ACTIVITY-BASED MODEL AND COMMERCIAL VEHICLE MODEL VALIDATION REPORT

RSG
the science of insight

10.31.2016


PREPARED FOR:
SAN DIEGO ASSOCIATION OF GOVERNMENTS

SUBMITTED BY:
RSG

600 B Street, Suite 2202
San Diego, CA 92101
619.501.0559
www.rsginc.com

IN COOPERATION WITH:
VRPA
KEITH LAWTON

CONTENTS

List of Figures

List of Tables

List of Equations

1.0 INTRODUCTION

The San Diego Association of Governments (SANDAG) activity-based (AB) model was completed in 2013 and has been used in the development of the 2050 Regional Plan. The AB model contains components for San Diego resident travel; Mexican resident travel in San Diego County, and across the US/Mexico border; air passenger travel; visitor travel; and internal-external, external-external, and commercial vehicle travel (PB Americas, Inc. 2009). The commercial vehicle model consists of an aggregate Heavy-Duty Truck Model (HDTM) with parameters that were estimated based upon Freight Analysis Framework (FAF) data and intraregional parameters borrowed from the Los Angeles region; the commercial vehicle model also includes an aggregate nonfreight commercial vehicle model. The person travel components were calibrated to local survey data, while the nonfreight commercial vehicle models parameters were asserted and then scaled to match system-wide vehicle miles traveled (VMT) estimates (PB Americas, Inc. 2012).

SANDAG completed development of a new, disaggregate commercial travel model (CTM) in 2014 (HBA Specto Incorporated 2014). This model was based upon a local commercial vehicle survey and replaces the aggregate intraregional HDTM and nonfreight commercial vehicle components of the original aggregate commercial vehicle model. The internal/external component of the HDTM was retained in the new model system, but was updated to FAF 3 data. Validation on the CTM was independently performed of the AB model personal travel results; a full model validation was not completed.

This report compares two versions of the SANDAG model – one with the original aggregate commercial vehicle model and one with the disaggregate commercial vehicle model - using independent observed data, including traffic counts, trip tables developed from cellular phone data, Global Positioning System (GPS)-based travel speed data, and other sources. This analysis will help to determine whether the new model system is suitable for use in regional forecasting and—if it is not suitable—make model system improvement recommendations to better match observed data prior to the model's use.

This report's authors have assumed that each model has been properly calibrated to the data upon which it was developed. For this reason, this report does not replicate the calibration summaries reported in the original calibration reports. Instead, this report's analysis has relied on the following independent datasets:

- Traffic counts (2012) from the following sources:
    - Caltrans Performance Measurement System (PeMS) traffic counts, collected via automated sensors spanning freeways in San Diego County, and short-term and quarterly counts at specific locations. A portion of the counts are vehicle classification counts (by axle); weight-based counts are also available at a limited number of weigh-in-motion (WIM) stations.
    - Arterial traffic counts from the 18 cities and county in SANDAG's region, and arterial classification counts from the City of San Diego. SANDAG Vehicle Occupancy and Classification Study data collected in 2012.

- Observed trip tables based upon cellular phone data provided by AirSage (2014).
- American Transportation Research Institute (ATRI) GPS data, a sample of large-truck-route data in San Diego County (2012).
- Observed travel speed data—provided by INRIX—covering all freeways and approximately 50% of arterials in San Diego County (October 2012).
- Statistics compiled from transport surveys performed in San Diego and other regions throughout the United States, but not utilized in model development, including:
  - Average tour and trip rates and average trip distances from the National Household Travel Survey (NHTS) conducted in 2009 and the California Household Travel Survey conducted in 2012;
  - Federal Highway Administration (FHWA) research on commercial vehicle trip rates and trip lengths based on data from 13 regions across the United States in 2002 (Cambridge Systematics, Inc. 2004).

This report compares both model systems to traffic counts; however, for efficiency's sake, only the new model system (with the disaggregate commercial vehicle model) is compared to AirSage, INRIX, ATRI, and independent survey data since this is SANDAG's preferred model system for moving forward.

This report comprises the following sections:

- Overview of the person-based and commercial vehicle model systems that were analyzed.
- Description of model systems' validation.
- Comparison of resident travel models to independent household travel survey data.
- Comparison of resident travel models to AirSage data.
- Comparison of disaggregate commercial vehicle model to ATRI data.
- Comparison of the assigned travel model speeds to INRIX data.
- Conclusions and Recommendations.

2.0 MODEL SYSTEM OVERVIEW

2.1 | PERSON-TRANSPORT MODELS

The SANDAG travel demand model system represents several travel markets, including typical daily resident travel, travel by Mexican residents, visitor travel, and internal-external travel, among others. Many of the models used to represent demand are simulation-based models (i.e., AB or tour-based model approaches). Other travel models are aggregate three- or four-step representations of travel. Table 1 lists the SANDAG travel markets along several key dimensions.

Table 1 lists two broad models and three specific models. Disaggregate models refer to models whose demand is generated via a stochastic simulation paradigm where discrete choices are made from a set of alternatives using Monte Carlo simulation. Both AB and tour-based models are simulation-based; in other words, these models use a synthetic population to generate travel and stochastic processes to choose alternatives and output disaggregate demand in the form of tour and trip lists. Disaggregate models represent all origins and destinations using a detailed geographic level referred to as the Master Geographic Reference Area (MGRA). MGRAs are smaller than Transportation Analysis Zones (TAZs); there are approximately 23,000 MGRAs in the SANDAG Series 13 land-use system, compared to approximately 5,000 TAZs.

Only the resident travel model is an AB model—all tours and activities are scheduled into available time windows across the entire day. This approach recognizes that a person can be in only one place at one time and a person's entire day is accounted for in the model. A tour-based model is used for other special travel markets. This approach does not attempt to model all travel throughout the day for each person; rather, once tours are generated they are modeled independently. For example, it is possible that a visitor generates two tours where the outbound and return time periods for each tour are scheduled consistently but the starting time of the second tour is before the return time of the first tour. A tour-based model does not attempt to schedule all travel into available time windows.

Aggregate models rely upon probability accumulation processes to produce travel demand and output trip tables. In the original model, the commercial vehicle model and certain external travel models are aggregate. In the revised model, internal commercial vehicle trips are disaggregate, and external commercial vehicle and person trips remain aggregate.

The disaggregate AB and tour-based models output trip lists with 30-minute temporal resolution, identifying the departure time period of each trip. Note that the four time periods from 3:00 a.m. to 4:59 a.m. are aggregated into one time-of-day alternative, and the six time periods from 12:00 a.m. to 2:59 a.m. are also aggregated, creating 40 explicit time periods. The implied duration of each out-of-home activity—found by subtracting the departure time from the previous activity from the departure time for the activity—implicitly includes the travel time to the activity.

After each model is run, the outputs are summarized into trip tables. All auto trip tables are collapsed by occupancy, toll\high-occupancy vehicle (HOV) lane eligibility and time period and assigned to the highway network using a multiclass equilibrium capacity-constrained assignment algorithm. Table 3 details person vehicle classes.

TABLE 1: TRAVEL MARKETS IN SANDAG TRAVEL DEMAND MODEL SYSTEM

SEGMENT	TRAVEL MARKET	DESCRIPTION	MODEL TYPE	TEMPORAL RESOLUTION	SPATIAL RESOLUTION	VEHICLE CLASSES
1	San Diego resident travel (internal)	Average weekday travel made by San Diego residents within San Diego County	Disaggregate AB	30-minute	MGRA	8 (classes 1 through 8 in Table 3)
2	San Diego resident travel (internal-external)	Average weekday travel by San Diego residents between San Diego County and other county/Mexico	Disaggregate tour-based	30-minute	Internal MGRA –external cordon TAZ	8 (classes 1 through 8 in Table 3)
3	Mexican resident travel (external-internal and internal-internal)	Average weekday travel by Mexican residents into, out of, and within San Diego County	Disaggregate tour-based	30-minute	Internal MGRA –external cordon TAZ	8 (classes 1 through 8 in Table 3)
4	Overnight visitor	Average weekday travel made by overnight visitors to San Diego County	Disaggregate tour-based	30-minute	MGRA	8 (classes 1 through 8 in Table 3)
5	Airport passenger	Average weekday travel made by air passengers and related trips such as taxis to/from airport	Disaggregate trip-based	30-minute	MGRA	8 (classes 1 through 8 in Table 3)

SEGMENT	TRAVEL MARKET	DESCRIPTION	MODEL TYPE	TEMPORAL RESOLUTION	SPATIAL RESOLUTION	VEHICLE CLASSES
6	Special events	Special event travel; not average but event-specific	Disaggregate trip-based	30-minute	MGRA	8 (classes 1 through 8 in Table 3)
7	External-external	Average weekday travel with neither origin nor destination in San Diego County	Aggregate trip-based	5 time periods	External cordon TAZ	2 (classes 1 and 2 in Table 3)
8	Other US-internal travel	Average weekday external-internal trips made by non-San Diego and non-Mexican residents	Aggregate trip-based	5 time periods	External cordon TAZ – Internal TAZ	2 (classes 1 and 2 in Table 3)

Aggregate models output demand in the five time periods, as shown in Table 2, which are also consistent with the periods used in highway and transit assignment. Trip tables from disaggregate model output are created by aggregating trip lists into these periods prior to assignment, and applying occupancy factors to convert person trips to vehicle trips.

TABLE 2: TIME PERIODS IN EQUILIBRIUM CAPACITY-RESTRAINT ASSIGNMENT

TIME PERIOD	DURATION
Early AM	3:00 a.m. to 5:59 a.m.
AM Peak	6:00 a.m. to 8:59 a.m.
Midday	9:00 a.m. to 3:29 p.m.
PM Peak	3:30 p.m. to 6:59 p.m.
Evening	7:00 p.m. to 2:59 a.m.

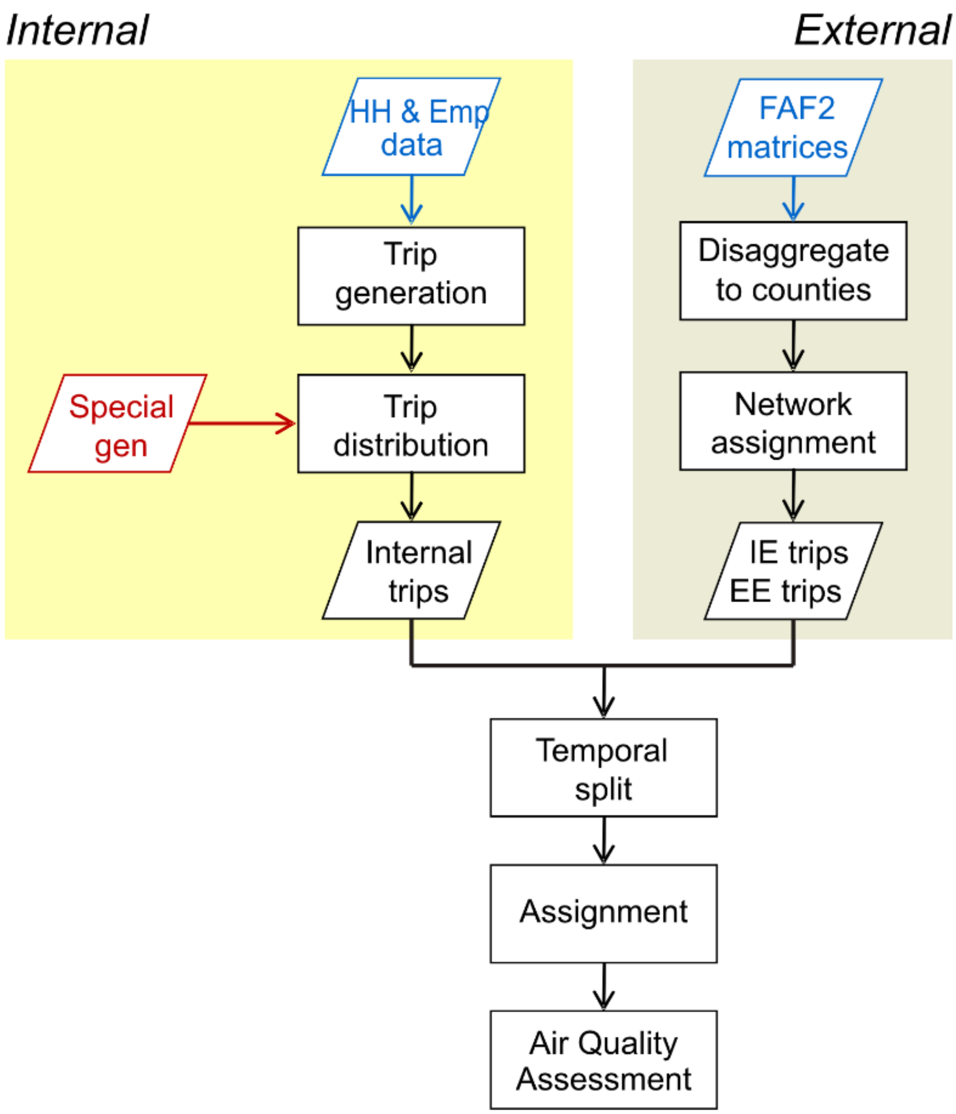
TABLE 3: VEHICLE CLASSES USED IN EQUILIBRIUM CAPACITY-RESTRAINT ASSIGNMENT

NUMBER	CLASS	DESCRIPTION	VALUE OF TIME	PASSENGER CAR EQUIVALENT
1	Drive-alone nontoll eligible	Single occupant vehicles not allowed to utilize toll facilities for their route choice	$30.00/hour	1.0
2	Drive-alone toll eligible	Single occupant vehicles allowed to utilize toll facilities for their route choice	$30.00/hour	1.0
3	Shared-2 nontoll, non-HOV eligible	Two occupant vehicles not allowed to utilize toll facilities or HOV lanes for their route choice	$30.00/hour	1.0
4	Shared-2 nontoll, HOV eligible	Two occupant vehicles not allowed to utilize toll facilities but allowed to utilize HOV lanes for their route choice	$30.00/hour	1.0
5	Shared-2 toll, HOV eligible	Two occupant vehicles allowed to utilize toll facilities and HOV lanes for their route choice	$30.00/hour	1.0
6	Shared 3+ nontoll, non-HOV eligible	Three or more occupant vehicles not allowed to utilize toll facilities or HOV lanes for their route choice	$30.00/hour	1.0
7	Shared 3+ nontoll, HOV eligible	Three or more occupant vehicles not allowed to utilize toll facilities but allowed to utilize HOV lanes for their route choice	$30.00/hour	1.0
8	Shared 3+ toll, HOV eligible	Three or more occupant vehicles allowed to utilize toll facilities and HOV lanes for their route choice	$30.00/hour	1.0

2.2 | AGGREGATE HEAVY-TRUCK MODEL

The initial aggregate heavy-truck model developed for SANDAG has two components: 1) an internal component; and 2) an external component. These components are illustrated in Figure 1. A multilayer model approach allows simulating internal trips and long-distance trips with increased granularity appropriate to each trip type.

FIGURE 1: HEAVY-TRUCK MODEL STRUCTURE



The internal truck model generates local truck trips and an external truck model simulates truck trips across the entire United States. The internal model creates truck trips that stay within San Diego County, or internal-internal (II). The external model generates flows between all counties in the United States. For San Diego County, the external model simulates trips that leave the county—or internal-external (IE). The external model also simulates trips that enter the county—or external-internal (EI)—and flows that go through the county, or external-external (EE). The latter are mostly truck trips that go into or originate in Mexico.

The internal model is a trip-based model with trip generation based on socioeconomic data, trip distribution, and a temporal split. A special generator module simulates trips from sources that cannot be captured by their employment only, such as cruise ship terminals or airports. The special generator module is based on a model developed by Cambridge Systematics, Inc. (Cambridge Systematics, Inc. 2008) for the Southern California Association of Governments (SCAG). Given the geographic proximity, truck-generation patterns were found to be comparable in San Diego County. Trucks are distinguished by three truck types defined by their weight class, as shown in Table 4. These truck types are compatible with vehicle types defined by Emfac2007, which is used by SANDAG to calculate vehicle emissions. An additional toll diversion component (not shown in Figure 1) was added to the model to split trips into toll-eligible and nontoll-eligible trucks prior to assignment.

TABLE 4: TRUCK WEIGHT CLASSES

TYPE	WEIGHT CLASS
Light-Heavy Duty (LHD)	8,500–14,000 lbs.
Medium-Heavy Duty (MHD)	14,000–33,000 lbs.
Heavy-Heavy Duty (HHD)	> 33,000 lbs.

The external model uses FAF3 data, disaggregates flows to counties, temporarily assigns these flows to the external network, and provides external trips by three time periods. Based on commodity-specific payload factors, FAF flows are transformed into annual truck trips. Based on assumptions about the empty-load rate and an average number of business days per year, daily truck trips are estimated between FAF regions. A national average of truck type distribution is used to distinguish three truck types defined for the internal truck model. Truck trips between FAF regions are further disaggregated to truck trips between counties. Truck flows are assigned to a US highway network. This assignment is only used to extract the number of trucks that enter or leave San Diego County at any given highway exit, including through trips and trips that have their origin or destination in San Diego County. Thus, the external model generates IE, EI, and EE trips. These trips are subjected to a toll diversion model and split into toll-eligible and nontoll eligible trip tables for assignment.

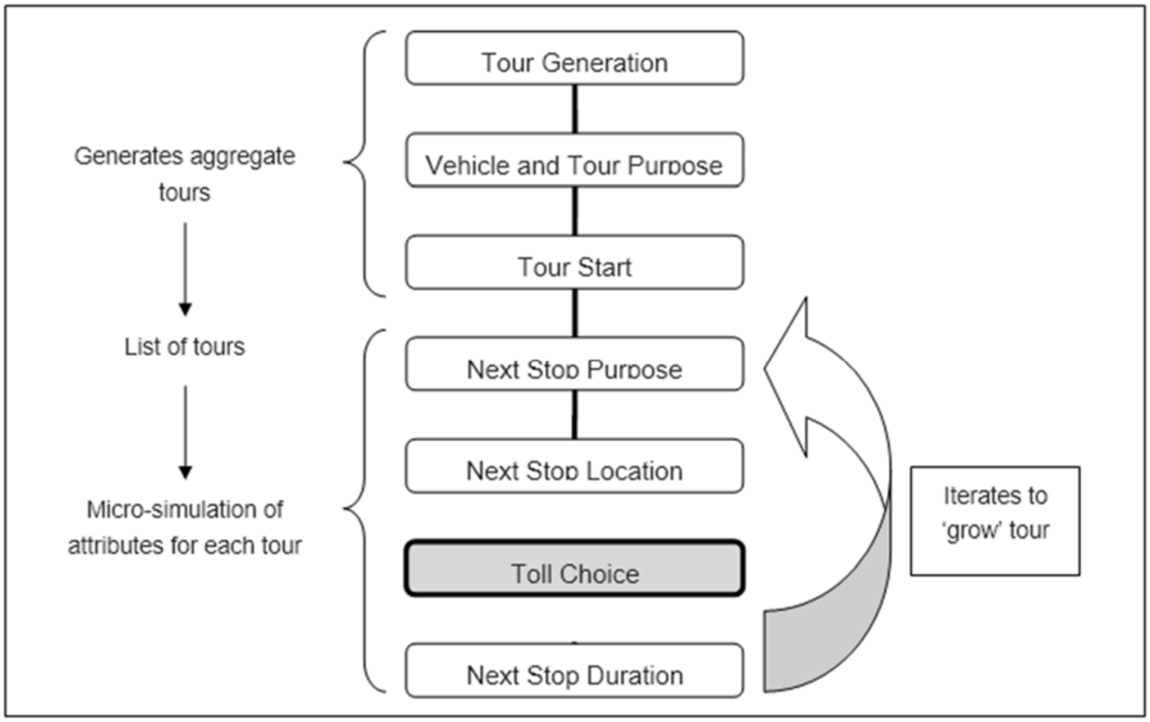
2.3 | AGGREGATE LIGHT COMMERCIAL VEHICLE MODEL

The aggregate light commercial vehicle model developed for SANDAG was initially borrowed from the Metropolitan Transportation Commission (MTC). The MTC model is a three-step quick-response model consisting of trip generation, trip distribution, and assignment. The trips in this model include service trips; sales calls; public services; construction-related travel; and other travel in passenger vehicles, light vans, trucks, and non-heavy-truck freight travel not covered in the HDTM (FedEx, other home deliveries, etc.). However, San Diego County lacked the requisite data for calibrating this model to light commercial vehicle counts. Therefore, commercial vehicle trip rates and lengths were iteratively adjusted during highway validation process to achieve a better match between estimated and observed total traffic counts for SANDAG's light commercial vehicle model.

2.4 | DISAGGREGATE COMMERCIAL VEHICLE MODEL

The disaggregate commercial vehicle model (CVM) is a tour-based model developed from work-related person and vehicle movement travel data, collected as part of the SANDAG Work-Related Travel Survey conducted between November 2012 and September 2013. A basic schematic of the models is shown in Figure 2. Tour generation quantities by vehicle type, tour purpose, and time of day are generated for each TAZ—using logit and regression equations applied with aggregate TAZ inputs and travel accessibilities—to create a list of tours.

FIGURE 2: DISAGGREGATE CVM STRUCTURE



Individual tours generated from each TAZ are then assigned a next-stop purpose, next-stop location, and next-stop duration using a microsimulation process. In this process, Monte Carlo techniques are used to incrementally "grow" a tour by having a "return-to-establishment" alternative within the next-stop purpose allocation. If the next-stop purpose is not "return-to-establishment," then the tour extends by one more stop. The location and duration of the next stop are then estimated. For each trip, a toll choice model determines whether a toll facility is used as part of the route-choice process. These steps are repeated until the "return-to-establishment" next-stop purpose is chosen.

Seven establishment types are considered, based on aggregations of NAICS categories:

- **Industrial (IN)**
  - NAICS 11, 21, 23, 31–33
- **Wholesale (WH)**
  - NAICS 42

- **Service (SE)**
  - NAICS 61, 62, 71, 72, 81
- **Government/Office (GO)**
  - NAICS 51, 52, 53, 54, 55, 56, 92
- **Retail (RE)**
  - NAICS 44, 45
- **Transport and Handling (TH)**
  - NAICS 22, 48, 49
- **Fleet Allocator (FA)**
  - All but military: a specific type of establishment that uses a large, coordinated fleet that tends to service an area rather than specific demands; examples include mail and courier, garbage hauling, newspaper delivery, utilities, and public works.

Four commercial vehicle types are used, defined in Table 5:

- **Light vehicle**
  - FHWA classes 1–3
- **Medium truck < 8.8 short tons (17,640 lbs.)**
  - FHWA classes 5, 6
- **Medium truck > 8.8 short tons (17,640 lbs.)**
  - FHWA classes 5, 6
- **Heavy truck**
  - FHWA classes 7–13

Medium trucks are collapsed and six trip tables (by three weight categories and two toll categories) are assigned to the auto network.

TABLE 5: CVM VEHICLE CLASSES BY FHWA CLASSIFICATION/CALTRANS COUNT PROGRAM

FHWA 13 CLASSES			CVM	CALTRANS
#	VEHICLE TYPE	AXLE RANGE		
1	Motorcycle	2	n/a	n/a
2	Passenger Car	2+	Light	n/a
3	Pickups, Vans, SUV	2+	Light	n/a
4	Buses	2+	n/a	n/a
5	2-axle, 6-tire, Single Unit Truck	2	Medium	2-axle
6	3-axle, Single Unit Truck	3	Medium	3-axle
7	4+ axle, Single Unit Truck	4+	Heavy	4-axle
8	4 or less axle, Single-Trailer truck	4 or less	Heavy	4-axle
9	5-axle, Single-Trailer truck	5	Heavy	5+ axle
10	6+ axle, Single-Trailer truck	6+	Heavy	5+ axle
11	5 or less axle, Multi-Trailer truck	5 or less	Heavy	5+ axle
12	6-axle, Multi-Trailer truck	6	Heavy	5+ axle
13	7+ axle, Multi-Trailer truck	7+	Heavy	5+ axle
14	5-axle, 3-axle tractor 2-axle trailer	5	Heavy	5+ axle

Note: Class 14 is Caltrans additional Class.

2.5 | AGGREGATE VERSUS DISAGGREGATE CVM

The model system used for the 2050 RTP includes all the person-transport models described in the Person-Transport Models section, the heavy-truck model described in the Aggregate Heavy-Truck Model section, and the *aggregate* light CVM described in the Aggregate Light Commercial Vehicle Model section. This model system is referred to as the "aggregate" model throughout this report.

The "disaggregate" model referred to throughout this document includes all of the person-transport models described in the Person-Transport Models section, the *disaggregate* CVM described in the Disaggregate Commercial Vehicle Model section, and only the EI and EE portions of the heavy-truck model described in the Aggregate Heavy-Truck Model section, since the disaggregate CVM covers all freight movement entirely within San Diego County. Through the course of this project, there were several disaggregate model runs, including the initial disaggregate model, a revised version that includes a slightly modified land-use input file; and another version that includes a slightly modified network and refined person-transport models. However, all versions of the disaggregate model result in nearly equivalent VMT and validation results; therefore, these versions are treated as the same disaggregate model here.

3.0 OVERALL MODEL VALIDATION

The following section quantifies the difference between estimated and observed traffic counts for each model system using a variety of statistical measures. Most traffic count data were obtained from Caltrans PeMS traffic count program, but this was supplemented by traffic counts obtained by the City of San Diego, other jurisdictions in San Diego County, and from traffic counts collected directly by SANDAG.

3.1 | TRAFFIC COUNT COMPARISON STATISTICS

Estimated versus observed traffic count comparisons utilize several statistical measurements suggested by FHWA in its Travel Model Validation and Reasonableness Checking Manual (Cambridge Systematics, Inc. 2010). Caltrans criteria for travel model validation for each of these statistics is shown in Table 6. The statistical measures are described below.

Volume-to-Count Ratio

The volume-to-count ratio is computed by dividing the volume assigned by the model (Flow [y]) by the actual traffic count (Observed [x]) for individual road segments across the model. This value provides a general context for the relationship (i.e., high or low) between the model estimated volumes and actual (observed) traffic counts.

EQUATION 1: VOLUME-TO-COUNT RATIO

$$\frac{Estimated\ (y)}{Observed\ (x)}$$

Volume-to-Count Deviation

The volume-to-count deviation is the difference between the model volume and the actual count, divided by the actual count. The Caltrans deviation thresholds recognize that allowances shrink as the count increases (i.e., lower tolerance for differences between the model volume estimates and counts).

EQUATION 2: VOLUME-TO-COUNT DEVIATION

$$\frac{Estimated\ (y) - Observed\ (x)}{Observed\ (x)}$$

Correlation Coefficient

The correlation coefficient (Pearson's product-moment correlation coefficient [R])—measured -1.0 to 1.0—estimates the correlation, or the strength and direction of the linear relationship, between the actual/observed counts and the estimated/predicted traffic volumes from the model.

EQUATION 3: CORRELATION COEFFICIENT

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

The coefficient of determination (R^2, or the square of R) is typically interpreted as the proportion of the variance in a dependent variable (y) attributable to the variance in an independent variable (x). For instance, 0% (zero) indicates that the model explains none of the variability around the mean, and 100% indicates that the model explains all the variability. R^2 can compare results between different model runs or iterations of the same model.

Percent Root Mean Square Error

Percent Root Mean Square Error (RMSE%) is the square root of the squared actual observed count minus the model's estimated (predicted) volume, divided by the number of counts. It is measured in a manner like Standard Deviation in that it assesses the assignment accuracy of the entire model. Lower percentages indicate better goodness of fit. The RMSE% is one of the most frequent methods of comparing different models to each other.

EQUATION 4: RMSE%

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left[(Count_i - Model_i)^2\right]}{N}}$$

and

$$\%RMSE = \frac{RMSE}{\left(\frac{\sum_{i=1}^{N} Count_i}{N}\right)} \times 100$$

TABLE 6: CALTRANS VALIDATION ACCEPTANCE CRITERIA

VALIDATION ITEM	CRITERIA FOR ACCEPTANCE
Estimated-to-observed deviation • Freeways • Principal arterials • Arterials • Collectors	< 25% overall • < 7% • < 10% • < 15% • < 25%
• Percent of links with estimated-to-observed deviations within Caltrans' deviation allowance	• At least 75%
• Correlation coefficient	• At least 0.88

VALIDATION ITEM	CRITERIA FOR ACCEPTANCE
RMSE% • > 60,000 daily volumes • 25,000–60,000 • 10,000–25,000 • 5,000–10,000 • 2,500–5,000 • 1,000–2,500 • < 1,000	Below 30–40% overall • < 21% • 21–22% • 22–25 % • 25–29% • 29–36% • 36–47% • 47–60% +

3.2 | ESTIMATED VERSUS OBSERVED AUTO VOLUMES (ALL VEHICLES)

The aggregate model (shown in Table 7) meets Caltrans model validation benchmarks. Flow ratios are acceptable for all facility types, though there is more error on HOV and toll lanes than across all other freeway segments. The correlation coefficient is also within Caltrans criteria. The aggregate model slightly over-estimates total traffic with a slope of 1.019. Freeway RMSE in the aggregate model is 13%, well within the acceptable target of 20%. Overall, the model is producing a RMSE of 23%. As expected, the lower road classes produce a higher percent RMSE. The aggregate model meets the Caltrans target correlation coefficient of greater than 0.88 for all link classes, except for HOV/toll, and the overall correlation coefficient is 0.98. Overall, 60% of the links fall within Caltrans' acceptable level of deviation. Freeways meet the deviation target while other facility types have more links outside the Caltrans deviation criteria.

Table 8 show the validation benchmarks for the disaggregate model. The model somewhat underestimates total traffic with a slope of 0.960. With the exception HOV/toll (1.187) class roads, the model underestimates traffic volumes for all other classes.

Freeway RMSE in the disaggregate model is 14%, well within the acceptable target of 20%; overall, the model is producing an RMSE of 25% (compared to the aggregate model's 23%). As expected, the lower road classes produce a higher percent RMSE. These results align with the aggregate model. However, the percent RMSE for principal arterials (88%) is substantially higher in this model than the aggregate CVM model (24%).

A correlation coefficient of 0.979 was achieved in the disaggregate model. All road classes (and the model overall) exceed the Caltrans correlation coefficient criteria of 0.88, except for HOV/toll. Overall, 58% of the links fall within Caltrans' acceptable level of deviation, which is slightly lower than the aggregate model (60%).

Scatterplots of modeled traffic volumes versus the observed traffic volumes are useful validation tools, especially when combined with R^2 summaries. The scatterplots in Figure 3 through Figure 6 provide a visual interpretation of the results included in Table 7 and Table 8.

TABLE 7: AGGREGATE MODEL ASSIGNMENT VALIDATION—MODEL ESTIMATED VOLUMES VS. OBSERVED COUNT SUMMARY BY ROAD CLASS

GAP RANGE	NUMBER OF LINKS WITHIN GAP RANGE BY LINK CLASSES							PERCENT OF LINKS WITHIN GAP RANGE BY LINK CLASSES						
	FREEWAY	RAMP	HOV/TOLL	PRINCIPAL ARTERIAL	ARTERIAL	COLLECTOR	ALL	FREEWAY	RAMP	HOV/TOLL	PRINCIPAL ARTERIAL	ARTERIAL	COLLECTOR	ALL
>=100%	1	18	2	0	9	24	54	0%	4%	13%	0%	2%	4%	2%
50%~100%	10	33	3	5	22	33	106	2%	7%	20%	5%	4%	5%	4%
30%~50%	22	37	1	6	23	32	121	3%	8%	7%	6%	4%	5%	5%
20%~30%	47	30	3	5	18	29	132	7%	6%	20%	5%	3%	5%	5%
10%~20%	122	27	2	10	53	42	256	19%	5%	13%	9%	10%	7%	11%
0%~10%	213	51	1	18	76	64	423	34%	10%	7%	17%	14%	10%	18%
0%~-10%	158	57	0	20	66	59	360	25%	12%	0%	19%	12%	9%	15%
-10%~-20%	39	57	0	10	70	60	236	6%	12%	0%	9%	13%	9%	10%
-20%~-30%	11	68	0	15	55	61	210	2%	14%	0%	14%	10%	10%	9%
-30%~-50%	6	79	1	13	103	99	301	1%	16%	7%	12%	19%	15%	12%
<-50%	2	36	2	5	34	137	216	0%	7%	13%	5%	6%	21%	9%
Total	**631**	**493**	**15**	**107**	**529**	**640**	**2415**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
-10%~+10%	371	108	1	38	142	123	783	59%	22%	7%	36%	27%	19%	32%
-20%~+20%	532	192	3	58	265	225	1275	84%	39%	20%	54%	50%	35%	53%
-30%~+30%	590	290	6	78	338	315	1617	94%	59%	40%	73%	64%	49%	67%
AVERAGE OF GAPS								**TOTAL SHARE OF POSITIVE/NEGATIVE LINKS**						
Average (+) Gaps	13%	48%	52%	20%	26%	68%	34%	66%	40%	80%	41%	38%	35%	45%
Average (-) Gaps	-8%	-28%	-69%	-22%	-27%	-38%	-28%	34%	60%	20%	59%	62%	65%	55%
Average of All	6%	2%	28%	-5%	-7%	-7%	0%	100%	100%	100%	100%	100%	100%	100%
RMSE														
RMSE Percent	13%	42%	70%	24%	32%	58%	23%							
TREND LINE SLOPE														
Slope	1.038	0.921	1.232	0.933	0.876	0.663	1.019							
R-SQUARED														
R_2	0.899	0.642	0.351	0.748	0.664	0.433	0.960							
CORRELATION COEFFICIENT														
Correlation Coef.	0.948	0.802	0.592	0.865	0.815	0.658	0.980							
% WITHIN CALTRANS ACCEPTABLE DEVIATION														
Total Valid Count	645	493	12	43	528	639	2412							
Link Within Deviation	484	295	8	23	266	357	1458							
Link Outside Deviation	161	198	4	20	262	282	954							
Acceptable Link %	75%	60%	67%	53%	50%	56%	60%							

Source: 160914 highway_assignment_Aggregate_CVM_Wildcat_BASE_run.xlsx, tab: summary_table_All Classes.

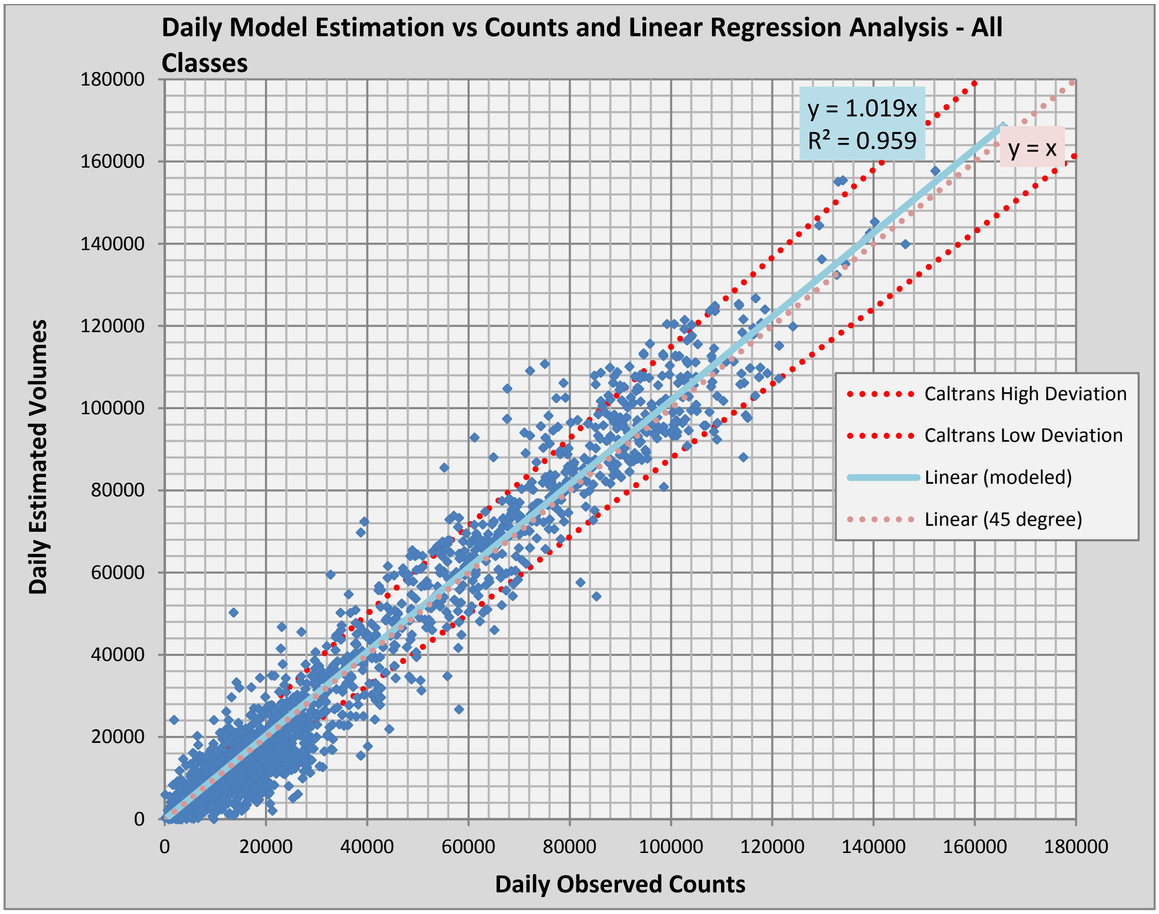
TABLE 8: DISAGGREGATE MODEL ASSIGNMENT VALIDATION—MODEL ESTIMATED VOLUMES VS. OBSERVED COUNT SUMMARY BY ROAD CLASS

GAP RANGE	NUMBER OF LINKS WITHIN GAP RANGE BY LINK CLASSES							PERCENT OF LINKS WITHIN GAP RANGE BY LINK CLASSES						
	FREEWAY	RAMP	HOV/TOLL	PRINCIPAL ARTERIAL	ARTERIAL	COLLECTOR	ALL	FREEWAY	RAMP	HOV/TOLL	PRINCIPAL ARTERIAL	ARTERIAL	COLLECTOR	ALL
>=100%	1	12	1	0	5	20	39	0%	2%	8%	0%	1%	3%	2%
50%~100%	5	18	3	1	16	26	69	1%	4%	25%	1%	3%	4%	3%
30%~50%	6	29	0	5	9	19	68	1%	6%	0%	5%	2%	3%	3%
20%~30%	15	21	3	2	13	20	74	2%	4%	25%	2%	2%	3%	3%
10%~20%	40	21	2	2	15	35	115	6%	4%	17%	2%	3%	5%	5%
0%~10%	114	40	1	6	31	40	232	18%	8%	8%	6%	6%	6%	10%
0%~-10%	232	42	1	20	78	71	444	37%	9%	8%	19%	15%	11%	18%
-10%~-20%	159	61	0	23	85	70	398	25%	12%	0%	22%	16%	11%	16%
-20%~-30%	49	81	0	15	84	69	298	8%	16%	0%	14%	16%	11%	12%
-30%~-50%	9	119	1	27	130	115	401	1%	24%	8%	25%	25%	18%	17%
<-50%	4	49	0	5	62	157	277	1%	10%	0%	5%	12%	24%	11%
Total	**634**	**493**	**12**	**106**	**528**	**642**	**2415**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
-10% ~ +10%	346	82	2	26	109	111	676	55%	17%	17%	25%	21%	17%	28%
-20% ~ +20%	545	164	4	51	209	216	1189	86%	33%	33%	48%	40%	34%	49%
-30% ~ +30%	609	266	7	68	306	305	1561	96%	54%	58%	64%	58%	48%	65%
AVERAGE OF GAPS								**TOTAL SHARE OF POSITIVE/NEGATIVE LINKS**						
Average (+) Gaps	11%	47%	46%	22%	32%	79%	42%	29%	29%	83%	15%	17%	25%	25%
Average (-) Gaps	-12%	-32%	-20%	-24%	-29%	-38%	-27%	71%	71%	17%	85%	83%	75%	75%
Average of All	-5%	-9%	35%	-17%	-19%	-19%	-10%	100%	100%	100%	100%	100%	100%	100%
RMSE														
RMSE percent	14%	42%	56%	88%	36%	58%	25%							
TREND LINE SLOPE														
Slope	0.930	0.813	1.187	0.797	0.762	0.663	0.910							
R-SQUARED														
R_2	0.890	0.649	0.299	0.730	0.678	0.443	0.958							
CORRELATION COEFFICIENT														
Correlation Coef.	0.943	0.806	0.547	0.855	0.823	0.666	0.979							
% WITHIN CALTRANS ACCEPTABLE DEVIATION														
Total Valid Count	634	493	12	106	528	639	2412							
Link Within Deviation	475	295	9	24	266	330	1399							
Link Outside Deviation	159	198	3	82	262	309	1013							
Acceptable Link %	75%	60%	75%	23%	50%	52%	58%							

Source: 160914 highway_assignment_Disaggregate_CVM_CTMInput_run.xlsx, tab: summary_table_All Classes.

Figure 3 shows a scatterplot of estimated versus observed link volumes for the aggregate model. The solid blue linear regression line shows that the model is estimating only slightly higher than the available counts overall. There are a few points well outside the Caltrans low- and high-deviation lines that could be considered outliers and should be investigated. However, an increasing percentage of points tend to fall within those lines as the counted volume increases, which is expected.

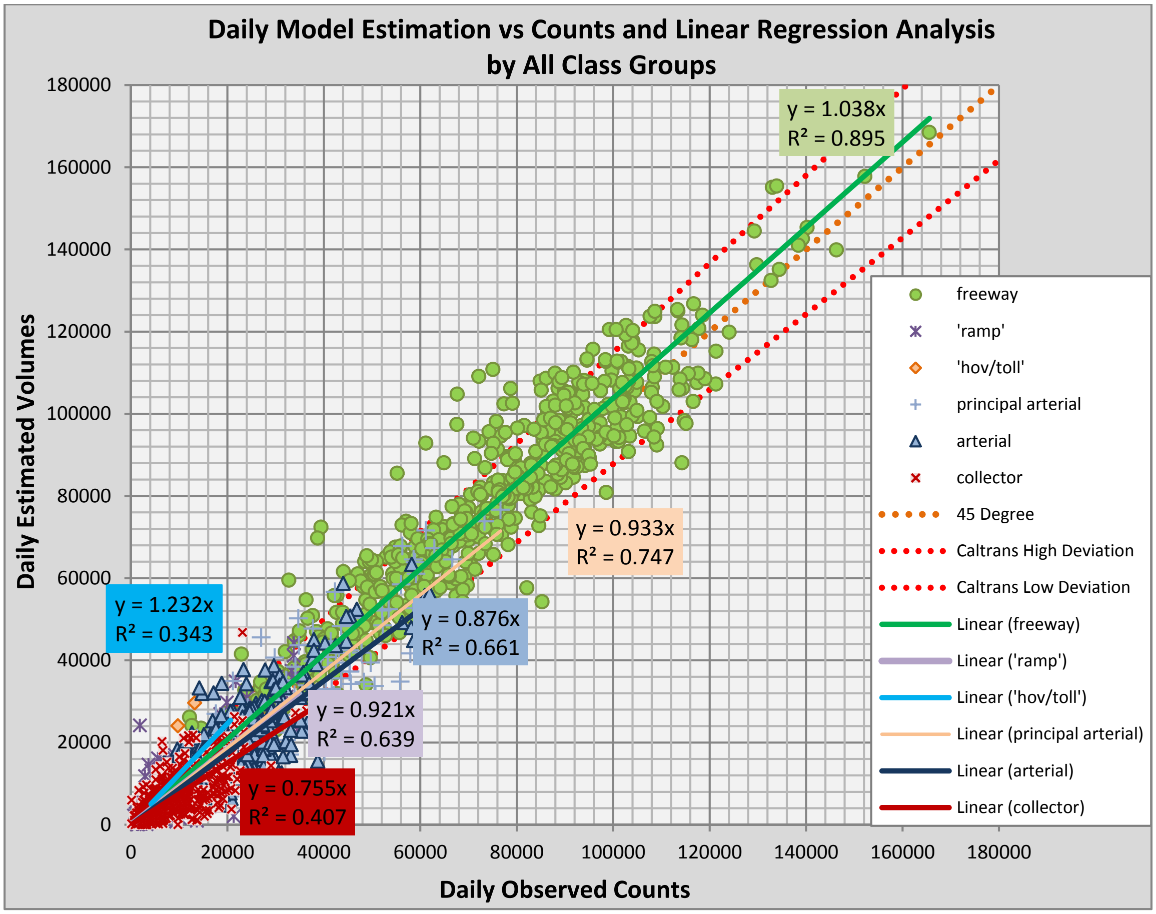
FIGURE 3: AGGREGATE MODEL ESTIMATED VS. OBSERVED VOLUME SCATTERPLOT, ALL LINK CLASSES COMBINED



Source: 160914 highway_assignment_Aggregate_CVM_Wildcat_BASE_run.xlsx, tab: scatterplot_plot_all_points

Figure 4 further examines the aggregate model by separating the points in Figure 3 by link class. The model slightly over-estimates freeways and underestimates lower-volume link classes such as arterials and collectors, and the variability tends to increase inversely with respect to volume. Although the model over-estimates HOV/toll link class (1.232 slope), the model has a relatively low ability (R^2=.343) to explain the volume variability.

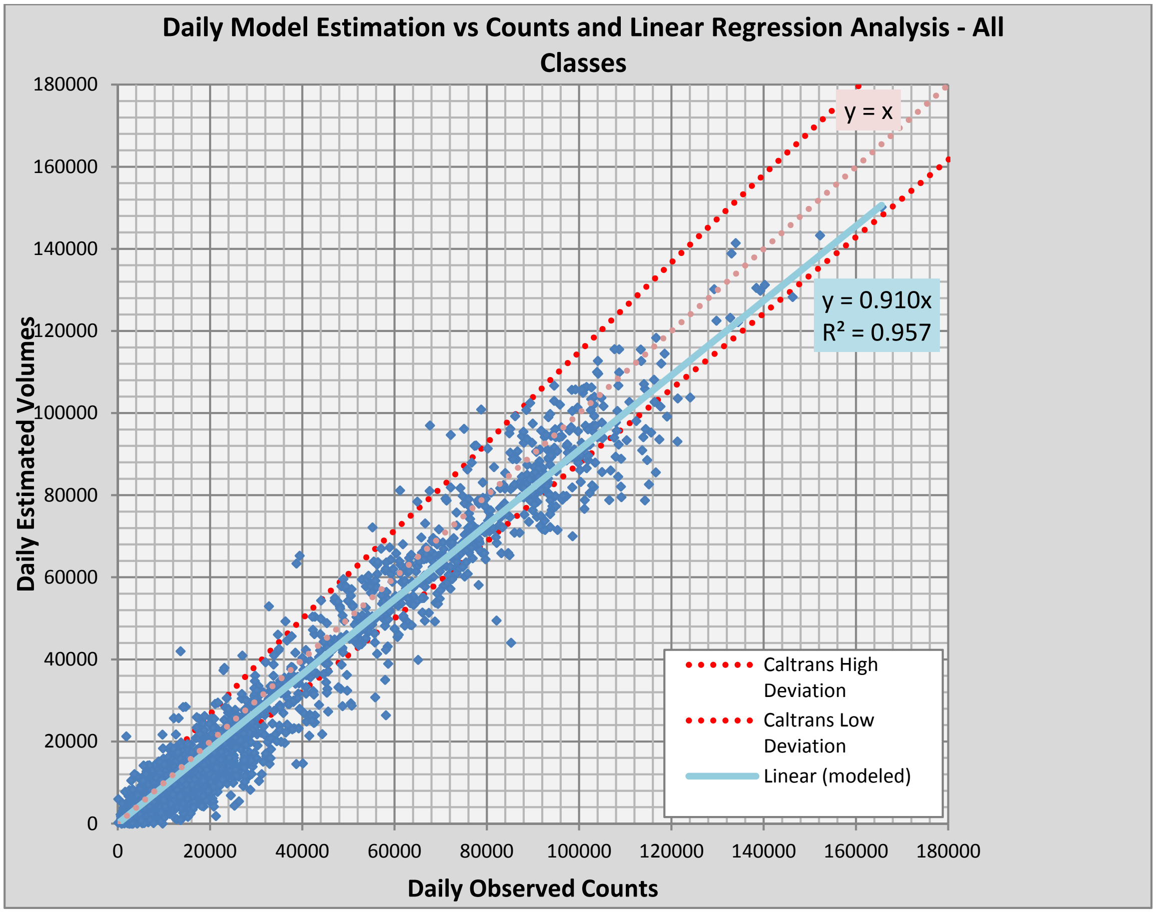
FIGURE 4: AGGREGATE MODEL ESTIMATED VS. OBSERVED VOLUME SCATTERPLOT, BY LINK CLASS



Source: 160914 highway_assignment_Aggregate_CVM_Wildcat_BASE_run.xlsx, tab: scatterplot_plot_All Classes

Figure 5 shows a scatterplot of estimated versus observed link volumes for the disaggregate model. This model is underestimating traffic as shown by the regression line slope lower than 1.0 (0.910), though the trend in the scatterplot with respect to variability by volume and outliers is like that shown in the aggregate model.

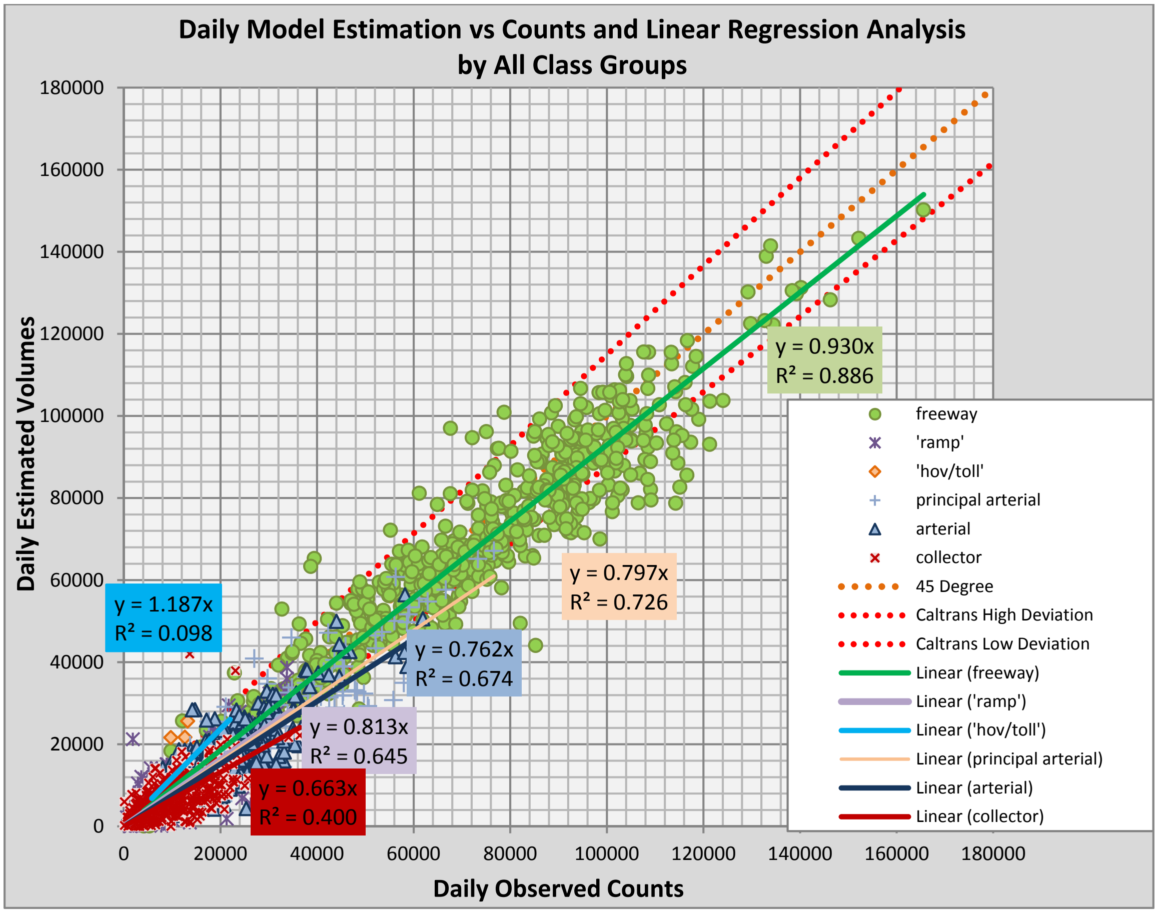
FIGURE 5: DISAGGREGATE MODEL ESTIMATED VS. OBSERVED VOLUME SCATTERPLOT, ALL LINK CLASSES COMBINED



Source: 160914 highway_assignment_Disaggregate_CVM_CTMInput_run.xlsx, tab: scatterplot_plot_all_points

Figure 6 breaks out the disaggregate model scatterplot by link class. The trends observed in the aggregate model are evident here, with generally lower volume estimates and higher variability as observed volume decreases.

FIGURE 6: DISAGGREGATE MODEL ESTIMATED VERSUS OBSERVED VOLUME SCATTERPLOT, BY LINK CLASS



Source: 160914 highway_assignment_Disaggregate_CVM_CTMInput_run.xlsx, tab: scatterplot_plot_All Classes

3.3 | ESTIMATED VERSUS OBSERVED VEHICLE MILES OF TRAVEL

The Highway Performance Monitoring System (HPMS) is a federally mandated planning study designed by the FHWA.[1] Caltrans collects HPMS data for cities, counties, Metropolitan Planning Organizations, and other partners and stakeholders. Recently, Caltrans has been collecting its own traffic counts along local roads. Public road mileage, lane miles, and VMT are reported to FHWA. Regional VMT are sampled from traffic counts for HPMS to estimate current needs and future performance of the system.

---

[1] *Highway Performance Monitoring System (HPMS)*, US Department of Transportation, Federal Highway Administration, Office of Highway Policy Information, http://www.fhwa.dot.gov/policyinformation/hpms.cfm

Table 9 compares the aggregate and disaggregate models against the HPMS count of total VMT for San Diego County. The HPMS VMT estimate is based on 2012 reported HPMS VMT of approximately 75.6 million, factored up by 1.05 to convert from a 7-day average to a 5-day average since the model represents average weekday travel. The comparison shows that the aggregate model is within 1% of HPMS VMT, while the disaggregate model system underestimates HPMS VMT by 10%.

TABLE 9: HPMS VALIDATION OF VEHICLE MILES OF TRAVEL, AGGREGATE AND DISAGGREGATE MODEL

SOURCE	VMT	PERCENT DIFFERENCE
HPMS[2]	79,585,925	NA
Aggregate Model	78,866,579	-0.9%
Disaggregate Model	71,455,854	-10.2%

Source: 160819 VersionAssignmentCompare for CVM PowerPt Tables.xlsx, tab: PowerPoint Slide.

Table 10 compares the aggregate versus disaggregate model VMT by vehicle class. For the sake of simplicity and consistency, all passenger vehicle classes (drive-alone, shared 2, and shared 3+) were collapsed into a single passenger cars and vans category. The light commercial vehicle category includes commercial vehicles from the aggregate CVM, and for the disaggregate CVM also includes an estimate of IE light-duty vehicles. The light and medium heavy truck category includes all light-heavy and medium-heavy trucks from the aggregate CVM and medium trucks from the disaggregate CVM. The heavy-heavy category only includes heavy-heavy trucks from the aggregate truck model and heavy trucks from the disaggregate CVM. The discrepancy between estimated and observed VMT in the disaggregate model is mostly due to the difference in light commercial vehicles; the disaggregate CVM predicts significantly lower (-8.6 million) light, non-heavy-duty commercial vehicle VMT than the aggregate model.

TABLE 10: VEHICLE MILES OF TRAVEL BY VEHICLE TYPE

MODE	AGGREGATE MODEL	DISAGGREGATE MODEL	DIFFERENCE	PERCENT DIFFERENCE
Passenger cars and vans	63,651,188	64,295,349	**(644,161)**	-1%
Light commercial vehicle	11,800,897	3,197,012	**8,603,885**	73%
Light and medium-heavy trucks	2,509,916	2,192,974	**316,942**	13%
Heavy-heavy trucks	904,579	1,770,519	**(865,940)**	-96%
Subtotal: commercial vehicles	15,215,392	7,160,504	**8,054,887**	53%

[2] *2012 California Public Road Data: Statistical Information Derived from the Highway Performance Monitoring System*, California State Transportation Agency, http://www.dot.ca.gov/hq/tsip/hpms/hpmslibrary/prd/2012prd/2012PRD_Updated.pdf.

MODE	AGGREGATE MODEL	DISAGGREGATE MODEL	DIFFERENCE	PERCENT DIFFERENCE
Total	**78,866,579**	**71,455,854**	**7,410,726**	9%

Source: 160819 VersionAssignmentCompare for CVM PowerPt Tables.xlsx, tab: BASE files.

Table 11 shows total commercial vehicle trips by type in the aggregate and disaggregate models. The table shows that the aggregate CVM has over one million more trips in the light commercial vehicle category. The aggregate model has less trips in the light-heavy, medium-heavy, and heavy-heavy vehicle categories, but the difference does not offset the difference in light commercial vehicles.

Table 12 shows the average trip length for each commercial vehicle category in miles. Light commercial vehicles have nearly the same trip length in each model. The trip length for heavy trucks is approximately twice as long in the aggregate model than the disaggregate model. This may be because the aggregate model borrowed parameters from the Southern California Association of Governments (SCAG) CVM, which is a larger region than San Diego County.

TABLE 11: TOTAL COMMERCIAL VEHICLE TRIPS BY VEHICLE TYPE

MODE	AGGREGATE MODEL	DISAGGREGATE MODEL	DIFFERENCE	PERCENT DIFFERENCE
Light commercial vehicle	1,420,313	390,105	1,030,208	73%
Light and medium-heavy trucks	130,264	221,629	(91,366)	-70%
Heavy-heavy trucks	30,455	110,281	(79,826)	-262%
Total	1,581,031	722,016	859,016	54%

TABLE 12: AVERAGE COMMERCIAL VEHICLE TRIP LENGTH (MI) BY VEHICLE TYPE

MODE	AGGREGATE MODEL	DISAGGREGATE MODEL	DIFFERENCE	PERCENT DIFFERENCE
Light commercial vehicle	8.3	8.2	0.1	1%
Light and medium-heavy trucks	18.8	9.7	9.1	48%
Heavy-heavy trucks	27.2	15.0	12.3	45%
Total	9.5	9.7	(0.2)	-2%

3.4 | CALIFORNIA DEPARTMENT OF MOTOR VEHICLES DATA

Commercial vehicle assignment results were also compared against data reported in the FHWA report on commercial vehicles (Cambridge Systematics, Inc. 2004). The report includes several commercial vehicle VMT estimates for San Diego County. Per 2002 California DMV data, commercial vehicles account for 6.3 million vehicle miles of travel, or 10.1% of total VMT for the San Diego Metropolitan Statistical Area (MSA). The population at the date that the data was collected was 2.65 million. The 2012 population was approximately 3 million, so scaling the VMT based on population growth results in an estimate of 7.1 million vehicle miles of travel for commercial vehicles. That aligns closely with the disaggregate model estimate of 7.1 million shown in Table 10, while the aggregate model predicts another 8 million vehicle miles of commercial vehicle travel beyond that.

Note that the DMV data has its shortcomings. There are some commercial vehicles that are registered outside of San Diego County that travel within the county. And, some commercial vehicles registered in San Diego County travel outside the county. It is not possible to measure the magnitude of each of these variables from the data. Furthermore, vehicles that are not registered as commercial vehicles are used for commercial purposes. For example, a person who uses their personal vehicle to deliver pizza or a real estate agent who uses their personal car to show properties. It is likely that the DMV data is somewhat downward biased in terms of measurement of total commercial vehicle travel, particularly in the light commercial vehicle category, because of this uncounted travel.

3.5 | ESTIMATED VERSUS OBSERVED COMMERCIAL VEHICLE ASSIGNMENT RESULTS

There are two observed data sources used in the comparison of commercial vehicle assignment results to observed data. One is the SANDAG Vehicle Classification and Occupancy Study, and the other is 2012 WIM data for 15 sites in San Diego County. The WIM data is available for all five assignment periods (Table 2) and is included in the heavy-truck assignment comparisons below.

SANDAG last conducted a vehicle classification and occupancy study at locations around San Diego County in 2012. The purpose of the study is to identify both the type of vehicle and the number of occupants. The one-day vehicle counts are performed manually by personnel stationed at key locations. Vehicles are classified by type:

- Bus (school, transit, tour, and shuttle are differentiated)
- Business vehicle (auto and light-duty truck are differentiated)
- Taxi
- Motorcycle
- Pedestrian
- Bike
- Heavy truck (2 axle, 3 axle, 4 axle, and 5+ axles are differentiated)

For purposes of comparison to CVM results, business vehicles (comprising autos and light-duty trucks) were combined into a light commercial vehicle category and all heavy-truck types were combined into a medium and heavy-truck category. The classification of a business versus nonbusiness vehicle is subjective. If the vehicle has a clear business logo it would be classified as a business vehicle. However, some portion of light commercial vehicles and trucks do not have business logos, and some vehicles with business logos may not actually be traveling for commercial purposes when observed.

Many of the commercial vehicle survey counts were collected in one of the two peak time periods (AM or PM) plus part of the midday period adjacent to the counted peak period (e.g., a location may have been counted for the AM peak period and from 9:00 a.m. to 12:00 p.m., while another location may have been counted from 12:00 p.m. to the end of the PM peak period). The counted midday period data were expanded—to maximize their use—to represent the entire midday period using a factor derived from those locations in which the entire midday period was counted.

Table 13 summarizes the comparison of assigned commercial vehicle volumes from the aggregate model for cars and light trucks to counts from the vehicle classification and occupancy study. The table summarizes both observed counts and estimated counts separately, and the total for each. It indicates that the aggregate CVM over-estimates observed light commercial vehicle traffic volumes from the occupancy study. Approximately 75% of the count locations are over-estimated by more than 40%. Only 5% of count locations are within +/- 10% of observed volume. The slope of the regression line is approximately 2.84; in other words, light commercial volumes are over-estimated by a factor of almost three. Figure 7 displays this data graphically in a scatterplot.

TABLE 13: AGGREGATE MODEL COMMERCIAL VEHICLE COUNT VALIDATION, CARS AND LIGHT TRUCKS

NUMBER OF LINKS BY VOLUME DEVIATION			
VOLUME DEVIATION	**OBSERVED**	**ESTIMATED**	**ALL**
>=40%	85	27	112
30%~40%	2	0	2
20%~30%	1	1	2
10%~20%	2	2	4
0%~10%	4	0	4
0%~-10%	2	1	3
-10%~-20%	7	0	7
-20%~-30%	1	0	1
-30%~-40%	1	0	1
<-40%	13	0	13
Total	**118**	**31**	**149**
PERCENT OF NUMBER OF LINKS BY VOLUME DEVIATION			
VOLUME GAP RANGE	**OBSERVED**	**ESTIMATED**	**ALL**
>=40%	72%	87%	75%
30%~40%	2%	0%	1%
20%~30%	1%	3%	1%
10%~20%	2%	6%	3%
0%~10%	3%	0%	3%
0%~-10%	2%	3%	2%
-10%~-20%	6%	0%	5%
-20%~-30%	1%	0%	1%
-30%~-40%	1%	0%	1%
<-40%	11%	0%	9%
Total	**100%**	**100%**	**100%**
-10%~10%	5%	3%	5%
-20%~20%	13%	9%	13%
-30%~30%	15%	12%	15%
AVERAGE VOLUME DEVIATION AND PERCENT OF (±) LINKS			
% of Links with (+) Deviation	80%	97%	83%
% of Links with (-) Deviation	20%	3%	17%
Avg. (+) Deviation	276%	227%	165%
Avg. (-) Deviation	-49%	-7%	-48%
Ave of All	**210%**	**220%**	**141%**
RMSE			
RMSE Percent	321%	265%	377%
SLOPE OF TREND LINE			
Slope	2.84	2.84	2.84

**Source: 161025 Commercial Vehicle Validation.xlsx, tab: DashBoard.*

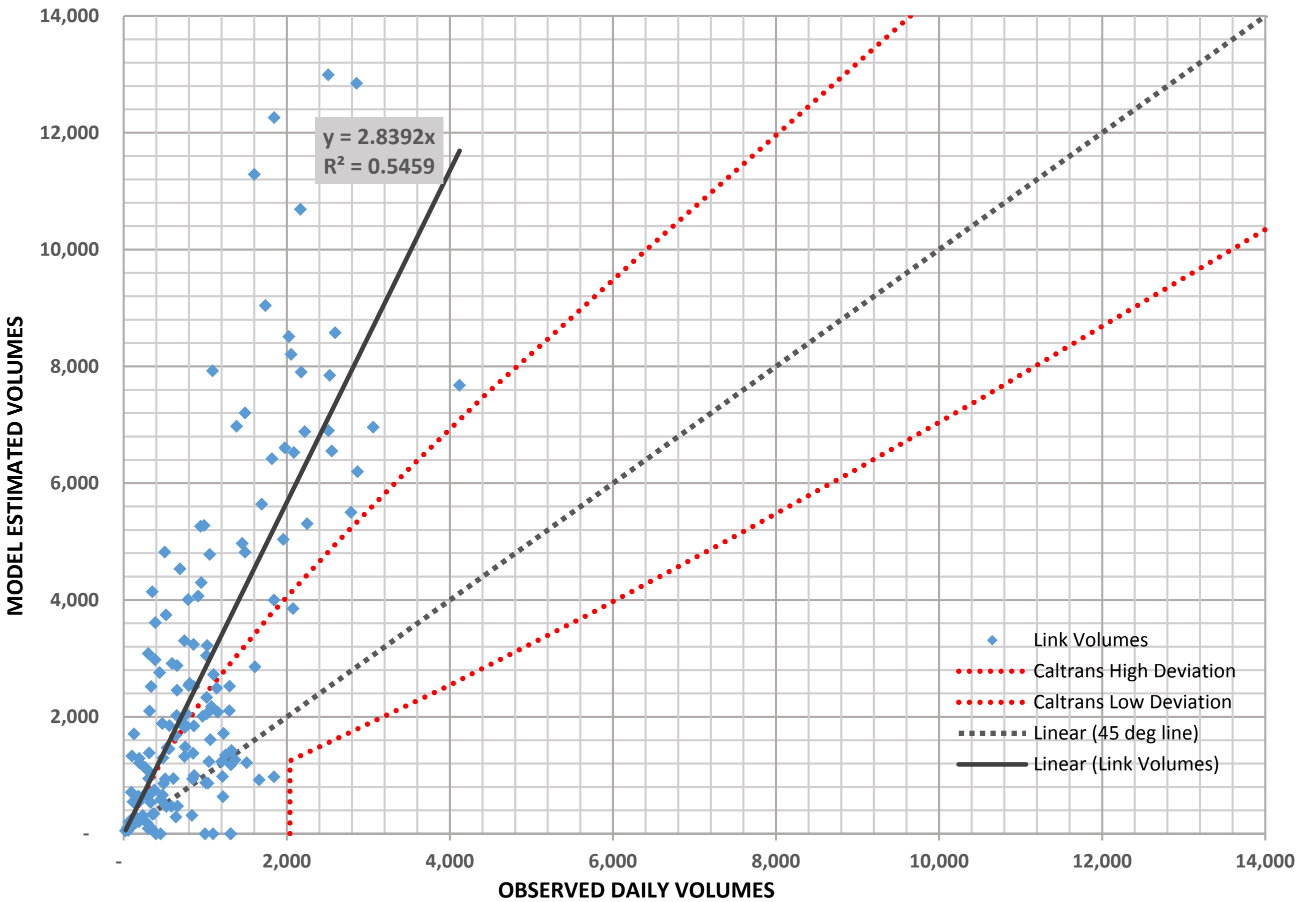
Table 14 shows the same comparison for the disaggregate model. The comparison of estimated versus counted light commercial volume is much closer for the disaggregate model; only approximately 11% of the links are over-estimated by 40% or greater. Twenty-three percent of the estimates are within +/- 20% of the counted volume. The RMSE is 62%, which is a significant improvement over the aggregate model (377%). The slope on the disaggregate model is 0.80, indicating that the disaggregate model underestimates light commercial volume by approximately 20%. These data are visualized in the scatterplot in Figure 8.

TABLE 14: DISAGGREGATE MODEL COMMERCIAL VEHICLE COUNT VALIDATION, CARS AND LIGHT TRUCKS

NUMBER OF LINKS BY VOLUME DEVIATION			
GAP RANGE	**OBSERVED**	**ESTIMATED**	**ALL**
>=40%	16	0	16
30%~40%	7	1	8
20%~30%	7	2	9
10%~20%	8	0	8
0%~10%	6	1	7
0%~-10%	9	3	12
-10%~-20%	5	3	8
-20%~-30%	8	3	11
-30%~-40%	10	5	15
<-40%	42	13	55
Total	**118**	**31**	**149**
PERCENT OF NUMBER OF LINKS BY VOLUME DEVIATION			
GAP RANGE	**OBSERVED**	**ESTIMATED**	**ALL**
>=40%	14%	0%	11%
30%~40%	6%	3%	5%
20%~30%	6%	6%	6%
10%~20%	7%	0%	5%
0%~10%	5%	3%	5%
0%~-10%	8%	10%	8%
-10%~-20%	4%	10%	5%
-20%~-30%	7%	10%	7%
-30%~-40%	8%	16%	10%
<-40%	36%	42%	37%
Total	**100%**	**100%**	**100%**
-10%~10%	13%	13%	13%
-20%~20%	24%	23%	23%
-30%~30%	37%	39%	36%
AVERAGE DEVIATION AND PERCENT OF (±) LINKS			
% of Links with (+) Deviation	37%	13%	32%
% of Links with (-) Deviation	63%	87%	68%
Avg. (+) Deviation	48%	23%	18%
Avg. (-) Deviation	-44%	-38%	-42%
Avg. of All	**-10%**	**-30%**	**-9%**
RMSE			
RMSE Percent	50%	50%	62%
SLOPE OF TREND LINE			
Slope	0.86	0.67	0.80

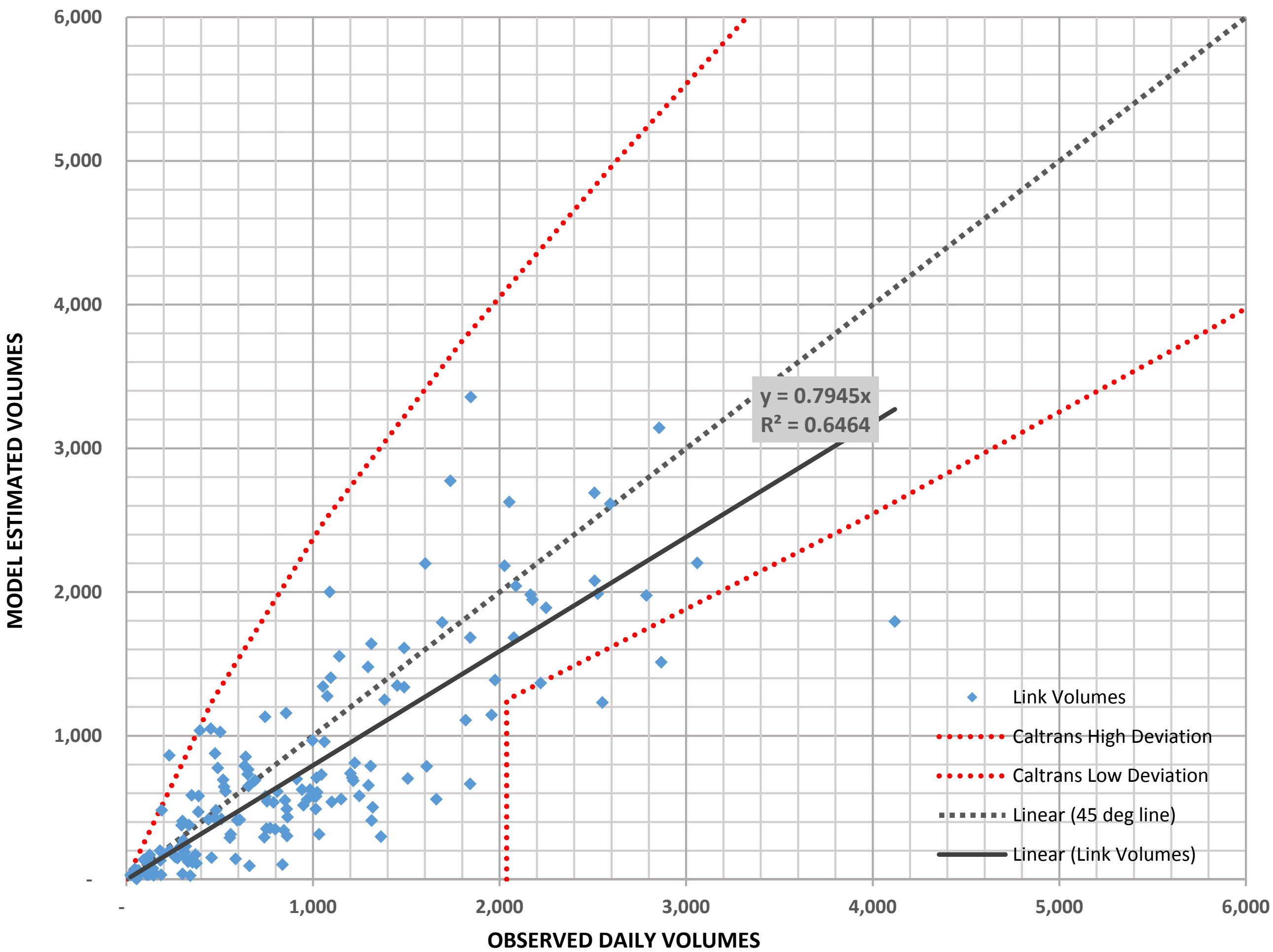
***Source: 161025 Commercial Vehicle Validation.xlsx, tab: Dashboard.*

FIGURE 7: AGGREGATE MODEL COMMERCIAL VEHICLE COUNT SCATTERPLOT, CARS, AND LIGHT TRUCKS



Source: 161025 Commercial Vehicle Validation.xlsx, tab:DashBoard.

FIGURE 8: DISAGGREGATE MODEL COMMERCIAL VEHICLE COUNT SCATTERPLOT, CARS, AND LIGHT TRUCKS



Source: 161025 Commercial Vehicle Validation.xlsx, tab: DashBoard.

Table 15 shows a comparison of estimated versus counted heavy-truck volumes based on the SANDAG Vehicle Classification and Occupancy Study data and the aggregate CVM. The table indicates that heavy trucks are over-estimated by approximately 30%, and that 56% of counted volumes are over-estimated by more than 40%. Only 24% of link volumes are within +/- 20%. The data are visualized in the scatterplot in Figure 9.

TABLE 15: AGGREGATE MODEL COMMERCIAL VEHICLE COUNT VALIDATION, HEAVY TRUCKS

NUMBER OF LINKS BY VOLUME DEVIATION			
GAP RANGE	**OBSERVED**	**ESTIMATED**	**ALL**
>=40%	112	13	125
30%~40%	6	2	8
20%~30%	7	1	8
10%~20%	10	4	14
0%~10%	19	2	21
0%~-10%	11	1	12
-10%~-20%	8	2	10
-20%~-30%	2	1	3
-30%~-40%	4	4	8
<-40%	14	1	15
Total	**193**	**31**	**224**
PERCENT OF NUMBER OF LINKS BY VOLUME DEVIATION			
GAP RANGE	**OBSERVED**	**ESTIMATED**	**ALL**
>=40%	58%	42%	56%
30%~40%	3%	6%	4%
20%~30%	4%	3%	4%
10%~20%	5%	13%	6%
0%~10%	10%	6%	9%
0%~-10%	6%	3%	5%
-10%~-20%	4%	6%	4%
-20%~-30%	1%	3%	1%
-30%~-40%	2%	13%	4%
<-40%	7%	3%	7%
Total	**100%**	**100%**	**100%**
-10%~10%	16%	9%	14%
-20%~20%	25%	28%	24%
-30%~30%	30%	34%	29%
AVERAGE DEVIATION AND PERCENT OF (±) LINKS			
% of Links with (+) Deviation	80%	71%	79%
% of Links with (-) Deviation	20%	29%	21%
Avg. (+) Deviation	89%	51%	84%
Avg. (-) Deviation	-33%	-32%	-33%
Avg. of All	**64%**	**27%**	**59%**
RMSE			
RMSE Percent	77%	57%	75%
SLOPE OF TREND LINE			
Slope	1.30	1.35	1.31

**Source: 161025 Commercial Vehicle Validation.xlsx, tab: Dashboard.*

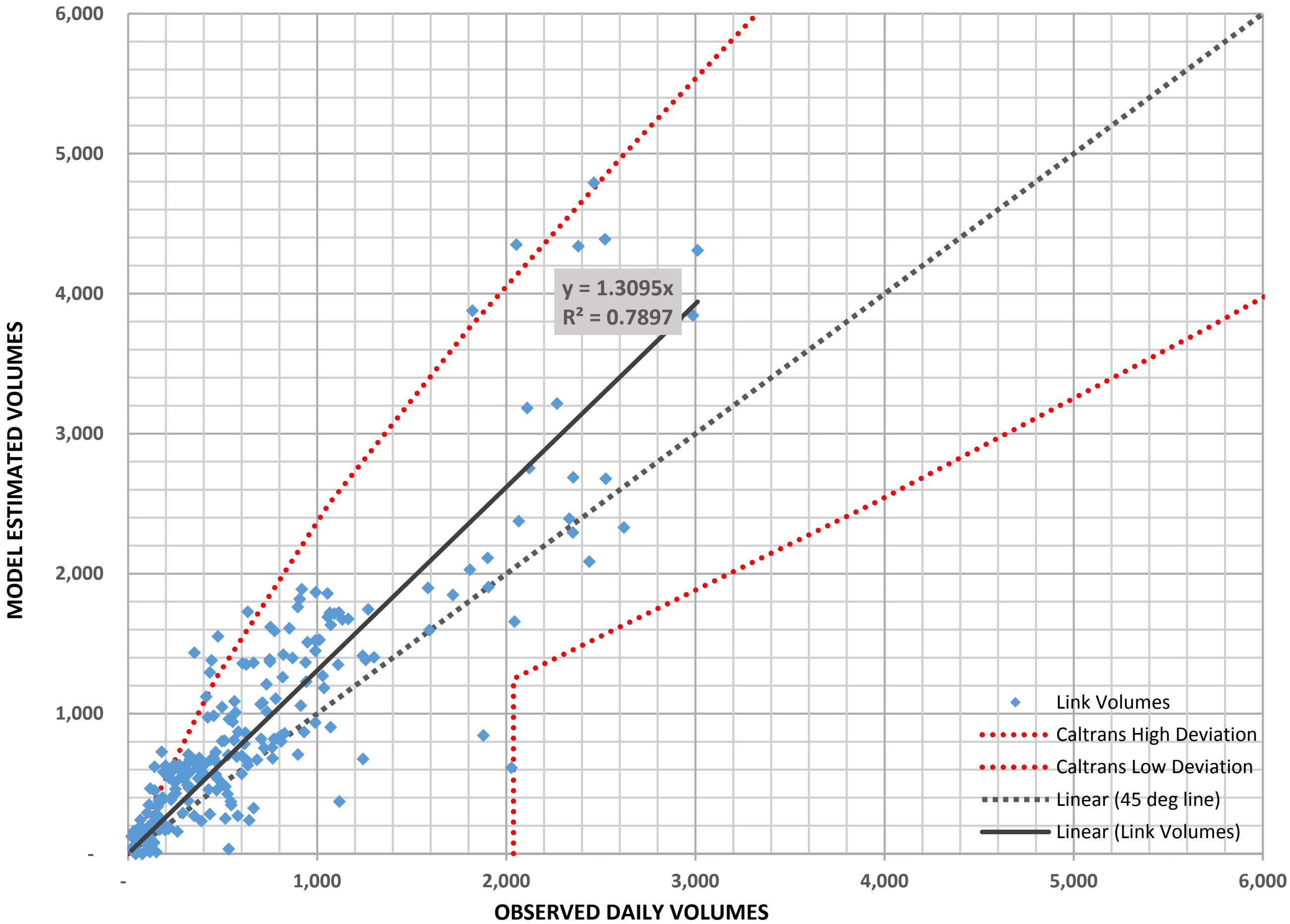
Table 16 shows the same comparison for the disaggregate model, which indicates a similar over-estimate of heavy-truck volumes; this is to be expected since the EI and EE heavy-truck trip tables are the same in both models. Figure 10 visualizes the disaggregate comparison in a scatterplot.

TABLE 16: DISAGGREGATE MODEL COMMERCIAL VEHICLE COUNT VALIDATION, HEAVY TRUCKS

NUMBER OF LINKS BY VOLUME DEVIATION			
GAP RANGE	**OBSERVED**	**ESTIMATED**	**ALL**
>=40%	98	14	112
30%~40%	11	3	14
20%~30%	8	6	14
10%~20%	10	3	13
0%~10%	13	0	13
0%~-10%	13	2	15
-10%~-20%	7	0	7
-20%~-30%	7	1	8
-30%~-40%	4	0	4
<-40%	22	2	24
Total	**193**	**31**	**224**
PERCENT OF NUMBER OF LINKS BY VOLUME DEVIATION			
GAP RANGE	**OBSERVED**	**ESTIMATED**	**ALL**
>=40%	51%	45%	50%
30%~40%	6%	10%	6%
20%~30%	4%	19%	6%
10%~20%	5%	10%	6%
0%~10%	7%	0%	6%
0%~-10%	7%	6%	7%
-10%~-20%	4%	0%	3%
-20%~-30%	4%	3%	4%
-30%~-40%	2%	0%	2%
<-40%	11%	6%	11%
Total	**100%**	**100%**	**100%**
-10%~10%	14%	6%	13%
-20%~20%	23%	16%	22%
-30%~30%	31%	38%	32%
AVERAGE DEVIATION AND PERCENT OF (±) LINKS			
% of Links with (+) Deviation	73%	84%	74%
% of Links with (-) Deviation	27%	16%	26%
Avg. (+) Deviation	98%	59%	92%
Avg. (-) Deviation	-35%	-39%	-36%
Avg. of All	**62%**	**43%**	**59%**
RMSE			
RMSE Percent	80%	67%	78%
SLOPE OF TREND LINE			
Slope	1.40	1.39	1.40

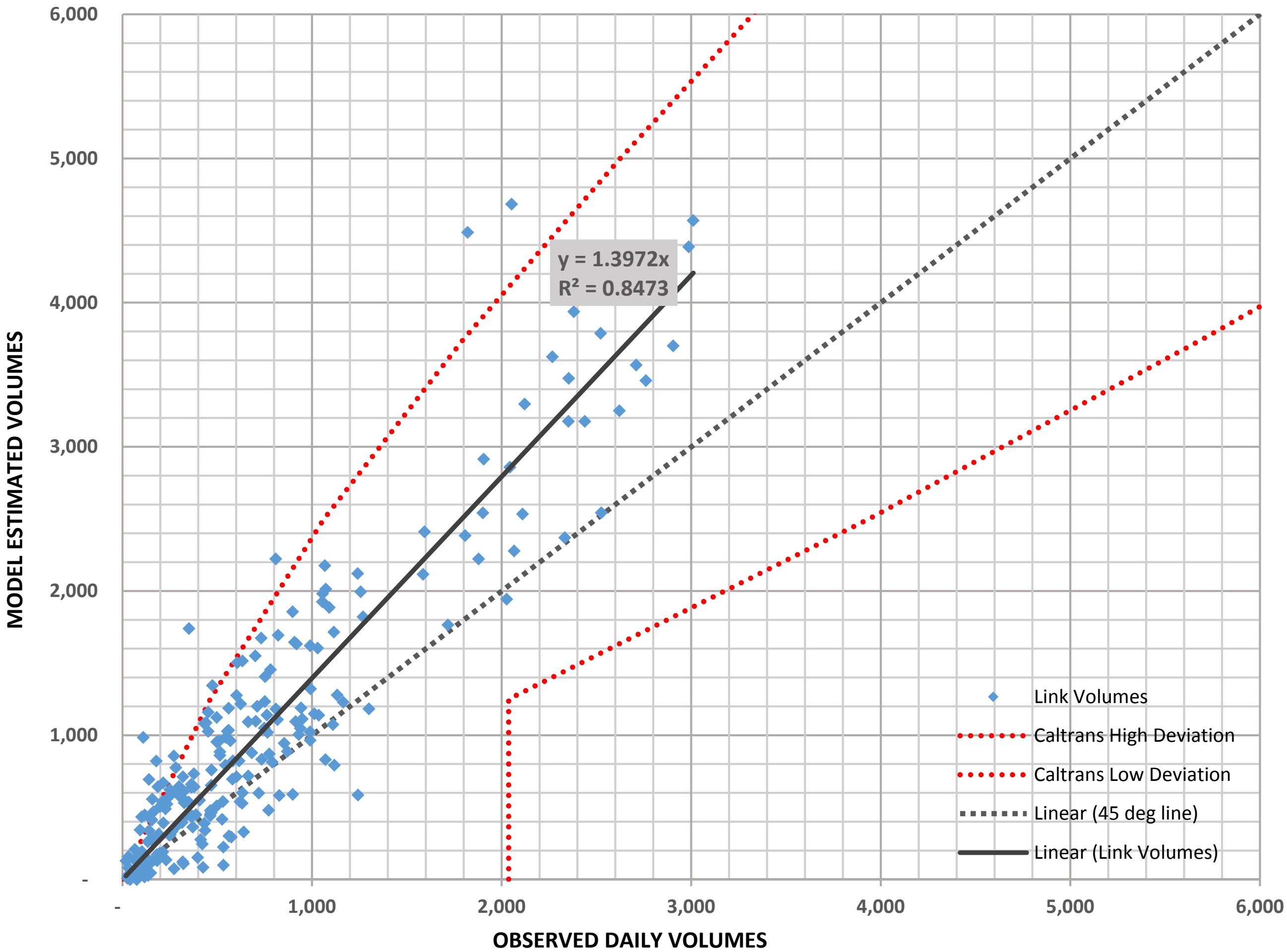
**Source: 161025 Commercial Vehicle Validation.xlsx, tab: DashBoard.*

FIGURE 9: AGGREGATE MODEL COMMERCIAL VEHICLE COUNT SCATTERPLOT, HEAVY TRUCKS



**Source: 161025 Commercial Vehicle Validation.xlsx, tab: DashBoard.*

FIGURE 10: DISAGGREGATE MODEL COMMERCIAL VEHICLE COUNT SCATTERPLOT, HEAVY TRUCKS



**Source: 161025 Commercial Vehicle Validation.xlsx, tab: DashBoard.*

3.6 | MODEL VALIDATION SUMMARY

The aggregate model more closely matches VMT reported in both the HPMS data and the traffic counts collected by Caltrans, SANDAG, and other agencies in San Diego County. The disaggregate model is approximately 8 million (or 10%) lower than the 79 million VMT reported by HPMS (Table 9). The amount of commercial travel predicted for passenger vehicles and light trucks was the primary differentiator in VMT between the disaggregate and aggregate models. The asserted aggregate CVM predicts 8 million more VMT for this market segment than the disaggregate model (Table 10).

Comparison to traffic counts provides supporting evidence that the disaggregate model underestimates VMT. Both the aggregate and disaggregate models have relatively low RMSE errors, while the slope of the regression line of disaggregate model estimated volumes to counts is 0.91; this indicates a count underestimate of approximately 9% (Table 8).

The SANDAG Vehicle Classification and Occupancy Study counts also suggest that the aggregate CVM over-predicts light commercial vehicles by a factor of 2.84 (Table 13). The FHWA study on commercial vehicles suggests that the VMT predicted by the disaggregate model is much closer to DMV estimates than the aggregate model. Comparisons to commercial vehicle counts suggests that the disaggregate model underestimates VMT for light commercial vehicles by approximately 20%. Comparisons to counts further suggests that heavy trucks are over-estimated by 30% and 40% for the aggregate and disaggregate models respectively.

4.0 PERSON-TRANSPORT MODEL ANALYSIS

This section compares the SANDAG resident AB travel model to data from the 2006 SANDAG household travel survey, the 2009 NHTS, and the 2012 California Household Travel Survey (CHTS) sample for San Diego County. The NHTS is a national household travel survey that was conducted from March 2008 to April 2009 (U.S. Department of Transportation n.d.). The NHTS data summaries rely upon work funded under the Strategic Highway Research Program (Resource Systems Group, Inc. 2014). The CHTS was last conducted by Caltrans starting in 2010 and ending in 2013 (California Department of Transportation 2013). Note that the comparisons of these data to the model focuses on the disaggregate model, though both aggregate and disaggregate model systems use the same person-transport models and therefore can be considered the same model for purposes of comparison.

The comparison focuses on measures of travel that are common across nearly all AB travel models and have the greatest impact on VMT. Transit and nonmotorized shares tend to be low in most metropolitan areas, including San Diego. Since the SANDAG AB model was developed and calibrated to the 2006 SANDAG household travel survey, it should closely match the survey data. The NHTS data provides a point of comparison to the SANDAG data and the model, though the following should be considered:

- Regional demographic differences (e.g., the share of retired persons, college students, and other travel markets with different travel characteristics than residents of San Diego County).
- Differences in urban form across regions (e.g., compact regions tend to have shorter trip lengths than regions consisting of multinodal urban centers).
- Differences in data collection methodologies (e.g., survey technology, survey length, and geocoding accuracy may impact tour and trip rates).
- Definition of a complete household, and household expansion factor methodology, can affect travel characteristics. NHTS defines a complete household as one in which person interviews were completed with at least 50% of the adult (age 18+) household members. It is not clear what biases might be introduced in the data as a result. For example, it is possible that adults who travel less frequently were more likely to participate in the survey, which would result in a downward bias on travel reported in the survey.

4.1 | TOUR AND TRIP RATE COMPARISON

Table 17 shows average tour rates per person for the NHTS data by region, the 2006 SANDAG household travel survey (HTS), the 2012 CHTS sample for San Diego, and the SANDAG AB model. The table also shows the sample size (person count) in each survey sample, the expanded persons in each region, and tour rates broken out by tour type. Home-Based Tours are tours that begin and end at home. Work-based subtours are tours that begin and end at work, and are largely composed of trips to lunch, personal errands, and work-related travel.

Average NHTS tours per person range from a low of 1.28 for Fresno, to a high of 1.41 for Sacramento and San Diego. The overall average for the NHTS is 1.36 tours per person. The SANDAG 2006 HTS reports an average of 1.34 tours per person, which is close to the NHTS average; however, the SANDAG model was calibrated to match a higher average tours per person

(1.39) than the survey. Because the observed share of nonworking adults staying home in the 2006 SANDAG HTS was lower than observed in other regions, the rate was calibrated up, which raised the average. Despite this increase, the tour rate remains 0.02 tours per person lower than the 2009 NHTS for San Diego. The 2012 CHTS has the lowest tour rate per person of any of the datasets listed, at 1.12 tours per person.

The person sample size for the 2009 NHTS for San Diego County is 1.7 times higher than that of the 2006 survey because of an add-on purchased for San Diego County. There were 6,050 households sampled in the NHTS compared to 3,000 in the 2006 SANDAG HTS. There were only 3,099 persons in the San Diego sample of the CHTS.

TABLE 17: AVERAGE TOUR RATES BY REGION, NHTS AND SAN DIEGO SURVEYS COMPARED TO MODEL

REGION	TOTAL SURVEYED PERSONS	TOTAL EXPANDED PERSONS	HOME-BASED TOURS PER PERSON	WORK-BASED SUBTOURS PER PERSON	AVERAGE TOURS PER PERSON
Sacramento, CA	3,096	2,172,272	1.36	0.06	1.41
Fresno, CA	978	847,548	1.26	0.02	1.28
San Joaquin, CA	1,711	1,446,537	1.26	0.04	1.30
Tampa, FL	5,275	2,686,603	1.31	0.05	1.36
Jacksonville, FL	2,624	1,196,099	1.27	0.05	1.32
Indianapolis, IN	2,049	1,649,705	1.35	0.04	1.39
Albany, NY	3,694	779,935	1.29	0.05	1.34
Rochester, NY	1,622	940,024	1.27	0.04	1.32
Piedmont, NC	11,578	1,173,920	1.25	0.05	1.30
Nashville, TN	1,353	1,382,416	1.26	0.08	1.35
San Antonio, TX	4,697	2,188,659	1.31	0.05	1.36
Dallas, TX	14,740	6,075,770	1.33	0.05	1.38
San Diego (NHTS)	14,727	2,774,615	1.37	0.04	1.41
Total/Average NHTS	68,144	25,314,104	1.31	0.05	1.36
San Diego (2006 HTS)	8,769	2,724,072	1.29	0.04	1.34
San Diego (2012 CHTS)	3,099	2,952,059	1.12	0.03	1.15
SANDAG Model	N/A	3,143,418	1.36	0.04	1.39

Table 18 shows average trip rates per person by region. The NHTS rates range from a low of 3.32 in Fresno to a high of 3.74 in San Diego. The average trip rate per person across all NHTS regions is 3.64. The CHTS has only 3.05 trips per person. The 2006 SANDAG HTS rate is 3.38 trips per person, which is 0.36 trips per person lower than the NHTS sample. The model produces 3.57 trips per person, which is higher than the survey upon which it is based due to the somewhat higher tour rate for nonworking adults. The model is slightly lower than the average for all NHTS regions, and 0.17 trips per person lower than the NHTS sample for San Diego.

TABLE 18: AVERAGE TRIP RATES BY REGION, NHTS AND SAN DIEGO SURVEYS COMPARED TO MODEL

REGION	TOTAL SURVEYED PERSONS	TOTAL EXPANDED PERSONS	AVERAGE TRIPS PER PERSON
Sacramento, CA	3,096	2,172,272	3.70
Fresno, CA	978	847,548	3.32
San Joaquin, CA	1,711	1,446,537	3.48
Tampa, FL	5,275	2,686,603	3.62
Jacksonville, FL	2,624	1,196,099	3.52
Indianapolis, IN	2,049	1,649,705	3.63
Albany, NY	3,694	779,935	3.56
Rochester, NY	1,622	940,024	3.64
Piedmont, NC	11,578	1,173,920	3.59
Nashville, TN	1,353	1,382,416	3.72
San Antonio, TX	4,697	2,188,659	3.62
Dallas, TX	14,740	6,075,770	3.70
San Diego (2009 NHTS)	14,727	2,774,615	3.74
Total NHTS	68,144	25,314,104	3.64
San Diego (2006 HTS)	8,769	2,724,072	3.38
San Diego (2012 CHTS)	3,099	2,952,059	3.05
SANDAG Model	N/A	3,143,418	3.57

4.2 | TRIP LENGTH COMPARISONS

Table 19 shows the average trip length by region in miles for NHTS regions, the 2006 SANDAG HTS, the 2012 CHTS for San Diego County, and the SANDAG AB model. Average trip length ranges from a low of 4.98 miles for San Joaquin to a high of 8.38 for Dallas, Texas. Not surprisingly, the ratio of the standard deviation of trip length to its mean (0.16) is much higher than the ratio of the standard deviation of tour or trip rate to their mean (0.03 each). The higher variance for trip length is likely due to differences in size and urban form across the regions.

The average trip length for the 2006 SANDAG household survey is 6.17 miles, which is 1.16 miles lower than the average across all NHTS regions and approximately 1 mile shorter than the NHTS data for San Diego. The CHTS survey average trip length (5.8 miles) is shorter than both the NHTS sample for San Diego County residents and the 2006 HTS.

TABLE 19: AVERAGE TRIP LENGTH BY REGION, NHTS AND SAN DIEGO SURVEYS COMPARED TO MODEL

REGION	TOTAL SURVEYED PERSONS	TOTAL EXPANDED PERSONS	AVERAGE TRIP LENGTH (MI.)
Sacramento, CA	3,096	2,172,272	6.53
Fresno, CA	978	847,548	5.79
San Joaquin, CA	1,711	1,446,537	4.98
Tampa, FL	5,275	2,686,603	6.99
Jacksonville, FL	2,624	1,196,099	8.27
Indianapolis, IN	2,049	1,649,705	6.94
Albany, NY	3,694	779,935	7.59
Rochester, NY	1,622	940,024	5.04
Piedmont, NC	11,578	1,173,920	7.18
Nashville, TN	1,353	1,382,416	7.82
San Antonio, TX	4,697	2,188,659	8.29
Dallas, TX	14,740	6,075,770	8.38
San Diego (2009 NHTS)	14,727	2,774,615[3]	7.18
Total	68,144	25,314,104	7.33

[3] The 2010 Census population estimate for San Diego County is 3,054,000 persons. The total expanded persons shown in this table is directly reported from the SHRP report cited above; it is possible that the data was expanded to an incomplete population estimate or there is some error in the weights with respect to household size.

REGION	TOTAL SURVEYED PERSONS	TOTAL EXPANDED PERSONS	AVERAGE TRIP LENGTH (MI.)
San Diego (2006 HTS)	8,769	2,724,072	6.17
San Diego (2012 CHTS)	3,099	2,952,059[4]	5.84
SANDAG Model	N/A	3,143,418	6.28

4.3 | PERSON-TRANSPORT MODEL ANALYSIS CONCLUSIONS

The preceding comparisons indicate that the AB model was well calibrated to the 2006 SANDAG HTS. However, 2009 NHTS data for San Diego County suggests that tour rates, trip rates, and trip lengths for San Diego residents are higher than those reported in the 2006 data. The NHTS survey predicts 0.17 more trips per person, and 0.9 more miles per trip, than the AB model.

It is possible to calculate the VMT that might result from calibrating the person-transport model to the NHTS data instead of the SANDAG HTS data. There are two components to this calculation; the VMT for existing trips which would result from the longer NHTS trip length, and the VMT for new trips that would result from the higher NHTS trip rate. For the trips that the model currently predicts whose average trip length is too short, the model would produce an additional 6.3M vehicle miles of travel, as follows:

> 3.143M persons * 3.57 trips per person * 0.9 additional miles per trip * 0.63 vehicles per person trip[5] = 6.3M VMT

And for the additional 0.17 trips per person, we might expect another 2.4M vehicle miles of travel, as follows:

> 3.143M persons * 0.17 additional trips per person * 7.18 miles per trip * 0.63 vehicles per person trip = 2.4M VMT

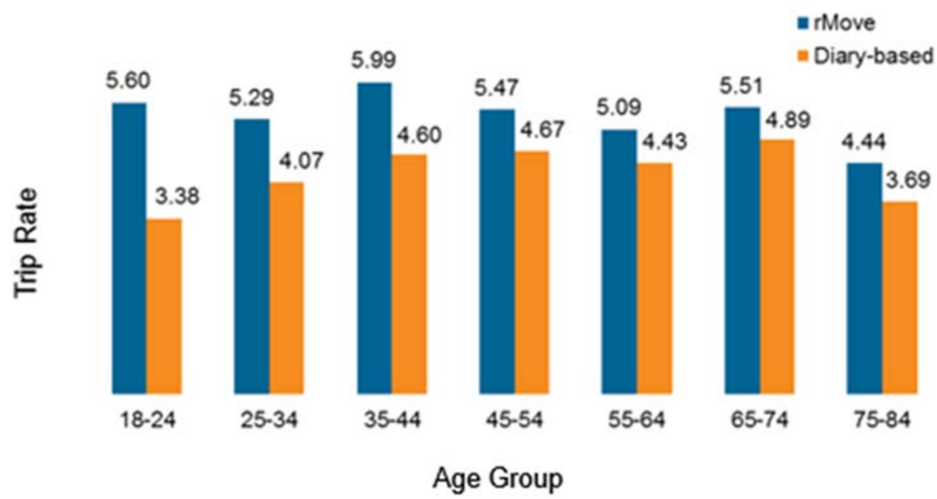
This results in total additional VMT of 8.7M.

[4] The 2012 Census population estimate for San Diego County is 3,176,000 persons. The total expanded persons shown in this table is taken from the household expansion factors attached to the CHTS data. Similar to the NHTS data, it is possible there is some bias in the expansion, though the difference is much less for the CHTS sample than for the NHTS sample.

[5] VMT per mile of person trip can be calculated by scaling person trips to vehicle trips using the ratio of vehicle trips to person trips from the model, which is 0.63. This factor considers auto occupancy and the share of person trips made by nonmotorized modes, school bus, and transit.

SANDAG is currently conducting a new HTS, to be completed in 2017. This survey relies upon GPS-enabled smartphone technology, rMove (Resource Systems Group, Inc. n.d.). The target number of completed households is 5,000 and a completed household is defined as one in which travel data is collected from all household members age five or more. Travel will be reported by proxy for children under 5, or otherwise imputed. When the data are available and coded, these data should be compared to the 2006 SANDAG HTS and the 2009 NHTS data to establish a solid foundation for the next round of model calibration. Initial research for GPS-enabled smartphone data collection are that under-reporting is occurring by as much as 26% higher than traditional diary-based HTSs and that this under-reporting is biased by age group (see Figure 11) as well as income group (Bradley and Greene 2016).

FIGURE 11: TRIP RATES BY AGE GROUP FROM SEATTLE 3-DAY rMOVE SURVEY



5.0 AIRSAGE DATA ANALYSIS

This section compares trips generated by the SANDAG models[6] to AirSage OD data for the San Diego region. The trips from the two data sources are compared by purpose and time of day. The comparisons are created for aggregate trips, district-to-district flows, and trip-length distributions. The rest of this section describes the two data sources (AirSage and model), data processing, and trip comparisons. In the end, findings are summarized and recommendations are provided.

AirSage Data

SANDAG obtained AirSage cell phone data collected in March 2015. The data are OD trips based on 539 AirSage land-use zones. The data are aggregated and expanded into an average weekday. AirSage synthesizes traveler type and trip purpose from habitual locations of the cell phone and these algorithms may not always align with reality (see Bindra paper). The trips are distinguished by traveler type, trip purpose, and time of day.

- **Traveler Type: resident or visitor**—this distinction is based on the vendor's algorithm for determining habitual home location of the cell phone. If the home location is inside of the San Diego County for at least two weeks, then the trip is considered a resident trip; otherwise it is considered a visitor trip.
- **Trip Purpose**—the ends of these trips are determined partially by the habitual activity location of the cell phone.
  - **Home-Based Work (HBW)**
    - School and university trips are often lumped into this category as it is not known whether the trip was to attend school or work at the school.
  - **Home-Based Other (HBO)**
  - **Non-Home-Based (NHB)**
- **Time of Day (TOD)**—aggregation to five time-of-day periods.
  - Early AM (3:00 a.m.–6:00 a.m.)
  - AM (6:00 a.m. to 9:00 a.m.)
  - Midday (9:00 a.m. to 3:00 p.m.)
  - PM (3:00 p.m. to 7:00 p.m.)
  - Evening (7:00 p.m. to 3:00 a.m.)

The preceding five time-of-day periods are aggregated to represent four time periods in the present analysis, Table 20.

[6] Model run "abmctm_ctminput" is used in the comparisons

TABLE 20: TIME-OF-DAY PERIODS

TOD	PERIOD
AM	6:00 a.m. to 9:00 a.m.
MD	9:00 a.m. to 3:00 p.m.
PM	3:00 p.m. to 7:00 p.m.
EV	7:00 p.m. to 6:00 a.m.

5.1 | DISAGGREGATE MODEL DATA

Table 1 shows travel markets in the SANDAG model system. The CT-RAMP resident travel model (1)[7], the IE model (2), the overnight visitor model (4) and the airport passenger model (5) are relevant to the present analysis. The Mexican resident travel model is excluded due to unavailability of equivalent trips in the AirSage data. The AirSage data includes EE trips, however, it is unknown whether they traversed through the San Diego region. For example, travel from Orange County to Imperial County has relevant paths through the San Diego County or around it, depending on traffic. The AirSage data is comprised of only passenger travel, so commercial vehicle trips are also excluded.

RELEVANT TRIPS

Table 21 summarizes model components relevant to the analysis. The table identifies model components by travel type (internal or external) and by traveler type (resident or visitor).

TABLE 21: MODEL COMPONENTS

COMPONENT	TYPE	Internal Travel		EXTERNAL TRAVEL	
		RESIDENT	VISITOR	RESIDENT	VISITOR
CT-RAMP	Disaggregate	•			
Visitor	Disaggregate		•		
Airport	Disaggregate	•	•		•
Internal-External	Disaggregate			•	
External-Internal	Aggregate				•

Internal travel is simulated in the CT-RAMP model, the visitor model, and the airport model. The airport model also constructs airport trips to/from outside the San Diego region (external). The external travel is primarily generated by IE and EI models for resident and visitors respectively. All models, except the EI model, simulate trips at disaggregate (MGRA) detail. The EI travel model generates trips at TAZ level, a more aggregate geometry. A list of trip files from the relevant model components is provided in Table 22.

[7] Number in the bracket represents travel market segment, as shown in Table 1.

TABLE 22: TRIP FILES FROM RELEVANT MODEL COMPONENTS

MODEL	TRIP FILES
CT-RAMP	indivTripData_3.csv and joinTripData_3.csv
Visitor	visitorTrips.csv
Airport	airport_out.csv
Int-Ext (D)*	internalExternalTrips.csv
Ext-Int (A)*	usSdWrk_[period].mtx and usSdNon_[period].mtx

** D-disaggregate and A-aggregate.*

The CT-RAMP resident travel model outputs separate trip files for individual and joint travel. A joint trip involves multiple persons making the trip together and is represented as one record in the join trip file. During this analysis, the joint trips are expanded by number of people involved in a joint trip.[8] Similarly, the visitor trips and airport trips are also expanded to represent individual trips.[9]

TRIPS BY PURPOSE

The AirSage data provide resident and visitor travel in three purpose categories: home-based other (HBO), home-based work (HBW), and non-home-based (NHB). The model trips are also aggregated into these three purposes. For model components that do not specify trip purpose (IE model and EI model), the trips are placed in "ALL" category.

The following definitions are used to aggregate model trips into the AirSage trip purposes:

- **HBW**—if either of the origin and destination purpose is "Home" and tour purpose is "Work", "School," or "University."
- **HBO**—if either of the original and destination purpose is "Home" and tour purpose is not "Work", "School," or "University."
- **NHB**—all other trips.
- **ALL**—no purpose information is available.

[8] Field "num_participants" contains information on number of participants.
[9] Fields "partySize" and "size" contains information on number of people involved in a visitor and an airport trip respectively.

Table 23 summarizes trip purposes available in the relevant model components.

TABLE 23: MODEL COMPONENTS AND TRIP PURPOSE

COMPONENT	RESIDENT				VISITOR			
	HBO	HBW	NHB	ALL	HBO	HBW	NHB	ALL
CT-RAMP (Indiv.)	●	●	●					
CT-RAMP (Joint)	●		●					
Visitor							●	
Airport	●		●				●	
Int-Ext (D)*				●				
Ext-Int (A)*				●				●

** D-disaggregate and A-aggregate.*

Internal Trips

The internal trips made by San Diego residents are constructed from three model outputs:

1. **Resident individual trips**—obtained from CT-RAMP output; the trips are categorized as HBW, HBO, and NHB.
2. **Resident joint trips**—obtained from CT-RAMP output; the trips are categorized as HBO and NHB. Note that joint trips to/from work (HBW) are individual trips in the model.
3. **Airport model resident trips**—obtained from airport ground access model output; the resident trips are categorized as "HBO+NHB" (HBW airport trips are modeled in CT-RAMP); a fixed split (Airport Survey 2012) of 88% to 12% is applied to distribute resident airport trips into HBO and NHB purposes respectively.

The visitor trips are obtained from two model outputs:

1. **Visitor model trips**—obtained from the visitor model output; the visitor trips are considered NHB.
2. **Airport model visitor trips**—obtained from the airport ground access model output; the airport visitor trips are considered NHB.

The model visitor trips are "NHB" and compared with "NHB" trips in the AirSage data.

External Trips

The IE trips are gathered from two model outputs:

1. **IE model (residents)**—no trip purpose information is available; the resident trips into and out of the San Diego County are aggregated into one purpose "ALL"
2. **EI model (visitors)**—no trip purpose information available; the visitor trips into and out of the San Diego County are aggregated into one purpose "ALL."

The AirSage data does not contain trips in or out of Mexico region, so the resident trips that are made to/from Mexico border external zones are removed. Also, since the airport model output does not specify external stations for the visitor airport trips, the airport external trips are excluded from the comparisons.

MODEL ZONES TO AIRSAGE ZONES

The AirSage data are OD trips based on 539 zones (internal = 505, and external = 34), whereas the model trips are based on 23,002 MGRAs and 4,996 TAZs. To compare trips at the same level of geography, the model trips are aggregated to AirSage zones.[10] For this, correspondences between model zones and AirSage zones are created.

Internal Zones

The following three correspondences are created for internal zones:

- **MGRA to AirSage Zone**—disaggregate internal trips represented at the MGRA level are aggregated to AirSage zones to compare with the AirSage trips.
- **TAZ to AirSage Zone**—aggregate IE trips represented at the TAZ level are aggregated to AirSage zones to compare with AirSage trips.
- **AirSage to TAZ**—AirSage trips are attached to model skims (trip lengths) that are available at TAZ detail.

External Zones

The model represents the zones outside the San Diego County area by 12 external stations (Figure 12); the first five stations represent Mexico zones. The AirSage data has 34 well-defined external zones. Therefore, a correspondence between the AirSage external zones and the model external stations is created. The mapping of the zones is shown in Table 24 and Figure 12.

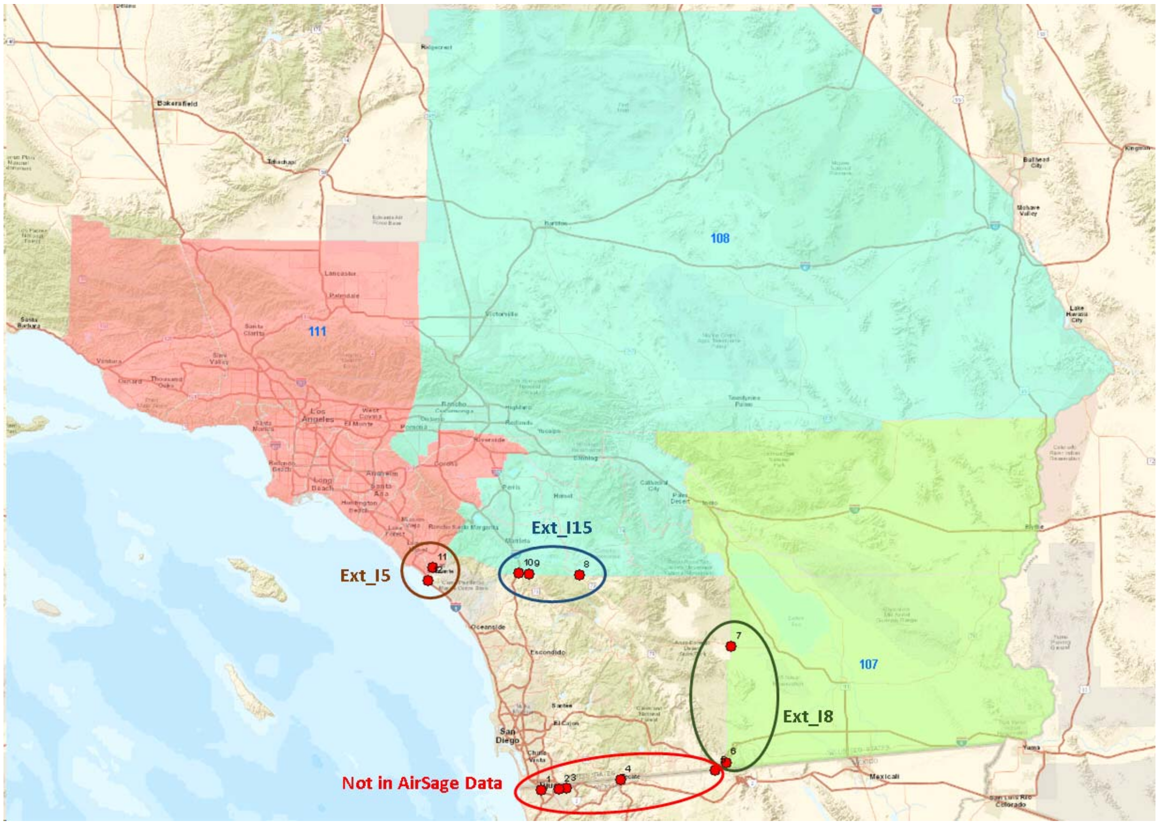
TABLE 24: EXTERNAL STATIONS

EXT STATION	ID	MODEL TAZ
Ext_I8	107	6, 7
Ext_I15	108	8, 9, 10
Ext_I5	111	11, 12
Ext_Mexico (excluded)		1,2,3,4,5

In all, the external zones are aggregated into three external stations: Ext_I8, Ext_I15, and Ext_I5. The model stations covering Mexico travel are excluded from the analysis. As shown in Figure 12, the AirSage zones within the same-colored polygon are assigned to the corresponding external station. For example, the AirSage zones in the blue polygon (ID=108) are represented as Ext_I15.

[10] In general, the AirSage zones are spatially bigger than the model zones (TAZ or MGRA).

FIGURE 12: EXTERNAL STATIONS



5.2 | AGGREGATE COMPARISONS

The AirSage and the model trips are first compared at an aggregate level. The total trips in the two datasets are compared by trip purposes (HBO, HBW, and NHB) and time of day (AM, MD, PM, and EV). The model trips are aggregated into the same time-of-day periods as in Table 20.

INTERNAL TRIPS

Table 25 and Table 26 summarize the resident trips made within (internal) the San Diego County. The model is producing trips that are comparable with the AirSage data. However, the model trips are a bit higher in the MD period and a bit lower in the PM period. The model closely matches AirSage HBO trips but produces more HBW trips and less NHB trips compared to the AirSage data. Differences in purpose may be due to error in the AirSage trip purpose imputation process. Overall, the AirSage data has approximately more trips (+5%) than the model (550,000). However, this may be partly due to differences in population between the model year (2012) and the year that AirSage data was collected (2015). Per US Census data, San Diego County grew 6.6% between April 2010 and July 2015.

TABLE 25: INTERNAL RESIDENT TRIPS (COUNT)

TOD	AIRSAGE				MODEL			
	HBO	HBW	NHB	TOTAL	HBO	HBW	NHB	TOTAL
AM	992,658	618,337	660,664	2,271,659	962,807	1,058,697	416,192	2,437,696
MD	1,984,782	593,938	1,724,152	4,302,872	1,856,918	672,469	985,210	3,514,597
PM	1,474,346	528,024	1,161,196	3,163,566	1,782,229	896,157	873,787	3,552,173
EV	1,295,445	326,646	445,563	2,067,654	986,983	331,987	425,148	1,744,118
Total	**5,747,231**	**2,066,946**	**3,991,575**	**11,805,752**	**5,588,937**	**2,959,310**	**2,700,337**	**11,248,584**

TABLE 26: INTERNAL RESIDENT TRIPS (SHARE)

TOD	AIRSAGE				MODEL			
	HBO	HBW	NHB	TOTAL	HBO	HBW	NHB	TOTAL
AM	8.4%	5.2%	5.6%	19.2%	8.6%	9.4%	3.7%	21.7%
MD	16.8%	5.0%	14.6%	36.4%	16.5%	6.0%	8.8%	31.2%
PM	12.5%	4.5%	9.8%	26.8%	15.8%	8.0%	7.8%	31.6%
EV	11.0%	2.8%	3.8%	17.5%	8.8%	3.0%	3.8%	15.5%
Total	**48.7%**	**17.5%**	**33.8%**	**100.0%**	**49.7%**	**26.3%**	**24.0%**	**100.0%**

All visitor trips produced by the model are considered as NHB and are compared only with the NHB visitor trips in the AirSage data (see Table 27).

TABLE 27: INTERNAL VISITOR TRIPS

TOD	COUNT		SHARE	
	AIRSAGE	MODEL	AIRSAGE	MODEL
	NHB	NHB	NHB	NHB
AM	149,387	45,161	13.6%	7.5%
MD	566,221	196,124	51.6%	32.4%
PM	268,754	143,903	24.5%	23.8%
EV	113,216	220,639	10.3%	36.4%
Total	**1,097,578**	**605,827**	**100.0%**	**100.0%**

The model generates approximately half as many visitor trips as in the AirSage data. There is also disagreement in the share of visitor trips by time period. The PM period trip shares are similar in the two datasets. However, the model predicts a higher share of trips in the evening (EV) period and a lower share in AM and MD periods.

To verify whether the visitor trip totals from the model look reasonable, we reviewed several sources. The San Diego Tourism Authority reported 16.4 million overnight visitors for 2013, for an average of 44,932 visitors per day (San Diego Tourism Authority 2013). Approximately 83% of visitors were in San Diego for leisure activities, with 17% visiting for business. Further, approximately 50% of leisure travelers, and 90% of business travelers, stayed in hotels; most of the rest stayed in homes. The average length of stay was approximately four nights. These statistics are like the values used in the application of the visitor model, which is based on 2009 San Diego Convention and Visitor Bureau data.

Per the visitor survey conducted as part of the visitor model development project, the average number of tours per day for business travelers is 1.2, and 1.1 tours per day for leisure travelers. If we assume that visitors spend one day out of the six average days they are in San Diego traveling to/from the airport, we might expect about 260k daily visitor tours (45000 visitors * 5 travel days per visitor * 1.15 tours per day); the model produces 247k daily visitor tours, which is very close to the number suggested by the Tourism Authority data.

This raises questions regarding the accuracy of the AirSage visitor trip estimates. Resident trips are scaled up based upon comparisons of sampled households to total households by Census Tract. However, it is unclear how the visitor trip sample was scaled to total visitor trips. On the one hand, not all visitors have smart phone service available when they are traveling away from home. On the other hand, the rate of visitors with services that have their phone activated may be higher than the rate of resident travelers, since they might be more likely to use their phone's GPS device for directions. Without a better understanding of the direction and magnitude of the bias, it is difficult to make solid conclusions with respect to the comparison of the AirSage and model visitor data.

5.3 | EXTERNAL TRIPS

Table 28 and Table 29 summarize resident and visitor trips that have either an origin or a destination outside of San Diego County. The trips are compared by time of day only as external trips from the model do not contain trip purpose information.

The total resident external trips are higher in the AirSage data (Table 28). The AirSage data reports approximately five times more trips than the model. The distribution of trips across time is similar in the two datasets. However, the model forecasts a higher share of trips in the PM period and lower share of trips in the EV period.

TABLE 28: IE RESIDENT TRIPS

TOD	COUNT		SHARE	
	AIRSAGE	MODEL	AIRSAGE	MODEL
AM	120,785	25,413	20.5%	22.1%
MD	205,460	38,197	34.9%	33.2%
PM	138,133	33,374	23.5%	29.0%
EV	124,672	17,946	21.2%	15.6%
Total	**589,049**	**114,930**	**100.0%**	**100.0%**

As with the resident trips, the total visitor external trips are also higher in the AirSage data (Table 29). The total model trips equal only approximately 37% of the total AirSage trips. The distribution across time periods is also different across the two datasets. The model produces lower trip shares in the MD period and higher trip shares in the other periods.

TABLE 29: IE VISITOR TRIPS

TOD	COUNT		SHARE	
	AIRSAGE	MODEL	AIRSAGE	MODEL
AM	57,519	24,767	18.5%	21.3%
MD	163,096	43,717	52.4%	37.6%
PM	52,181	26,347	16.8%	22.7%
EV	38,604	21,430	12.4%	18.4%
Total	**311,399**	**116,261**	**100.0%**	**100.0%**

Overall, the AirSage data observes more external trips (900,448) than predicted by the model (231,191), see Table 31.

5.4 | DISTRICT FLOWS

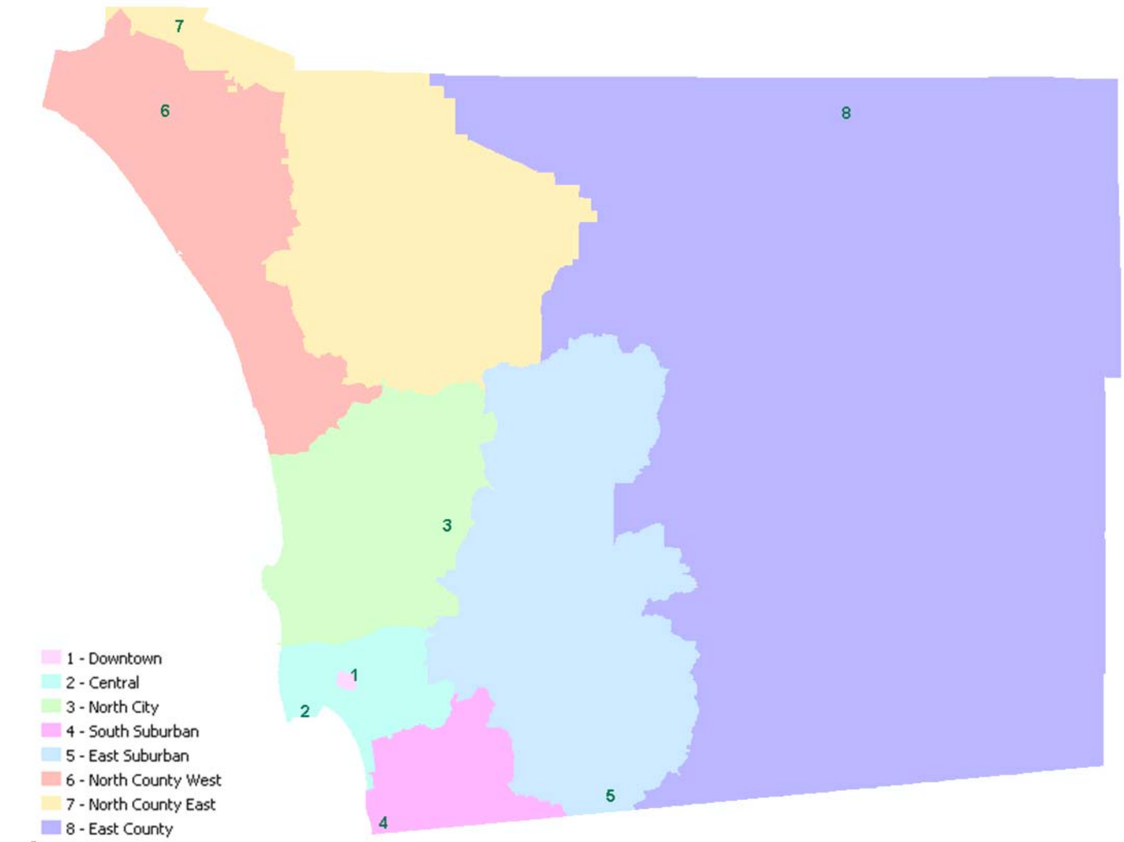
The San Diego County region is classified into eight districts,[11] as shown in Table 30 and Figure 13.

TABLE 30: DISTRICTS

DISTRICT	LABEL
1	Downtown
2	Central
3	North City
4	South Suburban
5	East Suburban
6	North County West
7	North County East
8	East County

[11] The MGRA-based input file contains a field "pseudoPMSA" providing district corresponding to an MGRA.

FIGURE 13: DISTRICTS IN SAN DIEGO COUNTY



Trips in the two datasets[12] are aggregated to these districts based on their origin and destination ends. The resulting district OD flows are summarized and compared. District flows by purpose and time of day look like the total trips summaries and are not discussed here. However, all summaries are available in the spreadsheet "AirSage_Model_Summary.xlsx" provided with the report.

INTERNAL TRIPS

The district-to-district flows of resident internal trips are presented in Table 32, Table 33, Table 34, Table 35, Table 36, Table 37. Overall, the flows closely match in the two datasets with differences ranging from -14% to 19% (Table 37). However, the model predicts slightly lower flows for East Suburban (model=13.8% and AirSage=15.0%) and North County West (model=12.5% and AirSage=13.4%), and slightly higher flows for South Suburban (model=11.5% and AirSage=10.1%).

The district-to-district flows of visitor internal trips are presented in Table 38, Table 39, Table 40, Table 41, Table 42, Table 43. As discussed in aggregate comparisons, there are about twice as many AirSage visitor trips as expected. Therefore, we discuss trip shares rather than absolute numbers of trips at a district level. The trip shares by district largely match the two datasets, except Downtown (model=13.8% and AirSage=4.7%) and North County West (model=10.8% and AirSage=17.4%), which receive higher and lower trip shares in the model, respectively.

[12] While assigning districts to AirSage zones, some AirSage zone boundaries did not match the district.

EXTERNAL TRIPS

Trips were compared with counts at external stations to further validate accuracy of the datasets (see Table 31).

TABLE 31: COUNTS AT EXTERNAL STATIONS

EXT. STATION	COUNTS		AIRSAGE		MODEL	
	IN	OUT	IN	OUT	IN	OUT
Ext_I8	7,090	7,179	46,425	42,618	3,967	3,953
Ext_I15	70,246	69,235	58,444	188,249	61,947	61,851
Ext_I5	62,867	62,223	105,661	230,387	47,871	47,723

Note that in the table above, the counts include both auto and truck trips, however, the model and AirSage numbers are only for auto trips. Also, the counts at Ext_I8 are at a few interchanges/intersections away from the external station, hence, the actual counts at the external station would be smaller.

The counts largely match the trips predicted by the model. In contrast, the AirSage trips are unreasonably higher (up to six times) than the counts. This pattern is consistent with the externals trips analysis (Table 28 and Table 29), where both resident and visitor trips in the AirSage data are five and two times as many as in the model respectively. This, once again, generates less confidence in the AirSage external travel data.

TABLE 32: INTERNAL RESIDENT DISTRICT FLOWS (COUNT)—AIRSAGE

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	TOTAL
Downtown	19,533	71,511	61,209	29,209	25,255	7,193	4,886	299	**219,096**
Central	71,665	967,584	460,800	235,946	217,862	33,572	26,193	2,636	**2,016,258**
North City	61,243	467,564	2,124,917	127,939	276,301	172,414	150,465	4,401	**3,385,243**
South Suburban	27,572	223,107	116,477	733,618	73,296	7,954	9,123	2,336	**1,193,483**
East Suburban	24,026	210,645	253,660	72,335	1,156,253	15,773	22,430	16,587	**1,771,710**
North County West	7,400	34,809	165,933	8,049	17,096	1,085,693	267,404	763	**1,587,146**
North County East	4,851	25,595	140,361	9,424	23,377	263,302	1,103,388	1,799	**1,572,097**
East County	295	2,635	4,208	2,530	14,974	790	1,859	33,427	**60,718**
TOTAL	**216,586**	**2,003,450**	**3,327,566**	**1,219,050**	**1,804,413**	**1,586,691**	**1,585,748**	**62,248**	**11,805,752**

TABLE 33: INTERNAL RESIDENT DISTRICT FLOWS (COUNT)—MODEL

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	TOTAL
Downtown	98,797	95,014	31,408	17,133	12,295	1,516	1,373	337	**257,873**
Central	95,848	1,187,107	331,449	147,973	155,107	11,406	9,212	1,816	**1,939,918**
North City	31,423	332,379	2,401,138	55,495	164,266	96,360	97,386	5,313	**3,183,760**
South Suburban	16,255	148,044	55,950	1,024,086	45,238	3,475	3,053	1,566	**1,297,667**
East Suburban	12,127	154,746	164,183	45,074	1,150,228	7,343	12,198	11,957	**1,557,856**
North County West	1,650	11,284	96,177	3,254	6,917	1,084,276	197,569	769	**1,401,896**
North County East	1,435	9,432	97,840	3,048	12,343	196,680	1,216,641	3,708	**1,541,127**
East County	304	1,946	5,615	1,604	11,462	840	3,695	43,021	**68,487**
TOTAL	**257,839**	**1,939,952**	**3,183,760**	**1,297,667**	**1,557,856**	**1,401,896**	**1,541,127**	**68,487**	**11,248,584**

TABLE 34: INTERNAL RESIDENT DISTRICT FLOWS DIFF (COUNT)— (MODEL-AIRSAGE)

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	TOTAL
Downtown	79,264	23,503	-29,801	-12,076	-12,960	-5,677	-3,513	38	**38,777**
Central	24,183	219,523	-129,351	-87,973	-62,755	-22,166	-16,981	-820	**-76,340**
North City	-29,820	-135,185	276,221	-72,444	-112,035	-76,054	-53,079	912	**-201,483**
South Suburban	-11,317	-75,063	-60,527	290,468	-28,058	-4,479	-6,070	-770	**104,184**
East Suburban	-11,899	-55,899	-89,477	-27,261	-6,025	-8,430	-10,232	-4,630	**-213,854**
North County West	-5,750	-23,525	-69,756	-4,795	-10,179	-1,417	-69,835	6	**-185,250**
North County East	-3,416	-16,163	-42,521	-6,376	-11,034	-66,622	113,253	1,909	**-30,970**
East County	9	-689	1,407	-926	-3,512	50	1,836	9,594	**7,769**
TOTAL	**41,253**	**-63,498**	**-143,806**	**78,617**	**-246,557**	**-184,795**	**-44,621**	**6,239**	**-557,168**

TABLE 35: INTERNAL RESIDENT DISTRICT FLOWS (SHARE)—AIRSAGE

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	TOTAL
Downtown	0.2%	0.6%	0.5%	0.2%	0.2%	0.1%	0.0%	0.0%	**1.9%**
Central	0.6%	8.2%	3.9%	2.0%	1.8%	0.3%	0.2%	0.0%	**17.1%**
North City	0.5%	4.0%	18.0%	1.1%	2.3%	1.5%	1.3%	0.0%	**28.7%**
South Suburban	0.2%	1.9%	1.0%	6.2%	0.6%	0.1%	0.1%	0.0%	**10.1%**
East Suburban	0.2%	1.8%	2.1%	0.6%	9.8%	0.1%	0.2%	0.1%	**15.0%**
North County West	0.1%	0.3%	1.4%	0.1%	0.1%	9.2%	2.3%	0.0%	**13.4%**
North County East	0.0%	0.2%	1.2%	0.1%	0.2%	2.2%	9.3%	0.0%	**13.3%**
East County	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.3%	**0.5%**
TOTAL	**1.8%**	**17.0%**	**28.2%**	**10.3%**	**15.3%**	**13.4%**	**13.4%**	**0.5%**	**100.0%**

TABLE 36: INTERNAL RESIDENT DISTRICT FLOWS (SHARE)—MODEL

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	TOTAL
Downtown	0.9%	0.8%	0.3%	0.2%	0.1%	0.0%	0.0%	0.0%	**2.3%**
Central	0.9%	10.6%	2.9%	1.3%	1.4%	0.1%	0.1%	0.0%	**17.2%**
North City	0.3%	3.0%	21.3%	0.5%	1.5%	0.9%	0.9%	0.0%	**28.3%**
South Suburban	0.1%	1.3%	0.5%	9.1%	0.4%	0.0%	0.0%	0.0%	**11.5%**
East Suburban	0.1%	1.4%	1.5%	0.4%	10.2%	0.1%	0.1%	0.1%	**13.8%**
North County West	0.0%	0.1%	0.9%	0.0%	0.1%	9.6%	1.8%	0.0%	**12.5%**
North County East	0.0%	0.1%	0.9%	0.0%	0.1%	1.7%	10.8%	0.0%	**13.7%**
East County	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.4%	**0.6%**
TOTAL	**2.3%**	**17.2%**	**28.3%**	**11.5%**	**13.8%**	**12.5%**	**13.7%**	**0.6%**	**100.0%**

TABLE 37: INTERNAL RESIDENT DISTRICT FLOWS DIFF (SHARE)— (MODEL-AIRSAGE)/AIRSAGE

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	TOTAL
Downtown	405.8%	32.9%	-48.7%	-41.3%	-51.3%	-78.9%	-71.9%	12.8%	**17.7%**
Central	33.7%	22.7%	-28.1%	-37.3%	-28.8%	-66.0%	-64.8%	-31.1%	**-3.8%**
North City	-48.7%	-28.9%	13.0%	-56.6%	-40.5%	-44.1%	-35.3%	20.7%	**-6.0%**
South Suburban	-41.0%	-33.6%	-52.0%	39.6%	-38.3%	-56.3%	-66.5%	-33.0%	**8.7%**
East Suburban	-49.5%	-26.5%	-35.3%	-37.7%	-0.5%	-53.4%	-45.6%	-27.9%	**-12.1%**
North County West	-77.7%	-67.6%	-42.0%	-59.6%	-59.5%	-0.1%	-26.1%	0.8%	**-11.7%**
North County East	-70.4%	-63.1%	-30.3%	-67.7%	-47.2%	-25.3%	10.3%	106.1%	**-2.0%**
East County	2.9%	-26.1%	33.4%	-36.6%	-23.5%	6.3%	98.8%	28.7%	**12.8%**
TOTAL	**19.0%**	**-3.2%**	**-4.3%**	**6.4%**	**-13.7%**	**-11.6%**	**-2.8%**	**10.0%**	**-4.7%**

TABLE 38: INTERNAL VISITOR DISTRICT FLOWS (COUNT)—AIRSAGE

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	TOTAL
Downtown	6,840	22,474	13,942	2,206	1,700	3,392	980	305	**51,839**
Central	23,465	110,282	72,624	15,552	13,136	15,211	5,748	1,493	**257,511**
North City	14,913	67,369	208,895	6,432	14,001	25,015	10,369	2,190	**349,184**
South Suburban	2,369	14,524	6,730	33,596	3,086	1,104	528	385	**62,322**
East Suburban	1,846	12,690	13,396	2,869	44,974	1,403	1,321	1,805	**80,303**
North County West	3,130	13,813	23,995	1,191	1,301	131,331	14,817	857	**190,436**
North County East	1,145	5,381	9,852	529	1,206	14,958	56,467	428	**89,966**
East County	584	2,736	3,176	468	1,848	1,179	575	5,451	**16,017**
TOTAL	**54,292**	**249,271**	**352,610**	**62,844**	**81,251**	**193,593**	**90,805**	**12,913**	**1,097,578**

TABLE 39: INTERNAL VISITOR DISTRICT FLOWS (COUNT)—MODEL

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	TOTAL
Downtown	59,162	14,586	6,435	1,557	1,161	557	271	20	**83,749**
Central	14,457	62,742	32,024	8,081	6,476	3,025	1,781	109	**128,695**
North City	6,580	31,922	135,158	4,960	8,119	7,610	5,507	238	**200,094**
South Suburban	1,559	8,054	5,052	20,433	2,138	372	353	38	**37,999**
East Suburban	1,116	6,423	8,147	2,167	20,733	680	696	202	**40,164**
North County West	567	3,057	7,558	416	635	46,273	6,808	39	**65,353**
North County East	290	1,779	5,493	339	716	6,781	30,808	471	**46,677**
East County	18	132	227	46	186	55	453	1,979	**3,096**
TOTAL	**83,749**	**128,695**	**200,094**	**37,999**	**40,164**	**65,353**	**46,677**	**3,096**	**605,827**

TABLE 40: INTERNAL VISITOR DISTRICT FLOWS DIFF (COUNT)— (MODEL-AIRSAGE)

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	TOTAL
Downtown	52,322	-7,888	-7,507	-649	-539	-2,835	-709	-285	**31,910**
Central	-9,008	-47,540	-40,600	-7,471	-6,660	-12,186	-3,967	-1,384	**-128,816**
North City	-8,333	-35,447	-73,737	-1,472	-5,882	-17,405	-4,862	-1,952	**-149,090**
South Suburban	-810	-6,470	-1,678	-13,163	-948	-732	-175	-347	**-24,323**
East Suburban	-730	-6,267	-5,249	-702	-24,241	-723	-625	-1,603	**-40,139**
North County West	-2,563	-10,756	-16,437	-775	-666	-85,058	-8,009	-818	**-125,083**
North County East	-855	-3,602	-4,359	-190	-490	-8,177	-25,659	43	**-43,289**
East County	-566	-2,604	-2,949	-422	-1,662	-1,124	-122	-3,472	**-12,921**
TOTAL	**29,457**	**-120,576**	**-152,516**	**-24,845**	**-41,087**	**-128,240**	**-44,128**	**-9,817**	**-491,751**

TABLE 41: INTERNAL VISITOR DISTRICT FLOWS (SHARE)—AIRSAGE

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	TOTAL
Downtown	0.6%	2.0%	1.3%	0.2%	0.2%	0.3%	0.1%	0.0%	4.7%
Central	2.1%	10.0%	6.6%	1.4%	1.2%	1.4%	0.5%	0.1%	23.5%
North City	1.4%	6.1%	19.0%	0.6%	1.3%	2.3%	0.9%	0.2%	31.8%
South Suburban	0.2%	1.3%	0.6%	3.1%	0.3%	0.1%	0.0%	0.0%	5.7%
East Suburban	0.2%	1.2%	1.2%	0.3%	4.1%	0.1%	0.1%	0.2%	7.3%
North County West	0.3%	1.3%	2.2%	0.1%	0.1%	12.0%	1.4%	0.1%	17.4%
North County East	0.1%	0.5%	0.9%	0.0%	0.1%	1.4%	5.1%	0.0%	8.2%
East County	0.1%	0.2%	0.3%	0.0%	0.2%	0.1%	0.1%	0.5%	1.5%
TOTAL	**4.9%**	**22.7%**	**32.1%**	**5.7%**	**7.4%**	**17.6%**	**8.3%**	**1.2%**	**100.0%**

TABLE 42: INTERNAL VISITOR DISTRICT FLOWS (SHARE)—MODEL

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	TOTAL
Downtown	9.8%	2.4%	1.1%	0.3%	0.2%	0.1%	0.0%	0.0%	13.8%
Central	2.4%	10.4%	5.3%	1.3%	1.1%	0.5%	0.3%	0.0%	21.2%
North City	1.1%	5.3%	22.3%	0.8%	1.3%	1.3%	0.9%	0.0%	33.0%
South Suburban	0.3%	1.3%	0.8%	3.4%	0.4%	0.1%	0.1%	0.0%	6.3%
East Suburban	0.2%	1.1%	1.3%	0.4%	3.4%	0.1%	0.1%	0.0%	6.6%
North County West	0.1%	0.5%	1.2%	0.1%	0.1%	7.6%	1.1%	0.0%	10.8%
North County East	0.0%	0.3%	0.9%	0.1%	0.1%	1.1%	5.1%	0.1%	7.7%
East County	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.3%	0.5%
TOTAL	**13.8%**	**21.2%**	**33.0%**	**6.3%**	**6.6%**	**10.8%**	**7.7%**	**0.5%**	**100.0%**

TABLE 43: INTERNAL VISITOR DISTRICT FLOWS DIFF (SHARE)—(MODEL-AIRSAGE)/AIRSAGE

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	TOTAL
Downtown	764.9%	-35.1%	-53.8%	-29.4%	-31.7%	-83.6%	-72.3%	-93.4%	**61.6%**
Central	-38.4%	-43.1%	-55.9%	-48.0%	-50.7%	-80.1%	-69.0%	-92.7%	**-50.0%**
North City	-55.9%	-52.6%	-35.3%	-22.9%	-42.0%	-69.6%	-46.9%	-89.1%	**-42.7%**
South Suburban	-34.2%	-44.5%	-24.9%	-39.2%	-30.7%	-66.3%	-33.1%	-90.1%	**-39.0%**
East Suburban	-39.5%	-49.4%	-39.2%	-24.5%	-53.9%	-51.5%	-47.3%	-88.8%	**-50.0%**
North County West	-81.9%	-77.9%	-68.5%	-65.1%	-51.2%	-64.8%	-54.1%	-95.4%	**-65.7%**
North County East	-74.7%	-66.9%	-44.2%	-35.9%	-40.6%	-54.7%	-45.4%	10.1%	**-48.1%**
East County	-96.9%	-95.2%	-92.9%	-90.2%	-89.9%	-95.3%	-21.2%	-63.7%	**-80.7%**
TOTAL	**54.3%**	**-48.4%**	**-43.3%**	**-39.5%**	**-50.6%**	**-66.2%**	**-48.6%**	**-76.0%**	**-44.8%**

TABLE 44: EXTERNAL TRIPS DISTRICT FLOWS (COUNT)—AIRSAGE

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	Ext_I8	Ext_I15	Ext_I5	TOTAL
Downtown	-	-	-	-	-	-	-	-	1,216	4,295	7,983	13,495
Central	-	-	-	-	-	-	-	-	7,088	19,583	26,574	53,244
North City	-	-	-	-	-	-	-	-	10,727	41,016	60,579	112,321
South Suburban	-	-	-	-	-	-	-	-	2,976	7,642	14,307	24,925
East Suburban	-	-	-	-	-	-	-	-	6,214	9,031	11,654	26,899
North County West	-	-	-	-	-	-	-	-	4,183	37,648	71,703	113,534
North County East	-	-	-	-	-	-	-	-	2,307	65,803	35,992	104,102
East County	-	-	-	-	-	-	-	-	7,907	3,232	1,595	12,733
Ext_I8	1,592	8,218	12,568	2,949	6,177	4,966	2,476	7,478	-	-	-	46,425
Ext_I15	4,070	16,726	35,570	7,306	8,181	32,256	57,098	2,898	-	-	-	58,444
Ext_I5	8,470	27,285	60,145	13,959	11,449	69,846	36,149	1,361	-	-	-	105,661
TOTAL	**14,132**	**52,229**	**108,282**	**24,214**	**25,808**	**107,068**	**95,724**	**11,737**	**42,618**	**188,249**	**230,387**	**900,449**

TABLE 45: EXTERNAL TRIPS DISTRICT FLOWS (COUNT)—MODEL

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	Ext_I8	Ext_I15	Ext_I5	TOTAL
Downtown	-	-	-	-	-	-	-	-	231	1,536	1,165	2,932
Central	-	-	-	-	-	-	-	-	690	5,795	4,664	11,149
North City	-	-	-	-	-	-	-	-	1,502	17,453	13,067	32,021
South Suburban	-	-	-	-	-	-	-	-	323	2,320	1,788	4,431
East Suburban	-	-	-	-	-	-	-	-	618	3,591	2,492	6,701
North County West	-	-	-	-	-	-	-	-	281	13,782	15,081	29,144
North County East	-	-	-	-	-	-	-	-	264	17,275	9,421	26,961
East County	-	-	-	-	-	-	-	-	46	99	44	189
Ext_I8	231	692	1,507	324	619	282	265	46	-	-	-	3,967
Ext_I15	1,540	5,806	17,487	2,323	3,597	13,800	17,293	99	-	-	-	61,947
Ext_I5	1,172	4,681	13,121	1,794	2,500	15,117	9,442	45	-	-	-	47,871
TOTAL	**2,944**	**11,179**	**32,115**	**4,441**	**6,717**	**29,199**	**27,000**	**190**	**3,953**	**61,851**	**47,723**	**227,313**

TABLE 46: EXTERNAL TRIPS DISTRICT FLOWS DIFF (COUNT)— (MODEL – AIRSAGE)

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	Ext_I8	Ext_I15	Ext_I5	TOTAL
Downtown	-	-	-	-	-	-	-	-	-986	-2,759	-6,818	**-10,563**
Central	-	-	-	-	-	-	-	-	-6,398	-13,788	-21,910	**-42,096**
North City	-	-	-	-	-	-	-	-	-9,225	-23,563	-47,512	**-80,300**
South Suburban	-	-	-	-	-	-	-	-	-2,653	-5,322	-12,519	**-20,494**
East Suburban	-	-	-	-	-	-	-	-	-5,597	-5,439	-9,162	**-20,198**
North County West	-	-	-	-	-	-	-	-	-3,902	-23,866	-56,622	**-84,390**
North County East	-	-	-	-	-	-	-	-	-2,043	-48,528	-26,571	**-77,141**
East County	-	-	-	-	-	-	-	-	-7,861	-3,133	-1,550	**-12,544**
Ext_I8	-1,361	-7,526	-11,061	-2,625	-5,558	-4,684	-2,211	-7,432	-	-	-	**-42,458**
Ext_I15	-2,530	-10,920	-18,082	-4,983	-4,584	-18,455	-39,805	-2,799	-	-	-	**3,502**
Ext_I5	-7,298	-22,604	-47,024	-12,166	-8,949	-54,729	-26,707	-1,317	-	-	-	**-57,789**
TOTAL	**-11,189**	**-41,050**	**-76,167**	**-19,773**	**-19,091**	**-77,869**	**-68,724**	**-11,547**	**-38,664**	**-126,398**	**-182,664**	**-673,136**

TABLE 47: EXTERNAL TRIPS DISTRICT FLOWS (SHARE)—AIRSAGE

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	Ext_I8	Ext_I15	Ext_I5	TOTAL
Downtown	-	-	-	-	-	-	-	-	0.1%	0.5%	0.9%	1.5%
Central	-	-	-	-	-	-	-	-	0.8%	2.2%	3.0%	5.9%
North City	-	-	-	-	-	-	-	-	1.2%	4.6%	6.7%	12.5%
South Suburban	-	-	-	-	-	-	-	-	0.3%	0.8%	1.6%	2.8%
East Suburban	-	-	-	-	-	-	-	-	0.7%	1.0%	1.3%	3.0%
North County West	-	-	-	-	-	-	-	-	0.5%	4.2%	8.0%	12.6%
North County East	-	-	-	-	-	-	-	-	0.3%	7.3%	4.0%	11.6%
East County	-	-	-	-	-	-	-	-	0.9%	0.4%	0.2%	1.4%
Ext_I8	0.2%	0.9%	1.4%	0.3%	0.7%	0.6%	0.3%	0.8%	-	-	-	5.2%
Ext_I15	0.5%	1.9%	4.0%	0.8%	0.9%	3.6%	6.3%	0.3%	-	-	-	6.5%
Ext_I5	0.9%	3.0%	6.7%	1.6%	1.3%	7.8%	4.0%	0.2%	-	-	-	11.7%
TOTAL	**1.6%**	**5.8%**	**12.0%**	**2.7%**	**2.9%**	**11.9%**	**10.6%**	**1.3%**	**4.7%**	**20.9%**	**25.6%**	**100.0%**

TABLE 48: EXTERNAL TRIPS DISTRICT FLOWS (SHARE)—MODEL

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	Ext_I8	Ext_I15	Ext_I5	TOTAL
Downtown	-	-	-	-	-	-	-	-	0.1%	0.7%	0.5%	1.3%
Central	-	-	-	-	-	-	-	-	0.3%	2.5%	2.1%	4.9%
North City	-	-	-	-	-	-	-	-	0.7%	7.7%	5.7%	14.1%
South Suburban	-	-	-	-	-	-	-	-	0.1%	1.0%	0.8%	1.9%
East Suburban	-	-	-	-	-	-	-	-	0.3%	1.6%	1.1%	2.9%
North County West	-	-	-	-	-	-	-	-	0.1%	6.1%	6.6%	12.8%
North County East	-	-	-	-	-	-	-	-	0.1%	7.6%	4.1%	11.9%
East County	-	-	-	-	-	-	-	-	0.0%	0.0%	0.0%	0.1%
Ext_I8	0.1%	0.3%	0.7%	0.1%	0.3%	0.1%	0.1%	0.0%	-	-	-	1.7%
Ext_I15	0.7%	2.6%	7.7%	1.0%	1.6%	6.1%	7.6%	0.0%	-	-	-	27.3%
Ext_I5	0.5%	2.1%	5.8%	0.8%	1.1%	6.7%	4.2%	0.0%	-	-	-	21.1%
TOTAL	**1.3%**	**4.9%**	**14.1%**	**2.0%**	**3.0%**	**12.8%**	**11.9%**	**0.1%**	**1.7%**	**27.2%**	**21.0%**	**100.0%**

TABLE 49: EXTERNAL TRIPS DISTRICT FLOWS DIFF (SHARE)— (MODEL-AIRSAGE)/AIRSAGE

	Downtown	Central	North City	South Suburban	East Suburban	North County West	North County East	East County	Ext_I8	Ext_I15	Ext_I5	TOTAL
Downtown	-	-	-	-	-	-	-	-	-81.0%	-64.2%	-85.4%	**-78.3%**
Central	-	-	-	-	-	-	-	-	-90.3%	-70.4%	-82.4%	**-79.1%**
North City	-	-	-	-	-	-	-	-	-86.0%	-57.4%	-78.4%	**-71.5%**
South Suburban	-	-	-	-	-	-	-	-	-89.1%	-69.6%	-87.5%	**-82.2%**
East Suburban	-	-	-	-	-	-	-	-	-90.1%	-60.2%	-78.6%	**-75.1%**
North County West	-	-	-	-	-	-	-	-	-93.3%	-63.4%	-79.0%	**-74.3%**
North County East	-	-	-	-	-	-	-	-	-88.6%	-73.7%	-73.8%	**-74.1%**
East County	-	-	-	-	-	-	-	-	-99.4%	-96.9%	-97.2%	**-98.5%**
Ext_I8	-85.5%	-91.6%	-88.0%	-89.0%	-90.0%	-94.3%	-89.3%	-99.4%	-	-	-	**-91.5%**
Ext_I15	-62.2%	-65.3%	-50.8%	-68.2%	-56.0%	-57.2%	-69.7%	-96.6%	-	-	-	**6.0%**
Ext_I5	-86.2%	-82.8%	-78.2%	-87.1%	-78.2%	-78.4%	-73.9%	-96.7%	-	-	-	**-54.7%**
TOTAL	**-79.2%**	**-78.6%**	**-70.3%**	**-81.7%**	**-74.0%**	**-72.7%**	**-71.8%**	**-98.4%**	**-90.7%**	**-67.1%**	**-79.3%**	**-74.8%**

5.5 | TRIP LENGTHS

Trip lengths are generated by mapping AirSage zones to model TAZs and obtaining distance skim values corresponding to zone pairs. As external trips in the model do not have exact locations of their external trip end, only internal trip lengths are compared with the AirSage data.

RESIDENT TRIPS

On average, the resident internal trips predicted by the model are shorter than the trips in the AirSage data (Table 50). As shown in Figure 14, the model generates more trips that are less than 6 miles in length and fewer trips that are longer.

Trip lengths for the HBW and the NHB trip purposes show significant differences. For the two purposes, the trips lengths in the model are approximately 40% shorter than the AirSage data. Trip length frequency distributions by purpose are shown in Figure 15, Figure 16, and Figure 17. The distributions show the same pattern of more trips of shorter lengths and fewer trips of longer lengths. It should be noted that the AirSage data reports a significantly longer trip length than any of the surveys of San Diego County resident trips reported in Chapter 4 of this report.

TABLE 50: AVG. TRIP LENGTH (MILES) OF RESIDENT INTERNAL TRIPS

	HBO	HBW	NHB	ALL
AIRSAGE	7.67	11.98	10.34	**9.32**
MODEL	6.12	7.85	6.15	**6.58**

FIGURE 14: TRIP-LENGTH FREQUENCY FOR RESIDENT INTERNAL TRIPS—ALL

FIGURE 15: TRIP-LENGTH FREQUENCY FOR RESIDENT INTERNAL TRIPS—HBW

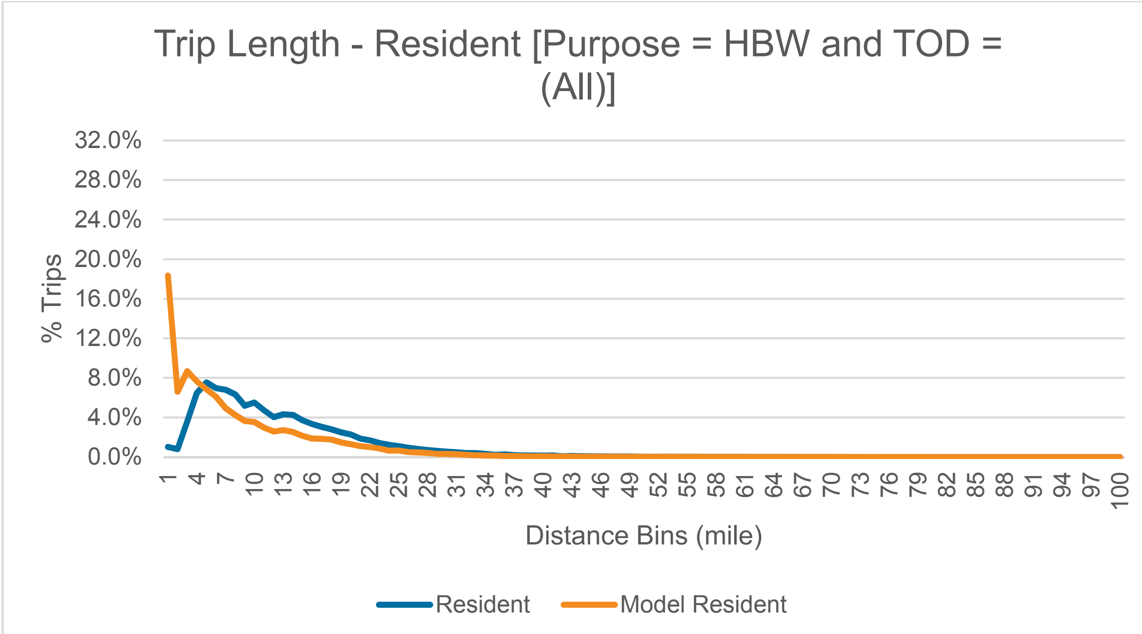


FIGURE 16: TRIP-LENGTH FREQUENCY FOR RESIDENT INTERNAL TRIPS—HBO

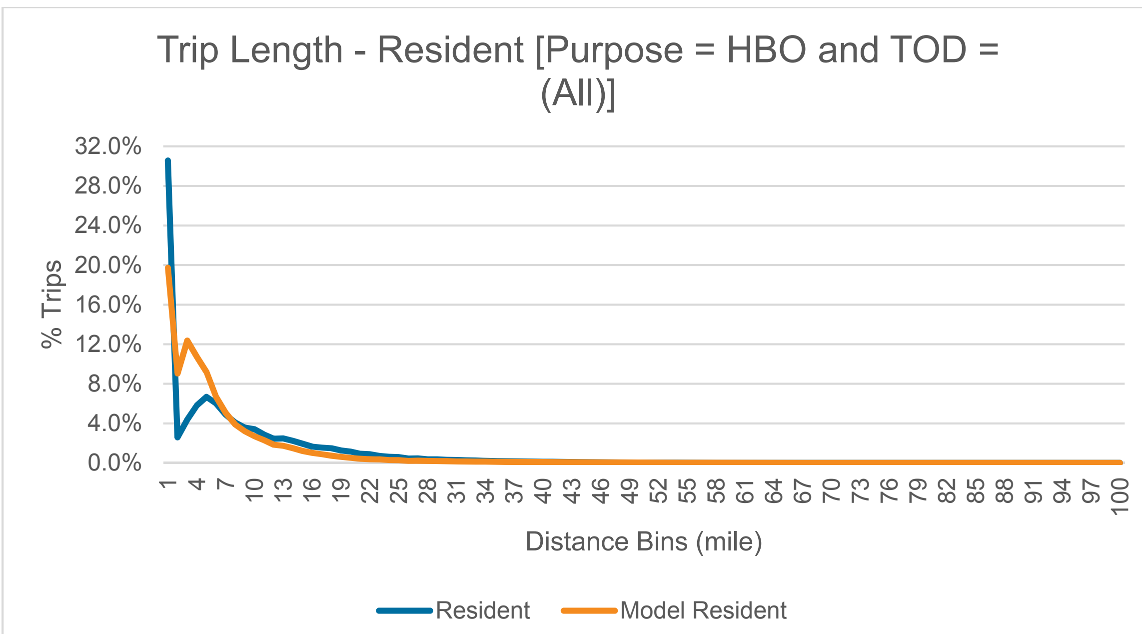
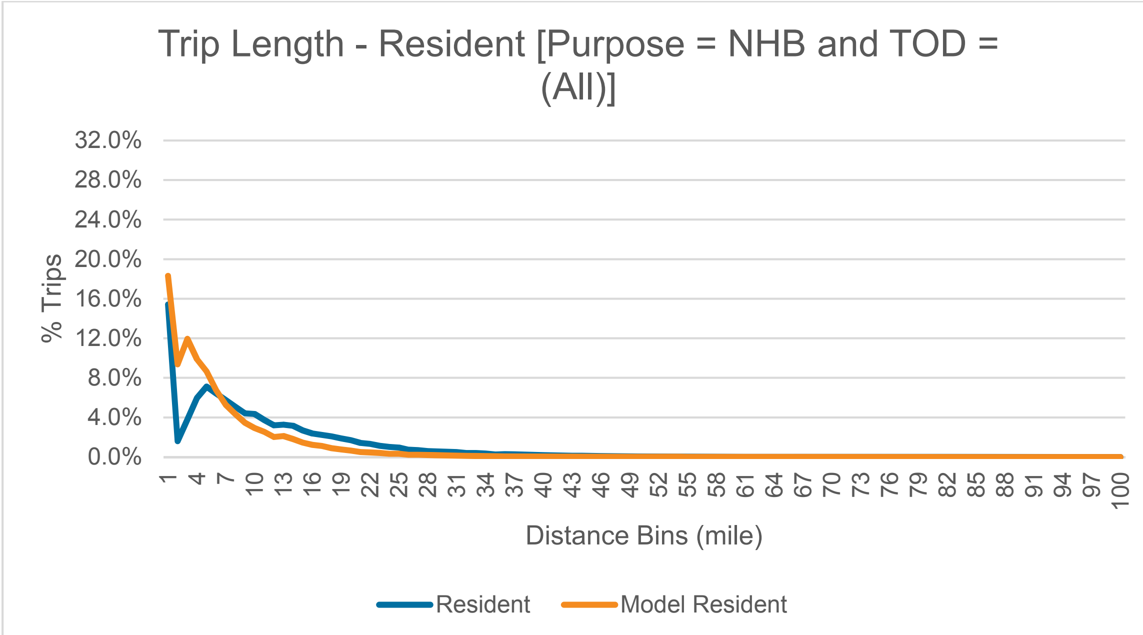


FIGURE 17: TRIP-LENGTH FREQUENCY FOR RESIDENT INTERNAL TRIPS—NHB



Trip-length frequency distributions across time periods were also examined as part of this analysis. However, the summaries by time period do not provide additional information and so are excluded from this report.

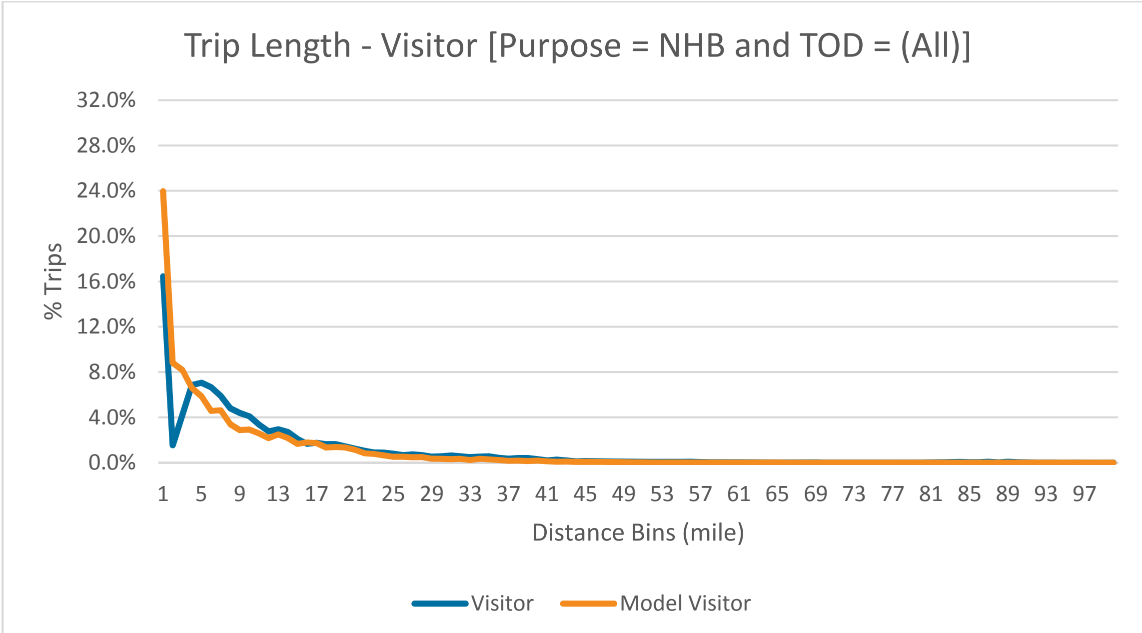
VISITOR TRIPS

As was the case with resident trips, visitor trip lengths are also shorter in the model (Table 51). As shown in Figure 18, this is due to high number of shorter trips (less than 4 miles) and fewer longer trips (more than 4 miles) in the model. Note that the model generates only NHB trips for visitors; therefore, the model trips are compared with the NHB trips reported in the AirSage data.

TABLE 51: AVG. TRIP LENGTH OF VISITOR INTERNAL TRIPS—NHB

	TRIP LENGTH (MILE)
AIRSAGE	11.23
MODEL	7.71

FIGURE 18: TRIP-LENGTH FREQUENCY DISTRIBUTION FOR VISITOR TRIPS—ALL



5.6 | SUMMARY AND CONCLUSIONS

The following observations are derived from the comparisons of internal trips made by residents.

- The total magnitude of trips and the distribution of trips by time of day match well in the two datasets, though the AirSage data has slightly more trips than the resident model predicts, considering differences in years between the model and the AirSage data.
- The HBO trips are similar in the two datasets; however, the disaggregate model predicts more and fewer trips for HBW and NHB purposes, respectively. This may be due to errors introduced in the AirSage trip imputation process.
- District-to-district flows compare well; the SANDAG model predicts slightly lower flows for East Suburban and North County West districts, and slightly higher flows for South Suburban district.
- On average, the SANDAG model generates more trips with shorter lengths. However, there may be bias introduced in the aggregation process from model geography to AirSage geography.
- Model trip lengths for HBW and NHB trip purposes are approximately 40% shorter than the trip lengths in the AirSage data.

Given the magnitude of differences between the model estimate of visitor and external trips compared to AirSage trips, we do not make any specific conclusions regarding these comparisons. We recommend further investigation into visitor and external travel in the CHTS.

6.0 ATRI ANALYSIS

6.1 | ATRI DATA DESCRIPTION

As part of the validation effort, the CVM output was compared with ATRI data obtained for the SANDAG model region. ATRI provided SANDAG both the processed and raw version of ATRI's Freight Performance Measures dataset; real-time anonymized freight truck data sourced through industry partnerships. These data include periodic time, location, speed and anonymous unique identification information for a subset of truck trips. The data for San Diego includes only trips that had an origin within San Diego County. Once a trip left the county, only the first point outside of the county was recorded and the appropriate exit station TAZ ID was attached to that point. There were 12 weeks of data included in the deliverable, with each week number corresponding to the weeks shown in Table 52.

TABLE 52: STUDY WEEKS

WEEK NUMBER	START DAY (12:00:00 A.M.)	END DAY (11:59:59 P.M.)
1	08/13/2012	08/17/2012
2	09/10/2012	09/14/2012
3	10/08/2012	10/12/2012
4	11/05/2012	11/09/2012
5	12/10/2012	12/14/2012
6	01/14/2013	01/18/2013
7	02/04/2013	02/08/2013
8	03/11/2013	03/15/2013
9	04/08/2013	04/12/2013
10	05/13/2013	05/17/2013
11	06/10/2013	06/14/2013
12	07/08/2013	07/12/2013

6.2 | IDENTIFICATION OF TRIPS FROM GPS TRACES

SANDAG provided a GIS file representing TAZs to ATRI. ATRI selects all truck GPS traces entering, exiting, traveling within or passing through the San Diego region for each of the sampled weeks. ATRI uses a data management and analysis software package to further prepare the dataset for integration into the truck trip table. Truck positions for each unique vehicle are sorted into a time series, and within each series each truck position was matched with the subsequent truck position to produce a set of truck position pairs. The geodetic distance between the first and second truck positions for each of the truck position pairs was then calculated. ATRI then replaces the precise

GPS location data in its records with the TAZ. In addition to supporting the ultimate development of a trip table, this process also offers some benefit of further ensuring the anonymity of the data by associating truck positions with geographic areas far more generalized than a discrete latitude/longitude position – which could allow for the development of an address-specific customer list. The dataset is then reformatted so that each record represents the movement of a truck between GPS 'pings'. ATRI then delivers a dataset containing an anonymous truck identifier, the distance between pings, the TAZ position of the beginning and ending ping and the timestamp of the beginning and ending ping.

In addition, ATRI further processed the data product described above with the goal to determine when a vehicle is stopping at a destination as opposed to merely stopping at a traffic signal or other interim stops. For this step, it was assumed by ATRI that a truck must be stationary for at least 30 minutes for a destination to be recorded. Once a vehicle is stopped and a trip has been classified as "ended", a new trip did not begin unless the vehicle travels at least one-quarter of a mile (this is done to filter out short movements, e.g. movement within a distribution facility). Experience with the ATRI data suggests that the 30-minute stopping window is too long to define a trip and probably misses several trips with shorter stopping times. It was thus decided that the initial dataset needed to be reprocessed with new thresholds to identify truck trips from the TAZ tagged GPS traces to represent trips between origin-destination pairs (Figure 19).

FIGURE 19: IDENTIFYING STOPS FROM GPS TRACE DATA

from TAZ	to TAZ	distance	time	elapsed time	speed	status1	status2
10	101032	66.0	57.7	57.7	68.6	moving	moving
101032	101033	16.3	14.3	72.0	68.6	moving	moving
101033	101015	26.8	27.9	99.9	57.5	moving	moving
101015	101015	0.0	5.0	5.0	0.0	stopped	stopped
101015	101015	0.2	2.7	7.7	5.2	stopped	stopped
101015	101015	0.3	9.8	17.5	2.0	stopped	stopped
101015	101015	0.1	0.3	0.3	28.2	moving	*stopped?*
101015	2035	37.1	60.0	60.3	37.1	moving	moving
2035	18099	67.8	65.4	125.7	62.2	moving	moving
18099	27006	5.9	5.4	131.1	65.3	moving	moving
27006	18023	10.0	15.9	147.0	37.8	moving	moving
18023	18023	0.0	5.0	5.0	0.0	stopped	stopped

Trip	O	D
1	10	101015
2	101015	18023

This is done in two steps, first identifying for each record whether the truck was in motion or stopped. This determination is made based on the set of criteria for minimum travel speed and a minimum elapsed time and/or distance in multiple steps described below:

- First all records where the speed is over 150 MPH are removed. The speed is calculated using the distance and time information provided by ATRI.
- Then a threshold of 8 MPH is used tag records as stopping or moving (status1)
- In the next pass through the dataset, all records that have been stopped for under five min as assumed to be moving and are grouped together with the previous records (status2).
- The records are then collapsed to get a trip table between OD pairs

The criteria suggested above are necessary to avoid including brief stops at traffic signals or brief repositioning movements within a single site (see Figure 20 for an illustration).

FIGURE 20: BRIEF REPOSITIONING MOVEMENTS TO BE DISTINGUISHED FROM TRIPS

Once the moving records and stopped (speed less than 8 MPH and stopped for over 5 minutes) records are identified, the records are processed to identify the origin and destination for each sequence of moving records. When a stop record was found in the list, it signified the destination of the trip and the origin for the subsequent trip (Figure 19). The result is a list of trips by origin-destination pair, which can be aggregated by origin-destination pair to produce a trip table in flat/list format, which, in turn, can be read into a matrix format file by most travel modeling software.

The resulting trip table still must be cleaned for several reasons. The largest issue was GPS positional errors, or "blips," where the GPS location jumps from one location to another in a way that could not possibly represent a real movement (e.g., a change in position of 50 miles in a span of 30 seconds). Given the size of the sample data, it was not necessary to invest large effort to correct these blips; rather, trips with such errors were simply identified and removed from the dataset. These records were less than 1% of the overall dataset. Moreover, a conservative speed of 150 mph was used to mark what qualified as a good trip.

Some trips at the very beginning and end of each study period were removed to avoid capturing trip fragments or partial trips in progress at the beginning or end of the period. If a truck is initially moving (no starting records of a stop) within the first hour of the start time, then those records were flagged. This time was determined to be a reasonable buffer by looking at the starting time distribution of all initial trips. Similarly, trucks that did not display a final stop and had movement within 3 hours of the end time were flagged. After a look at ping length and ending trip distribution, it was assumed that if a truck displayed no pings for over 3 hours that its trip had ended.

For each trip, its GPS calculated length was also compared to a centroid-to-centroid geodetic distance. Trips were flagged if the ratio was outside the bounds of 0.25 & 3. This was used to catch both blips that slipped through the initial filter as well as undetected stops and helped to confirm "clean" probable trips. All these filters combined removed approximately 0.5% of the overall dataset.

A small number of trips appeared to start and end in the same zone (an intrazonal trip), but with unreasonable VMT. Considering individual pings, it appears that these trips went through several zones and made a large circular trip. This seemed be from either a brief trip outside the model or in many cases an undetected stop. Intrazonal trips greater than 30 miles were generally flagged and removed. No records were found or removed due to this criterion.

When comparing the data processed by ATRI and RSG at the end of this reprocessing step, the number of truck trips increased from approximately 99k to 113k.

6.3 | SCALING ATRI DATA TO COUNTS

The final effort is on the expansion of this resulting raw trip table. The simplest method is to scale the trip table to reflect the total number of truck trips or truck VMT. However, this fails to account for differences in the portion of the universe of trucks represented in the sample. For example, it is known that short-haul movements, while present in the data, are under-represented; without correcting for this, it is not possible to produce accurate information regarding average trip lengths, etc. There was no initial hypothesis as to whether there were any geographic biases in the data, but it was considered a possibility, since the sample is not randomly drawn. If there were geographic biases it would be important to correct for this to avoid distorting the spatial distribution of trips.

The approach taken in this effort begins by simply scaling the raw ATRI trip table to represent the proper amount of truck trips. Studies were done in Iowa and Tennessee to compare the ATRI trip-length frequency distribution with an OD matrix produced via a seed OD matrix and counts and applying the ODME process. When comparing the two Trip-Length Frequency Distributions (TLFD), it was observed that trips needed to be factored up or down based on the trip-length bin and weighting scheme (presented in Table 53) was developed by comparing the ATRI and ODME values in each distance bin. For example, all trips that were between zero to 10 miles long were weighted by a factor of 3.83 and trips that were between 60 and 100 miles long were weighted by a factor of 1.37. For the ATRI matrix the trip originating and destined to a zone were not equal creating truck sinks/sources. Thus, the trip table was transposed and averaged to ensure that row and column marginals are equal.

It is also important to note that the ATRI matrix provided by SANDAG did not include any EE trips or external to internal trips to/from Mexico. To fix this gap in data, the EE trips and the trips between Mexico and the SANDAG region (external stations 1, 2, 3 and 4) were extracted from the CVM output from the SANDAG model and added to the ATRI matrix before assignment.

TABLE 53: WEIGHTING SHORTER TRIPS

DISTANCE BIN	WEIGHT
10	3.83

DISTANCE BIN	WEIGHT
20	2.69
30	2.31
40	1.96
60	1.61
100	1.37
140	1.33
180	1.24
220	1.19
260	1.19
300	1.18
340	1.15
420	1.11
500	1.12
600	1.09
1000	1.21

As a final scaling step, the weighted and balanced ATRI matrix was then assigned to the SANDAG network and adjusted to link counts. The trip table was assigned using an all or nothing assignment algorithm based upon midday congested travel time. Unfortunately, there are relatively few commercial vehicle counts available for San Diego compared to other urban areas (their locations are shown in Figure 21). Truck counts are classified by axle type; medium trucks have 2 axles while heavy trucks have 3+ axles. However, the heavy-truck model classifies vehicles based on their weight (Table 4), making a comparison to truck counts challenging. The conversion table shown in Table 54 was used to make appropriate comparisons.

FIGURE 21: SANDAG COUNT LOCATIONS

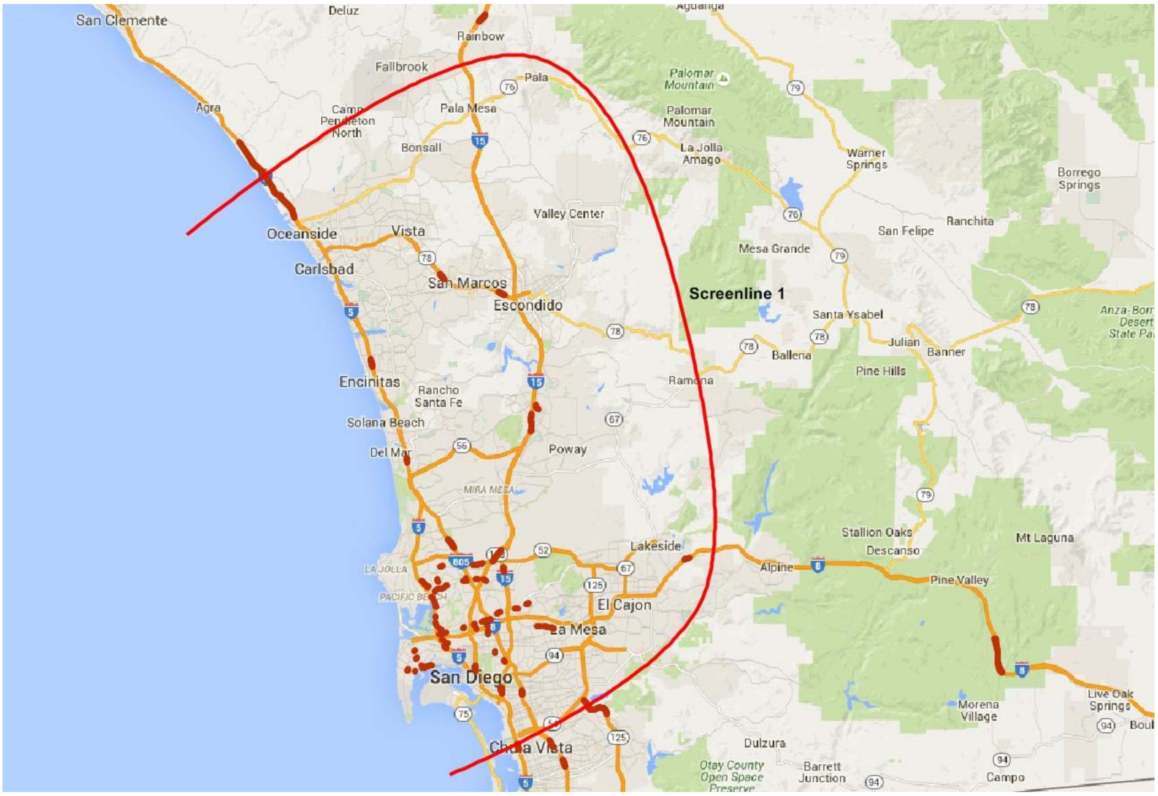


TABLE 54: SANDAG CVM VEHICLE CLASSIFICATION

MODEL MODE NAME	DESCRIPTION	INTERNAL/ EXTERNAL TRIPS	REFERENCE
lhdn	Light Heavy-Duty Nontoll	External	Heavy-Duty Truck
mhdn	Medium Heavy-Duty Nontoll	External	Heavy-Duty Truck
hhdn	Heavy Heavy-Duty Nontoll	External	Heavy-Duty Truck
lhdt	Light Heavy-Duty Toll	External	Heavy-Duty Truck
mhdt	Medium Heavy-Duty Toll	External	Heavy-Duty Truck
hhdt	Heavy Heavy-Duty Toll	External	Heavy-Duty Truck
CVM:LT	Light-Duty Commercial Toll	Internal	Light-Duty
CVM:MT	Light & Medium Heavy-Duty Toll	Internal	Heavy-Duty Truck
CVM:HT	Heavy Heavy-Duty Toll	Internal	Heavy-Duty Truck
CVM:LN	Light-Duty Commercial Nontoll	Internal	Light-Duty
CVM:MN	Light & Medium Heavy-Duty	Internal	Heavy-Duty Truck
CVM:HN	Heavy Heavy-Duty Nontoll	Internal	Heavy-Duty Truck
ExtLDN	Light-Duty Commercial Nontoll	External	Light-Duty
ExtLDT	Light-Duty Commercial Toll	External	Light-Duty

Two different factoring methods were used to scale the ATRI trip table to counts. One method relied on only the ratio between total estimated versus observed volumes on counted links. The second method utilized a screenline shown in red on Figure 21 in addition to the counts. In the second method, three ratios were calculated using the assignment flows from the ATRI matrix and counts on the SANDAG network:

1. All links with counts
2. Counts that are close to the screenline and travel from inside to outside the screenline
3. Counts that are close to the screenline and travel from outside to inside the screenline.

These ratios were applied to the appropriate cells of the ATRI matrix, namely: (1) to all cells, (2) to cells that have an origin inside the screenline and a destination outside the screenline, and (3) to cells that have an origin outside the screenline and a destination inside the screenline. Note that (2) and (3) are mutually exclusive sets. The adjusted matrix was reassigned to the network and the process was repeated until the ratio between total counted versus estimated volume from one iteration to the next changed by less than 0.0001.

The factoring process was applied to a heavy-truck trip table compared to 3+ axle counts, and to a trip table of both medium and heavy trucks, compared to 2+ axle truck counts. Tables below present the results from the four combinations of truck trips and scaling method. The tables show that the assignment of medium and heavy trucks compared to the relevant commercial vehicle counts results in lower error than just heavy trucks compared to 3+ axle counts. They also show that the inclusion of screenline counts as an additional factoring step improves the goodness of fit for medium and heavy trucks but not for heavy trucks alone.

TABLE 55: COUNT VS MODEL VOLUME—MEDIUM + HEAVY TRUCKS ALL REGION SCALING

ITEM	NUMBER OF OBS	TOTAL COUNT VOLUMES	TOTAL MODEL VOLUMES	PERCENT ERROR	PERCENT RMSE
FREEWAY	47	135,945	153,755	13.10	55.32
PRIME ARTERIAL	9	18,048	14,539	-19.44	147.30
MAJOR ARTERIAL	29	22,327	12,945	-42.02	161.59
COLLECTOR	5	3,170	879	-72.28	117.59
0 TO 1000 AADT	2	364	2,467	577.71	592.40
1001 TO 5000 AADT	13	4,003	4,398	9.87	383.84
5001 TO 9,999 AADT	17	12,842	8,715	-32.13	152.64
10,000 TO 19,999 AADT	20	12,374	5,037	-59.30	115.95
20,000 TO 49,999 AADT	18	30,141	33,776	12.06	135.71
OVER 49,999 AADT	35	122,817	128,137	4.33	48.00

TABLE 56: COUNT VS MODEL VOLUME—MEDIUM + HEAVY TRUCKS ALL REGION SCALING + SCREENLINE

ITEM	NUMBER OF OBS	TOTAL COUNT VOLUMES	TOTAL MODEL VOLUMES	PERCENT ERROR	PERCENT RMSE
FREEWAY	47	135,945	159,792	17.54	49.50
PRIME ARTERIAL	9	18,048	10,720	-40.61	116.67
MAJOR ARTERIAL	29	22,327	11,192	-49.87	162.94
COLLECTOR	5	3,170	538	-83.02	120.20
0 TO 1000 AADT	2	364	2,207	506.19	521.47
1001 TO 5000 AADT	13	4,003	4,211	5.20	384.84
5001 TO 9,999 AADT	17	12,842	8,359	-34.91	151.20
10,000 TO 19,999 AADT	20	12,374	4,446	-64.07	125.69
20,000 TO 49,999 AADT	18	30,141	28,551	-5.28	110.09
OVER 49,999 AADT	35	122,817	134,767	9.73	43.08

TABLE 57: COUNT VS MODEL VOLUME—HEAVY TRUCKS ALL REGION SCALING

ITEM	NUMBER OF OBS	TOTAL COUNT VOLUMES	TOTAL MODEL VOLUMES	PERCENT ERROR	PERCENT RMSE
FREEWAY	47	97,701	94,657	-3.12	71.69
PRIME ARTERIAL	9	2,258	6,191	174.16	579.28
MAJOR ARTERIAL	29	5,821	6,157	5.76	276.51
COLLECTOR	5	1,207	308	-74.47	136.55
0 TO 1000 AADT	2	122	1,200	883.28	914.84
1001 TO 5000 AADT	11	2,041	2,263	10.88	333.19
5001 TO 9,999 AADT	17	6,060	4,924	-18.75	144.96
10,000 TO 19,999 AADT	20	2,082	2,570	23.45	347.01
20,000 TO 49,999 AADT	18	7,936	16,556	108.61	247.85
OVER 49,999 AADT	35	89,213	79,941	-10.39	66.04

TABLE 58: COUNT VS MODEL VOLUME—HEAVY TRUCKS ALL REGION SCALING + SCREENLINE

ITEM	NUMBER OF OBS	TOTAL COUNT VOLUMES	TOTAL MODEL VOLUMES	PERCENT ERROR	PERCENT RMSE
FREEWAY	47	97,701	90,678	-7.19	75.07
PRIME ARTERIAL	9	2,258	8,716	286.03	817.77
MAJOR ARTERIAL	29	5,821	7,314	25.65	279.80
COLLECTOR	5	1,207	533	-55.86	137.31
0 TO 1000 AADT	2	122	1,372	1024.44	1052.56
1001 TO 5000 AADT	11	2,041	2,382	16.71	330.96
5001 TO 9,999 AADT	17	6,060	5,160	-14.86	147.86
10,000 TO 19,999 AADT	20	2,082	2,961	42.20	311.15
20,000 TO 49,999 AADT	18	7,936	20,009	152.13	342.75
OVER 49,999 AADT	35	89,213	75,570	-15.29	68.60

In summary, since ATRI's sample rate of the actual commercial vehicle fleet is unknown, it was necessary to scale the trips based on traffic counts. While this ensures the truck travel patterns from the ATRI GPS data more accurately represent truck traffic in San Diego, there may still be biases in the ATRI data related either to the representativeness of the trucks in their sample or detection issues related to the identification of trips from GPS traces. These biases could be further reduced by utilizing more screenlines or Origin-Destination Matrix Estimation, but each of these would require more counts than were available for this effort. We were only able to construct one screenline for matrix partitioning using the available counts and the use of this screenline only modestly improved goodness of fit. The comparison of these results with the CVM output is presented below.

6.4 | ATRI COMPARISON TO CVM TRIP TABLES

We compare two sets of factored ATRI data to model output. We compare the ATRI medium + heavy-truck matrices scaled to counts with screenline adjustments. Since the inclusion of screenline factors did not improve the goodness of fit to traffic counts for heavy trucks, heavy-truck ATRI matrices scaled using the simple scaling method without screenline adjustments are compared directly to CVM outputs. Table 59 and Table 60 presents district flow comparisons between ATRI and CVM normalized by total number of trips in the CVM output for heavy trucks and medium+heavy trucks respectively. Each cell in Table 59 and Table 60 presents the difference between CVM and ATRI flow (CVM minus ATRI) for the given district pair divided by the total number of trips from CVM model output. The districts in these tables are based upon MSAs shown in Figure 13.

It can be noted from Table 59 and Table 60 that CVM predicts 46.9% more Heavy trips (+52k) and 60.2% more Heavy + Medium trips (+214k) compared to the ATRI data. However, much of this over-prediction is for inter-MSA flows, indicating that model has more short-distance intra-MSA trips versus the ATRI data. As stated in the previous section, the ATRI data under-represents short-distance travel. Although short-distance bias factors were applied to the data, it is possible that they did not fully compensate for this bias, especially since the factors were borrowed from other urban areas. Furthermore, most of the truck counts used for matrix expansion are on freeways as opposed to arterials where there are likely to be more short-distance truck trips. Therefore, an over-prediction of short-distance trips in Table 59 and Table 60 is expected.

Figure 22 and Figure 23 presents TLFD comparisons for heavy and medium+heavy trucks respectively. The TLFDs show a good match in overall pattern between CVM and ATRI but a clear bias in trip length can also be observed. The average trip length in ATRI data for Heavy trucks is 21.93 miles compared to 15.85 miles in the CVM output. Similarly, average trip length for Heavy + Medium trucks in ATRI data is 16.17 miles compared to 13.26 miles in CVM output. The bias in average trip lengths further support the observations from the MSA-MSA comparison tables.

To make observations about effects in the model other than the short trip bias, comparison tables were adjusted to exclude the diagonals and North County East and West interactions, which represents much of the short-distance travel. Table 61 and Table 62 presents the adjusted comparisons, where each cell is still normalized by the total number of trips in the CVM output. It can be noted that model still over-predicts truck trips for both Heavy and Heavy + Medium truck cases, but at a much better range of 9 to 18%. Much of the over-prediction in the model output is for trips to/from North City MSA.

TABLE 59: ATRI VS. CVM DISTRICT COMPARISON, HEAVY TRUCKS

	CENTRAL	NORTH CITY	SOUTH SUBURBAN	EAST SUBURBAN	NORTH COUNTY WEST	NORTH COUNTY EAST	EAST COUNTY	EXTERNAL	TOTAL
CENTRAL	1.7%	1.0%	0.4%	0.7%	0.0%	0.0%	0.0%	0.1%	**3.7%**
NORTH CITY	1.0%	11.1%	0.0%	1.3%	0.7%	0.8%	0.0%	0.7%	**15.4%**
SOUTH SUBURBAN	0.4%	-0.1%	0.2%	0.3%	-0.1%	-0.1%	0.0%	-0.2%	**0.4%**
EAST SUBURBAN	0.6%	1.3%	0.3%	3.1%	0.0%	0.0%	0.1%	0.2%	**5.7%**
NORTH COUNTY WEST	0.0%	0.7%	-0.1%	0.0%	5.5%	3.1%	0.0%	-0.4%	**8.8%**
NORTH COUNTY EAST	0.0%	0.8%	-0.1%	0.0%	3.2%	9.1%	0.0%	-0.4%	**12.5%**
EAST COUNTY	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.8%	-0.2%	**0.5%**
EXTERNAL	0.2%	1.0%	-0.2%	0.1%	-0.5%	-0.7%	-0.2%	0.0%	**-0.3%**
TOTAL	**3.8%**	**15.7%**	**0.5%**	**5.7%**	**8.8%**	**12.3%**	**0.5%**	**-0.3%**	**46.9%**

NOTE: COMPARISON REPRESENTS CVM MINUS ATRI

TABLE 60: ATRI VS. CVM DISTRICT COMPARISON, MEDIUM + HEAVY TRUCKS

	CENTRAL	NORTH CITY	SOUTH SUBURBAN	EAST SUBURBAN	NORTH COUNTY WEST	NORTH COUNTY EAST	EAST COUNTY	EXTERNAL	TOTAL
CENTRAL	2.7%	0.7%	1.5%	1.0%	0.1%	0.1%	0.0%	0.5%	**6.6%**
NORTH CITY	0.7%	9.0%	0.5%	1.2%	0.9%	1.1%	0.0%	1.3%	**14.7%**
SOUTH SUBURBAN	1.5%	0.4%	1.7%	0.5%	-0.1%	-0.1%	0.0%	-0.2%	**3.6%**
EAST SUBURBAN	0.9%	1.2%	0.5%	4.6%	0.1%	0.1%	0.0%	0.3%	**7.7%**
NORTH COUNTY WEST	0.1%	0.9%	-0.1%	0.1%	6.7%	3.4%	0.0%	0.0%	**11.1%**
NORTH COUNTY EAST	0.1%	1.1%	-0.1%	0.1%	3.4%	10.3%	0.0%	-0.2%	**14.7%**
EAST COUNTY	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	-0.2%	**0.4%**
EXTERNAL	0.6%	1.5%	-0.2%	0.3%	-0.1%	-0.3%	-0.2%	0.0%	**1.5%**
TOTAL	**6.5%**	**14.8%**	**3.7%**	**7.7%**	**11.1%**	**14.6%**	**0.4%**	**1.6%**	**60.2%**

NOTE: COMPARISON REPRESENTS CVM MINUS ATRI

FIGURE 22: TLFD COMPARISON—ATRI VS. CVM, HEAVY TRUCKS

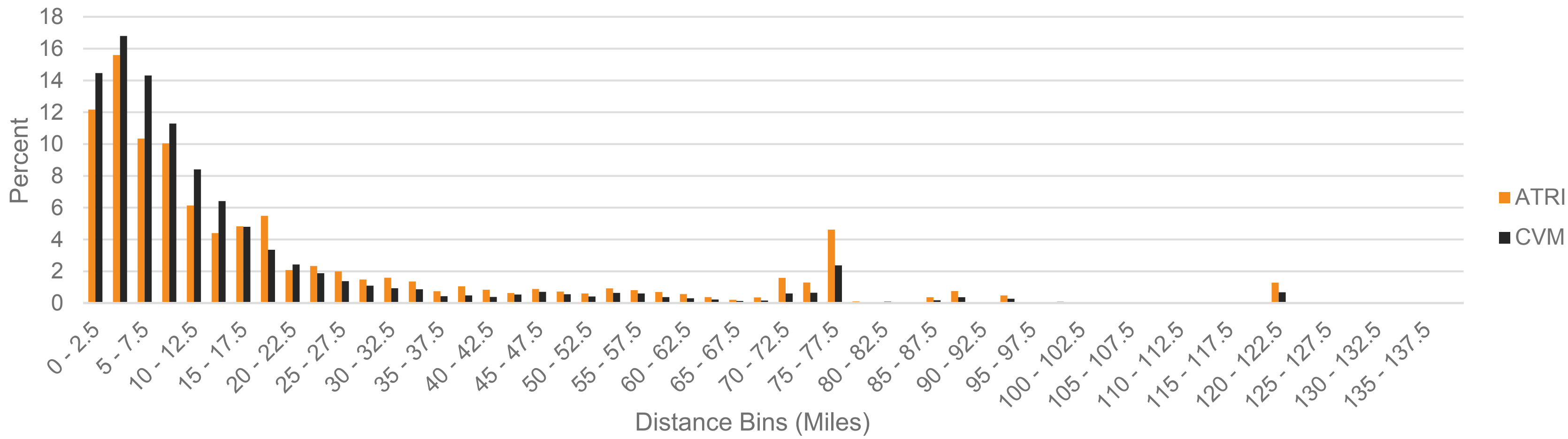


FIGURE 23: TLFD COMPARISON—ATRI VS. CVM, MEDIUM + HEAVY TRUCKS

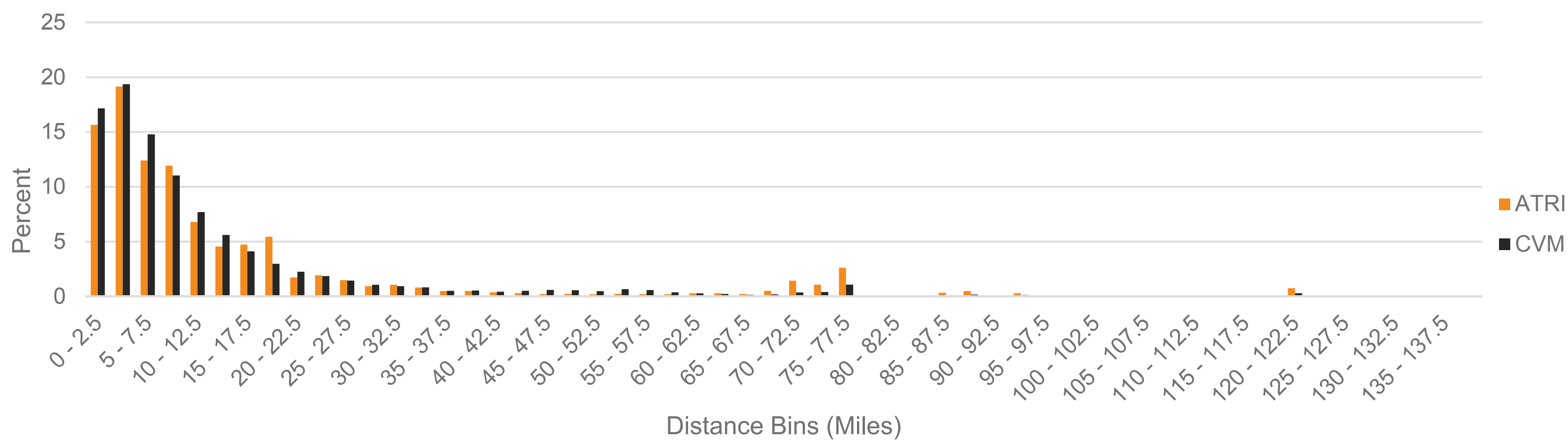


TABLE 61: ADJUSTED ATRI VS. CVM DISTRICT COMPARISON, HEAVY TRUCKS

	CENTRAL	NORTH CITY	SOUTH SUBURBAN	EAST SUBURBAN	NORTH COUNTY WEST	NORTH COUNTY EAST	EAST COUNTY	EXTERNAL	TOTAL
CENTRAL		1.0%	0.4%	0.7%	0.0%	0.0%	0.0%	0.1%	**2.0%**
NORTH CITY	1.0%		0.0%	1.3%	0.7%	0.8%	0.0%	0.7%	**4.4%**
SOUTH SUBURBAN	0.4%	-0.1%		0.3%	-0.1%	-0.1%	0.0%	-0.2%	**0.2%**
EAST SUBURBAN	0.6%	1.3%	0.3%		0.0%	0.0%	0.1%	0.2%	**2.6%**
NORTH COUNTY WEST	0.0%	0.7%	-0.1%	0.0%			0.0%	-0.4%	**0.1%**
NORTH COUNTY EAST	0.0%	0.8%	-0.1%	0.0%			0.0%	-0.4%	**0.3%**
EAST COUNTY	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%		-0.2%	**-0.2%**
EXTERNAL	0.2%	1.0%	-0.2%	0.1%	-0.5%	-0.7%	-0.2%		**-0.3%**
TOTAL	**2.1%**	**4.7%**	**0.2%**	**2.5%**	**0.0%**	**0.1%**	**-0.2%**	**-0.3%**	**9.1%**

TABLE 62: ADJUSTED ATRI VS. CVM DISTRICT COMPARISON, MEDIUM + HEAVY TRUCKS

	CENTRAL	NORTH CITY	SOUTH SUBURBAN	EAST SUBURBAN	NORTH COUNTY WEST	NORTH COUNTY EAST	EAST COUNTY	EXTERNAL	TOTAL
CENTRAL		0.7%	1.5%	1.0%	0.1%	0.1%	0.0%	0.5%	**3.8%**
NORTH CITY	0.7%		0.5%	1.2%	0.9%	1.1%	0.0%	1.3%	**5.7%**
SOUTH SUBURBAN	1.5%	0.4%		0.5%	-0.1%	-0.1%	0.0%	-0.2%	**2.0%**
EAST SUBURBAN	0.9%	1.2%	0.5%		0.1%	0.1%	0.0%	0.3%	**3.1%**
NORTH COUNTY WEST	0.1%	0.9%	-0.1%	0.1%			0.0%	0.0%	**1.0%**
NORTH COUNTY EAST	0.1%	1.1%	-0.1%	0.1%			0.0%	-0.2%	**1.0%**
EAST COUNTY	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		-0.2%	**-0.2%**
External	0.6%	1.5%	-0.2%	0.3%	-0.1%	-0.3%	-0.2%		**1.5%**
TOTAL	**3.8%**	**5.8%**	**2.0%**	**3.0%**	**1.0%**	**0.9%**	**-0.2%**	**1.6%**	**17.9%**

6.5 | ATRI COMPARISON CONCLUSIONS

The following conclusions can be made based upon the comparison of ATRI data to the CVM model:

1) The comparison provides further evidence that the CVM does not under-represent travel for medium and heavy trucks. The CVM consistently estimates more travel than the expanded ATRI data, for both medium and heavy trucks.
2) The ATRI data appears to be biased against short trips.
3) The model comparisons show a better match for heavy trucks alone. It can be concluded that much of the mismatch in CVM versus ATRI can be attributed to under-representation of short-haul truck trips in the ATRI data. Despite this, the shape of the CVM TLFDs compare favorably to the ATRI data.

7.0 INRIX ANALYSIS

Observed vehicle speeds historically have been time consuming and expensive to collect for regional model validation. Data often is only available for freeways from 24/7 loop detector measurements or floating car runs for specific corridors. INRIX speed data has enabled a more cost effective approach to model speed validation. With increased collection and compilation of GPS device readings, INRIX provides continuous observed speed data coverage that can aid in validation of traffic models.

This section compares model link speed from the disaggregate model traffic assignment results to the INRIX observed speeds. Comparisons include an investigation into the regional differences between model and observed travel speeds as well as freeway corridor comparisons. Recommendations for potential model changes and continued study are provided.

7.1 | INRIX DATA BACKGROUND

INRIX is a worldwide supplier of real-time and historic traffic data with specific emphasis on the conversion and compilation of GPS device geolocation traces into speed and travel time along roadway segments. Typical geolocation devices include fleet enabled GPS trackers and in-vehicle navigation devices. Earlier years of INRIX data were more fleet dominant and potentially had small biases built into driving characteristics such as following posted speed limits. Each year more personal vehicles are tracked which increases the number of trace paths along segments throughout the day and ultimately the confidence in the speed information. For roads with less volume, and thus potentially fewer GPS traces, off-peak speed data often reverts to a reference speed or speed limit. This may limit the use of the speed data outside of peak periods and midday time periods for validation.

SANDAG has an INRIX data subscription for 2011 to current year travel times with data provided by traffic message channel (TMC) segmentation. TMCs are maintained by an independent organization, the Traveller Information Services Association, which can be slow to make modifications and additions. The 2012 TMC segmentation does not include notable new roads such as the SR125 toll road, SR52 extension through Santee, or the I-5 local bypass lanes. These facilities were added in a more recent released edition of TMC segments. INRIX's process still collected GPS traces on those segments, but without the TMC reference they had no method for conveyance of the information. INRIX has since developed a new, more detailed segmentation scheme, but it was unavailable to SANDAG at the time of this study. The TMC segments also do not include most ramps or connectors and tend to only cover higher volume roadways. Despite these deficiencies, the data covers 1,850 miles of roadway throughout San Diego County with 727 miles on freeways and ramps and 1,123 on arterials as shown in Figure 24. Figure 25 shows the North County West area and the coverage of higher classification roads in the TMC segmentation and lack of local arterials.

FIGURE 24: SAN DIEGO COUNTY TMC COVERAGE

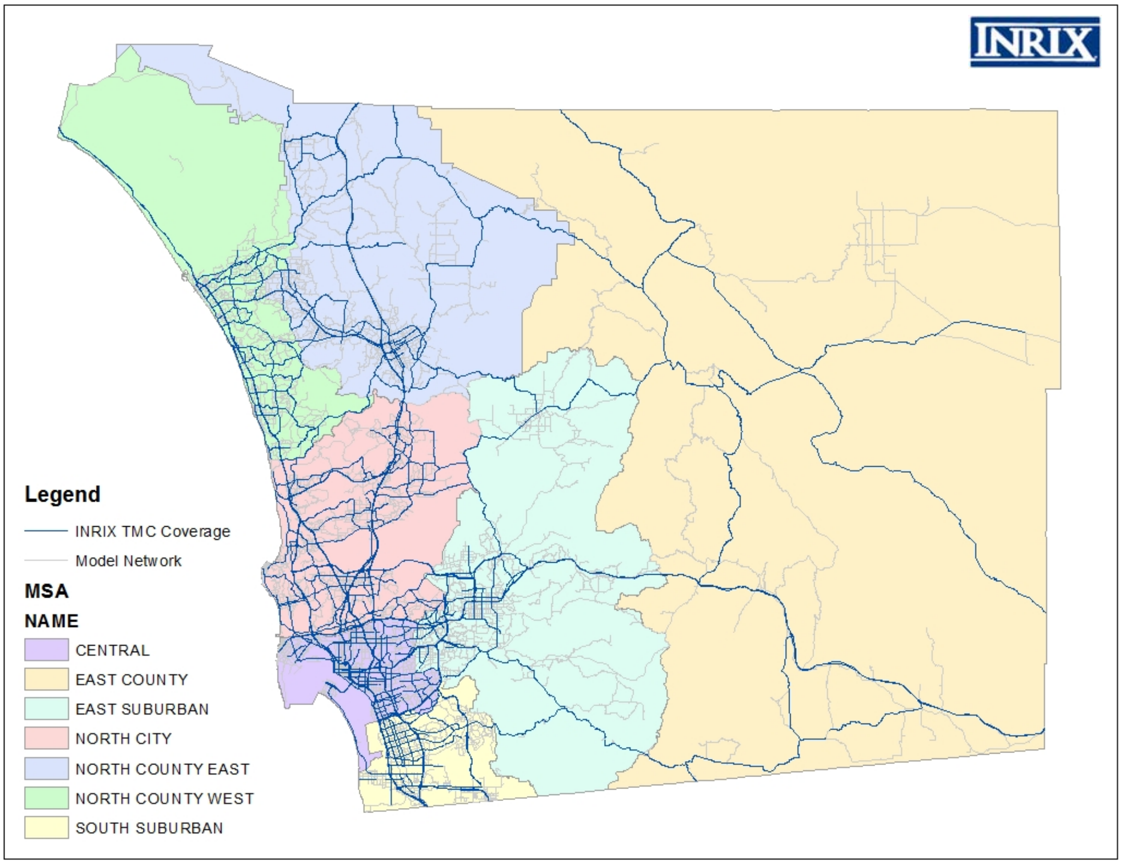
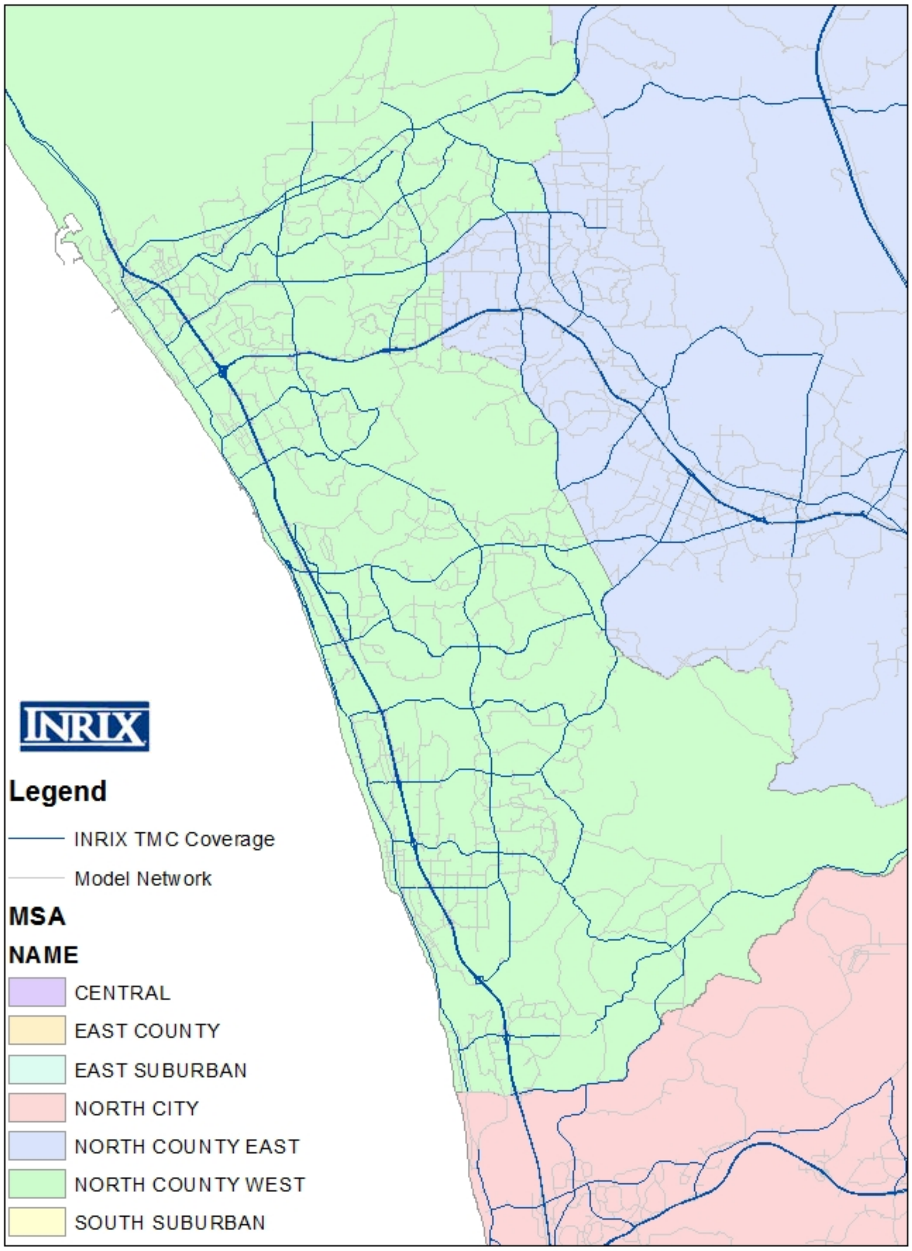


FIGURE 25: NORTH COUNTY WEST TMC COVERAGE



For the purposes of this study, October 2012 was chosen as the comparison time frame with speed data aggregated to compare to the models AM (6:00 a.m.–9:00 a.m.) and PM (3:30 p.m.–7:00 p.m.) peak time period. The INRIX speed data has not been compared to other data sources to prove the validity of the information. Further investigation is warranted to compare to other traditional data sources such as loop detectors and floating car surveys.

7.2 | INRIX AND MODEL DATA PREPARATION

Comparing INRIX and model speed data is not a simple process. The two datasets have different temporal resolutions, the street length segmentation is different, and arterial streets in INRIX are mostly 1-way links versus 2-way in the model.

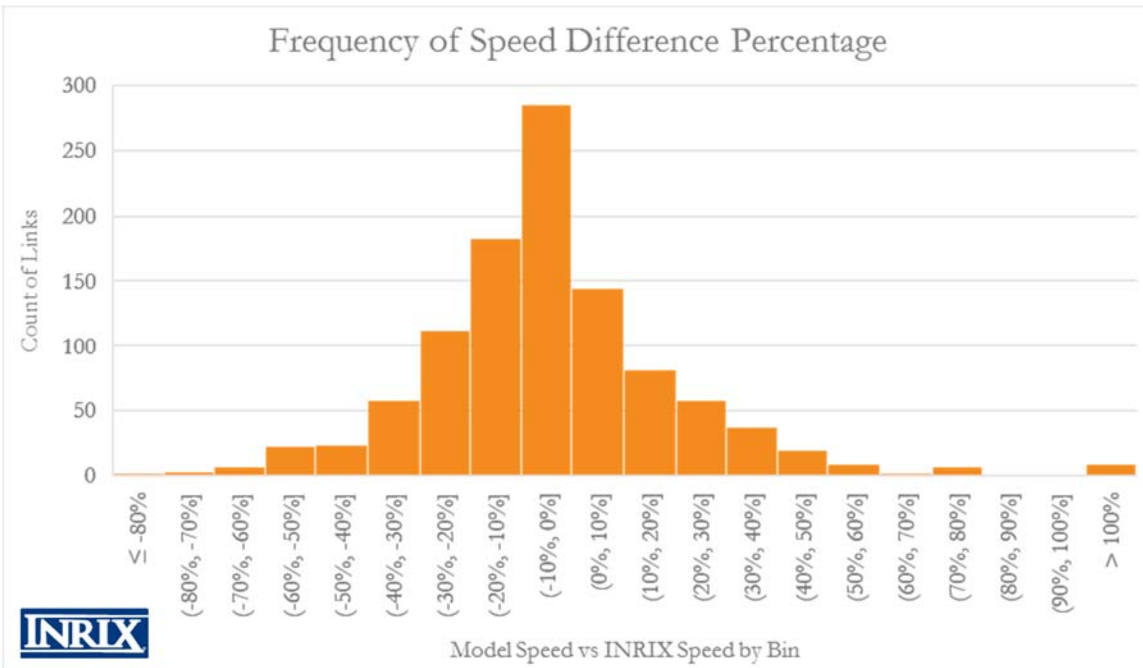
Temporal resolution issues can impede the comparison of disaggregate INRIX speed data with aggregate travel model speed data. The travel model is considered an average spring or fall weekday, with five assigned aggregate times of day (Table 2). INRIX is a continuous data monitoring system with 1-minute resolution speed data. For creating a validation dataset INRIX data was pulled for all weekdays in October 2012 and averaged to the corresponding time ranges. Consideration was made for using a diurnal distribution of volumes to weight the speed data for averaging. Early tests indicated there was not much difference and the lack of observed arterial volume distributions may bias the data.

The TMC segmentation may not always correspond to the segmentation of the model network. The correspondence process between networks accounts for this by outputting the length of the model link to TMC overlap, and the proportion of that overlap to both model link and TMC link. Most arterials in the model network are 2-way links meaning that information is coded for both directions on a single link. TMC segments tend to be 1-way links. Extra processing is required to match the correct model direction to the correct TMC segment. This process was not ready for this validation comparison. Instead the modeled speed was averaged between the two directions and used to compare to the TMC. Before any changes are made to arterials based on the results below, the 2-way vs 1-way correspondence should be updated.

7.3 | REGIONAL COMPARISON

The modeled traffic assignment results for the disaggregate model were compared to the INRIX speed for all network links with a matching TMC code. The comparisons provided below are a summarization of more detailed data and graphs. The model vs observed difference ranges have been collapsed into three categories 1) model speed is faster than INRIX speed (>10%), 2) model speed is within plus or minus 10% of INRIX speed, and 3) model speed is slower than INRIX speed (<-10%). When investigating outliers and reviewing the distribution of speed difference the data is viewed at greater disaggregation, such as the example in Figure 26 where the speed difference in the South Suburban MSA is shown with 10% bins.

FIGURE 26: EXAMPLE OF SPEED DIFFERENCE PERCENTAGE FREQUENCY, AM PERIOD, SOUTH SUBURBAN MSA



The disaggregate model was shown in Table 9 to underestimate VMT by 10%. Despite having less auto travel, the model speeds are slower than those reported by INRIX as shown in Figure 26 (AM period speed comparison) and Table 63 (all periods speed comparison). Percentage difference was calculated as (Model – INRIX) / INRIX. Table 63 shows that each time period has a similar pattern of category percentages. The AM and PM periods tend to show greater congestion have the highest percentages for the model speed being slower, 49% and 50% respectively. Likewise, the less congested early AM and evening periods show the highest percentage of within +/- 10% of any period at 39% and 38% respectively.

FIGURE 27: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, OVERALL

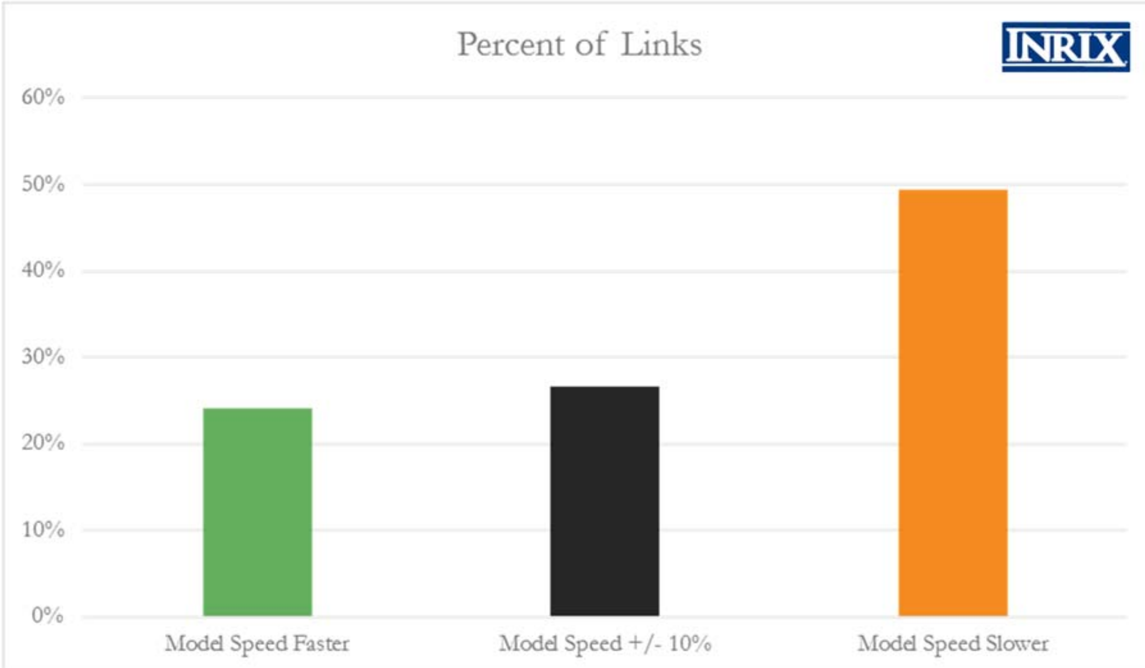


TABLE 63: MODEL VS INRIX SPEED COMPARISON, OVERALL

TIME PERIOD	MODEL SPEED FASTER	MODEL SPEED +/- 10%	MODEL SPEED SLOWER
EARLY AM	19%	39%	42%
AM	24%	27%	49%
MIDDAY	26%	33%	42%
PM	23%	27%	50%
EVENING	20%	38%	42%

Figure 28 shows a scatterplot of INRIX vs modeled travel speeds for the AM time period. There is no real discernable pattern and a considerable number of outliers for investigation. The trendline shows a considerable underestimate of travel speed and a low fit of the data. A well calibrated model for speed should show a linear line at a 45-degree angle with intercept at 0.

FIGURE 28: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, SCATTERPLOT

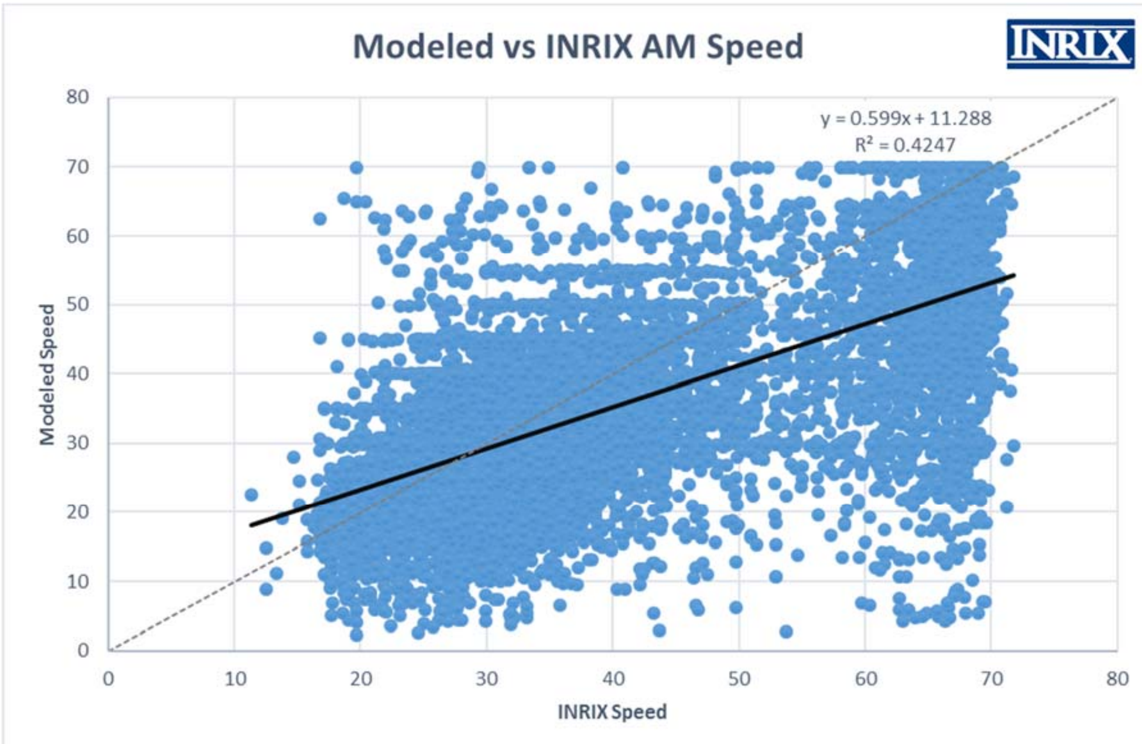
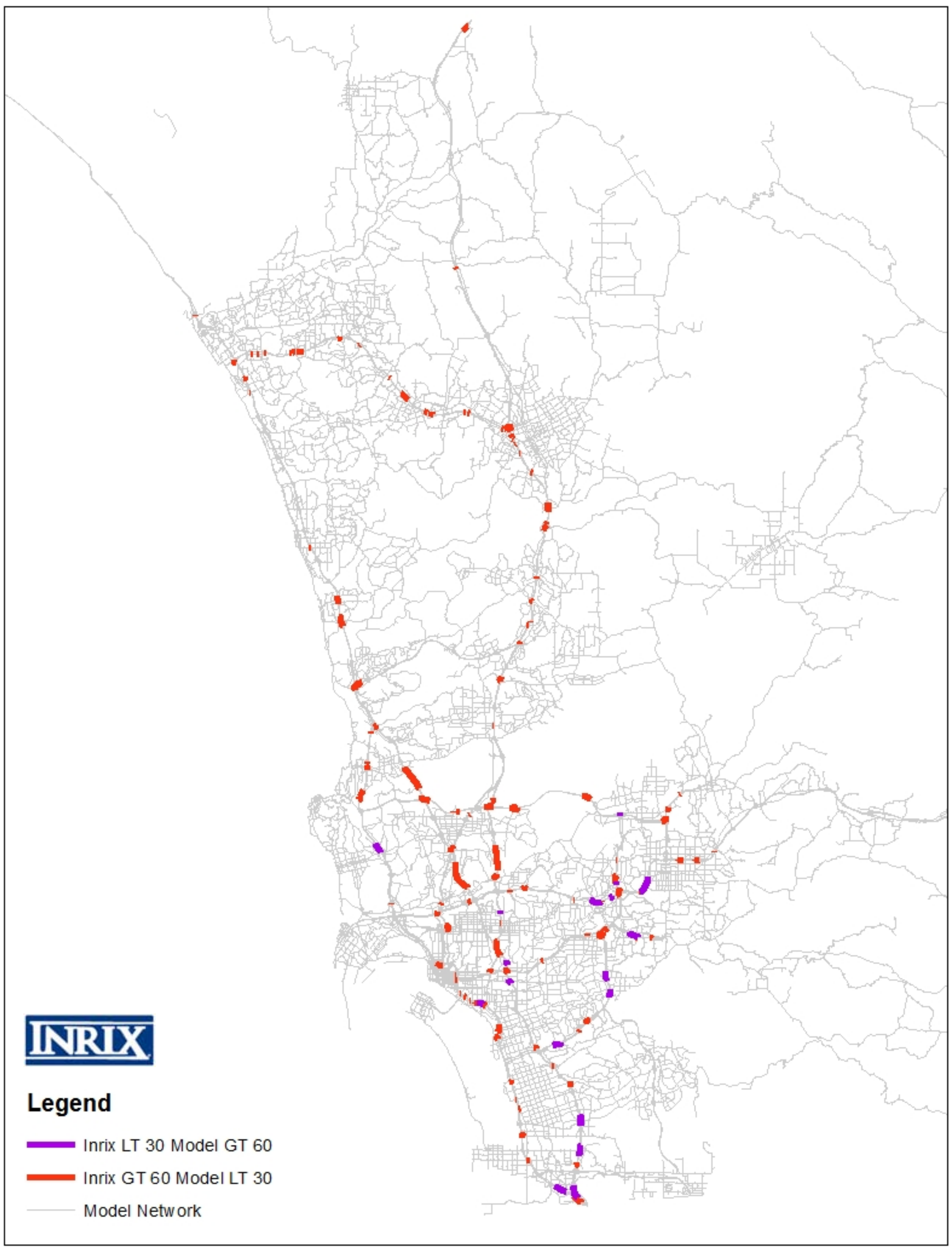


Figure 29 shows an outlier analysis for two conditions: 1) where modeled speed is greater than 60 mph and INRIX speed is less than 30 mph and 2) where modeled speed is less than 30 mph and INRIX speed greater than 60 mph. As expected, locations where modeled speed is greater than 60 mph and INRIX speed is less than 30 mph are all freeway locations. While the outliers all occur in the southern half of the county, there is not a common reason for the discrepancy. Four freeway segments near the San Ysidro border crossing could be due to border crossing volumes. For locations where INRIX speed is greater than 60 mph and the modeled speed is less than 30 mph, the majority (64%) are on freeway ramps and connectors. Freeway ramp speed underestimation could be due to either ramp meter delays or low speed limits.

FIGURE 29: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, OUTLIERS



There are several potential changes to the traffic assignment model that could be investigated. First, the model assumes free-flow speed is equal to the posted speed limit of the roadway. This assumption may not hold true and could be adjusted. Further investigation of the INRIX data may reveal a better assumed free-flow speed value for different facilities or areas. Second, the volume-delay function (VDF) may be over-predicting congestion resulting in slower modeled speeds. In VDF equations as the volume on the link increases so does the associated travel time delay. If the function increases delay too quickly that could result in poor speed comparisons. Third, where modeled speeds are faster than INRIX speeds there could be an incorrect speed limit coded or not enough assigned volume. Low volumes could result from several reasons and can be investigated corridor by corridor for any systematic biases. Fourth, some previous speed adjustments to handle diversion may no longer be valid. Speed adjustments were typically made for local calibration to encourage underutilized paths. An example of this practice is on the I-5 local bypass between the I-805 merge and SR52 where volumes were too low on the bypass and too high on the main general purpose lanes. Speed adjustments were made to attract users to the bypass to replicate behavior in the field. These local speed adjustments can be systematically identified and reviewed to ensure they are still pertinent. Lastly, the INRIX data is assumed to be the true baseline. A further examination of the data may be warranted to verify whether the INRIX speeds are too high, the link correspondence is incorrect, or looking at a peak period instead of a peak hour is resulting in poor results.

The following subsections investigate the overall data by several key attributes.

METROPOLITAN STATISTICAL AREA (MSA)

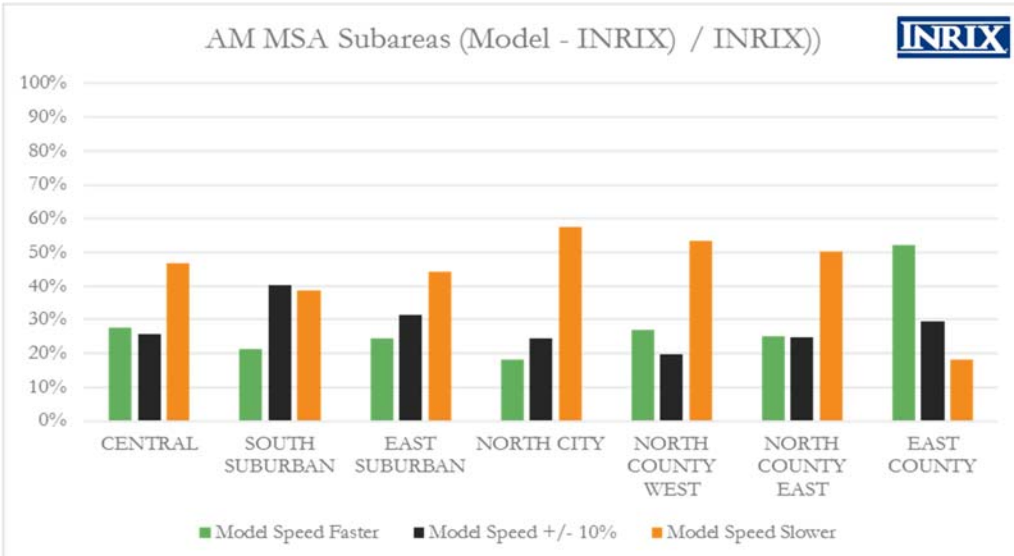
INRIX speed data is based on GPS traces so the more heavily populated the area or the higher the volume of roadway, the more likely it will be reported in the data. Table 64 breaks down the number of TMC links by area and demonstrates that the less populated East County has relatively few links reporting while the Central and North City MSAs contain the largest number of matched TMC links.

TABLE 64: DISTRIBUTION OF LINKS BY MSA

MSA	COUNT OF LINKS
CENTRAL	2,606
EAST COUNTY	280
EAST SUBURBAN	1,430
NORTH CITY	3,342
NORTH COUNTY EAST	1,575
NORTH COUNTY WEST	1,444
SOUTH SUBURBAN	1,068
TOTAL	11,745

Results in Figure 30 show that the pattern of faster or slower is not uniform across areas. In South Suburban, the largest category is of model speed matching relatively closely and in East County, though with very few links, the modeled speeds are generally too fast.

FIGURE 30: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, MSA SUBAREAS



FUNCTIONAL CLASS

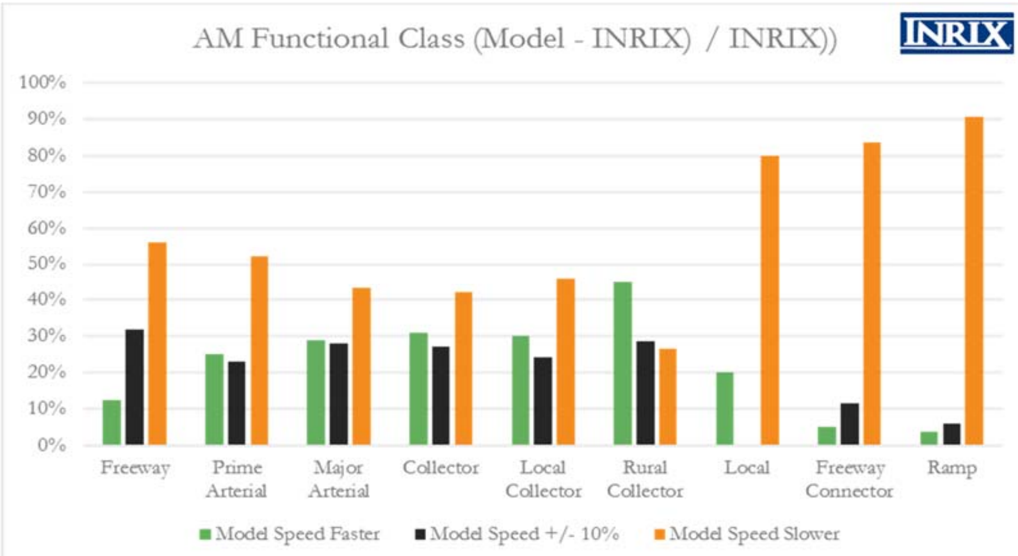
Comparing by functional class in Table 65 shows, as expected, most the links are on higher functional roads such as freeways and prime or major arterials with relatively few local or rural collector links. The local links should be checked for correspondence issues or miscoding of functional class.

TABLE 65: DISTRIBUTION OF LINKS BY FUNCTIONAL CLASS

FUNCTIONAL CLASS	COUNT OF LINKS
FREEWAY	2,293
PRIME ARTERIAL	1,764
MAJOR ARTERIAL	5,274
COLLECTOR	1,180
LOCAL COLLECTOR	424
RURAL COLLECTOR	140
LOCAL	15
FREEWAY CONNECTOR	304
RAMP	351
TOTAL	11,745

Overall Figure 31 shows the same pattern of model speeds being slower other than rural collectors. Rural collectors tend to be in East County which also showed in the MSA comparison a tendency for the model speeds to be too fast. Most freeway links were too slow possibly pointing back to too slow free-flow speeds. Additionally, ramps and freeway connector speeds need to be reviewed.

FIGURE 31: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, ROADWAY FUNCTIONAL CLASS



ROADWAY POSTED SPEED LIMIT

This subsection reviews the speed differences by posted speed limit. As suggested earlier, a possible issue in the model could be the reliance on using posted speed as the free-flow speed. The link counts in Table 66 show 3 links with 20mph speed limits. This speed limit is usually reserved for driveway entrances which suggest a correspondence issue. 60mph is also not a typical speed limit and should be reviewed.

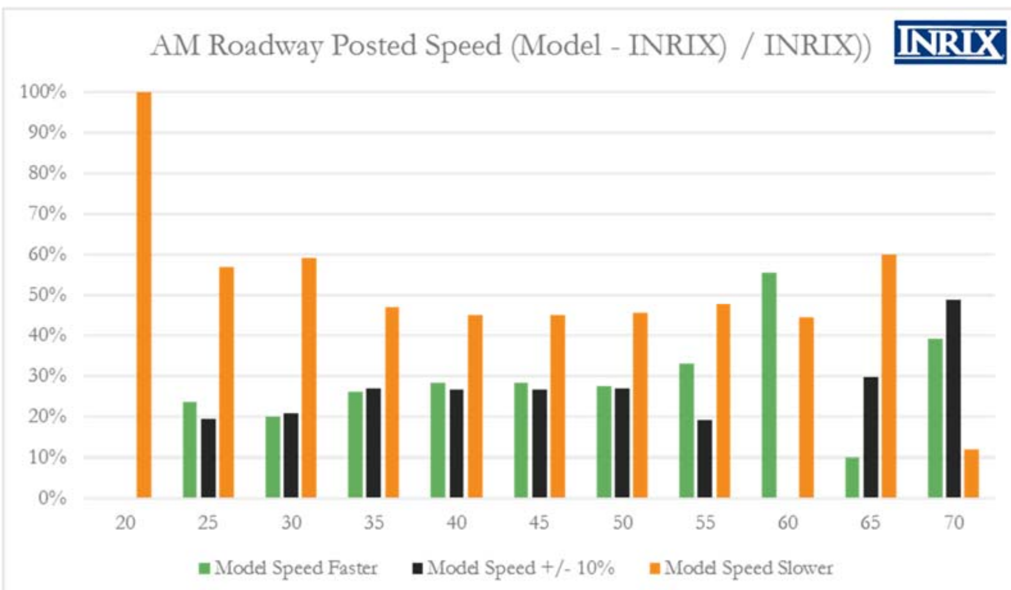
TABLE 66: DISTRIBUTION OF LINKS BY ROADWAY POSTED SPEED LIMIT

POSTED SPEED LIMIIT	COUNT OF LINKS
20	3
25	331
30	929
35	2,546
40	1,616
45	1,946
50	1,412
55	617

POSTED SPEED LIMIIT	COUNT OF LINKS
60	18
65	2,143
70	184
TOTAL	11,745

Figure 32 breaks down the speed comparison by posted speed limit of the link. While speed limits 25 and 30 mph have fewer link counts they show a large majority of links being modeled slower than INRIX observed. The only links that have speeds over-estimated are freeways with 70mph limits which is a reversal of the 65mph freeway links that are modeled slower overall. One possibility is that there could be a larger proportion of fleet/truck vehicles making up the speed observations for these links.

FIGURE 32: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, ROADWAY POSTED SPEED



INTERSECTION CONTROL

Since the correspondence had issues linking the correct directionality for arterials this review identified the highest control in either direction. So, if a link had a traffic signal in the A to B direction and a stop sign in the B to A direction, the predominant control was selected to be a traffic signal. Traffic signals being the predominant control, see Table 67, was expected as arterial links in the INRIX dataset tend to be the higher volume roads.

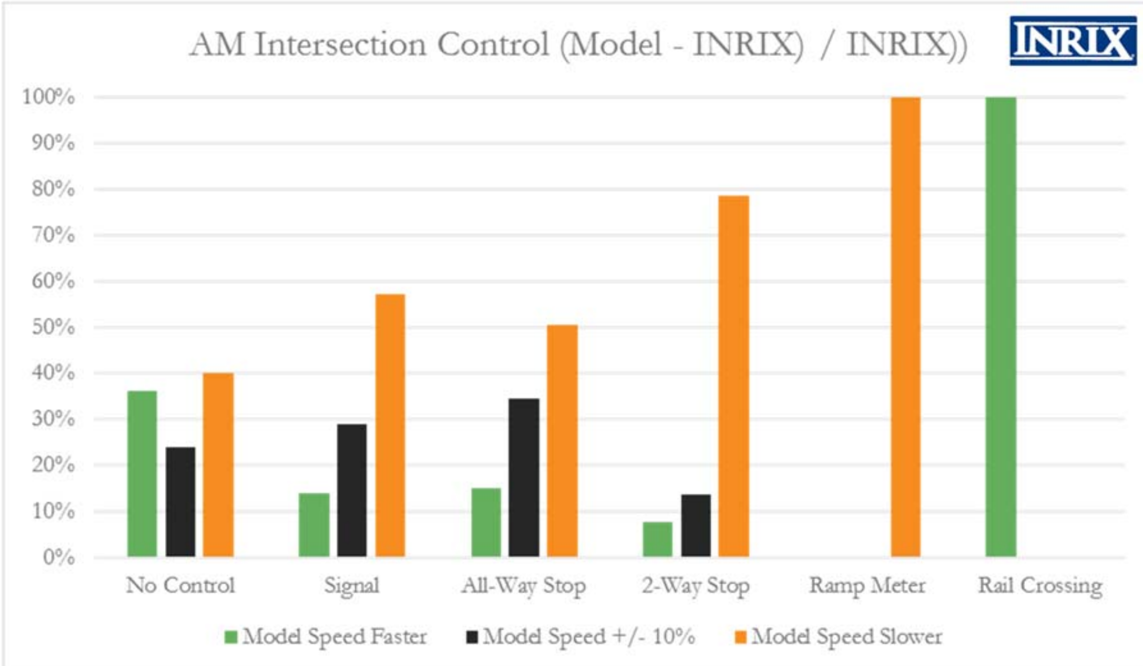
TABLE 67: DISTRIBUTION OF LINKS BY INTERSECTION CONTROL

INTERSECTION CONTROL	COUNT OF LINKS
NO CONTROL	5,264
TRAFFIC SIGNAL	6,132
ALL-WAY STOP	229
2-WAY STOP	66
RAMP METER	25
RAIL CROSSING	29
TOTAL	11,745

While the sample size of ramp meters and rail crossings are small they are both 100% slow or fast. Ramp meter links could be impacted because all model delay is being applied at the node and first connected link whereas the resulting queue might spill back beyond the corresponding link. Other possibilities could include the duration of the ramp meter being different from the modeled period, AM/PM meter designation is not identified in the model, too much delay is being applied, meters are incorrectly coded, or ramp posted speeds could be slow. With only 25 ramp meters being tagged, further investigation can identify quickly the potential issues. Rail crossings are reversed of ramp meters with all links being too fast. Possible reasons could include too little delay being applied to the nodes or that many rail crossings are near signalized intersections and the signal delay does not get carried back upstream to the next link in the model.

Controlled intersections, in general, are slower in the model than observed. 2-way stops, though having few samples, are predominately slower in the model compared to signals and all-way stops. This may mean that VDF parameters in the traffic assignment could be causing too much delay or that intersection capacities are too small. Links with no control are balanced between the fast and slow categories.

FIGURE 33: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, INTERSECTION CONTROL



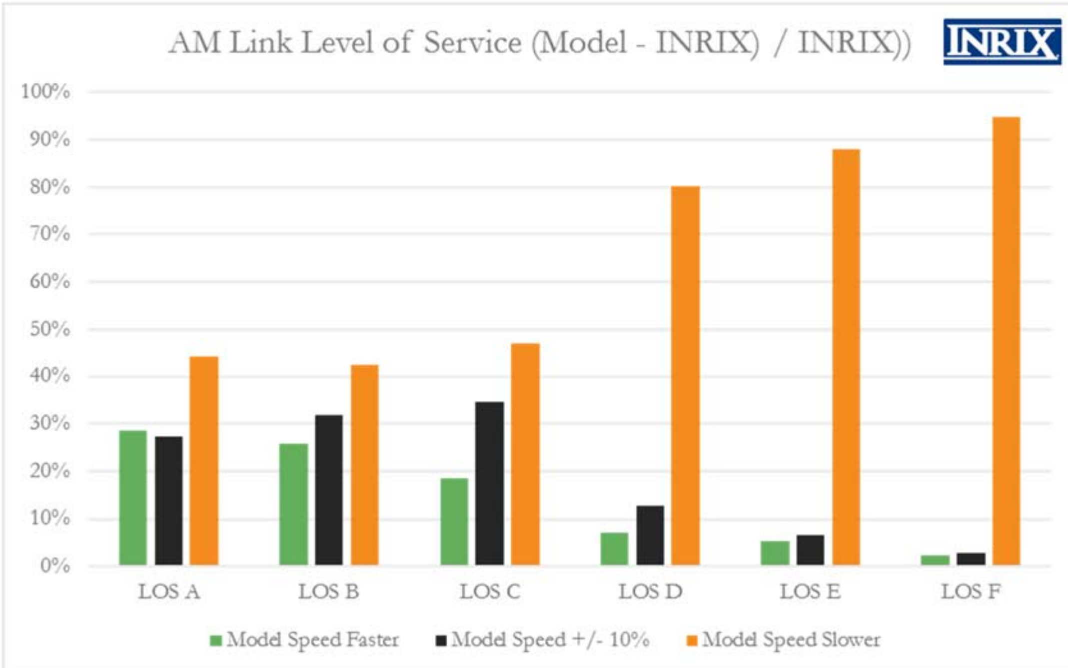
LEVEL OF SERVICE (LOS)

Level of service is calculated by taking the modeled link volume divided by the link capacity and assigning ranges of values letter grades, see Table 68 for the ranges used. LOS A is considered free flow and LOS F is congested. The disaggregate model has relatively low congestion levels, as discussed earlier, being 8M VMT lower than HPMS. This is evidenced in Table 68 where over 87% of the links with TMC correspondence are at LOS C or better. When links are at LOS A, the facility should be operating close to free-flow conditions. Figure 34 shows the model does not replicate free-flow speeds well with only 27% falling with +/- 10% of the INRIX speed. As the congestion increases on the link the model speeds grow disproportionately slower. This suggests the VDF equation might be too steep.

TABLE 68: DISTRIBUTION OF LINKS BY LEVEL OF SERVICE

LINK LEVEL OF SEVICE	VOLUME / CAPACITY RANGE	COUNT OF LINKS
LOS A	0 to 0.3	6,430
LOS B	0.3 to 0.5	2,705
LOS C	0.5 to 0.7	1,139
LOS D	0.7 to 0.85	576
LOS E	0.85 to 1	549
LOS F	> 1	346
TOTAL		11,745

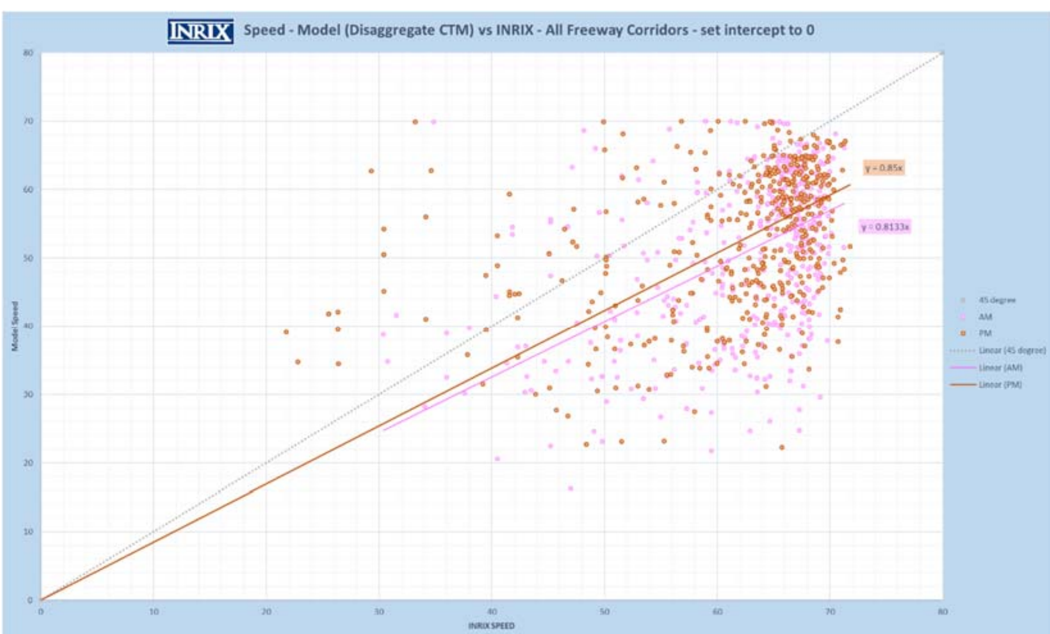
FIGURE 34: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, LEVEL OF SERVICE



7.4 | CORRIDOR COMPARISON

Freeway corridors were examined for the entire San Diego region. Overall model volumes were shown to be lower than expected, Figure 35, with the linear regression line slope for the AM time period at 0.8133 and PM time period at 0.85. The plot shows there is no discernable pattern with numerous outliers to both sides of the 45-degree line.

FIGURE 35: MODEL VS INRIX SPEED COMPARISON, AM AND PM TIME PERIODS, ALL FREEWAY LOCATIONS



Freeway corridor specific diagrams were created for every freeway in the region. Five major corridors are shown below: I-5 (Figure 36 and Figure 37), I-15 (Figure 38 and

Figure 39), and I-805 (Figure 40) with INRIX data in the solid lines and model data in the dashed lines. In general, the plots show that where the corridor has congestion, the model is also showing congestion. The I-5 and I-15 plots do show the same resounding theme of the model being too sensitive to congestion with speeds lower than reality. I-805 NB was one of the few locations where model speed was too high. From Telegraph Canyon to Plaza the model was generally 5-10mph higher than the INRIX speed. This section of roadway had HOVs under construction during this time period and it is possible that the construction had a negative impact on real speed which would not be present in the model. This could be confirmed by checking INRIX speeds over time but specifically before construction activity occurred.

FIGURE 36: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, I-5 SB CORRIDOR

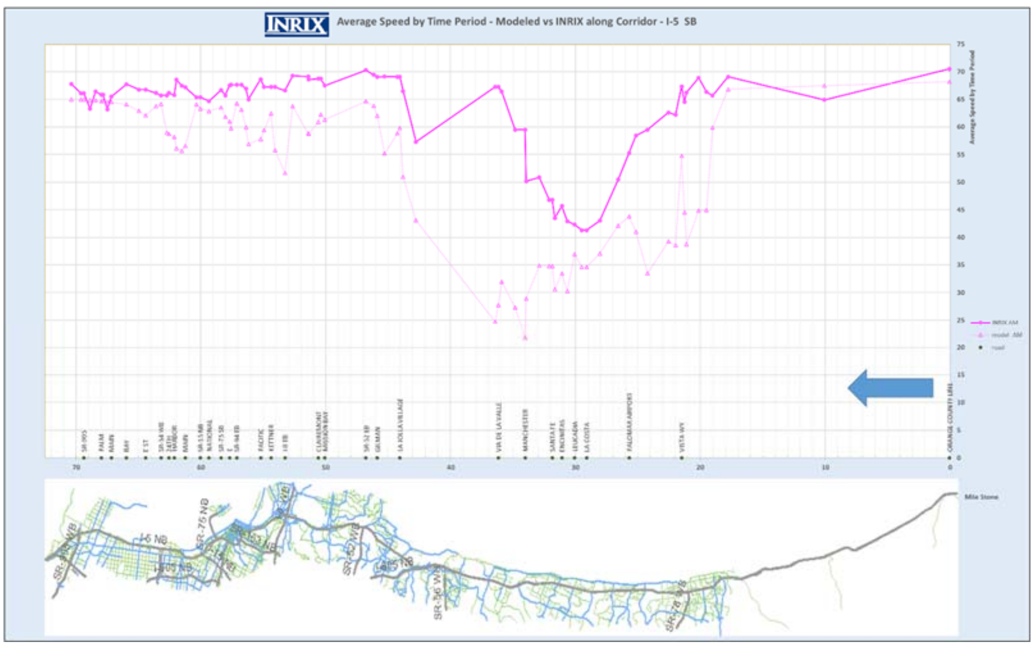


FIGURE 37: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, I-5 NB CORRIDOR

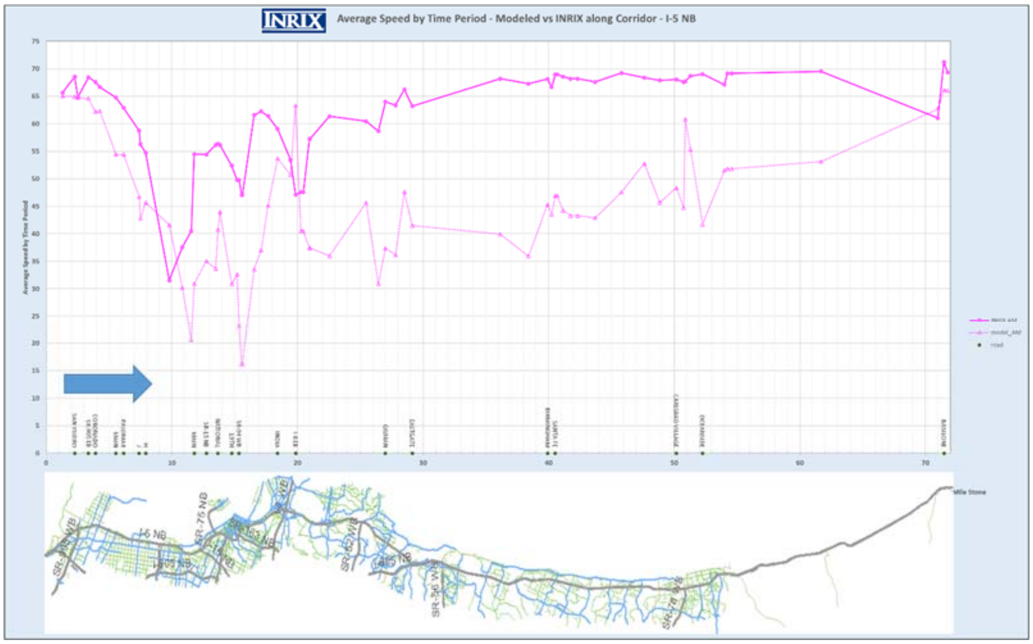


FIGURE 38: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, I-15 SB CORRIDOR

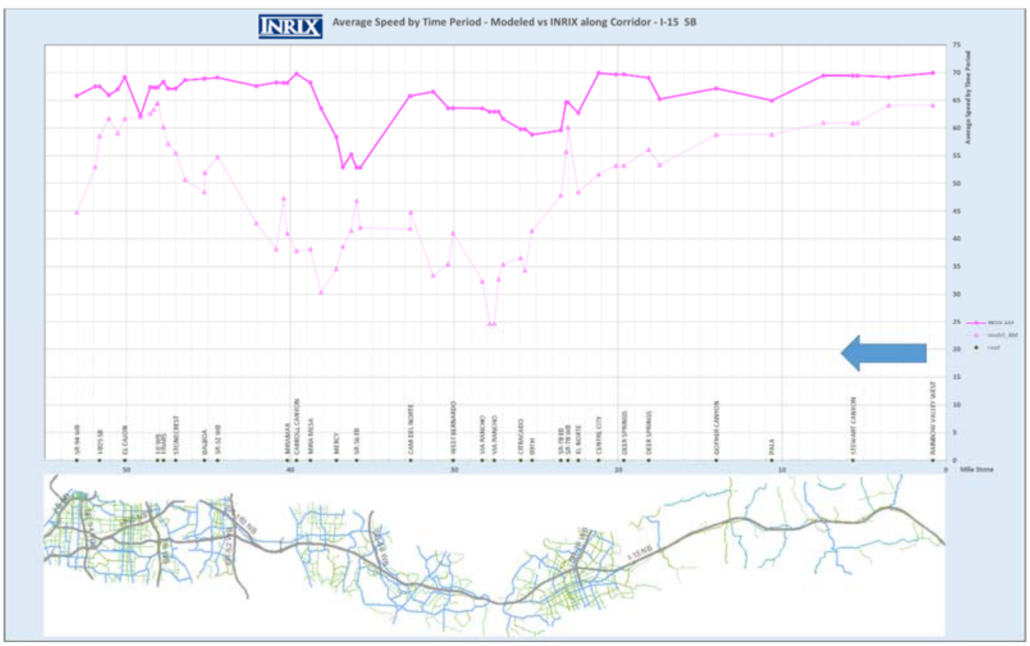


FIGURE 39: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, I-15 NB CORRIDOR

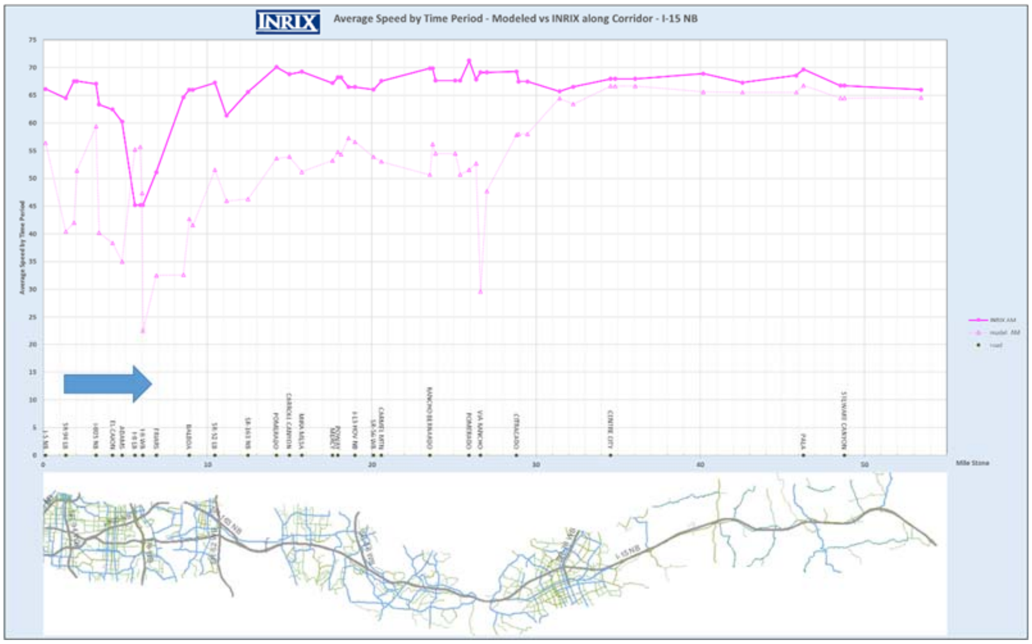
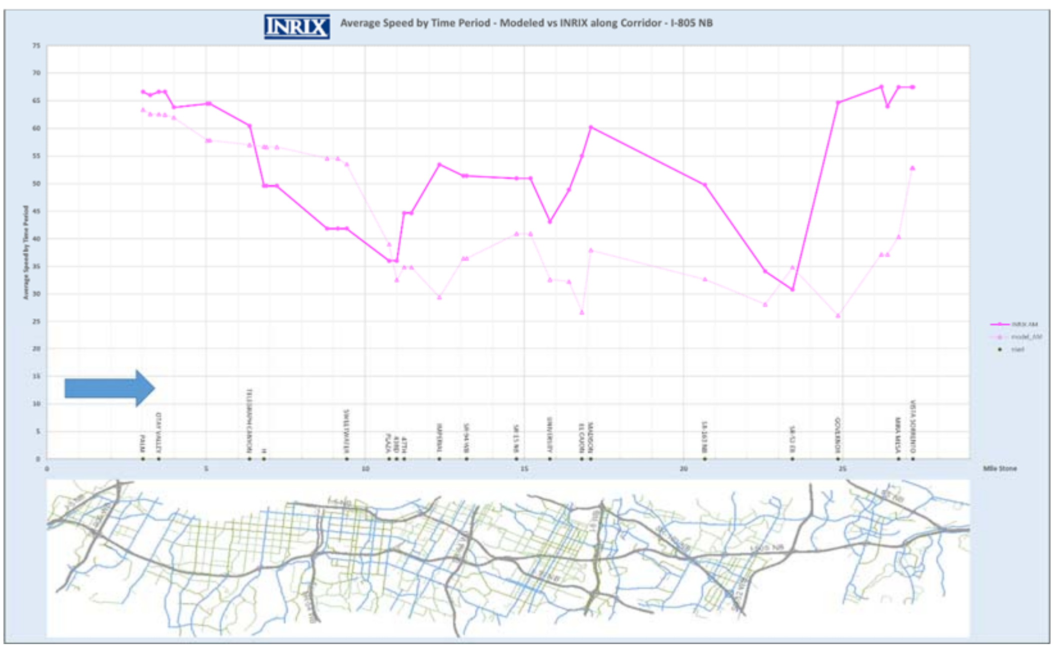


FIGURE 40: MODEL VS INRIX SPEED COMPARISON, AM TIME PERIOD, I-805 NB CORRIDOR



7.5 | INRIX COMPARISON CONCLUSIONS

INRIX speed data allows for an extensive speed validation for regional models, especially for arterials where little or no validation data exists. The process for aggregating the data and linking to model networks is also not straightforward and can be time consuming. Recommendations from the matching process included:

- Completing a more detailed review of the correspondence process to ensure the correct model link to TMC link correspondence is being created
- Investigating averaging techniques by weighting speeds by vehicle distributions by area and time of day
- Updating the current match process to handle the directionality of arterial links
- Comparing INRIX speeds to other observed data sources to validate the baseline data source

Model results compared to INRIX speed were found to be generally slower with 40-50% of the links falling into that category, and only a third of the links were within a reasonable, +/- 10%, percent difference. Slower speeds were consistent across detailed comparisons with a few exceptions such as the East County area which was generally too fast. Future review and testing work includes:

- Reviewing point to point travel times on freeways and arterials in addition to link by link comparisons
- Testing adjustments to posted speed limits by facility or area types
- Testing of VDF and/or capacity changes especially for controlled intersections and volume to capacities ratios close to 1.
- Reviewing network coding for posted speed limits and previous speed adjustments
- Reviewing arterial corridor diversion issues in conjunction with volume validation results
- Reviewing ramp meter coding for:
    - time duration of the ramp meter being different from the modeled period
    - meter node delay for single and HOVs
    - meter designation is correctly coded
    - ramp posted speeds
    - adding AM/PM meter designation to the model

8.0 SUMMARY AND RECOMMENDATIONS

This report compares two model systems developed for San Diego County to a number of data sources. The two model systems are referred to as the 'aggregate' model and the 'disaggregate' model. The key difference between the two systems is that the aggregate model includes an aggregate CVM that consists of two components; an aggregate internal truck model that was borrowed from SCAG and an aggregate light CVM that was largely asserted and calibrated to match total traffic counts. The disaggregate model replaces these two components with a tour-based CVM whose parameters were calibrated to match data obtained from a commercial vehicle survey in San Diego County. Both model systems include the same person-transport models for San Diego residents and non-residents traveling into, out of, within, and through San Diego County.

We compare these systems to a number of data sources, including Caltrans and local jurisdiction traffic counts, vehicle occupancy and classification study data, data from weigh-in motion stations, HPMS data, California Department of Motor Vehicles data, NHTS and CHTS data, AirSage data, ATRI truck GPS data, and INRIX travel time data.

The comparisons of model results to traffic counts and to HPMS data clearly demonstrate that the aggregate model matches observed VMT relatively well compared to the disaggregate model, which underestimates VMT by approximately eight million miles, or between 9% and 10% depending on whether one compares the results to traffic counts or HPMS respectively. Further analysis of the differences between the aggregate and disaggregate models show that the difference in VMT is largely due to the aggregate CVM which was asserted as part of the model development effort in order to better match traffic counts. The biggest impact of the difference is that in the disaggregate model, the error on principal arterials is much higher (88% RMSE) than in the aggregate model (24%). It is unclear from the analysis exactly why principal arterials are impacted the most by the difference.

Comparisons to vehicle classification and occupancy study and WIM data suggest that the aggregate model over-estimates business autos and light trucks by a factor of 2.8, while the aggregate model underestimates business autos and light trucks by approximately 20%. Heavy trucks in both models are over-estimated by approximately 30% to 40% for the aggregate and disaggregate models respectively. The disaggregate CVM also estimates more trips than expanded ATRI data, for both medium and heavy trucks. Biases in trip lengths in the ATRI data confound our ability to make further useful observations about these vehicle types in the disaggregate model.

California DMV data reinforces the finding from the commercial vehicle count data comparison; that the disaggregate model closely matches the approximately 7.1M vehicle miles of travel for commercial vehicles registered in San Diego County. However, because vehicles not registered with DMV as commercial also are used for commercial vehicle travel, we suspect that the DMV data is downward biased. This provides some evidence that the commercial vehicle count source is a better target to match than VMT reported by DMV, and that the **trip rates for commercial vehicles from the disaggregate model should be adjusted to better match counts before adopting the model for use in applications.**

The person-transport models were validated against two sources; travel survey data and AirSage data. The person-transport comparisons indicate model closely replicates the tours per person, trips per person, and trip lengths from the 2006 SANDAG Household Travel Survey. However, the NHTS trip rate is slightly higher, and the NHTS trip length nearly 1 mile longer than 2006 survey. The AirSage data comparison confirms that the model slightly underestimates total trips. The AirSage data is clearly biased in terms of trip length, so we are unable to confirm the trip length reported from the NHTS sample. However, SANDAG is currently fielding a new HTS with a larger sample size (5,000 households) than the 2006 survey that the model was calibrated to. If confirmed by the current 2016 SANDAG Household Travel Survey, the trip rate and length differences could explain much of the VMT underestimate in the disaggregate model. **We recommend that SANDAG carefully analyze and compare the modeled tour rates, trip rates, and trip lengths against the new survey data as soon as it is available to determine whether adjustments to model parameters will improve the match to counts in the disaggregate model.**

Finally, we compare estimated speeds to observed speeds obtained from averaging INRIX travel time data for each assignment period. We find that the model tends to underestimate speeds across all facility types. However, issues in linking the INRIX data to the model network, and the method used for calculating an average speed from the INRIX data, introduce unknown biases into the comparison. **We recommend that SANDAG refine the method to link INRIX speeds to the transport network and average those speeds within each time period. We further recommend that SANDAG eliminate links with volume-to-capacity ratios over 1.0 from the comparison since these conditions do not exist in the real world. We suggest that SANDAG calibrate free-flow speeds and VDF parameters to better match observed INRIX speeds.**

9.0 REFERENCES

Bradley, Mark, and Elizabeth Greene. 2016. "The Transition from Diary-Based to Smartphone-Based Travel Survey Data: Implications for Travel Demand Modeling." *Transportation Research Board* (Innovations in Travel Modeling Conference).

California Department of Transportation. 2013. "2010-2012 California Household Travel Survey Final Report."

Cambridge Systematics, Inc. 2004. "Accounting for Commercial Vehicles in Urban Transportation Models Summary Report."

Cambridge Systematics, Inc. 2008. "SCAG Heavy-Duty Truck Model Update."

Cambridge Systematics, Inc. 2010. "Travel Model Validation and Reasonableness Checking Manual - Second Edition."

HBA Specto Incorporated. 2014. "SANDAG - Commercial Vehicle Model (CVM) Final CVM Model Development and Calibration."

PB Americas, Inc. 2012. "Activity-Based Travel Model Calibration/Validation: Coordinated Travel Regional Activity-Based Modeling System (CT-RAMP) for San Diego County."

PB Americas, Inc. 2009. "Activity-Based Travel Model Specifications: Coordinated-Travel Regional Activity-Based Modeling Platform (CT-RAMP) for San Diego County."

Resource Systems Group, Inc. n.d. *rMove*. https://rmove.rsginc.com/ accessed on 10/7/2016.

Resource Systems Group, Inc. 2014. *Transferability of Activity-Based Model Parameters*. Washington, D.C.: Strategic Highway Research Program, Transportation Research Board.

San Diego Tourism Authority. 2013. "Visitor Profile Summary."

U.S. Department of Transportation, Federal Highway Administration, 2009 National Household Travel Survey. URL: http://nhts.ornl.gov. n.d.