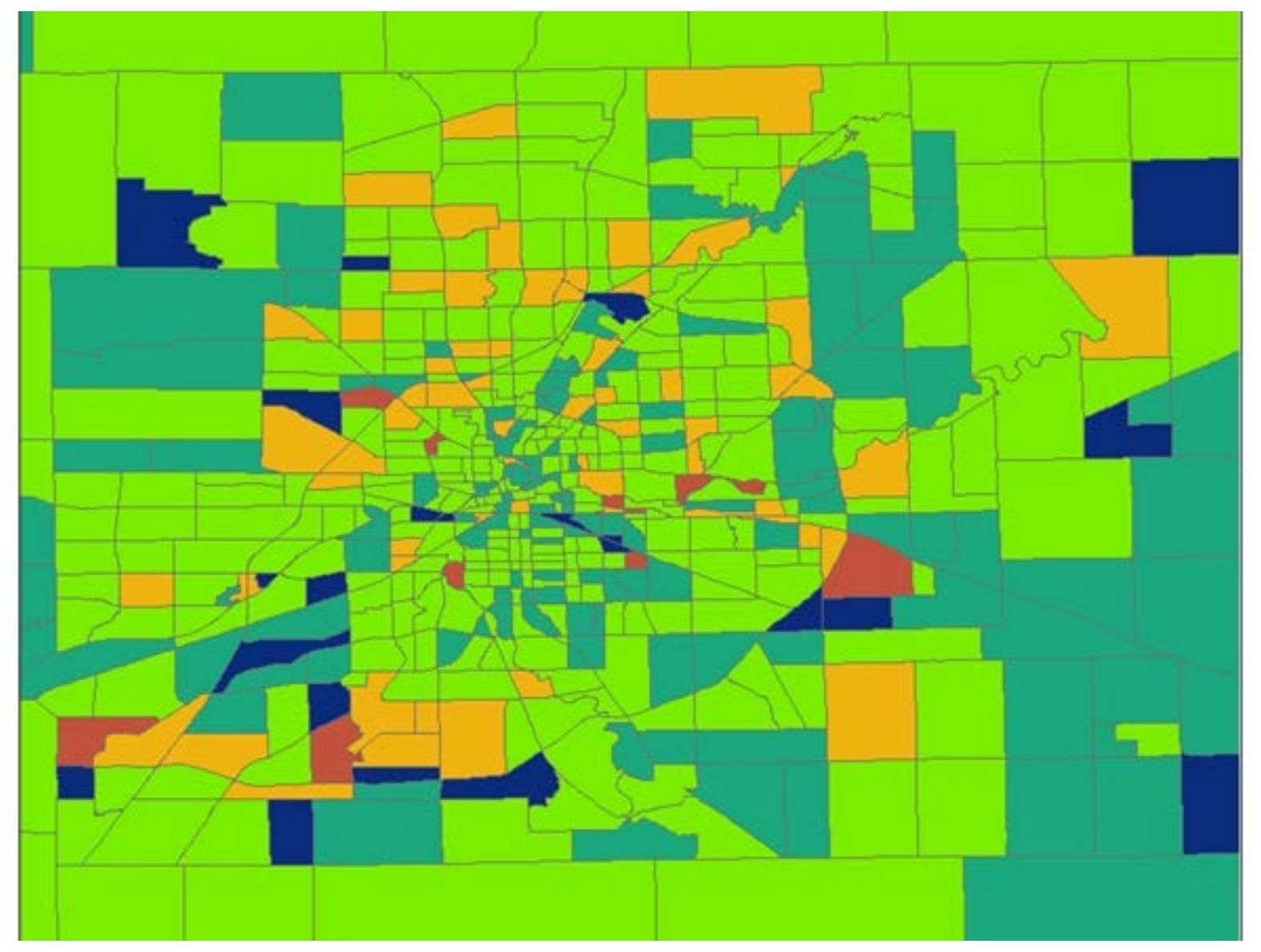


FIGURE 6. Screenlines for Iterative Screenline Fitting of Cellular Data in Chattanooga, TN

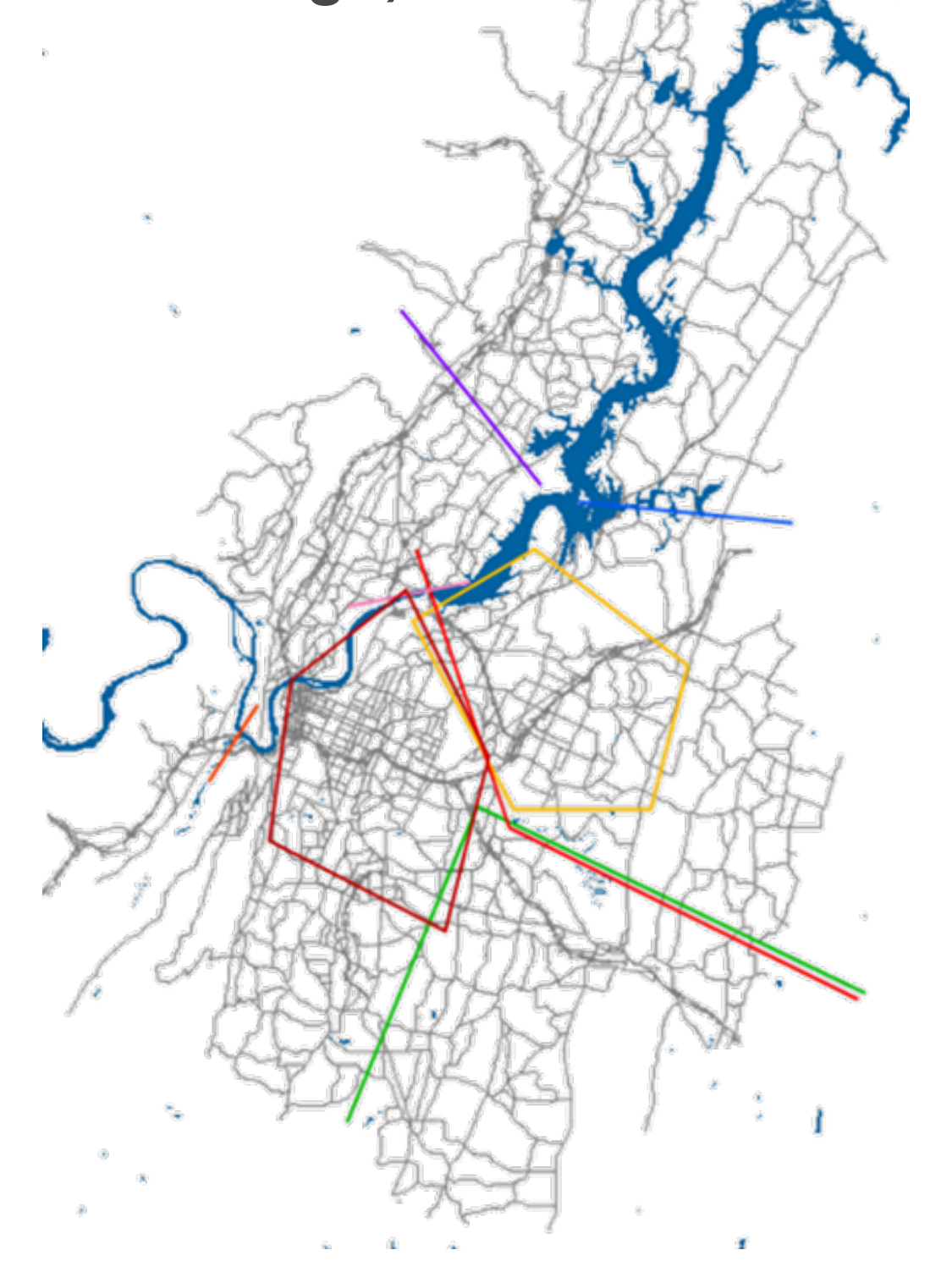


FIGURE 7. Truck Trip Length Based GPS Expansion Factors from Indirect ODME in Iowa

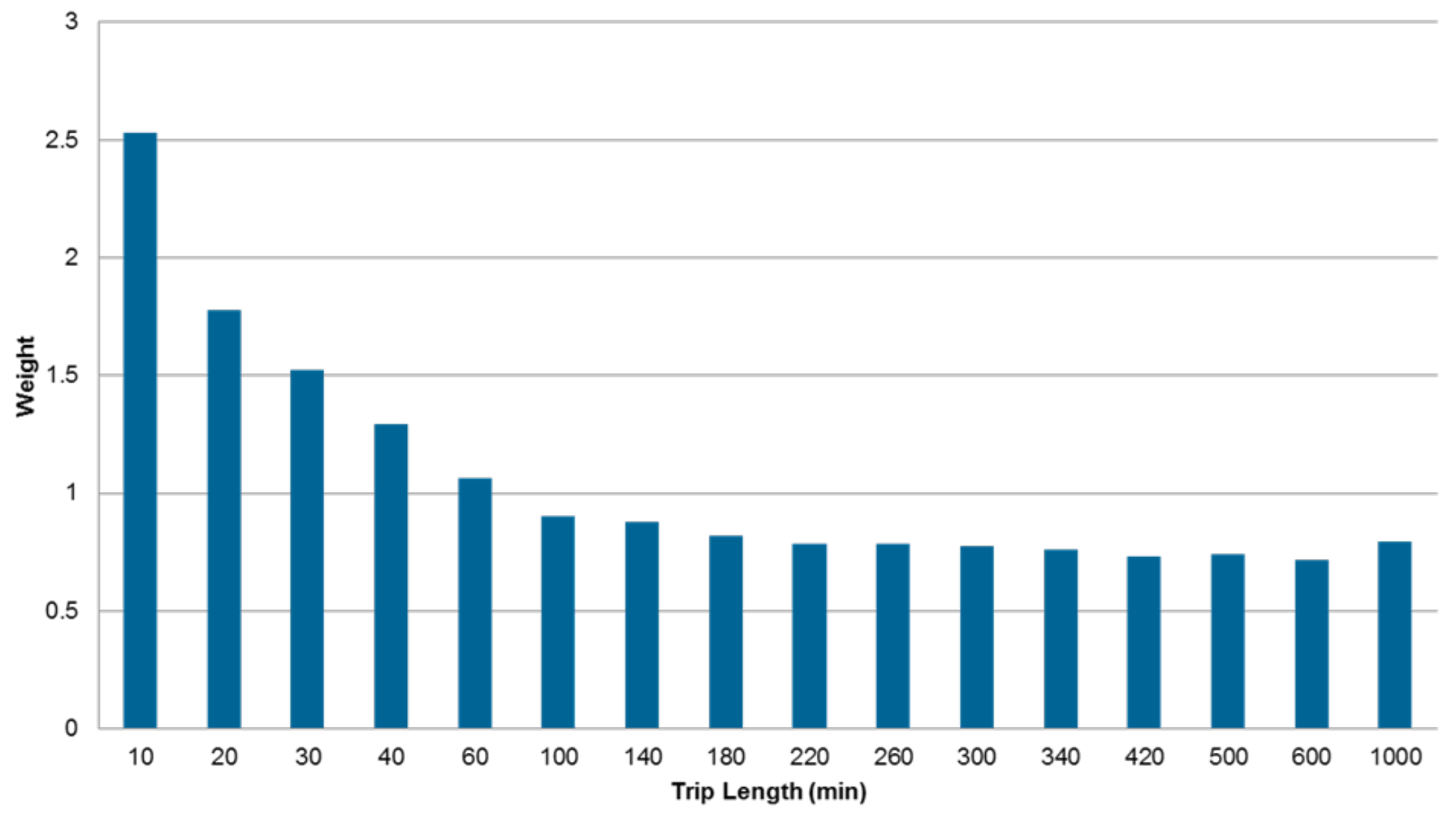
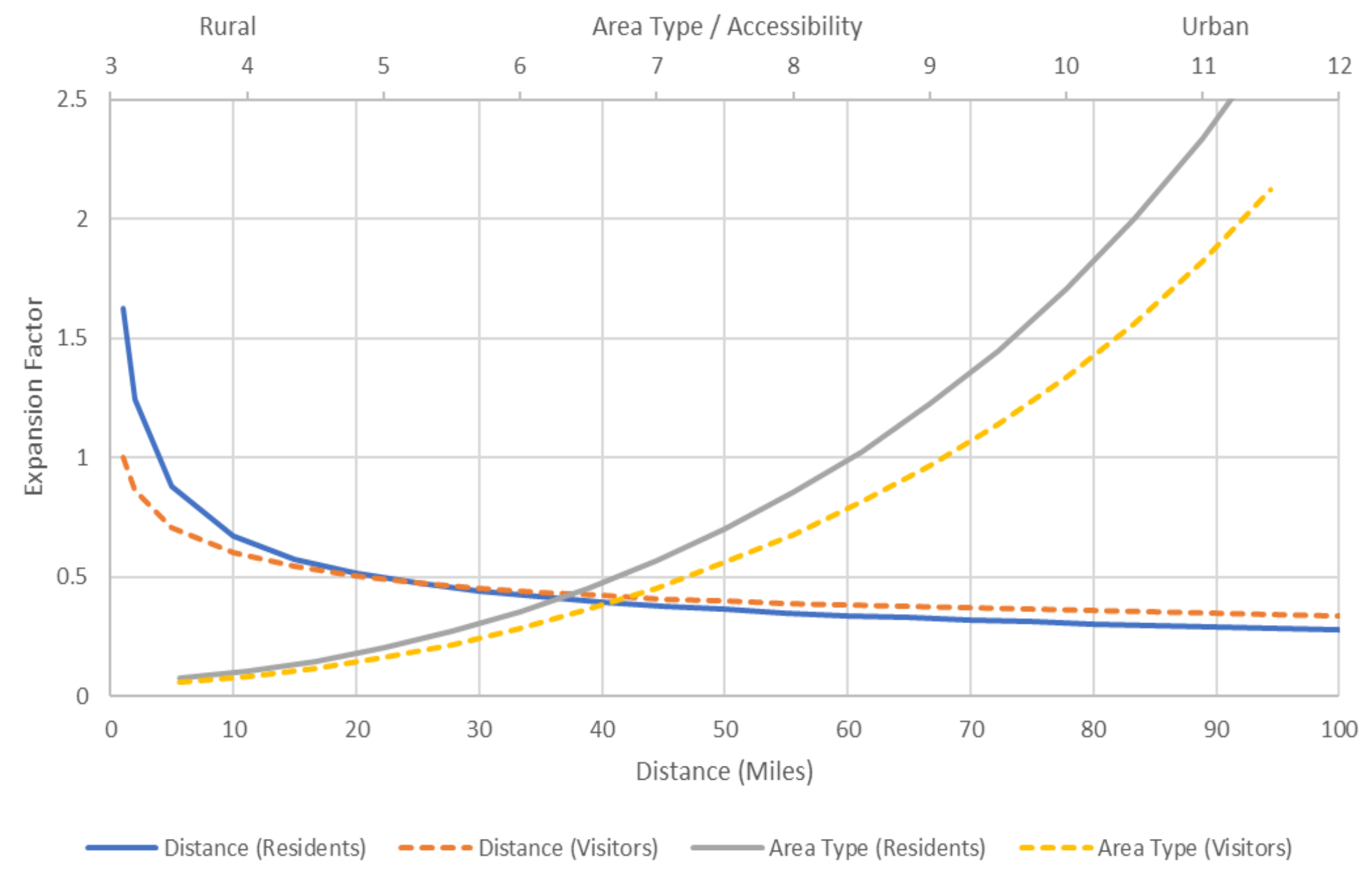


FIGURE 8. Parametric Expansion Factor Curves for Cellular Data for Charlotte, NC


Rural
Area Type / Accessibility
Urban
Expansion Factor
Distance (Miles)
Distance (Residents)
Distance (Visitors)
Area Type (Residents)
Area Type (Visitors)