MINIMUM TRAVEL DEMAND MODEL CALIBRATION and VALIDATION GUIDELINES FOR STATE OF TENNESSEE UPDATED 2012

By:

Dr. Fred Wegmann
Dr. Jerry Everett

Center for Transportation Research
University of Tennessee
Knoxville, Tennessee

THE UNIVERSITY of TENNESSEE
KNOXVILLE

Table of Contents

Calibration and Validation Guidelines

Introduction

The calibration and validation of travel demand models is essential to accurately model current/future travel for a metropolitan area. A major shortcoming of many travel demand forecasting (TDF) models is the lack of attention and effort placed on the model evaluation and reasonableness-checking phase. There is a trade off between increasing the level of accuracy of a model and the cost associated with additional data collection, calibration, and validation. There are, however a number of inexpensive evaluation and reasonableness checks that can be performed to enhance TDF model's forecasting ability.[(1)] Since the initial guidelines were published, new data and approaches have become available. The 2012 version of guidelines attempts to update and incorporate the new information.

It is important that reasonability checks be performed after each step of the four-step modeling process. Too often the modeler proceeds through four steps and then conducts tests based on the overall results of the travel model. The preferred approach is to apply evaluation and reasonableness checks during the process of calibrating each individual modeling step. After each step has been validated, the overall model is validated. If this incremental validation approach is not followed the problem of "error propagation" will occur. Error propagation occurs when errors made in each step are compounded and thereby increase the overall modeling error.

Travel modes rely on several primary data sources for model calibration and validation.

1. Accurate estimates of base year traffic analysis zone (TAZ) household characteristics and employment information
2. An accurate representation of the base year highway and transit network
3. An accurate base year travel survey and
4. Accurate base year ground counts are all needed for model calibration

The household survey data are used to develop mathematical relationships or functions, which relates observed travel behavior to variables that can be more easily and directly forecasted. In general, the model parameters that rely on travel survey information are assumed to remain constant between the model year and the model forecast year.

Travel survey information is sometimes lacking, as financial resources are often a barrier in conducting a full set of travel surveys. When this is the case, part of the model may have to be developed using "borrowed" travel survey data from another similar urban area that has conducted a travel survey, or by relying on information from a potentially outdated travel survey, or by having to prepare model specifications with limited or incomplete travel survey information.

Network information for the base year is generally available and with the GIS and other software now available, base year highway and transit networks can be developed with a high degree of accuracy, provided care and quality checks are employed during preparation.

Base year ground counts in sufficient numbers of locations and accuracy are required for base year model validation. The MPO or city traffic-engineering department often collects these data locally. Often, the local traffic count data will be supplemented by data collected by the state Department of Transportation.

Calibration and validation occurs **after** the original "estimation" or development of the model set. This report is NOT intended to be a manual on travel demand model development. This report assumes that the final model structure including the selection of relevant variables and the specification of the initial parameters has already been determined. Even the techniques used for model calibration are complex and for the most part are beyond the scope of this report. The guidelines only highlight some of the more fundamental calibration and validation checks that can be performed by the analyst. Special attention will be given to the validation of the traffic assignment step.

Major resources utilized in the preparation of the guidelines include the following reports:

> Ismart, D. "Calibration and Adjustment of System Planning Models." Federal Highway Administration. Washington, DC. December 1990. (1)
>
> "Model Validation and Reasonability Checking Manual." Barton-Aschman Associates, Inc. and Cambridge Systematics, Inc. February 1997. (2)
>
> "FSUTMS-Cube Framework Phase II – Model Calibration and Validation Standards." Cambridge Systematics, Inc. prepared for the Florida Department of Transportation System Planning Office. Tallahassee, Florida. October 2008.* (3)

Since the initial publication of these guidelines, a comprehensive report on Calibration and Validation has been published by the Federal Highway Administration. Pertinent comments are included as Appendix B. The document serves to solidify the guidelines as presented in the current document.

*Also see the companion report FSUTMS-Cube Framework Phase I Default Model Parameters. Cambridge Systematics, Inc. prepared for the Florida Department of Transportation System Planning Office. Tallahassee, Florida. October 2006.

Calibration and Validation

Once the model parameters have been estimated the process of calibration and validation begins. Model calibration adjusts parameter values until the predicted travel matches the observed travel within the region for the base year. For purposes of forecasting it is assumed these parameters will remain constant over time. Calibration is conducted in all four steps of the modeling process and normally occurs after establishing model parameters.[(4)]

Model validation tests the ability of the model to predict future behavior. Validation requires comparing the model predictions with information other than that used in estimating the model. Validation is typically an iterative process linked to calibration. If the analyst finds that the model output and the independent data are in acceptable agreement, the model can be considered validated. [(5)]

There are normally two types of validation checks - reasonableness checks and sensitivity checks. Reasonableness checks are tests that include the comparison of rates, checking of the total regional values and logic tests, etc. The analyst evaluates the models in terms of acceptable levels of error, ability to perform according to theoretical and logical expectations, and consistency of model results with the assumptions used in generating the results. [(6)]

Sensitivity checks are tests that ascertain the responses to transportation system, socioeconomic, or political changes. Sensitivity often is expressed as the elasticity of a variable. For example, one might examine this impact on travel demand if parking or toll fees are doubled. Work by Schiffer provides a comprehensive application of a sensitivity analysis for the Wasatch Front Regional Council (Salt Lake City) [(7)] on major socioeconomic changes. Elasticity is the concept used to describe the degree of sensitivity.

In practical application the meanings of calibration and validation have changed over the years. This is especially true in areas where funding limitations preclude the conduct of travel surveys. In many locations the processes of calibration and validation have become a single exercise. Without current Origin Destination (O-D) studies, many of the models are calibrated by using default values derived from other studies and transferred to the local environment. In this process, calibration and validation are merged since an independent database is not available for calibration. In order to develop a good match between model link volumes and ground counts, model parameters are modified to provide “reasonable” agreements. This is clearly the situation in many Tennessee MPOs where there has traditionally been a high reliance on transferable parameters for many planning variables.

The ultimate test of a travel demand model set is its ability to accurately predict traffic volumes on the transportation system. Therefore, in many areas traffic counts are the primary data parameter used for model validation. As will be discussed in this report, a number of checks are used to compare the model’s simulated link values with the traffic counts. In any accuracy check it must be recognized there are discrepancies associated with the ground counts. These discrepancies may be due to daily variations as well as equipment issues, the inappropriate use of daily and seasonal factors to estimate Annual Average Daily Totals (AADT), and the absence of good classification data to correct axle counts to vehicles. [(5)] Likewise, while validation involves the results from the assignment phase of the four-step travel demand process, the errors can be attributed to all phases in the modeling process. Errors from previous steps may either compensate or be additive.

Various accuracy checks have been established as part of validation. The guidelines proposed by FHWA indicate that the model needs to be accurate enough as to not affect the number of lanes required.[(5)] However, given the scrutiny that may accompany conformity analyses, MPOs are encouraged to strive for a higher degree of accuracy when resources permit. Recent work by Schiffer et al has recalculated the "one lane" guidelines based on recent changes to the Highway Capacity Manual. [(7)]

The remainder of this report will discuss some of the basic steps that can be taken to calibrate and validate the individual steps of the travel demand model set. Given the complexity of mode split models and their limited application in Tennessee no discussion of their calibration is provided in this report. However, selected references can be consulted for further information on the topic of mode split analysis. [(3)] Appendix A describes reasonableness checks for each step in the modeling process that the analyst can use as part of the calibration/validation process. [(4)]

Trip Generation Calibration/Validation

Trip generation calibration/validation involves three components: evaluation of the base year zonal socioeconomic estimates, development and application of a trip production (P) model, and the development and application of trip attraction (A) model. In the trip generation step, the trip production rate or model is applied using zonal socioeconomic estimates to produce trip productions. Similarly, the trip attraction rate or model is applied using zonal socioeconomic estimates to produce an estimate of zonal trip attractions.

The analyst is concerned with the reasonableness of the base year zonal socioeconomic estimates, the appropriate functions or mathematical expressions used for the trip production model and trip attraction model, and the parameters associated with these models. The analyst is also concerned with the trip Production and Attraction (P&A) balance.

The Census Transportation Planning Package (CTPP) provided (now the American Community Survey) provides a breakdown of households by TAZ, household size, automobile ownership, and income group. These household and automobile ownership data and the income data can be used to verify the reasonableness of the base year zonal estimates. A comparison of observed (CTPP data) and estimated (base year zonal estimates) households, by size subgroup, will help identify serious bias in the base year zonal estimates.

Household travel surveys are used to estimate trip rates. Surveys do not provide direct estimates of zonal trip ends. Therefore, there is not an observed number of trip ends to compare with model-estimated trip ends. The analyst can compare summary statistics such as vehicle trips per household, person trips per household by trip purpose, and the percentage of person trips by trip purpose for the study area with values for similar study areas.

As stated by Cambridge Systematics for the Florida Department of Transportation (FDOT): [(3)]

- **Aggregate Trip Rates**

"Table 1 depicts typical ranges for aggregate trip generation rates that can be obtained from model outputs. As with other benchmarks provided in this section, acceptability ranges were mostly derived from national and state guidance documents. It would be expected that most validated models would fall somewhere between the low and high values in this table. If the model is reporting statistics outside these ranges, additional verification and adjustments of socioeconomic data, trip production rates, trip attraction rates, dwelling unit weights, and/or special generators is likely warranted.

Table 1

Aggregate Trip Rate Benchmarks (3)

Statistic	Benchmarks[a] Low	High
Person Trips / TAZ	N/A	15,000
Person Trips / Person	3.3	4.0
Person Trips / DU (or HH)	8.0	10.0
HBW Person Trips / Employee	1.20	1.55

[a] Generally excludes nonmotorized trips; including nonmotorized trips could increase person trips per DU up to 11.5.

As indicated above, zones with more than 15,000 person trip productions or attractions should be reviewed to identify zones for further splitting. Ideally, any TAZ with greater then 15,000 person trip ends should be split into multiple zones; however, some uses (e.g., a shopping mall) might be difficult or impossible to split out." (3)

- **Percent Trips By Purpose**

"Table 2 provides typical ranges of percent trips by each trip purpose. Some of these ranges are quite large, so it is recommended that the modeler also review statistics on trip purpose from travel surveys, previous models of the same area as well as comparable regions in terms of population size and dominate employment types. Rules of thumb to consider are that HBW trips are usually 15 to 20 percent of regional trips and nonhome-based (HNB) trips generally comprise 25 to 33 percent of trips." (3)

Table 2
Percent Trips by Purpose (3)

Statistic	Benchmarks Low	High
Percent Trips by Purpose – HBW[a]	12%	24%
Percent Trips by Purpose – HBSH[b]	10%	20%
Percent Trips by Purpose – HBSR[c]	9%	12%
Percent Trips by Purpose – HBSC[d]	5%	8%
Percent Trips by Purpose – HBO[e]	14%	28%
Percent Trips by Purpose – HBNW[f]	45%	60%
Percent Trips by Purpose – NHB[g]	20%	33%

[a]HBW – Home based work
[b] HBSH – Home based shopping
[c]HBSR- Home based social recreational
[d]HBSC- Home based School
[e]HBO - includes a variety of special trip purposes depending on the model (e.g., airport, college, and shop).
[f] HBNW - accounts for all home-based trip purposes except HBW.
[g] NHB - includes combined purposes for NHB Work and NHB Nonwork, where appropriate.

- **Unbalanced Attractions versus Productions**

"Most trip generation models balance the total number of home-based trip attractions to the total number of home-based productions by each purpose. The reason for balancing to productions is because of the greater confidence in socioeconomic estimates of population and dwelling units versus employment. Most trip generation models will end up with a larger number of home-based attractions than home-based productions, prior to balancing. Even though balancing might help resolve these regional differences, it is still good practice to review the ratio between unbalanced attractions and productions as a large difference might indicate problems with employment estimates, trip rates, etc. Most literature on best practices recommends that the difference between unbalanced regional attractions and productions be kept to +/-10 percent for each purpose; although a review of model validation reports shows many models that exceed this standard. Upwards of +/-50 percent difference at the regional level might be considered acceptable under certain conditions and trip purposes." (3)

- **Percent External-External (EE) Trips**

"The amount of regional EE trips varies considerably depending on the size of a region, the proximity of neighboring urbanized areas, and the type of facilities that enter the region at external boundaries. It has been documented by three different sources in the literature review that the percent EE trips range between 4 percent (large regions) and 21 percent (smaller urbanized areas). While this range represents a decent rule of thumb, it should be recognized that the percent EE trips is quite variable, by region and by facility type of each external zone." (3)

Trip Distribution Calibration/Validation

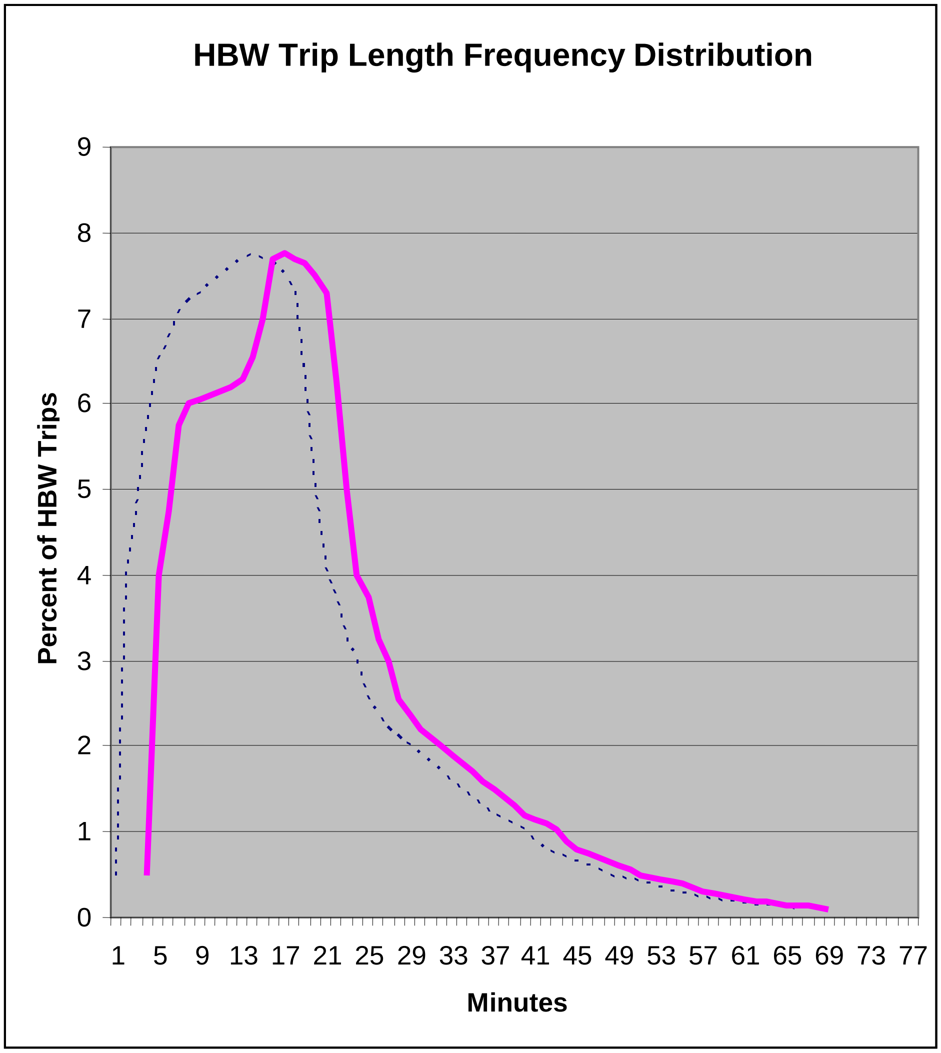
Calibration of the trip distribution model is often performed using two methods:

- Comparison of regional trip length frequency distribution (TLFD) and mean trip lengths by trip purpose, and

- Evaluation of the area-to-area flows of trips

A TLFD is a plot of the percent of total trips that occur for each separation in miles or in minutes as shown in Figure 1, which will be discussed in more detail later in this section.

Figure 1 – Trip Length Frequency Distribution (TLFD)



— Gravity Model TLFD ---- CTPP TLFD

Household travel survey data are used to estimate TLFD and mean trip lengths in miles and minutes for each trip purpose. A sample as small as 500 households will provide a stable estimate for the TLFDs and mean trip lengths for the three internal trip purposes. These observed TLFDs and mean trip lengths from the travel survey are compared with the TLFDs and mean trip lengths from the application of the gravity model.

The Census Journey to Work (CTPP), now the American Community Survey, can be used to develop an observed TLFD for the HBW trip purpose. Note that there are differences in the definition of trips between travel surveys and Journey to Work data that an analyst must consider.

Calibration of the gravity model continues by adjusting the friction factor values until a satisfactory agreement is achieved between the modeled and observed mean trip lengths and TLFDs for each trip purpose.

Figure 1 shows a typical comparison between the TLFDs for HBW trips, as estimated by the gravity model with observed trip TLFDs as reported in the survey data. The analyst will consider differences in the definitions of work trip and trip chaining conventions in evaluating the differences between the two curves. Trip chaining refers to how intermediate stops made as part of a home-to-work trip are considered in the analysis of travel survey data. For example, a stop to purchase coffee on the way to work may be considered as one or two trips.

Figure 1 shows separations in minutes on the X axis. Similar graphs are prepared comparing the TLFDs for all internal trip purposes (HBW, HBO, and NHB) as estimated by the gravity model with observed TLFDs estimated from the household travel survey.

To calculate a TLFD, the number of trips made at one minute of travel time, two minutes of travel time, three minutes of travel time, etc., are counted. The number of trips at each minute of separation is divided by the total number of trips, and the result is multiplied by 100. The resulting percentages are plotted with the minutes of separation on the X-axis and the percent of trips on the Y axis. To check trip length distribution data trip survey data are typically required. (4)

For work prepared by FDOT, the following trip distribution guidelines were prepared (3):

- **Average Trip Length by Purpose**

"Table 3 provides a summary of calibration and validation standards and benchmarks related to trip length, as derived from the literature review." (3)

Table 3
Average Trip Length and Frequencies by Purpose(3)

	Benchmarks	
Statistic	**Low**	**High**
Average Trip Length – HBW (minutes)	12	35
Average Trip Length – HBSH (minutes)	9	19
Average Trip Length – HBSR (minutes)	11	19
Average Trip Length – HBSC (minutes)	7	16
Average Trip Length – HBO[a] (minutes)	8	20
Average Trip Length – NHB[b] (minutes)	6	19
Average Trip Length – IE (minutes)	26	58
Statistic	**Standards**	
Mean Trip Length, Observed Total Trips	+/-3%	
Trip Length Frequency Distribution versus observed	+/-5%	

[a] HBO includes a variety of special trip purposes depending on the model (e.g., airport, college, and school).
[b] NHB includes combined purposes for NHB Work and NHB Nonwork, where appropriate.
[C] See Table 2 for a definition of trip purposes

- **Percent Intrazonal Trips**

"Measuring the percentage of intrazonal trips by purpose can be a means to identifying issues with friction factors, intrazonal time calculations, and even zone size. The percent of intrazonal trips is very dependent on zone size and composition. It is typical that HBW trips would have the smallest percentage of intrazonal trips while home-based nonwork (HBNW) trips, including shop, social recreational, and school, generally have the highest intrazonal activity. Table 4 depicts a benchmark range of percent intrazonal trips by purpose and standards for comparing observed and estimated intrazonal activity. The best comparisons are against observed geocoded survey data. Since intrazonal percentages vary considerably, it is recommended that comparisons be made against percentages from other similar models where household surveys have not recently been completed." (3)

Table 4
Percent Intrazonal Trips (c) (3)

Statistic	Benchmarks Low	Benchmarks High
Percent Intrazonal – HBW	1%	4%
Percent Intrazonal – HBSH	3%	9%
Percent Intrazonal – HBSR	4%	10%
Percent Intrazonal – HBSC	10%	12%
Percent Intrazonal – HBO[a]	3%	7%
Percent Intrazonal – NHB[b]	5%	9%
Percent Intrazonal – Total Trips	3%	5%

Statistic	Standards Acceptable	Standards Preferable
Percent Intrazonal, Observed Total Trips	+/-5%	

[a]HBO includes a variety of special trip purposes depending on the model (e.g., airport, college, and school).
[b] NHB includes combined purposes for NHB Work and NHB Nonwork, where appropriate.
[C] See Table 2 for a definition of trip purposes

Trip Assignment Calibration and Validation

Calibration of the trip assignment steps builds upon the successful calibration of each of the previous steps. If significant errors persist in these earlier steps, calibration of trip assignment will be meaningless. The analyst should proceed to a comparison of assigned volumes with counted volumes only after being satisfies that:

- The base year TAZ household characteristics and employment estimates are reasonable;
- The trip production and trip attraction estimates by trip attraction estimates by trip purpose are reasonable and in balance;
- The gravity model is distributing the trips appropriately and the resulting mean trip lengths and TLFDs by trip purpose are agreeing with observed values and;
- The mode split model, if utilized, is providing a reasonable share of trips to each mode by trip purpose

Trip assignment validation involves comparing model generated link volumes compared to traffic counts. These comparisons can be made at various levels:

1. Geographic Areas of the Study Area
2. Functional Classification of Roadway
3. Screenlines, Cutlines, Cordon lines
4. Link Specific Comparison

Statistics used to make the comparisons include [(2) (1)]:

1. "Absolute difference: Calculated as the actual difference, i.e. Estimated – Observed. The sign (positive or negative) may be an important indicator of performance.
2. Relative difference: Values are normalized to remove scaling effects. Can be expressed as a percentage difference (e.g. acceptable range might be +/-10%) or as a ratio (e.g. 0.9 to 1.1).
3. Correlation: In regression analysis, an equation is estimated which relates a dependent (or unknown) variable to one or more independent variables. Correlation analysis determines the degree to which the variables are related, i.e. how well the estimating equation actually describes the relationship. In the case of model validation, we determine the degree to which observed and estimated values are related. The most commonly used measure of correlation is the coefficient of determination R^2, which describes the amount of variation in the dependent variable which is explained by the regression equation. R^2 can range from 0 to 1, with a value of 0 for no correlation and 1 for perfect correlation. Acceptable values of R^2 can vary depending on the type of comparison being made, but it would ideally explain more than half of the variation $R^2 > 0.5$). Note that as aggregation increases, the amount of correlation will increase.
4. Variance: Statistical measures can be calculated which measure the variance between observed and estimated values. The most common measure for validation purposes is the Percent Root Means Square Error (RMSE)".

The specific measures utilized are:

- Volume to Count Ratios Percent Error by functional classification
- Volume to Count Ratios Percent Error by link volume groups
- Volume to Count Ratios Percent Error by screenlines and cut lines
- Aggregate VMT Statistics
- Coefficient of Determination
- Percent Root Mean Square Error

Additionally, for specialized analyses such as emissions estimation comparisons the quantity of estimated VMT can be made. Here model estimated link volumes multiplied by distance are compared to link estimates of VMT produced from the Highway Performance Monitoring System (HPMS) database.

For both highways and transit, the analyst must recognize the day-to-day inherent variability in traffic volumes and transit ridership. The analyst must be comfortable with the reliability in traffic volumes and transit ridership used for travel model validation. As a part of this, the analyst must verify how the highway counts were derived and whether the counts are AADT, annual average weekday traffic (AAWT), or some other measure.[3]

Guidelines provided by various agencies and the achieved level of accuracy are presented in this section. The information refers to standards rather than benchmarks. The first item to consider is the number of observations to be included in the verification, it is suggested that:

"The regional agency should strive to obtain traffic counts on 10 percent or more of the regionwide highway segments being analyzed, if resources allow. This 10 percent goal also applies to the distribution of counts in each functional classification (freeways and principal arterials, (at a minimum). Validation for groups of links in a screenline should include all highway segments crossing the screenline."[4]

- **Volume To Count Ratios and Percent Error**

Table 5 depicts recommended validation standards for volume-to-count ratios summed by category (e.g., facility type).

Table 5
Volume-to-Count Ratios and Percent Error [3]

	Standards	
Statistic	**Acceptable**	**Preferable**
Freeway Volume-to-Count	+/-7%	+/-6%
Arterial Volume-to-Count	+/-15%	+/-10%
Collector Volume-to-Count	+/-25%	+/-20%
One way/Frontage Road Volume-to-Count	+/-25%	+/-20%

- External model cordon lines should achieve +/-1 percent
- Screenlines with greater than 70,000 AADT should achieve +/-10 percent
- Screenlines with 35,000 to 70,000 AADT should achieve +/-15 percent
- Screenlines with less than 35,000 AADT should achieve +/-20 percent
- Cutlines +/-15 percent

The percent deviation traditionally is expressed as a function of link volumes or functional classification. A range in the desirable percent deviation for various link volumes has been developed by FHWA. [10] In tabular form they are provided by volume group in Table 6. For comparison purposes, values provided by the Michigan DOT are also included in Table 6. Also, functional classification is utilized to ensure the model is loading trips onto the functionally classified systems in a reasonable manner. The percent error by functional classification is the total assigned traffic volumes divided by the total traffic volumes (ground counts) for all links that have counted volumes, disaggregated by functional classification. [3]

Table 6
Percent Difference Targets for Daily Volumes Groupings

Average Annual Daily Traffic	Desirable Percent Deviation FHWA	Desirable Percent Deviation Michigan
< 1,000	200	60
1,000 – 2,500	100	47
2,500 – 5,000	50	36
5,000 – 10,000	25	29
10,000 – 25,000	20	25
25,000 – 50,000	15	22
> 50,000	10	21

Source: (8) (9) (10)

Recent work by Cambridge Systematics [3] has revised the percent errors based on five value groups. Table 7 depicts a range of acceptable and preferable accuracy ranges for five volume groups. The revisions were made upon the plus or minus one lane criteria considering revisions to the Highway Capacity Manual.

Table 7
Percent Error by Volume Group and Roadway Designs[3]

Statistic	Standards: Acceptable	Standards: Preferable
Percent Error: LT 10,000 Volume (2L road)	50%	25%
Percent Error: 10,000-30,000 (4L road)	30%	20%
Percent Error: 30,000-50,000 (6L road)	25%	15%
Percent Error: 50,000-65,000 (4-6L freeway)	20%	10%
Percent Error: 65,000-75,000 (6L freeway)	15%	5%
Percent Error: GT 75,000 (8+L freeway)	10%	5%

- **Aggregate Vehicle Miles of Travel (VMT) Statistics**

Cambridge Systematics has observed typical ranges of VMT per HH ranges from 60 to 75 while VMT per person is in the range of 24 to 32. It is suggested the assigned VMT should agree with the estimated VMT within +/-5% desirable and +/-12% acceptable.[3] VMT can also be compared by functional classification as noted in Table 8 for selected Metropolitan Area populations.

Table 8[11]
Urban Area VMT by Facility Type

Facility Type	**Urban Area Population**		
	Small (50-200K)	**Medium (200K-1M)**	**Large (>1M)**
Freeways/Expressways	**18-23%**	**33-38%**	**40%**
Principal Arterials	**37-43%**	**27-33%**	**27%**
Minor Arterials	**25-28%**	**18-22%**	**18-22%**
Collectors	**12-15%**	**8-12%**	**8-12%**

Source (11)

Work with the Memphis Metropolitan Planning Organization, Ohio Department of Transportation, and Virginia Department of Transportation Travel Demand Model viewed VMT by functional classification as a suggested percent difference as noted in Table 9.

Table 9[12]
Modeled Versus Observed VMT

Stratification		
Functional Class	**Acceptable**	**Preferable**
Freeways/Expressways	**±7%**	**±6%**
Principal Arterials	**±15%**	**±10%**
Minor Arterials	**±15%**	**±10%**
Collectors	**±25%**	**±20%**
All Links	**±5%**	**±2%**

- **Coefficient of Determination (R^2)**

The coefficient of determination estimates the correlation between the actual ground counts and the estimated traffic volumes and is produced by most software packages. A suggested regionwide coefficient of determination of more than 0.88 has been proposed.[8]

A sample scattergram is provided below (Figure 2). Further, it is suggested any links that lie outside of a reasonability boundary should be reviewed in more detail.

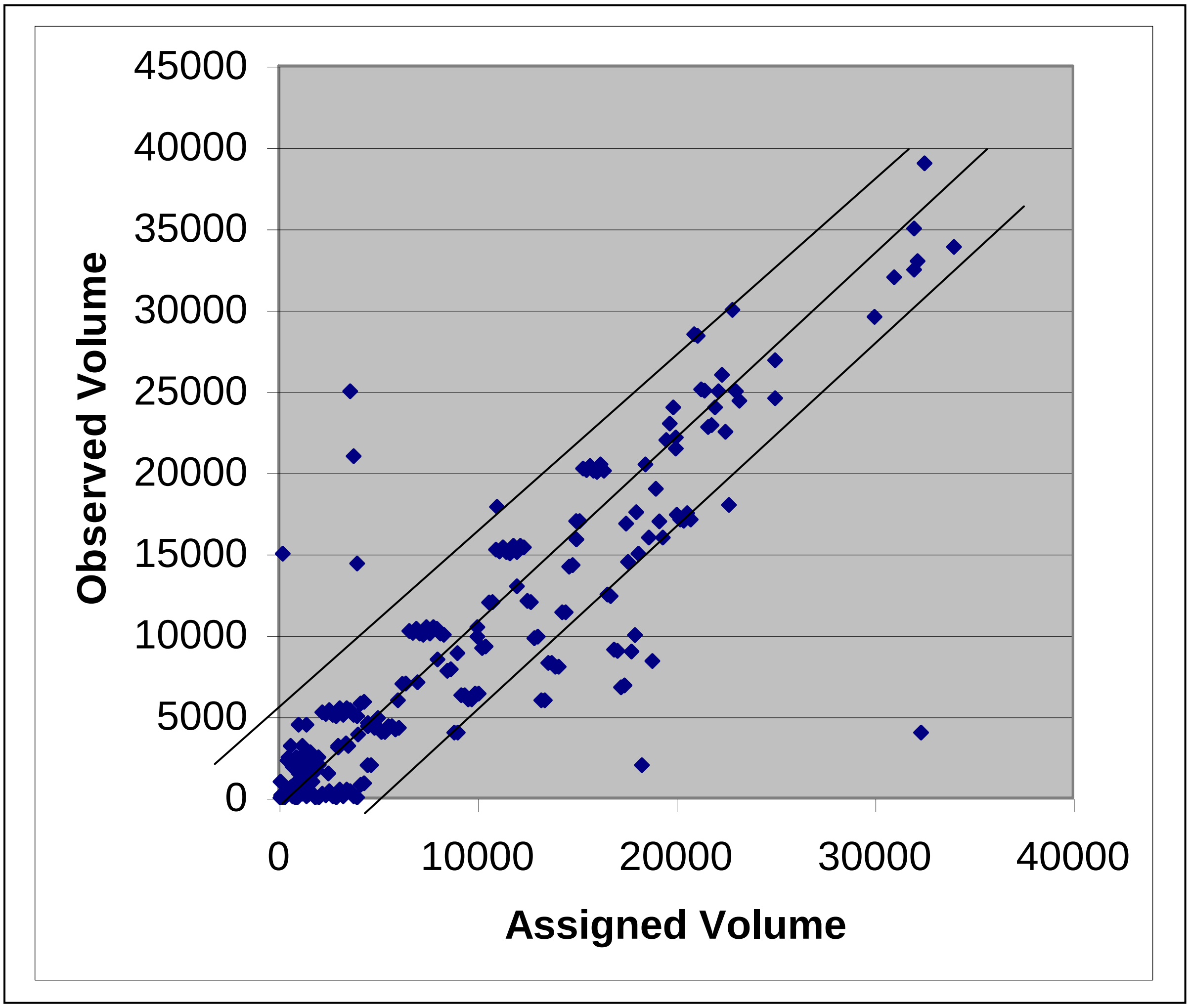


Figure 2- Sample Scatter Plot of Assigned vs. Observed LinkTraffic Volume

- **Percent Root Mean Square Error (RMSE) Counts**

The computation of RMSE is as follows [11]:

$$RMSE = \frac{\sqrt{\dfrac{\sum_j (Model_j - Count_j)^2}{(NumberofCounts - 1)}}}{\left((\sum_j Count_j)/(\text{Number of Counts}) \right)}$$

Cambridge Systematics[3] has provided RMSE as a function of lane volumes and an overall area wide performance measure as noted in Table 10.

Table 10
Root Mean Square Error (RMSE) By Volume Group[3]

Statistic	Standards Acceptable	Standards Preferable
RMSE: LT 5,000 VPD	100%	45%
RMSE: 5,000-9,999 VPD	45%	35%
RMSE: 10,000-14,999 VPD	35%	27%
RMSE: 15,000-19,999 VPD	30%	25%
RMSE: 20,000-29,999 VPD	27%	15%
RMSE: 30,000-49,999 VPD	25%	15%
RMSE: 50,000-59,999 VPD	20%	10%
RMSE: 60,000+ VPD	19%	10%
RMSE Areawide	45%	35%

The Virginia Department of Transportation has proposed guidelines concerning the Percent RMSE by Functional Roadway Type as presented in Table 11.

Table 11
Root Mean Square Error (RMSE) By Functional Class[13]

Functional Type	Small Regions	Large* Regions
Freeways	20%	20%
Principal Arterials	30%	35%
Minor Arterials	40%	50%
Collectors	70%	90%

*****Large regions are defined as Metropolitan Statistical Areas of population greater than 500,000 or have at least 200,000 population and are part of a metropolitan area with a population of more than 500,000.

- **Peak Hours/Period Validation Targets**
  Some agencies have focused on a model's peak hour volume estimate. Some typical validation criteria, as provided by Cambridge Systematics are [3]: The total AM and PM peak-hour volumes crossing each cordon or screenline should be within 10 percent of the AM and PM peak-hour counts, respectively. On an individual link basis, the following tolerances should be applied: [3] [10]
    - 75% of freeway link volumes within +/-20% acceptable
    - 50% of freeway link volumes within +/-10% desirable
    - 75% of major arterial/principal arterials link volumes with 10,000 vehicles per day within +/-30% acceptable
    - 50% of major arterial/principal arterials link volumes with 10,000 vehicles per day within +/-15% desirable
    - 75% of minor arterial link volumes within +/-40%
    - 50% of minor arterial link volumes within +/-20%

- **Automobile Occupancy**
  Table 12 provides a summary of suggested auto occupancy rates for Long Range "Highway Only" travel demand models:

Table 12
Suggested Auto Occupancy Rates [c] [3]

Statistic	Benchmarks/Settings Low	High
Auto Occupancy Rates – HBW	1.05	1.10
Auto Occupancy Rates – HBSH	1.50	1.80
Auto Occupancy Rates – HBSR	1.70	1.90
Auto Occupancy Rates – HBO[a]	1.65	1.95
Auto Occupancy Rates – NHB[b]	1.60	1.90

[a]HBO includes a variety of special trip purposes depending on the model (e.g., airport, college, and school).
[b] NHB includes combined purposes for NHB Work and NHB Nonwork, where appropriate.
[C] See Table 2 for a definition of trip purposes

Other Modeling Environments

Reference should be made to FSUTMS-Cube Framework Phase II – Model Calibration and Validation Standards[3] for a discussion of validation of models for use as sub area models, corridor models and models for traffic impact studies. Also an extensive discussion of mode choice model validation criteria is included for long range plans that have a mode choice element. Planners interested in a presentation of calibration and validation best practices should consult Chapter 3 of this document and the Travel Model Validation and Reasonability Checking Manual. [12]

Recommendations:

For application in Tennessee, it is suggested the following tests be conducted:

1. Percent difference in value for screenlines
2. Percent difference in value for link volumes (Table 5)
3. Percent difference in volume by classification(Table 6)
4. Correlate coefficient by link volumes
5. Root mean square for link volumes(Table 10)

Where FHWA and Michigan criteria are provided it is suggested the FHWA criteria be selected.

Tennessee MPO's and TPO's in the preparation of their next long range plan are encouraged, where practical, to consider the following test which will be considered for inclusion in the future draft of the calibration/validation guidelines.

1. Modeled versus observed VMT by functional classification (Table 9)
2. Root mean square by functional classification (Table 11)
3. Peak hour validation targets
4. Sample size documentation

References

1. Ismart, D. "Calibration and Adjustment of System Planning Models." Federal Highway Administration, Washington, DC, 1990.

2. "Model Validation and Reasonability Checking Manual." Barton Aschman Associates, Inc. and Cambridge Systematics, Inc. February 1997.

3. "FSUTMS – Cube Framework Phase II – Model Calibration and Validation Standards." Cambridge Systematics, Inc. prepared for the Florida Department of Transportation System Planning Office. Tallahassee, Florida. October 2008*.

4. FHWA, NHI, "Introduction to Urban Travel Demand Forecasting, Participant Workbook." NHI course number 152054, Publication # FHWA-NHI-02-040, March 2002.

5. N. Pedersen and D. Sandahl. "Highway Traffic Data for Urbanized Area Project Planning and Design." NCHRP Report 255. Transportation Research Board. Washington, DC 1982.

6. Harvey, G. et al. "A Manual of Regional Transportation Modeling Practice for Air Quality." Analysis for the Natural Association of Regional Councils, Washington, D.C. July 1993.

7. Robert Schiffer, et al. "Comparative Evaluation on the Elasticity of Travel Demand." Presented at 2005 TRB Meeting. Washington, DC 2005.

8. "Model Validation and Reasonableness Checking Manual 7.0 Assignment Procedures." Travel Model Improvement Program. U.S. Department of Transportation.

9. J.H.K. and Associates. "Travel Forecasting Guidelines for California Department of Transportation." Emeryville, California. November 1992.

10. L. Tekiele. "Michigan Department of Transportation Urban Model Calibration Target." Michigan Department of Transportation. Lansing, Michigan. June 1993.

11. P. Baso et al. "Travel Forecasting Guidelines." Texas Transportation Institute. Texas A&M University System. College Station, Texas. October 1994.

12. TMIP Travel Model Validation and Reasonability Checking Manual – Second Edition Cambridge Systematics Inc. Sept. 2010).

13. "Virginia Transportation Modeling Policies and Procedures Manual." Virginia Department of Transportation. Richmond, Virginia. May 2009.

*Also see the companion report FSUTMS-Cube Framework Phase I Default Model Parameter. Cambridge Systematics prepared for the Florida Department of Transportaion System Planning Office, Tallahassee, Florida. October 2006.

APPENDIX A

REASONABLENESS CHECKS
Source (1)

Socioeconomic Data Reasonableness Checks

- Review sources of TAZ base year estimates and forecast year projections.
- Graph 2000to 2010 county or metro population data from census: add forecast population used for travel model and evaluate trend for reasonableness.
- Graph 2000 to 2010 average household size from Census: add forecast household size used for travel model and evaluate trend for reasonableness.
- Graph 2000to 2010 regional median household income from Census after converting to constant dollars for the same year. Add forecast median household income used for travel model to the graph and evaluate trend reasonableness. The Census household income is for the year preceding the Census.
- Determine the constant dollar year that was used to develop the trip production rates, and verify that the forecast median household income is for that same constant dollar year.
- If average automobile availability per household was used in place of household income for trip production, or if automobile availability was used for the mode split model, plot average automobile availability against time and evaluate trend for reasonableness.
- Calculate the distribution of employment by basic, retail, and service employment, and compare the base and forecast year distributions. Do the changes in the distribution of employees by type of employment agree with what is known about the changing character of the urban area? Nationally, service employment is increasing relative to basic and retail employment.
- Plot average employees per household per capita for the region against time; evaluate for reasonableness. Nationally, employees per household have been increasing over time, but the rate of increase is decreasing.
- Plot average employees for the region against time checks reasonableness.
- Compare base year and forecast year number of households by zone or by sector (groups of zones), and evaluate for reasonableness. Do the zones or sectors with growth agree with your expectations for the urban area?
- Compare base year and forecast year employment by zone or sector and evaluate for reasonableness. Do the zones or sectors agree with your expectations for the urban area?
- Ask for or plot the area types for the base year and for each analysis year. The area type is a measure of urban density. Have the area type boundaries been revised to reflect increasing density as the urban area grows? Do the changes in area types appear reasonable?

Travel Survey Data Reasonableness Checks

- Determine the source of the travel survey information used for model calibration. Were travel surveys conducted? If yes, in what year? What kinds of travel surveys were conducted (household, workplace, external station, truck, special generator, etc.)?

- If surveys were not conducted, what was the source of the trip rates, mean trip lengths, and trip length frequency distributions used in the model calibration? If the survey data were "borrowed" from another urban area, is that urban area similar in terms of geographical area, population and household characteristics, employment characteristics, urban density, and transportation system characteristics?

- What definition was used for trip rates by trip chaining? These definitions are important as they directly affect the trip rates by purpose. It will be difficult to compare trip rates by trip purpose with those from other urban areas if the trip chaining definitions are different.

- How do the trip rates by trip purpose compare with those for other urban areas with similar characteristics? Were person or vehicle trip rates used? How do the rates compare with earlier rates for the same urban area?

- How do the mean trip lengths by trip purpose compare with those from other urban areas with similar characteristics? Are the HBW mean trip lengths the longest and the HBO mean trip lengths the shortest?

- How do the TLFDs by trip purpose compare with other urban areas with similar characteristics?

Network Data Reasonableness Checks

- Examine the plot of a tree from a major activity center, such as the regional airport, the CBD, or a regional shopping mall. A tree is a minimum time path from one zone to all other zones. Evaluate the routes for reasonableness. Are these the routes that a knowledgeable traveler would probably take? Repeat for several major attractors in the region.

- Examine the definitions of the facility types used to code the networks. What is the level of detail?

- Examine the speed and capacity look-up table. Was LOS C, D, or E used to code the capabilities? The link volume that must be achieved before the trip assignment algorithm starts decreasing the speed on the link is a function of the coded link capacity. And, the interpretation of the link volume to capacity ratio in the trip assignment output is directly dependent on the LOS capacity used for coding the network. The numerical values selected for the speed and capacity used for coding the network. The numerical values selected for the speed and capacity tables are part of the model calibration process. Once calibrated, these values do not change between the base and forecast year networks.

- The travel model should include a narrative or table description of all significant transportation projects added to the network for each analysis year. Examine this narrative or table description, and evaluate if these are still viable projects.

- Determine the major additions to the forecast networks in each analysis year. Are these transportation system improvements consistent with the most recent Long Range Plan (LRP)? Are the project scopes reflected in the network (location and number of additional lanes, location of LRT, location of HOV, etc.) consistent with the LRP?

- Most travel demand software packages have plotting capability tied to a GIS platform. Plots of facility types, number of lanes, freeflow speeds, and area types can be used to detect coding errors in facility type codes, number of lanes, freeflow speeds, and link area type. The software can also be used to compare the coded link area types. The software can also be used to compare the coded link lengths with lengths computed from the digitized node coordinates.

Trip Generation Reasonableness Check

- Examine the form of the trip production rate models. Is a cross classification or regression model used? Are the production rate models sensitive to household size, household income, and household automobile availability?

- Examine the form of the trip attraction rate models. Is a cross classification or regression model used? Are the attraction rate models sensitive to the total employment and retail employment? Are the models sensitive to basic, service, and retail employment and the number of households?

- What trips purposes were used? Were HBNW trips divided into additional trip purposes, such as retail or school trips?

- How were external-through and external-local trips accommodated in the travel model?

- Were the trip P&A rates for vehicle trips or for person trips?

- What was the P&A balance by trip purpose before scaling? Scaling factors of 0.9 to 1.1 are considered good. The scaling factor for the work trip purpose should be better than 0.9 to 1.1 and the scaling factor for NHB trips may be worse.

- Were special generators used in the travel model? If yes, what was the basis for the special generator trip rates? Were additional special generators of the same type added to the future year analysis? For example, if regional shopping malls were treated as a special generator were new shopping malls added to the forecast year? If yes, in what TAZ?

Trip Distribution Reasonableness Checks

- Examine the mean trip length for each trip purpose for the base year and for each forecast year. Is the mean trip length for HBW trips stable or increasing? Why? Is the change reasonable? The mean trip length for HBO trips is generally stable. Is it? If not, why not?

- Examine the base and forecast year plots of the TLFDs by trip purpose. A TLFD plot shows the percentage of trips occurring for each one minute separation in trip time. Do the plots look reasonable? Are there any significant changes in the shape of these distributions between the base and forecast years? If yes, why?

- Were the friction factors used in the gravity model kept constant between the base and forecast years? This is a general practice. If the friction factors were changed, what was the basis for the change?

- Were terminal times used for the trip distribution step? What logic was used to calculate terminal times? Was the same logic used for the forecast years?

- Were socioeconomic adjustments factors (k-factors) used in the base year calibration? If yes, what was the basis for calculating the k-factors? What distribution model deficiency did the k-factors correct?

- How does the HBW mean trip length resulting from the trip distribution step compare with the 2000 and 2010 Census journey-to-work (JTW) mean trip length?

- How do the sector interchange volumes or percentages resulting from the trip distribution step compare with the JTW sector interchange volumes or percentages? It is practical to conduct this comparison only at a sector level and not at the zone level.

One topic that receives a great deal of discussion is the use of K-factors in the calibration of the gravity model. Following is a discussion of the use of K-factors as presented in the Model Validation and Reasonability Checking Manual (2). The comments are presented in their entirety.

"K-factors are sector to sector factors which correct for major discrepancies in trip interchanges. These factors are computed as the ratio between observed and estimated trip interchanges. K-factors are typically justified as representing socioeconomic characteristics that affect trip making but are not otherwise represented in the gravity model. Physical barrier, such as a river crossing, may also result in differences between observed and modeled trip patterns. For example, trip movements between zones separated by a bridge may not be as great as would be expected using only quantifiable measures. In that case, the planner can use either k-factors or artificial times on the bridge links to match the actual interchange of travel."

A specific problem with trip distribution occurs when low income households are matched with high income jobs in the central business district, particularly for large metropolitan areas. Although there are certainly trips between low income residences and downtown business districts, trip distribution models can have a tendency to overstate these trips. This error can have an even greater impact on travel projections since low income riders tend to be more transit dependent and transit is usually more competitive with the automobile downtown.

The uses of K-factors are generally discouraged and are seen as a major weakness with traditional gravity models when used to correct for socioeconomic factors. Since K-factors represent characteristics of the population which change over time, the assumption that K-factors stay constant in the future can introduce a significant amount of error in predictions of future trip distributions.

A preferred approach is to stratify trip productions and attractions by income class (or auto ownership) and perform separate distributions of trips by class. Each model can reflect the different distributions of employment types throughout the region, as well as the unique sensitivities of different classes of travelers to travel time."

Trip Assignment Reasonableness Checks

- Examine the all-or-nothing assignment. These are the volumes that would result if there were enough capacity for everyone to travel on their minimum time path. A comparison of the all-or-nothing volume with the equilibrium assignment volume for a given link provides an indication of the amount of traffic that might divert to that link if additional capacity were provided.

- Examine the equilibrium assignment. This assignment is the travel model's estimate of the amount of traffic forecasted to occur on a particular network, given the zonal household and employment forecast (the demand of transportation) and the coded transportation network (the supply of transportation).

- What volume delay function was used? Was the BPR function? If not, what function was used? At what volume to capacity ratio does diversion begin to occur with the function used?

- For model validation, how did the assigned screen line volumes compare with the counted screen line volumes?

- For model validation, how did the assigned cut line volumes compare with the counted cut line volumes?

- Was a time-of-day assignment performed? If yes, what were the time periods? What was the source of directional split factors by time-of-day? What was the source of the time-of-day factors? Were time-of-day factors kept constant between the base and the forecast year or were peak spreading factors used? If peak spreading factors were used, what was the source of the factors?

- If time-of-day modeling was not performed, how were the 24-hour volumes converted to hourly volumes for the traffic assignment step? What conversion factor(s) was used? Or, what was the relationship between peak hour volume and 24-hour volume?

- Intrazonal and centroid connector trips can be used to estimate local travel not assigned to the transportation network. What fraction of total VMT was assigned to intrazonal and centroid connector trips? Is this value reasonable?

APPENDIX B

ADDENDUM TO GUIDELINES (Source 12)

ADDENDUM TO GUIDELINES

Since the proposed Tennessee Model User's Group (TNMUG) "Minimum Travel Demand Model Calibration and Validation Guidelines for the State of Tennessee" was developed, the Federal Highway Administration (FHWA) has published a comprehensive report entitled "Travel Model Validation and Reasonability Checking Manual Second Edition." The report prepared by Cambridge Systematics was released on September 2010 and is an update to the 1997 publication of the same title.

Contained in the report is a detailed discussion of the validation process with emphasis on the application of data resources, reasonability checks, and sensitivity analyses. Specific elements of the model validations process covered include:

1. Model Validation Plan Specification;
2. Collection and Assessment of Validation Data;
3. Validation of Model Components;
4. Validation of Model System, and
5. Documentation of Validation Results.

An interesting philosophy introduced is the concept of a standard versus threshold in the validation process. As stated:

> The term "threshold" rather and "standard" will generally be used throughout this manual. The term standard connotes a formal definition of acceptance: "The standard has been met, therefore the model is valid." While it is important to match base year observations for validation, simple matching of traffic counts, for instance, is not sufficient to establish the validity of a travel mode. Quality model validation must test all steps of the travel model and also should test model sensitivity. If standards are set for models by agencies or model reviewers, it is beneficial that they not convey a formal definition of acceptance but rather to help set boundaries or levels of confidence regarding the use of travel forecasts for studies.

In general, the following guidelines should be used to determine acceptable methods for achieving improved match between modeled and observed travel characteristics:

- The adjustments should reflect transportation supply or traveler behavior rather than simple arithmetic;
- The adjustments should be reproducible; and
- The reasons for adjustments should be clearly documented.

The manual is a valuable resource providing troubleshooting strategies for key steps in the modeling process and well as providing default values based on the analysis of the 2001 National Household Travel survey (NHTS) data. Included are the following default tables based on 2001 NHTS data.

- Trip Rates by Purpose Stratified by Number of Persons by MSA Population;
- Trip Rates by Purpose Stratified by Number of Workers by MSA Population;
- Trip Rates by Purpose Stratified by Number of Autos by MSA Population;
- Trip Rates by Purpose Stratified by Income Level by MSA Population;
- Motorized Trip Percentages by Urban Area Population, and
- Time of Day Percentage for Urban Area of Approximately 1 Million Population by Trip Purpose.

Also included are:

- Trip Distribution Gamma Factor Perimeters, and
- Example Friction Factors Based on Gamma Functions by Trip Purpose.

Finally, the assignment validation included discussions on the role of scatter plots, and speed versus volume/capacity ratio comparison plots. Example guidelines presented largely reflect what is presented in the Minimum Travel Demand Calibration and Validation Guidelines. A chapter on modal choice modeling is included which might be of interest to the larger Metropolitan Planning Organizations (MPOs).