MAG/WFRC BIKE MODEL CALIBRATION USING PASSIVE DATA

January 2023

Prepared for Mountainland Association of Governments

Report Title:

MAG/WFRC Bike Model Calibration Using Passive Data

Report Prepared by:

RSG

Report Prepared for:

Mountainland Association of Governments

For additional information regarding this report, or for questions about permissions or use of findings contained therein, please contact:

RSG (Headquarters)
55 Railroad Row
White River Junction, VT 05001
(802) 295-4999
www.rsginc.com

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

In cooperation with MAG, WFRC developed a bike demand model ("the model") that estimates bike trips across the Salt Lake City region. The model is a trip-based model with three steps: trip generation, trip distribution, and trip assignment. Lacking recent local survey data with a sufficient sample of bicycle trips, the model was initially estimated using coefficients derived from a household travel survey in Sacramento, CA with oversampling of bicycle trips.

Recently, MAG and WFRC received StreetLight origin-destination (OD) data for the region, including estimated trip flows by mode including bicycle. These data are summarized monthly from January 2019 through December 2021. Yearly and "bicycle month" data summaries are also available for each year from 2019 to 2021.

This project explores how StreetLight data in the region can be used to calibrate the bike model to better represent local conditions. The remainder of this memo is organized as follows: Section 2.0 describes foundational work to enable comparisons between the model and available StreetLight data, including several revisions made to the model to resolve implementation issues in the existing model. This section also documents performance of the baseline model relative to StreetLight OD data. Section 3.0 describes the model calibration process and Section 4.0 details performance of the calibrated model. Finally, Section 5.0 compares link-level model results to StreetLight pass-through data and Section 6.0 discusses current and future model applications.

2.0 BASELINE MODEL COMPARISON

Initially, the model was run using the most recent model code in the WFRC repository.[1] Review of the results obtained using this model yielded unexpected results in two areas: trip length distributions generated by the mode and the distribution of non-home-based trips. Additional exploration of these anomalies revealed existing implementation issues, detailed below.

2.1 EXISTING MODEL IMPLEMENTATION ISSUES

Skim top-coding

The trip length distribution generated by the initial model clustered around 5 miles, with an average trip distance of approximately 5.5 miles (Figure 1). Across all purposes, the trip length distribution differed substantially from the SACOG survey (Table 1). The dramatic differences between trip lengths produced by the model and those in the survey used to estimate the model indicated an issue with the existing model implementation. Specifically, the trip distribution submodel tended to assign trips to attractions near the maximum allowable trip distance for each trip category. Trip distance is introduced in this submodel through the generalized cost term, which is calculated when model skims are generated.

Because the skim matrix for the model is very large, a maximum trip distance is imposed on the matrix, and any cells the skim matrix exceeding this value are set to zero to reduce storage requirements. Then, in the trip distribution submodel, trips are allocated only to cells above this threshold. However, in the existing implementation of the model, this was calculated based on the distance skim, and the generalized cost function was calculated using the generalized cost skim. Generalized costs are nearly always greater than distance skims, so any cells with a distance skim less than the maximum trip distance but a generalized cost greater than the maximum trip distance to be included as possible destinations. Critically, for this set of cells, the generalized cost value was being set to zero since it exceeded the maximum trip distance threshold. As a result, the model was allocating a large share of trips to cells with artificially reduced generalized costs and, typically, trip distances near the maximum distance threshold.

This implementation issue was fixed by replacing generalized costs with zero only in cases where the distance skim exceeded the maximum trip distance threshold. This maintained the performance and storage gains achieved by storing skims in this way, while avoiding the cases described above. Using the same parameters as before the model produced a much more realistic trip length distribution, with improved alignment with the SACOG survey across all trip types (Figure 1).

[1] https://github.com/WFRCAnalytics/utah_bike_demand_model

TABLE 1. REVISED BASELINE MODEL AVERAGE TRIP LENGTH BY PURPOSE (MILES)

TRIP PURPOSE	BASELINE MODEL		REVISED BASELINE MODEL		SACOG SURVEY
	HOME-BASED TRIPS	NON-HOME-BASED TRIPS	HOME-BASED TRIPS	NON-HOME-BASED TRIPS	
Long-distance rec	7.15	7.66	6.77	7.12	13.3
Mountain bike rec	5.33	-	4.55	-	-
Family recreation	3.72	3.63	0.95	0.95	1.31
Other recreation	3.65	3.51	1.17	1.04	1.74
Work	5.41	5.47	1.56	1.40	2.44
Grade school	3.69	3.74	1.31	0.99	1.27
University	1.00	3.38	0.81	2.21	-
Maintenance	3.41	3.61	0.96	0.90	1.36
Discretionary	3.45	3.50	0.95	0.91	1.95

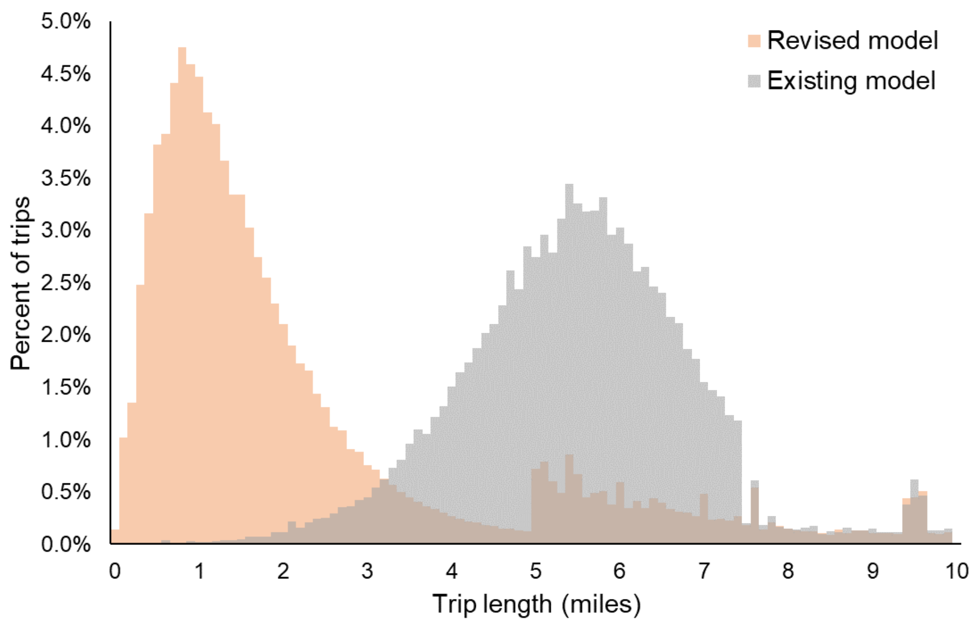

FIGURE 1. EXISTING MODEL TRIP LENGTH DISTRIBUTION (GREY) COMPARED TO REVISED MODEL TRIP LENGTH DISTRIBUTION (ORANGE)

Microzone indexing

A second implementation issue was discovered when analyzing the distribution of non-home-based trips generated by the model. The five microzones generating the highest number of non-home-based trips were examined, and the location and characteristics of these zones did not seem aligned with the trips they were producing. This was the result of an indexing issue within the model implementation: non-home-based trips are generated using non-indexed numpy arrays, which are then joined to a pandas-indexed dataframe used to sort trip productions. If the index of this pandas dataframe is the same as the numpy array (that is, starting at 0 and monotonically increasing by 1) the non-home-based trips are assigned to the correct zone; however, if the pandas dataframe is indexed in some other way the values from the numpy array will not be assigned to the correct zone. In the existing implementation of the model, the index of the *zones.csv* file was not monotonically increasing by 1; rather, five values were skipped in the *zone_id* column. This resulted in non-home-based trips being joined to the wrong zone. Because five ids were skipped, these non-home-based trips were joined to the zone with a *zone_id* value exactly five less than the correct zone (i.e., an off-by-five error).

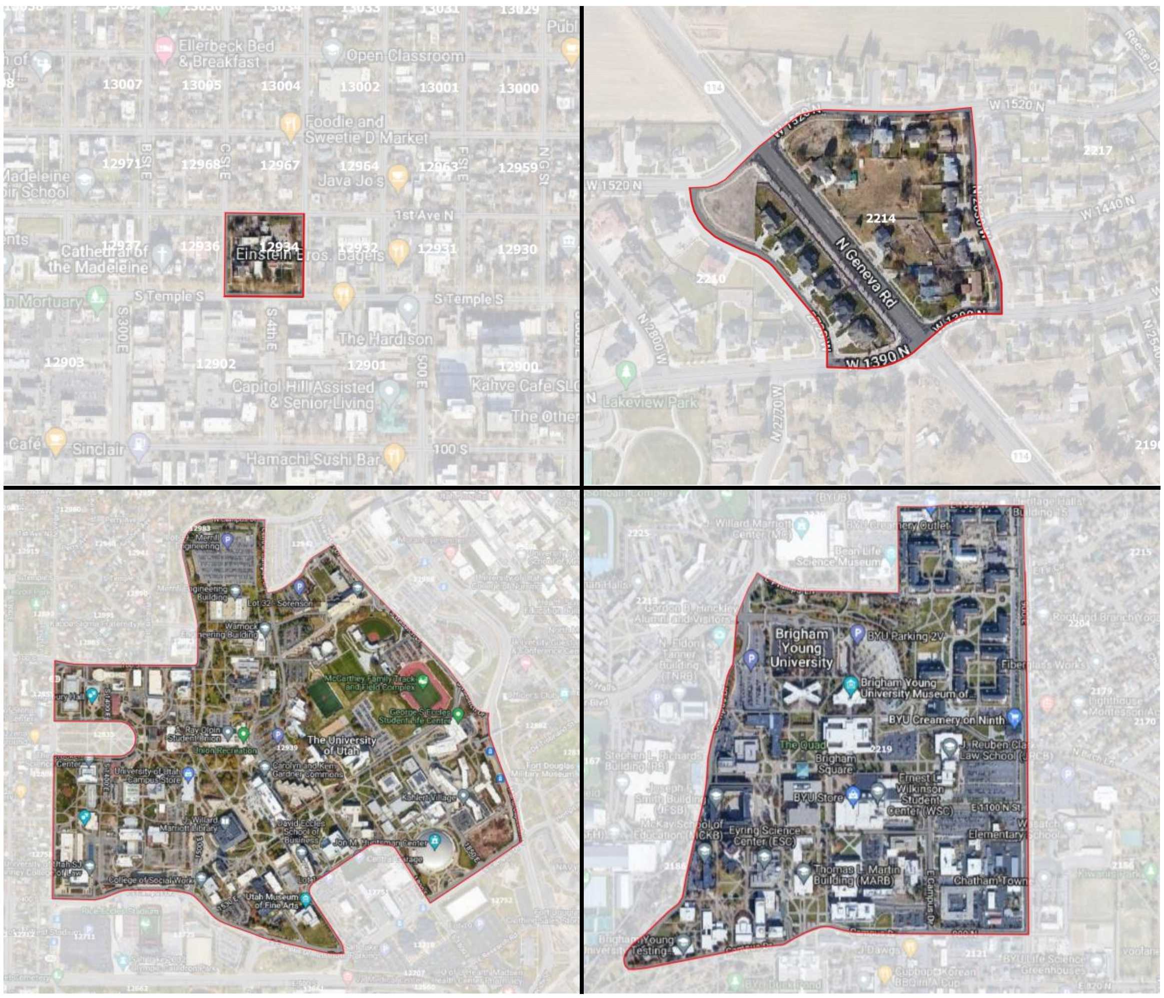

FIGURE 2. THE TWO ZONES PRODUCING THE HIGHEST NUMBER OF NON-HOME-BASED TRIPS BEFORE MODEL REVISIONS (TOP) AND AFTER (BOTTOM). ZONE_IDS ARE LABELED IN WHITE REFLECTING THE OFF-BY-FIVE INDEXING ERROR.

This issue was fixed by reindexing the *zones.csv* file. After this revision, the five microzones generating the highest number of non-home-based trips were much more intuitive. With the indexing issue present, the two zones producing the highest number of non-home-based trips were a small zone northeast of downtown Salt Lake (Figure 2; zone id: 12,934) and a small residential zone west of the BYU campus (zone id: 2,214). After fixing the issues, the two zones producing the highest number of non-home-based trips were the University of Utah (zone id: 12,939) and the BYU campus (zone id: 2,219), reflecting the off-by-five error described above.

2.2 ALIGNING MODEL GEOMETRY

To compare results from the model to StreetLight, a foundational task was to assign model microzones to the larger “StreetLight zones” that were used to obtain the OD data. However, the model does not nest cleanly within the StreetLight zones (Figure 3). To align these two geometries, the StreetLight zones were trimmed to match the extent of the model. This resulted in removing the zones representing the canyon routes east of the region as well as the polygon representing Antelope Island State Park and other zones in the southwest portion of the region.

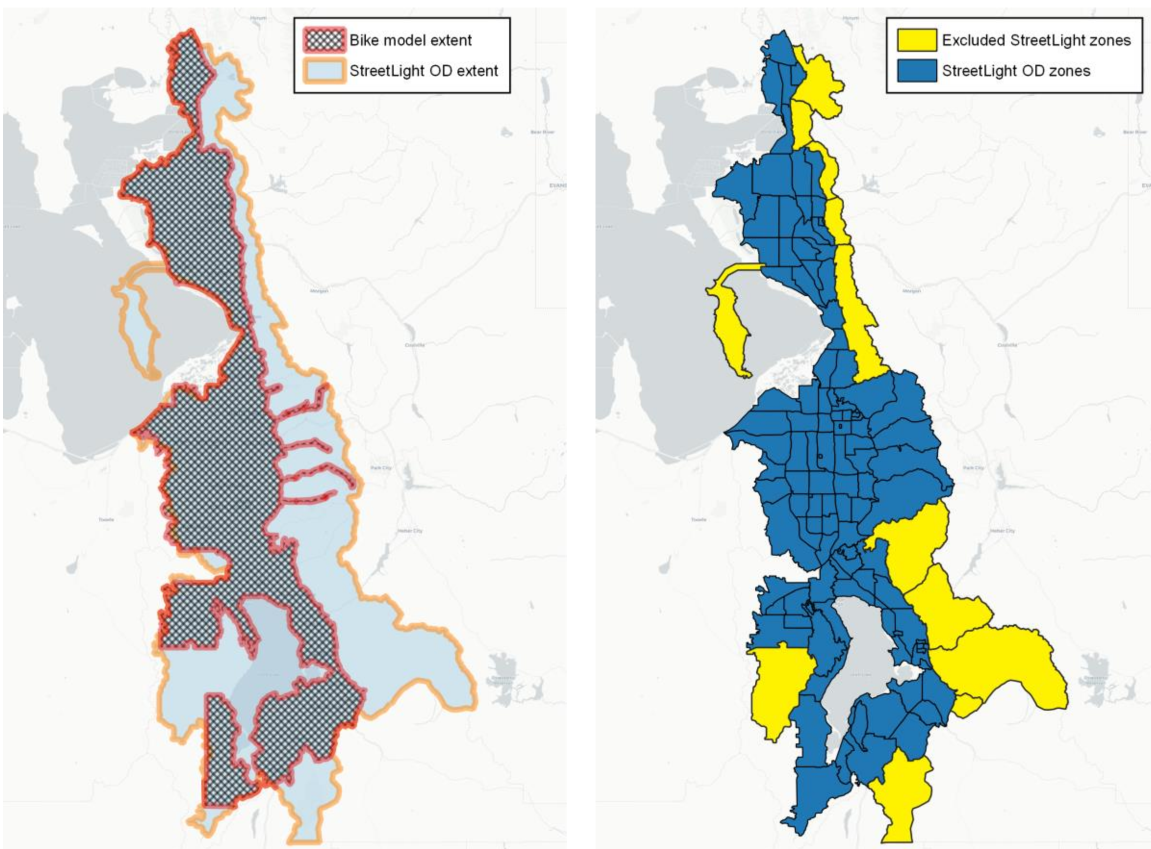

FIGURE 3. EXTENTS OF MODEL AND STL ZONES (LEFT) AND EXCLUDED STL ZONES (RIGHT)

Next, model microzones were assigned to the StreetLight zone in which they were contained. To do this, a union was performed between the two zone systems and microzones were assigned to the StreetLight zone containing the largest portion of the microzone's area. The borders of microzones aggregated to their assigned StreetLight zone match the original StreetLight zones well in most cases; however, there are several instances where the model microzone boundaries do not align with StreetLight zone boundaries (Figure 4).

2.3 MISSING DATA AND OUTLIERS

Once the geometry of the StreetLight zones and model zones were aligned, additional QAQC revealed two issues: missing data in model microzones within Box Elder County and outliers in the StreetLight data.

Box Elder County data

Microzones in Box Elder County were missing data for variables derived from the Real Estate Market Model (REMM). In the microzone data, these records were represented by zeroes—resulting in low trip generation rates across the county. In the StreetLight data, these model zones are represented by five zones—11, 21, 31, 32, and 41 Figure 4).

StreetLight outliers

StreetLight zones representing downtown Salt Lake City, the airport, and the industrial area south of the airport had higher than expected bicycle trips in the StreetLight data. Notably, the StreetLight data contains over 11,000 daily trips from downtown and over 2,000 daily bicycle trips from the airport. This may reflect limitations in the methods used to identify bicycle trips in areas with high transit usage or stem from underlying bias in the LBS data used by StreetLight. Regardless, these zones were ignored when calibrating the model to avoid overfitting (Figure 4).

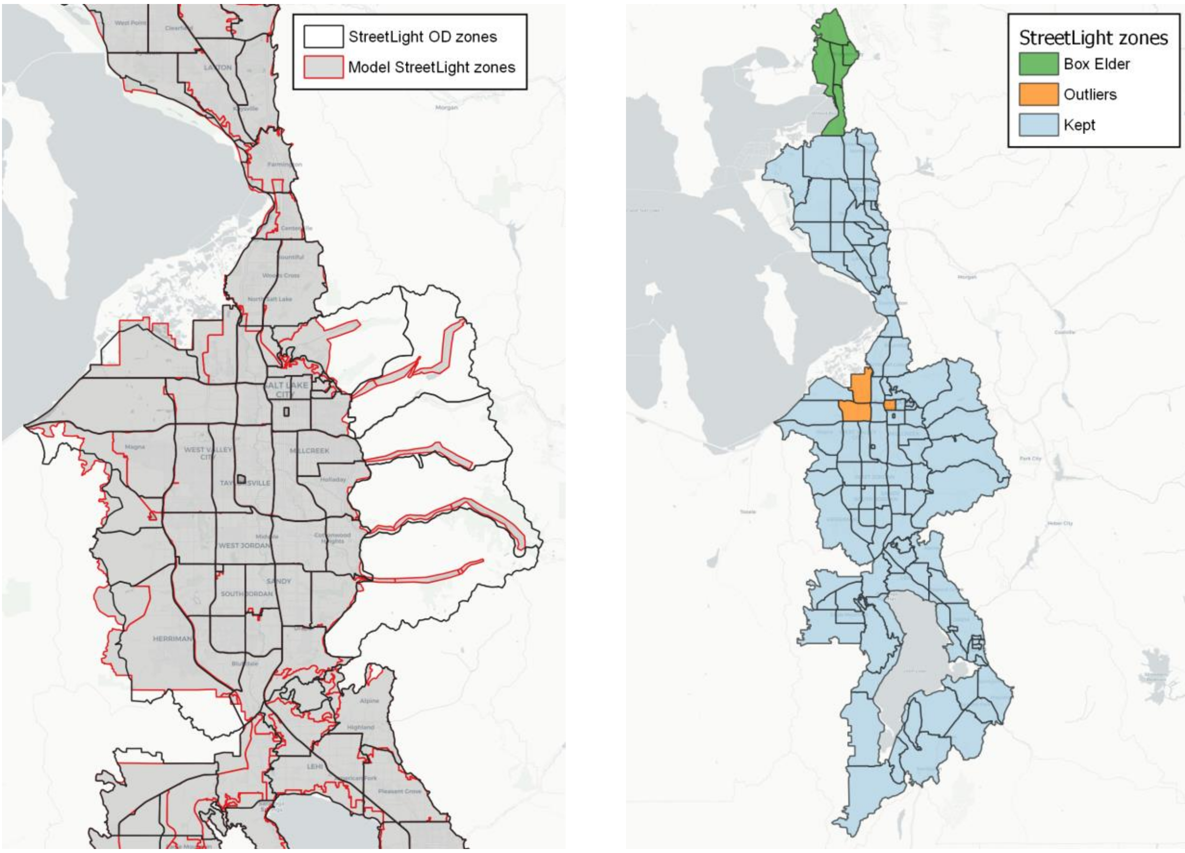

FIGURE 4. MODEL MICROZONES (RED OUTLINES) AGGREGATED TO STREETLIGHT ZONES (BLACK OUTLINES) AND EXCLUDED BOX ELDER ZONES

2.4 BASELINE MODEL PERFORMANCE RELATIVE TO AVAILABLE STREETLIGHT DATA

StreetLight bike OD data the region were available for all months from January 2019 through December 2021, including yearly and “bike month” aggregations of the monthly data[2]. These data exhibit expected seasonal variation, as well as a general increasing trend over time (Table 2). The last two months (November and December 2021) indicate a dramatic upswing in bike trips, with volumes increasing roughly 7-fold over previous months. These outliers may be th result of changes in StreetLight mode imputation methods and/or changes in underlying data and were ignored in this analysis. Spatially, the StreetLight data demonstrate a similar pattern in comparison with the model, though more trips are more concentrated in downtown areas, and notably sparser in the neighborhoods south of downtown Salt Lake City.

[2] Bike months include April-October of each year

TABLE 2. MODEL COMPARISON TO MONTHLY STREETLIGHT DATA

DATASET		TOTAL TRIPS	MODEL RELATIVE TO STREETLIGHT	TRIP ORIGINS R^2	OD PAIRS R^2	PERCENT OD PAIRS WITH ZERO TRIPS
Model (ref)		220,390				76.2%
2019	January	51,802	425%	0.32	0.41	93.8%
	February	61,383	359%	0.27	0.34	93.5%
	March	70,305	313%	0.35	0.43	93.4%
	April	94,760	233%	0.45	0.52	93.2%
	May	100,597	219%	0.48	0.54	92.7%
	June	108,257	204%	0.50	0.55	92.5%
	July	122,028	181%	0.45	0.52	92.6%
	August	122,468	180%	0.49	0.57	92.9%
	September	122,273	180%	0.46	0.52	92.8%
	October	97,285	227%	0.47	0.54	92.9%
	November	70,418	313%	0.47	0.54	93.4%
	December	58,536	377%	0.34	0.43	93.9%
2020	January	77,579	284%	0.29	0.40	94.1%
	February	88,379	249%	0.30	0.39	94.0%
	March	107,397	205%	0.45	0.55	93.6%
	April	127,962	172%	0.58	0.60	93.1%
	May	165,364	133%	0.57	0.60	92.9%
	June	176,338	125%	0.59	0.62	93.0%
	July	159,184	138%	0.59	0.63	92.9%
	August	165,233	133%	0.62	0.64	93.4%
	September	151,470	146%	0.64	0.66	93.1%
	October	143,667	153%	0.64	0.65	93.7%
	November	106,564	207%	0.60	0.64	94.2%
	December	89,016	248%	0.37	0.51	94.6%
2021	January	90,564	243%	0.29	0.43	95.1%
	February	68,529	322%	0.29	0.44	95.0%
	March	103,402	213%	0.26	0.41	94.5%
	April	102,403	215%	0.47	0.57	94.2%
	May	123,578	178%	0.60	0.65	93.8%
	June	141,365	156%	0.57	0.63	93.9%
	July	132,696	166%	0.55	0.63	93.9%
	August	145,792	151%	0.55	0.62	93.8%
	September	170,375	129%	0.56	0.62	93.7%
	October	134,399	164%	0.51	0.59	94.8%
	November*	661,730	33%	0.53	0.61	95.2%
	December*	655,725	34%	0.56	0.62	95.2%

**StreetLight data estimates much higher than expected*

To perform an initial comparison of the model results to the available StreetLight data, several measures were calculated. First, aggregate trip counts were compared. Overall, the model estimated 220,390 daily bike trips across the region—tending to over-predict bike trips relative to the StreetLight, though less so during prime bicycling months (Table 2Table 2). Aggregate trip volumes were relatively close for several months in 2020 (May, June, and August) and 2021 (September). For November and December 2021, the model dramatically underpredicts trips relative to the StreetLight data—though these are the months where the StreetLight data seem unreliable.

Next, trip origins were compared between the model and StreetLight at the zone level. Trip origins from the model and the StreetLight data were aggregated by StreetLight origin zone and the R^2 was calculated to compare model results to all StreetLight time periods. R^2 values vary dramatically over time, from as low as 0.27 in February 2019 to as high as 0.64 in September/October 2020. Finally, trip volumes at the OD level were also compared between the model and available StreetLight data. R^2 values for origin-destination ranged from 0.34 in February 2019 to 0.66 in September 2020. Across all time periods, the StreetLight data contain a much higher percentage of OD pairs with no trips than the model: around 94% in the StreetLight data versus around 76% in the model. Notably, the StreetLight data also contains a considerably larger share of interzonal trips—up to 80% of all trips—compared to only 40% in the model. The same comparison was also performed for the yearly and bike month aggregated datasets with similar results, though there are notably fewer OD pars with zero trips in the yearly and bike month aggregated datasets (Table 3).

TABLE 3. MODEL COMPARISON TO AGGREGATED STREETLIGHT DATA

DATASET		TOTAL TRIPS	MODEL RELATIVE TO STREETLIGHT	TRIP ORIGINS R^2	OD PAIRS R^2	PERCENT OD PAIRS WITH ZERO TRIPS
Model (ref)		220,390				76.2%
Bike months	2019	109,584	201%	0.47	0.54	88.5%
	2020	155,446	142%	0.62	0.64	89.0%
	2021	136,327	162%	0.57	0.64	89.9%
All year	2019	90,101	245%	0.45	0.53	89.8%
	2020	129,640	170%	0.61	0.64	88.7%
	2021	215,047	102%	0.64	0.67	90.1%

2.5 SELECTING THE CALIBRATION TIME PERIOD

Through this initial comparison of the bike model and available StreetLight data, three key themes emerged:

- The model tends to over-predict trips relative to the StreetLight data
- In the most comparable months, the model tends to under-predict trips in certain geographies, such as downtown Salt Lake City, and tends to over-predict trips in other geographies, most notably south of downtown Salt Lake City
- The StreetLight data are highly variable over time, reflecting both seasonality but also changes in methodology and/or underlying data

Taking these findings into account, the project team decided to use the Bike Months StreetLight dataset for model calibration. This dataset minimizes the impact of the pandemic on travel behavior while avoiding pre-pandemic time periods where StreetLight data compare less favorably to the baseline model. Critically, this aggregation also avoids the two outlier months at the end of 2021 (November and December) where StreetLight estimates spike unexpectedly.

3.0 MODEL CALIBRATION

Model calibration was performed in two stages (Figure 5). First, a sensitivity analysis was performed on all model parameters in the trip generation and trip distribution submodels and initial adjustments to model coefficients were made based on these findings. Next, the partially calibrated model was run and the model residuals (the difference between model estimates and the StreetLight data) were included as the outcome variable in a regression model with zonal data as exploratory variables. If significant associations were found between the model residual and zonal data, coefficient adjustments were made in the opposite direction of the association and the process was repeated until satisfactory model performance was achieved. To avoid overfitting to the StreetLight data, which has its own limitations and does not represent a true "ground truth" dataset, only mode coefficients were adjusted within reasonable ranges. Finally, the sensitivity of the trip assignment submodel was estimated be specifying 10 combinations of coefficients for the model generalized cost function, running the model once for each combination, and comparing to the model, using coefficients for the trip generation and trip assignment models form the first model calibration. These steps are described in turn below.

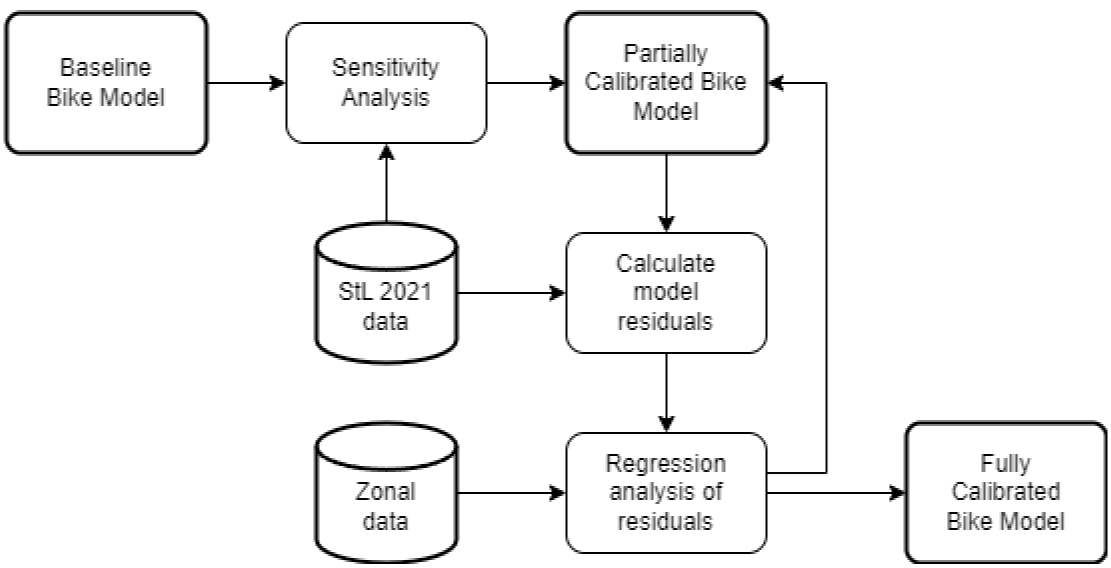

FIGURE 5. MODEL CALIBRATION PROCESS

3.1 SENSITIVITY ANALYSIS: TRIP GENERATION AND DISTRIBUTION SUBMODELS

To explore the sensitivity of the model to coefficients in the trip generation and trip distribution sub models, parameters in these submodels were iteratively set to zero and a model run was performed. For parameters appearing in both submodules, one model run was performed with the parameter set to zero in trip generation submodel, one with parameter set to zero in the trip distribution submodel, and one with the parameter set to zero in both submodels. This approach resulted in 48 model runs.

The sensitivity of the model to each variable was then calculated for several metrics. First, the difference in trip counts by purpose relative to the baseline model was calculated to test how each model parameter impacted trip counts estimated by the model. The same calculation was performed for average trip length. In general, the model is highly sensitive to trip generation constants, especially for work trips, the count of jobs in sectors 3 and 6, the percent of high-income households, and measure of agricultural land. In terms of average trip lengths, the model is less sensitive overall, with only a handful of parameters having a meaningful effect (Table 4).

To better understand the sensitivity of the model to the parameters summarized above, Tornado diagrams were also generated, depicting the sensitivity of the model to the top-20 parameters for both trips and mean trip length. Because of the limited sensitivity of model trip length distributions, only the trip count plot is shown below. The model is highly sensitive to the work trip generation constant, percent agricultural land variables, and several of the employment variables (Figure 6). For reference, a list of variables used in the model is provided in Appendix A.

TABLE 4. TOTAL TRIP AND MEAN TRIP DISTANCE MODEL SENSITIVITIES

PARAMETER		TRIP GENERATION		TRIP DISTRIBUTION		TRIP GENERATION & TRIP DISTRIBUTION	
		Total trips	Mean distance	Total trips	Mean distance	Total trips	Mean distance
Employment variables	Jobs3	-33,439	0.299	-17	-0.021	-33,456	0.240
	Jobs4	-3,118	-0.169	0	0.021	-3,118	-0.150
	Jobs5	4,570	-0.032				
	Jobs6	22,151	-0.081	0	0.000	22,151	-0.113
	Jobs7			0	0.001		
	Jobs9			0	-0.001		
	Jobs_total			-188	0.072		
Demographic variables	Enrol_elem	-309	0.000	-11	-0.001	-320	-0.014
	Enrol_midl	-90	0.000	0	0.000	-90	0.000
	Enrol_high	-120	-0.001	-1	0.000	-121	-0.001
	Coll_enrol	-2,632	0.015	-176	-1.497	-2,632	0.015
	Households	-3,326		-15	0.000	-3,341	-0.076
	Inc1	-17,755	0.022				
	Inc4	-44,162	-0.062				
	Pct_poplc2	44,481	-0.061				
	Pct_poplc3	40	0.106				
Built environment variables	Mixed_use	-7,705	0.097				
	Park_score	-36,275	0.122	0	0.028	-36,275	0.256
	Pct_ag1	125,660	-0.454				
	Pct_ag3	13,759	0.251				
	Ldr_score			-28,353	-0.858		
	Mtbn_score			-3,593	0.758		
Trip generation constants	Disc	-251	-0.014				
	Maint	15,371	0.037				
	Rec_farm	7,360	-0.020				
	Rec_long	-11,848	-0.278				
	Rec_other	-4,011	0.311				
	Sch_grade	8,488	-0.041				
	Sch_univ	1,028	0.025				
	Work	-151,866	1.501				

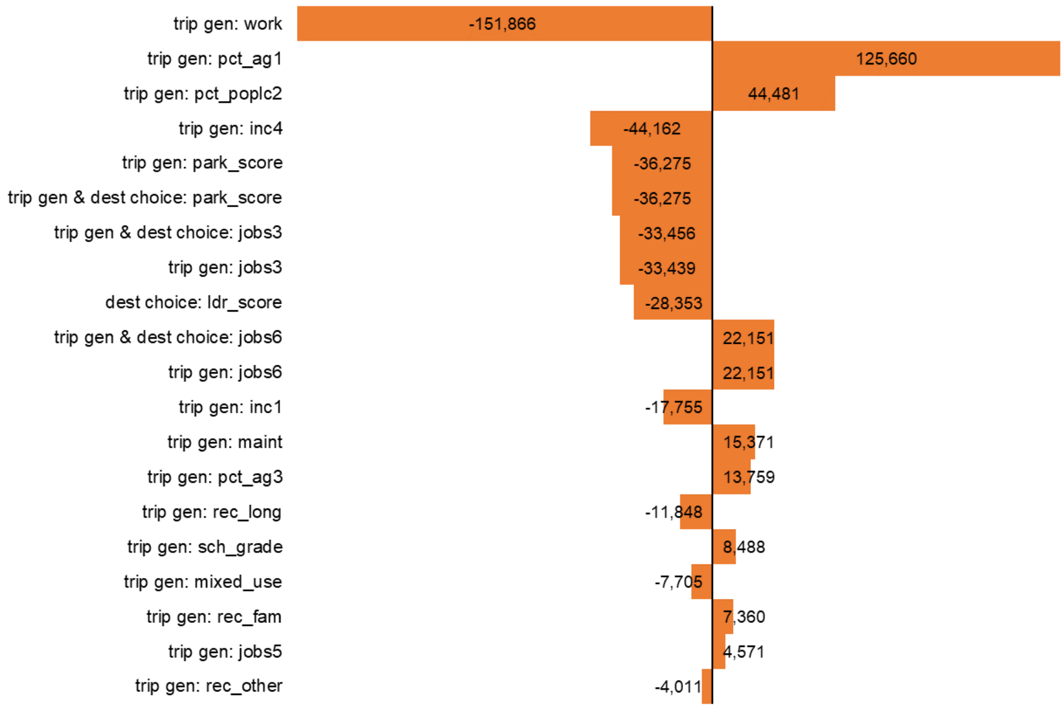

FIGURE 6. TORNADO PLOT ILLUSTRATING TRIP COUNT SENSITIVITIES TO MODEL PARAMETERS

Next, each model run was then compared to the 2021 StreetLight data and two performance metrics were calculated: the root mean square error (RMSE) and coefficient of determination (R^2). For example, an R^2 value of 0.03 in the Table below indicates that the model with that parameter set to zero improved the R^2 of the baseline model by 0.03. Conversely, a value of -0.02 indicates that setting that parameter to zero reduced the R^2 value by 0.02. This comparison was performed for both trip origins aggregated at the zonal level and origin-destination flows.

At the trip origin level, setting several model coefficients to zero improved model fit substantially. For example, setting the *jobs6* trip generation parameter to zero improved the R^2 measure by nearly 0.10 and reduced RMSE by over 20 (Table 5). Interestingly, some of the most influential parameters available to reduce total trips produced by the model, such as the work trip generation constant, negatively impact model performance as measured by the R^2 and RMSE (Figure 7). Further, some variables have mixed impacts across different submodels—for example, setting the *jobs3* coefficient to zero in the trip generation submodel has a substantial negative impact on model performance; however, doing the same in the trip distribution submodel marginally improves mode performance (Figure 7).

TABLE 5. MODEL PERFORMANCE METRIC SENSITIVITIES, TRIP ORIGINS

	PARAMETER	TRIP GENERATION		TRIP DISTRIBUTION		TRIP GENERATION & TRIP DISTRIBUTION	
		RMSE	R^2	RMSE	R^2	RMSE	R^2
Employment variables	Jobs3	66.13	-0.184	-10.45	0.037	64.22	-0.174
	Jobs4	2.52	-0.005	-0.10	0.000	2.41	-0.005
	Jobs5	-3.98	0.007				
	Jobs6	-20.12	0.091	0.02	0.000	-20.12	0.091
	Jobs7			1.90	-0.005		
	Jobs9			0.25	-0.001		
	Jobs_total			18.19	-0.078		
Demographic variables	Enrol_elem	0.29	0.000	0.42	-0.000	0.68	-0.000
	Enrol_midl	0.10	0.000	0.02	0.000	0.12	0.000
	Enrol_high	0.07	0.000	0.00	0.000	0.06	0.000
	Coll_enrol	10.52	-0.034	-0.06	-0.001	10.52	-0.034
	Households	0.96	0.009	-0.15	0.001	0.81	0.010
	Inc1	26.33	-0.038				
	Inc4	53.70	0.027				
	Pct_poplc2	-17.68	-0.022				
	Pct_poplc3	0.01	0.000				
Built environment variables	Mixed_use	10.41	-0.023				
	Park_score	47.13	-0.036	0.14	0.000	47.13	-0.036
	Pct_ag1	-4.64	-0.063				
	Pct_ag3	-7.28	-0.008				
	Ldr_score			0.36	0.059		
	Mtbn_score			1.90	0.003		
Trip generation constants	Disc	0.25	-0.000				
	Maint	-10.42	0.003				
	Rec_fam	-7.20	-0.002				
	Rec_long	5.82	0.030				
	Rec_other	5.09	-0.004				
	Sch_grade	-6.90	-0.004				
	Sch_univ	0.20	0.003				
	Work	728.50	-0.028				

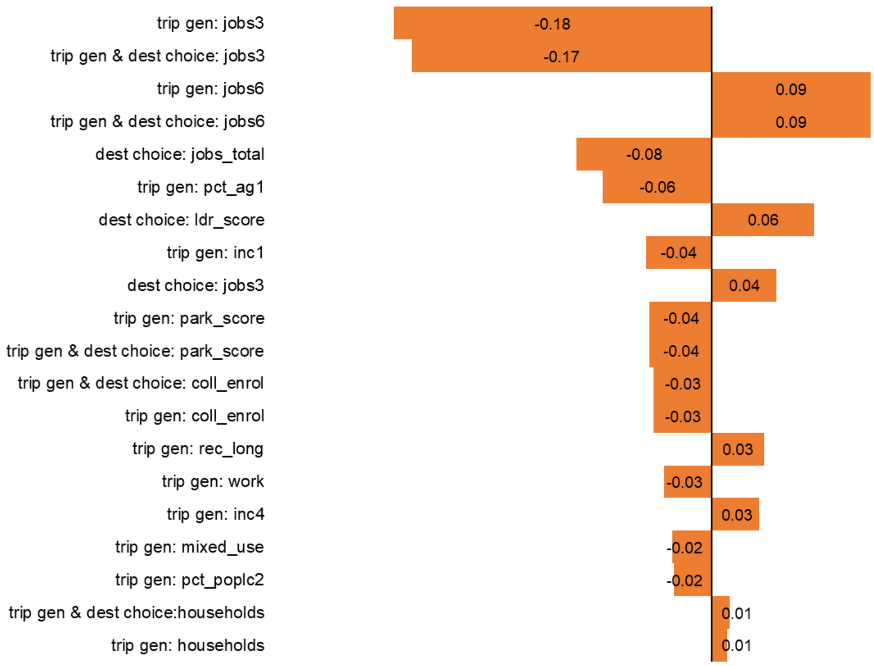

FIGURE 7. TORNADO PLOT ILLUSTRATING ORIGIN R^2 SENSITIVITIES TO MODEL PARAMETERS

Sensitivity analysis results at the origin-destination flow level were roughly consistent with the origin-level analysis, though a handful of parameters displayed different behavior (Table 6Table 6). For example, setting the trip generation constant for work trips to zero has a marginal negative impact on model performance at the origin level but a much stronger negative impact on model performance at the origin-destination flow level (Figure 8).

TABLE 6. MODEL PERFORMANCE METRIC SENSITIVITIES, ORIGIN-DESTINATION FLOWS

PARAMETER		TRIP GENERATION		TRIP DISTRIBUTION		TRIP GENERATION & TRIP DISTRIBUTION	
		RMSE	R^2	RMSE	R^2	RMSE	R^2
Employment variables	Jobs3	66.13	-0.086	-10.45	0.039	64.22	-0.078
	Jobs4	2.52	-0.001	-0.10	0.000	2.41	-0.001
	Jobs5	-3.98	0.002				
	Jobs6	-20.12	0.033	0.02	0.000	-20.12	0.033
	Jobs7			1.90	-0.006		
	Jobs9			0.25	-0.001		
	Jobs_total			18.19	-0.061		
Demographic variables	Enrol_elem	0.29	0.000	0.42	-0.001	0.68	-0.001
	Enrol_midl	0.10	0.000	0.02	0.000	0.12	0.000
	Enrol_high	0.07	0.000	0.00	0.000	0.06	0.000
	Coll_enrol	10.52	-0.026	-0.06	0.000	10.52	-0.026
	Households	0.96	0.005	-0.15	0.001	0.81	0.006
	Inc1	26.33	-0.030				
	Inc4	53.70	-0.003				
	Pct_poplc2	-17.68	0.016				
	Pct_poplc3	0.01	0.000				
Built environment variables	Mixed_use	10.41	-0.013				
	Park_score	47.13	-0.018	0.14	0.000	47.13	-0.018
	Pct_ag1	-4.64	-0.023				
	Pct_ag3	-7.28	-0.001				
	Ldr_score			0.36	0.025		
	Mtbn_score			1.90	0.003		
Trip generation constants	Disc	0.25	0.000				
	Maint	-10.42	0.009				
	Rec_fam	-7.20	0.009				
	Rec_long	5.82	0.015				
	Rec_other	5.09	-0.006				
	Sch_grade	-6.90	0.005				
	Sch_univ	0.20	-0.003				
	Work	728.50	-0.109				

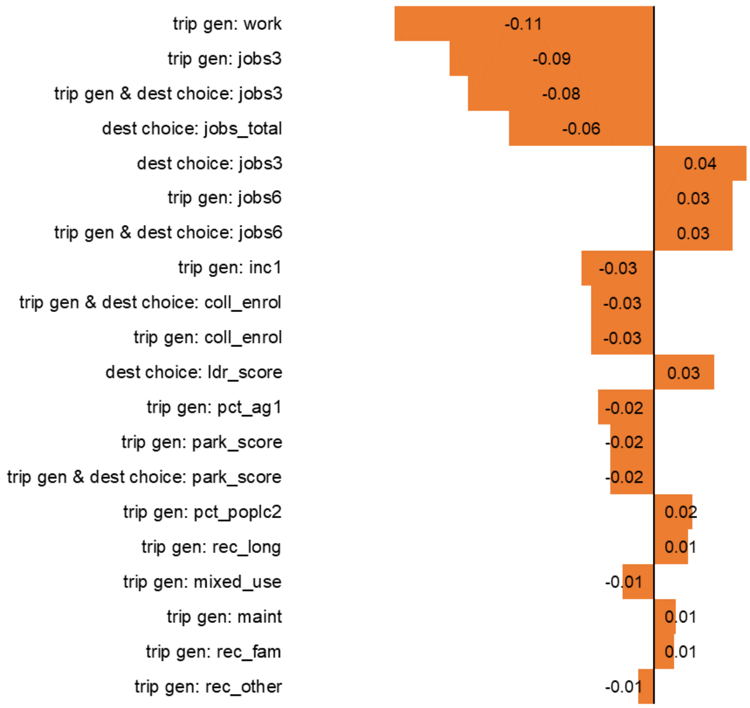

FIGURE 8. TORNADO PLOT ILLUSTRATING ORIGIN-DESTINATION R^2 SENSITIVITIES TO MODEL PARAMETERS

Overall, the sensitivity analysis demonstrated that the model is particularly sensitive to several parameters, including the trip generation constant for work trips, the trip generation zonal coefficient for the *jobs3* variable, and the trip generation and destination choice coefficients for the *jobs6* variable. However, many of these variables reflect tradeoffs when balancing multiple calibration objectives. Several of the parameters with the strongest influence on total trips produced by the model, including the work trip generation constant and the percent of agricultural land, negatively impact model performance when adjusted in the direction that reduces trips.

Informed by this sensitivity analysis, several model specifications were tested to serve as a basis for the residuals-based calibration process. Model parameters that positively impacted model performance across all objectives (trip count, R^2, and RMSE) were adjusted first, then parameters with strong influences on total trips were adjusted to aggregate model trip productions with the StreetLight data. This resulted in a partially calibrated model with adjustments to 16 model parameters, most of which improved model performance across all dimensions (Partially Calibrated Model Coefficients in Table 8). Some parameters were adjusted to reduce the total number of trips despite marginal negative impacts on other measures of model performance (e.g., the trip generation work coefficient).

3.2 SECONDARY REGRESSION OF MODEL RESIDUALS

To further refine the initially calibrated mode, an iterative procedure was employed. First, the model was run, and the model residuals were calculated for each StreetLight zone in the region. Next, a regression model was estimated using the calculated residuals as the outcome variable and zonal attributes aggregated to the StreetLight zone geometry as explanatory variables:

$$\boldsymbol{\gamma}(\boldsymbol{i}) = \boldsymbol{\alpha} + \boldsymbol{\beta 1} * \boldsymbol{\chi}(\boldsymbol{id}) + \boldsymbol{\beta 2} * \boldsymbol{\chi}(\boldsymbol{il})$$

where $\boldsymbol{\gamma}(\boldsymbol{i})$ is the model residual in zone i, $\boldsymbol{\chi}(\boldsymbol{id})$ is a matrix of demographic measures of households living in zone i, $\boldsymbol{\chi}(\boldsymbol{il})$ is a matrix of buffered land use measures of the attraction characteristics of microzones within three different buffers around microzone $\boldsymbol{m}$ aggregated to zone i, and $\boldsymbol{\alpha}$, $\boldsymbol{\beta 1}$, and $\boldsymbol{\beta 2}$ are model parameters that model associations between zonal attributes and model residuals.

This model provides insight on zonal characteristics associated with error in estimates from the partially calibrated model—that is, a positive and significant association between a zonal attribute and the model residual indicates that model over-predictions are associated with that zonal attribute. In the first iteration, the regression model revealed associations between model predictions and several zonal attributes, including parameters in the trip generation and/or trip distribution models and parameters appearing in neither model (Table 7).

TABLE 7. INITIAL RESIDUALS REGRESSION MODEL RESULTS

VARIABLE	COEFFICIENT	T-VALUE
Job4	0.1901	6.463***
Enrol_high	0.5466	2.689**
Bkpath_len	-2.664	-3.142**
Pct_ag1	-1,913	-3.416***
Pct_ag3	4,291	3.860***
Light_rail	-1,415	-2.434*
Constant	69.38	0.549

***p<0.001 **p<0.01 *p<0.05

The results of this regression model were used to further calibrate the model. If a significant positive association was found, then the coefficient for that variable was reduced in the model and vice versa. For parameters not already present in the partially calibrated model, model coefficients were introduced in the trip generation model again with a sign opposite of the regression coefficient to reduce model errors. For each model variable, this process was repeated (bike model estimation, calculation of residuals, estimation of regression model, coefficient adjustments) until no significant associations were found between model residuals and zonal attributes or additional adjustments did not yield significant improvements in model performance.

This process resulted in adjustments to several model parameters, including the trip generation submodel constants for work and maintenance trips and the trip generation zone coefficients for the *pct_ag3* parameter. Additionally, a new model parameter (*bkpath_len*) was added to the trip generation buffer coefficients model (Table 8Table 12).

TABLE 8. ADJUSTMENTS MADE TO CALIBRATED MODEL AND MODEL PERFORMANCE METRICS

SUBMODEL	EQUATION	PARAMETER	TRIP PURPOSE	BASELINE MODEL COEFFICIENT	PARTIALLY CALIBRATED MODEL COEFFICIENT	CALIBRATED MODEL COEFFICIENT
Trip generation	constants	constant	work	0.123	0.0903	0.11685
			rec_long	0.012	0.003	0.003
			maint	-0.029	-0.029	-0.058
	zone_coefs	Inc1	rec_long	0.033	0.033	0.00475
			rec_fam	0.011	0.011	0.00300
			rec_other	0.011	0.011	0.00375
			work	0.071	0.071	0.0003775
			sch_grade	0.048	0.048	0.00175
			maint	0.124	0.124	0.02675
			disc	0.092	0.092	0.01925
		Inc4	rec_long	0.033	0.00825	0.00825
			rec_fam	0.011	0.00275	0.00275
			rec_other	0.011	0.0055	0.0055
			work	0.071	0.01775	0.00125
			sch_grade	0.048	0.012	0.012
			maint	0.124	0.031	0.031
			disc	0.092	0.023	0.023
		pct_ag1	work	-0.068	-0.068	-0.114
			maint	-0.047	-0.047	-0.0945
			dsic	-0.047	-0.047	-0.107
		pct_ag3	work	-0.068	-0.24	-0.16
			maint	-0.047	-0.213	-0.142
			dsic	-0.047	-0.251	-0.167
	buffer_coefs	bkpath_len	rec_long	n/a	n/a	0.00015
			rec_fam	n/a	n/a	0.00015
			rec_other	n/a	n/a	0.00015
			work	n/a	n/a	0.00015
			sch_grade	n/a	n/a	0.00015
			sch_univ	n/a	n/a	0.00015
			maint	n/a	n/a	0.00015
			disc	n/a	n/a	0.00015
		households	rec_long	$8.58x10^{-7}$	$2.15x10^{-7}$	$4.27x10^{-7}$
		jobs6	rec_work	$-1.49x10^{-5}$	$-3.73x10^{-6}$	$-3.73x10^{-6}$
Destination choice	zone_coefs	jobs3	rec_other	1.0	0.25	0.5
			work	2.69	0.6725	0.6725
			sch_grade	0.31	0.0775	0.155
			disc	0.35	0.0875	0.175
		ldr_score	rec_long	1.0	0.25	0.5
			Total trips	220,390	143,642	139,334
			Origins RMSE	75.7	60.0	46.8
			Origins R^2	0.62	0.77	0.79
			Origin-destination flows R^2	0.66	0.70	0.85

Overall, regression analysis of model residuals yielded in modest performance improvements over the initially calibrated model (Table 8). While no formal targets exist for the calibration of a bike model to StreetLight data, the performance metrics for the calibrated model are reasonable and represent substantial improvement over the uncalibrated model.

3.3 SENSITIVITY ANALYSIS: TRIP ASSIGNMENT SUBMODEL

While performing sensitivity analysis for the trip generation and trip assignment sub models, coefficients in the bike generalized cost equation (used for skimming and assignment) were left unchanged. To explore the sensitivity of the model to changes in these parameters, the project team first grouped these parameters into 5 categories:

- Benefit of bike paths/bike boulevards (*bike_blvd* and *bike_path parameters*)
- Benefits of bike lanes (*bike_lane_medium_aadt* and *bike_lane_heavy_aadt*)
- Turn penalties (*turn*, *signal*, *left_or_straight_light_cross*, *left_or_straight_med_cross*, *left_or_straight_heavy_cross*, *right_heavy_cross*, *left_med_parallel*, and *left_heavy_parallel*)
- slope thresholds (*slope_levels*)
- AADT thresholds (*aadt_levels*, *aadt_cross*, and *aadt_parallel*)

Next, parameters in each group were varied in the *network.yml* file, both in isolation and in combination, and generated new skims with the resulting generalized cost function (Table 9).

TABLE 9. TRIP ASSIGNMENT SENSITIVITY RUN SPECIFICATIONS

SKIM VERSION	BIKE PATH/BIKE BOULEVARD	BIKE LANE	TURN PENALTIES	SLOPE	AADT
Base					
S1	Benefit Doubled				
S2		Benefit Doubled			
S3	Benefit Doubled	Benefit Doubled			
S4			Cost Doubled		
S5	Benefit Doubled		Cost Doubled		
S7	Benefit Doubled	Benefit Doubled	Cost Doubled		
S8	Benefit Doubled	Benefit Doubled	Cost Doubled	Thresholds halved	
S9	Benefit Doubled	Benefit Doubled	Cost Doubled		Thresholds halved
S10	Benefit Doubled	Benefit Doubled	Cost Doubled	Thresholds halved	Thresholds halved

The project team then re-ran the model once for each new skim using the model coefficients informed by calibration of the first two submodels (Table 9), and computed model performance metrics as before. Model performance metrics varied slightly when different skims were used, with skim S9 yielding the best fit relative to StreetLight data estimates (Table 10Table 10).

TABLE 10. TRIP ASSIGNMENT SENSITIVITY RUNS MODEL PERFORMANCE METRICS

SKIM VERSION	TOTAL TRIPS	ORIGINS RMSE	ORIGINS R^2	ODS R^2
S0	139,318	46.2	0.79	0.85
S1	139,294	46.5	0.79	0.86
S2	139,334	46.2	0.79	0.85
S3	139,314	46.2	0.79	0.85
S4	139,330	46.1	0.79	0.85
S5	139,289	46.1	0.79	0.86
S7	139,299	45.9	0.79	0.86
S8	139,285	46.0	0.79	0.86
S9	139,295	45.9	0.80	0.86
S10	139,291	46.4	0.79	0.86

In addition to differences in model performance metrics, model trip length distributions varied when different skims were used. As expected, the skim with the lowest generalized cost (S3) generated the highest average trip distance, e.g., 1.56 miles for work commute, whereas the skim with the highest generalized cost (S10) produced the lowest average trip distance, e.g., 1.19 miles for work commute (Table 11).

TABLE 11. TRIP ASSIGNMENT SENSITIVITY RUNS TRIP LENGTHS BY PURPOSE (MILES)

	TRIP PURPOSE	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Home-based	Long- distance rec	6.73	6.74	6.73	6.75	6.71	6.72	6.71	6.73	6.73	6.74	6.74
	Mountain bike rec	4.67	4.69	4.67	4.69	4.65	4.67	4.65	4.67	4.66	4.68	4.67
	Family rec	0.95	0.98	0.95	0.99	0.72	0.74	0.72	0.74	0.70	0.70	0.66
	Other rec	1.15	1.18	1.15	1.19	0.90	0.91	0.90	0.92	0.87	0.87	0.82
	Work	1.52	1.55	1.53	1.56	1.31	1.33	1.31	1.33	1.26	1.26	1.19
	Grade school	1.09	1.13	1.09	1.13	0.87	0.89	0.87	0.89	0.86	0.86	0.82
	University	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
	Maintenance	0.95	0.97	0.95	0.98	0.74	0.75	0.74	0.75	0.69	0.71	0.67
	Discretionary	0.89	0.92	0.90	0.92	0.71	0.72	0.71	0.72	0.68	0.68	0.64
Non-home-based	Long-distance rec	7.04	7.04	7.04	7.05	7.01	7.02	7.02	7.02	6.99	7.00	6.97
	Family rec	0.95	0.98	0.95	0.98	0.72	0.74	0.72	0.74	0.69	0.70	0.66
	Other rec	1.03	1.07	1.03	1.07	0.80	0.82	0.80	0.82	0.78	0.78	0.73
	Work	1.34	1.37	1.35	1.38	1.13	1.15	1.13	1.15	1.09	1.08	1.03
	Grade school	0.95	0.99	0.95	1.00	0.73	0.75	0.73	0.75	0.71	0.71	0.67
	University	2.21	2.23	2.21	2.23	2.07	2.07	2.07	2.07	2.11	2.11	2.14
	Maintenance	0.90	0.92	0.90	0.93	0.69	0.70	0.69	0.70	0.65	0.66	0.62
	Discretionary	0.87	0.89	0.87	0.90	0.68	0.69	0.68	0.69	0.65	0.65	0.61
All trips		**1.92**	**1.95**	**1.93**	**1.96**	**1.75**	**1.77**	**1.76**	**1.78**	**1.73**	**1.73**	**1.68**

Differences in trip length distributions were also notable, especially for specific trip purposes. For example, the trip length distribution for work trips is skewed to the left using the S10 skims, whereas the S3 skims generate longer work trips (Figure 9).

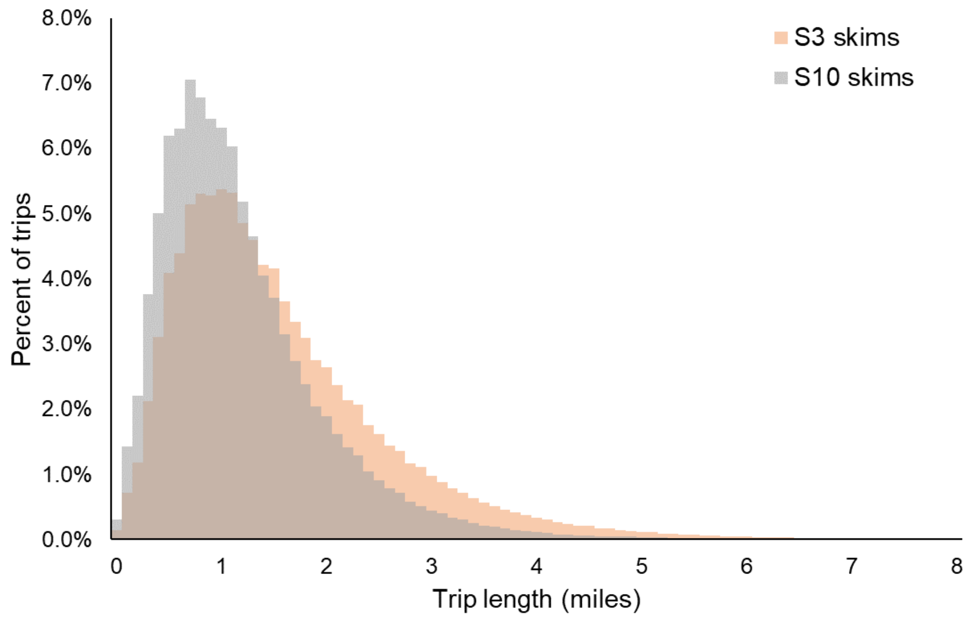

FIGURE 9. WORK TRIP LENGTH DISTRIBUTION USING S10 SKIMS (GREY) AND S3 SKIMS (ORANGE)

Overall, exploring the sensitivity of the model to parameters in the bicycle generalized cost function revealed two important findings. First, the destination choice model responds relatively strongly to changes in the generalized cost function, resulting in relatively large changes in the trip length distributions produced by the model. Second, the total number of trips produced by the model is much less sensitive to changes in the generalized cost function—i.e., the destination choice submodel simply assigns trips to closer zones in most cases. These two findings together are important, as changes to the generalized cost of travel may change how trips are assigned but not whether these trips will still occur. As a result, the sensitivity of the model to investments such as new bike facilities will be limited to trip assignment (routing) and destination choice versus trip generation. Thus, the model will likely miss important mode shifts that may occur with such investment. Future work could explore feedback mechanisms to increase this sensitivity (e.g., including accessibility measures in the trip generation sub-model).

4.0 FINAL CALIBRATED MODEL

The final calibrated model combines the model described in Table 8 with the highest-performing skim identified in section 3.3. The full specification and performance of this model are described below.

4.1 MODEL PARAMETERS

This final calibrated model includes parameters adjustments to the trip generation, destination choice, and trip assignment sub-models. Parameters adjustments to the trip generation and destination choice submodels are presented in Table 8; additional adjustments to trip assignment parameters are detailed below (Table 12).

TABLE 12. FINAL CALIBRATED MODEL PARAMETER ADJUSTMENTS

SUBMODEL	EQUATION	PARAMETER	TRIP PURPOSE	BASELINE MODEL	CALIBR-ATED MODEL
Assignment	aadt_levels	aadt_levels	light	10,000	5,000
			medium	20,000	10,000
			heavy	30,000	15,000
		aadt_cross	light	5,000	2,500
			medium	10,000	5,000
			heavy	20,000	10,000
		aadt_cross	light	5,000	2,500
			medium	10,000	5,000
			heavy	20,000	10,000
	network_coef	bike_commute	bike_blvd	-0.108	-0.216
			bike_path	-0.16	-0.32
			bike_lane_medium_aadt	0.25	0.125
			bike_lane_heavy_aadt	1.65	0.825
		bike_non_ commute	bike_blvd	-0.179	-0.358
			bike_path	-0.26	-0.52
			bike_lane_medium_aadt	0.6	0.25
			bike_lane_heavy_aadt	3.3	1.65
	fixed_costs	bike_commute	turn	0.034	0.068
			signal	0.017	0.034
			left_or_straight_light_cross	0.048	0.096
			left_or_straight_med_cross	0.05	0.10
			left_or_straight_heavy_cross	0.26	0.52
			right_heavy_cross	0.031	0.062
			left_med_parallel	0.073	0.146
			left_heavy_parallel	0.18	0.36
		bike_non_ commute	turn	0.074	0.148
			signal	0.033	0.066
			left_or_straight_light_cross	0.072	0.144
			left_or_straight_med_cross	0.10	0.20
			left_or_straight_heavy_cross	0.55	1.1
			right_heavy_cross	0.06	0.12
			left_med_parallel	0.15	0.3
			left_heavy_parallel	0.4	0.8

4.2 CALIBRATED MODEL PERFORMANCE

The final calibrated model has notably improved performance relative to the baseline model across all performance measures. The baseline model dramatically over-predicted trips south and west of downtown Salt Lake City, while the calibrated model is much closer to StreetLight estimates in these neighborhoods. The calibrated model still tends to under-predict trips in StreetLight outlier zones (Figure 10).

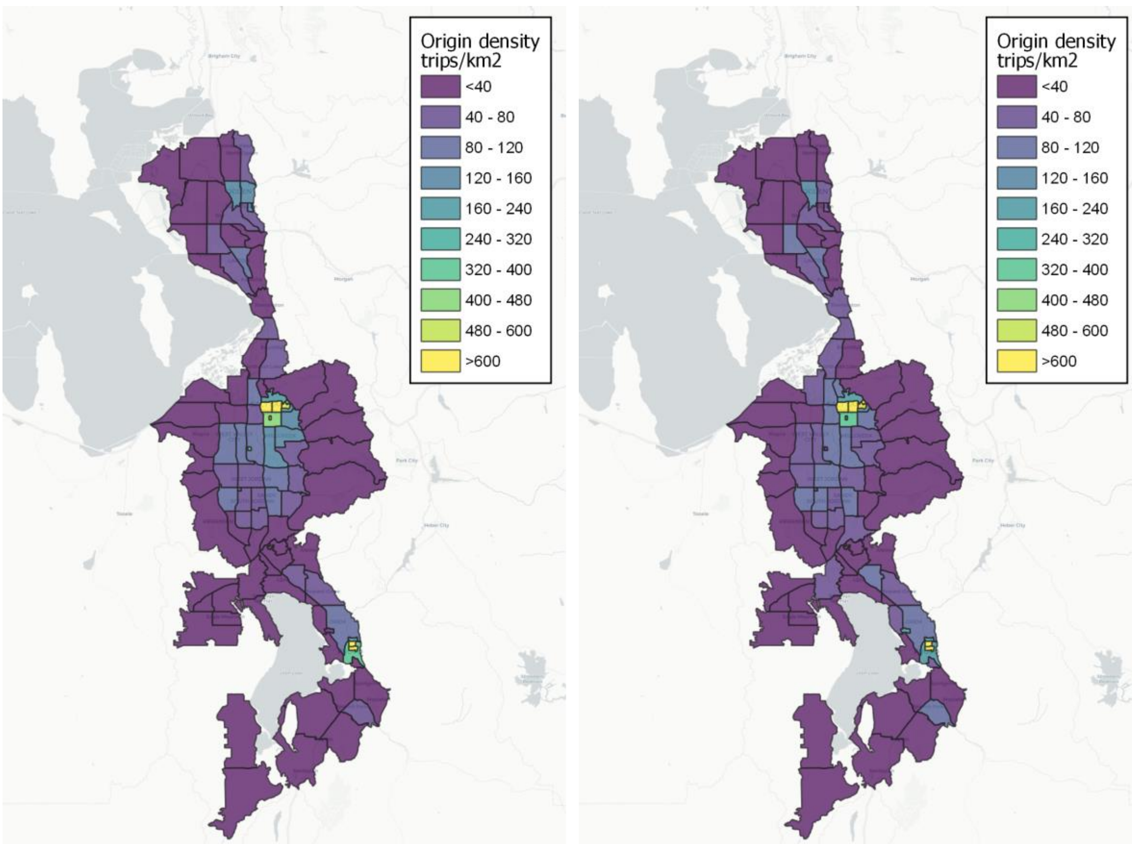

FIGURE 10. CALIBRATED MODEL ORIGIN DENSITY (LEFT) AND STREETLIGHT 2021 BIKE MONTHS ORIGIN DENSITY (RIGHT)

Visualizing model residuals for the calibrated and baseline models highlight how the calibration process addressed spatial bias in the baseline model. Over-predictions south of downtown Salt Lake City and north of Provo were largely addressed through calibration (Figure 11). Over-prediction in the foothills east of Ogden was substantially reduced as well. However, the calibrated model still tends to over-predict bicycle trips in a handle of zones southeast of downtown Salt Lake City. Finally, the only zones where the calibrated model substantially underpredicts bicycle trips relative to StreetLight are downtown Salt Lake City and the airport—two of the outlier zones identified previously. Overall, 45% of zones in the baseline model were within +/- 500 trips of the StreetLight estimates and 64% were within +/- 1,000 trips. For the calibrated mode, these percentages increase to 66% and 85%, respectively.

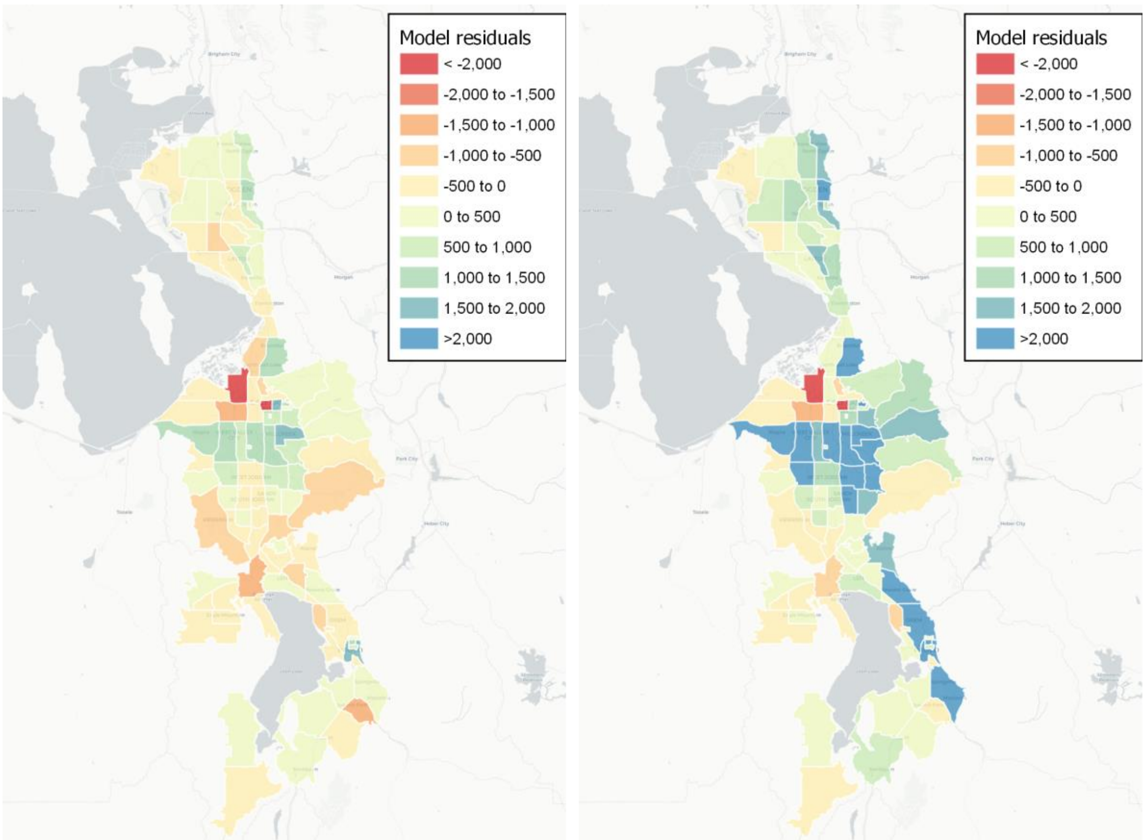

FIGURE 11. CALIBRATED MODEL RESIDUALS (LEFT) AND BASELINE MODEL RESIDUALS (RIGHT)

Overall, model fit improves substantially because of calibration, though largely a result of calibrating the trip generation and destination choice submodels (Table 13Table 13). Further calibration of the assignment submodel yields only marginal improvement. Most importantly, the baseline model over-estimated trips by more than 60%; conversely, the calibrated model predicts only ~2% more trips than are present in the StreetLight data.

TABLE 13. MODEL CALIBRATION PERFORMANCE SUMMARY

	STREETLIGHT 2021 BIKE MONTHS	BASELINE MODEL	CALIBRATED MODEL (TRIP GEN, DISTRIBUTION)	CALIBRATED MODEL (TRIP GEN, DISTRIBUTION, AND ASSIGNMENT)
Trip count	137,769	220,390	139,334	139,297
Average trip length (mi)	3.07	2.30	1.92	1.73
Origins R^2	-	0.63	0.79	0.80
OD flows R^2	-	0.70	0.85	0.86
Origins RMSE	-	75.2	46.2	45.9

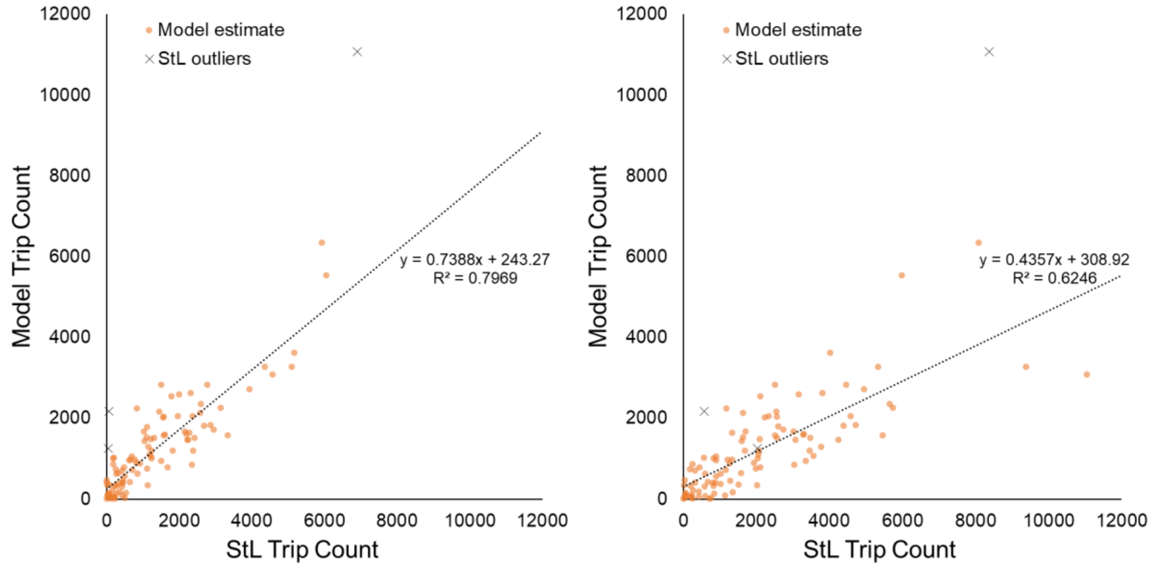

FIGURE 12. CALIBRATED MODEL ZONAL ORIGIN COUNTS VS STREETLIGHT DATA (LEFT) AND BASELINE MODEL ZONAL ORIGIN COUNTS VS STREETLIGHT DATA (RIGHT)

Finally, the calibrated model resulted in a slightly smoother trip length distribution than the baseline model. This is especially noticeable in the distribution of long recreational trips (greater than 5 miles) which is particularly spiky in the baseline model. This is likely a result of reducing model coefficients in the trip generation and destination choice sub models that were generating high numbers of trips between specific zones in the baseline model that were better scaled in the calibrated model (Figure 13).

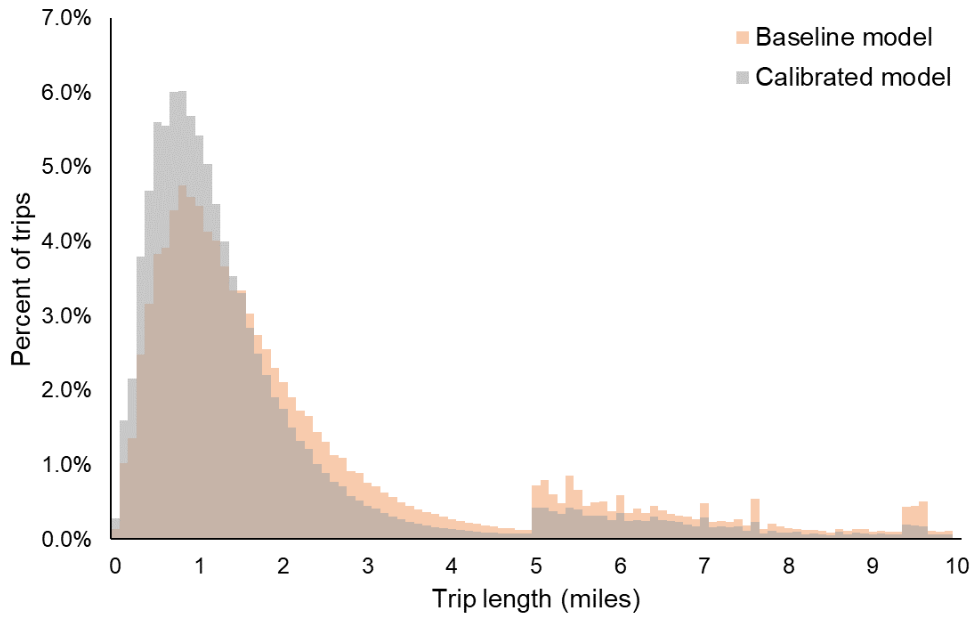

FIGURE 13. CALIBRATED MODEL TRIP LENGTH DISTRIBUTION (GREY) AND BASELINE MODEL TRIP LENGTH DISTRIBUTION (ORANGE)

5.0 STREETLIGHT MIDDLE FILTER COUNT COMPARISONS

5.1 MIDDLE FILTERS

Assigned network trips were compared to the results of a StreetLight middle filter query performed for the region. This middle filter analysis contained 46 zones distributed across the region, with a higher concentration near downtown Salt Lake City and several zones located along major recreational trails in the region (Figure 14). For each of these zones, StreetLight provides total pass-through volume as well as pass-through trip by trip OD combinations.

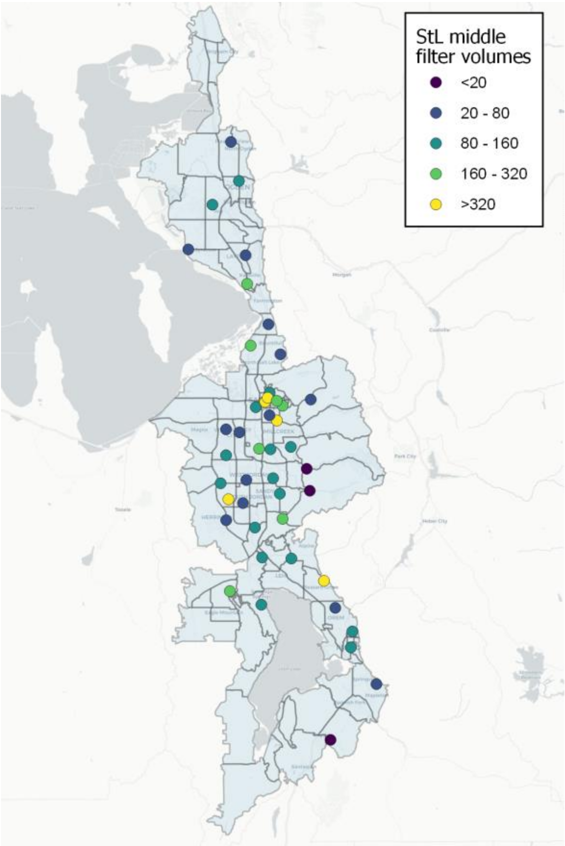

FIGURE 14. STREETLIGHT MIDDLE FILTER VOLUMES

Generally, StreetLight volumes are highest near downtown and lowest along recreational trails, in particular trails into the canyons on the east side of the region and the trail system to the north of the airport. Interestingly, there is one exception to low StreetLight volumes on recreational trails: the middle filter zone located on the Murdock Canal Trail north of Provo is among the five highest-volume zones in the region.

5.2 PERFORMING MIDDLE FILTER COMPARISON

Because the model is a trip-based model, disaggregate comparisons to these data are not possible. Instead, assigned trips volumes were compared to pass-through counts for each middle filter zone. The first step in comparing model assigned trips to StreetLight middle filter pass-through trip volumes was to join model network links to the middle filter zones. Because the zones are represented by 30-meter radius circles, the spatial join performed to link these two datasets resulted in several cases:

- Simple case (*n=19*): spatial join resulted in a one-to-one relationship, where one network link intersects one middle filter zone. In these cases, network volumes (in both directions) were compared directly to total pass-through volume.
- Parallel routes (*n=5*): spatial join resulted in a one-to-many relationship, where multiple parallel links intersect a middle filter zone. In these cases, network volumes (in both directions) were summed to compare to total pass-through volume. If any links began/ended within the middle filter zone, only one link along the same path was included to avoid double-counting.
- Three-way intersection (*n=4*): spatial join yielded three network links, forming a T-intersection. Any trip that enters the pass-through zone must exit on another link, so all trips will be double counted (unless the trip ends at the node in the center of the intersection). In these cases, network volumes were summed then divided by 2 to compare to pass-through volumes.
- Four-way intersection (*n=10*): spatial join yielded four network links, forming a 4-way intersection. As before, any trip that enters the pass-through zone must exit on another link, so the same logic was applied: network volumes were summed then divided by 2 to compare to pass-through volumes.
- Routes with internal nodes (*n=8*): spatial join resulted in multiple links representing the same route (i.e., links that start and/or end inside the middle filter zone). In these cases, each trip through the zone will be represented multiple times. In such cases, average link volume was used for comparison with the StreetLight middle filter data.

5.3 MIDDLE FILTER COMPARISON RESULTS

After joining middle filter zones to the model network, bicycle volumes were compared to StreetLight estimates. Overall, agreement between the two datasets was poor: while the StreetLight zones show the highest volumes near downtown and the lowest along recreational trails, the model tends to have the highest volumes along recreational trails and low volumes on many surface streets, including those near downtown (Figure 15).

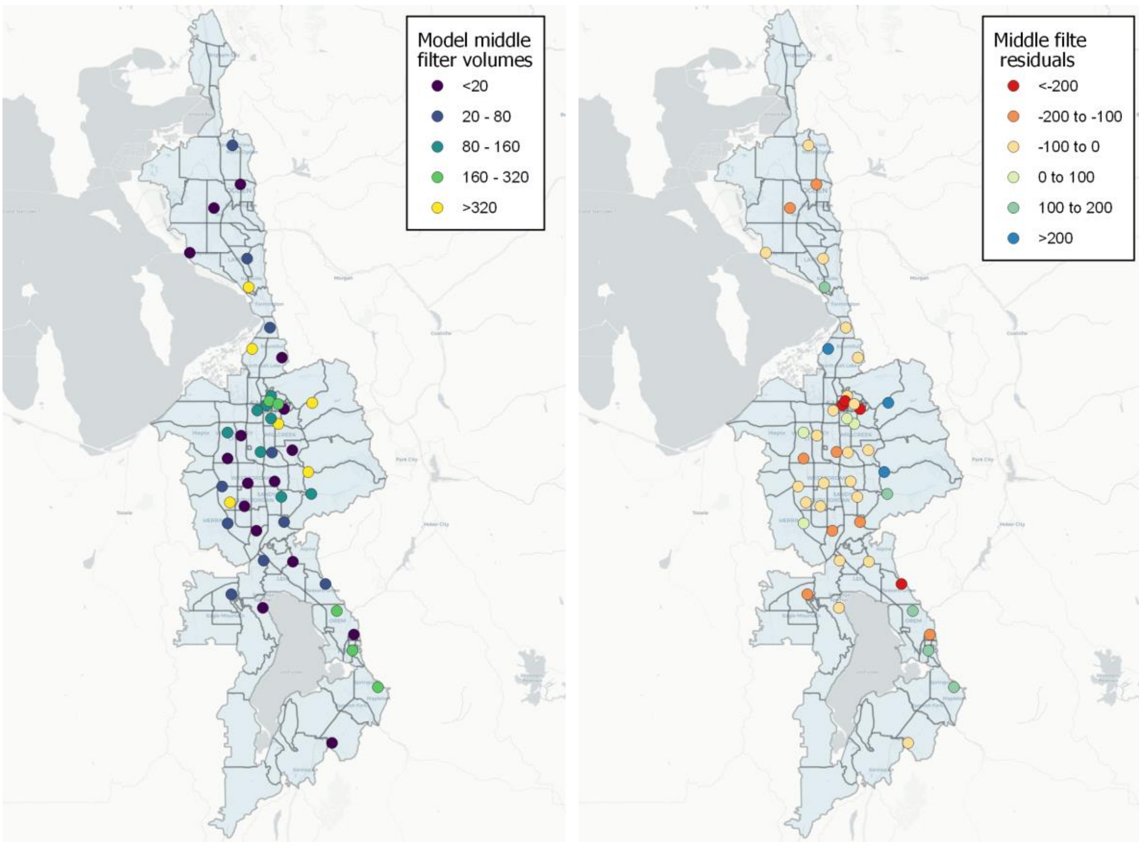

FIGURE 15. MODEL VOLUMES (LEFT) AND DIFFERENCE FROM STL ESTIMATES (RIGHT)

While the calibrated model is well-aligned with StreetLight OD data, agreement at the link-level is poor. However, it is unclear how reliable StreetLight middle filter estimates are for bikes. For example, low volumes along most trails in the region provide reason to be skeptical. Conversely, modeling bicycle route choice is also highly uncertain, and the model produces lower than expected volumes in certain cases (e.g., along the Murdock Canal Trail). Further, biases in the StreetLight data may contribute to this disagreement—namely, the high number of bicycle trips downtown and the tendency to underrepresent long recreational trips.

Overall, this middle filter comparison highlights the complexity of using StreetLight data to calibrate a bike model: uncertainty in the StreetLight estimates combined with uncertainty in bike trip assignment make it difficult to determine the why disagreements between the two datasets exists. This difficulty highlights the continued need for high-quality bicycle count data, both to use directly for model calibration and validation but also as a source of ground truth to perform reasonableness checks on imputed volume data. Finally, this difficulty underscores the need for high-quality local bike trip and route choice data from travel surveys to estimate models based on local observed behavior.

6.0 MODEL APPLICATIONS AND FUTURE DIRECTIONS

Calibration of the MAG/WFRC bike model using StreetLight bike OD data was largely successful in localizing the model, which was initially developed with model coefficients borrowed from Sacramento, CA. Additionally, the calibration process identified and fixed two important model implementation issues. Critically, this project also revealed a key limitation of the existing model: without an accessibility term in the trip generation submodel, the model does not predict additional bicycle trips that may be generated when new bicycle infrastructure is built. Nonetheless, the calibrated model provides MAG and WFRC staff with a clearer picture of current bicycle behaviors in the Salt Lake City region, and can provide decision support in several key areas outlined below.

6.1 PRACTICAL APPLICATIONS OF THE BIKE MODEL

By providing better understanding of how and where cyclists are traveling in the Salt Lake City region, the calibrated bike model provides decision support in important areas like safety and equity.

- **Safety.** The calibrated model provides a more accurate picture of trip origins and destinations and relative volumes on the bicycle network across the region. These volumes can support safety analysis by providing a better measure of exposure—i.e., where people are cycling today—to identify high-risk locations in the bicycle network more proactively.
- **Equity.** Because the calibrated model represents existing bicycle demand, it can be a useful tool in equity analysis focused on bicycle infrastructure. For example, the model could help MAG and WFRC staff understand if there are OD pairs that are underserved by existing bicycle infrastructure in the community and where there may be key gaps in the bicycle network.

It is also important to identify applications which the model does not currently support. Critically, the model does not include an accessibility term in the trip generation submodel, meaning that the model will not estimate additional bike trips given investments in bicycle infrastructure. This limits the ability of the model to provide decision support in decision processes such as project prioritization where estimates of new bike trips generated by investments may be important. However, the model is well-positioned to incorporate new features to increase such sensitivities in future work.

6.2 FUTURE WORK

The application of StreetLight bike OD data to calibrate the MAG/WFRC bike model has a handful of important limitations. Frist, StreetLight estimates tend to over-estimate biking in places with high transit use and under-estimate long recreational trips. Second, limitations of the model framework itself limit the ability of the model to capture some phenomena that may be present in observed data (e.g., the importance of accessibility in generating bicycle trips). Finally, uncertainty in the future of location-based services data and products derived from these data (such as StreetLight) given recent privacy concerns related to these data.

The upcoming Utah statewide travel survey provides a timely opportunity to address many of the issues noted above. Assuming a sufficient sample of bicycle trips is obtained via this survey, two activities will address limitations of model calibration using StreetLight data and broaden the potential applications of the model:

- **Re-estimate model using local data.** The StreetLight data used for this model calibration has several important limitations. Further, when Sacramento data were used to estimate trip generation and trip distribution submodels, accessibility measures were not included due to data limitations related to the bicycle network in Sacramento. Re-estimating the model using results from the Utah statewide household travel survey will provide a more localized behavioral basis for the model and explore relationships between bicycle trip-making and accessibility, providing information needed to incorporate accessibility measures into the trip generation submodel.
- **Incorporate an accessibility measure in the trip generation submodel.** An important use case of the bike model is understanding how residents respond to changes in bicycle infrastructure. A common approach to incorporate such sensitivities into travel models is to include an accessibility term in the trip generation mode, such as logsums from the destination choice model. Concurrent with re-estimating the model using local data, an accessibility measure should be incorporated into the trip generation model to support broader use cases, such as prioritizing investments in bicycle infrastructure.

APPENDIX A. LIST OF VARIABLES IN WFRC MICROZONE DATABASE

Households: Number of households (residential units that are occupied)
Residential: Number of residential units
Population: Total population
jobs1: Number of Accommodation, Food Services Jobs
jobs3: Number of Government and Education jobs
jobs4: Number of Health Care jobs
jobs5: Number of Manufacturing jobs
jobs6: Number of Office jobs
jobs7: Number of Other Jobs (non-typical commuting/travel patterns)
jobs9: Number of Retail Trade jobs
job10: Number of Wholesale, transport jobs
jobsT: Total number of jobs
ENROL_ELEM: Elementary school enrollment / population
ENROL_MIDL: Middle school enrollment / population
ENROL_HIGH: High school enrollment / population
ENROL_UNIV: University enrollment / population
POP_LC1: Total Population LC1 (households with no children and seniors)
POP_LC2: Total Population LC2 (households with children and no seniors)
POP_LC3: Total Population LC3 (households with seniors and may have children)
HHSIZE_LC1: Mean household size LC1 (households with no children and seniors)
HHSIZE_LC2: Mean household size LC2 (households with children and no seniors)
HHSIZE_LC3:Mean household size LC3 (households with seniors and may have children)
PCT_INC1:Household Percentage INC1 (Under $35,000)
PCT_INC2:Household Percentage INC2 ($35,000-$49,999)
PCT_INC3:Household Percentage INC3 ($50,000-$99,999)
PCT_INC4:Household Percentage INC4 ($100,000 or more)
PCT_AG1:Population Percentage AG1 (Children - 0 to 17)
PCT_AG2:Population Percentage AG2 (Adults - 18 to 64)
PCT_AG3:Population Percentage AG3 (Seniors - 65 +)
PARK_SCORE: Presence of desirable park spaces. 1) Acreage > 10, 2) 5 < Acreage < 10, 3) Acreage < 5
TRAIL_HEAD: Presence of a trailhead or other common ride starting point 1) yes, 0) no
LIGHT_RAIL: Presence of a light rail station 1) yes, 0) no
COMM_RAIL: Presence of commuter rail station 1) yes, 0) no