Recommendations for Expanded Validation Metrics for CFRPM v6.2

Draft Version

July 21, 2016

#### Acronyms and Abbreviations

AADT Annual Average Daily Traffic

ACS American Community Survey

AIP CFRPM Application Improvement Plan

AM Ante meridiem

APC Automated Passenger Counter(s)

AT Area type

CFRPM Central Florida Regional (Transportation) Planning Model

CFX Central Florida Expressway

CTPP Census Transportation Planning Program

D5 Florida Department of Transportation, District 5

DoD United States Department of Defense

EE External to External trips

FDOT Florida Department of Transportation

FHWA Federal Highway Administration

FT Facility type

FTA Federal Transit Administration

HBW Home-based Work trips

HBSH Home-based Shopping trips

HBSR Home-based Social/Recreation trips

HBSC Home-based School trips

HBO Home-based Other trips

HBNW Home-based Non-work trips (typically the sum of HBSH, HBSR, HBSC and HBO)

I-4 BtU Interstate 4 Beyond the Ultimate Project

IE Internal to External trips

IVTT In-vehicle Travel Time

LRTP Long-Range Transportation Plan

MPO Metropolitan Planning Organization

NHB Non-home-based trips

NHTS National Highway Travel Survey

NL Number of lanes

OCCC Orange County Convention Center

OIA Orlando International Airport

OVTT Out-of-vehicle Travel Time

PM Post meridiem

%RMSE Percent Root Mean Squared Error

TAZ Traffic Analysis Zone

VHT Vehicle-hours of Travel

VMT Vehicle-miles of Travel

VOT Value of time

Table of Contents

[Introduction 4](#_Toc456798326)

[Proposed Validation Tests and Metrics 5](#_Toc456798327)

[Input Data 6](#_Toc456798328)

[Travel Generation 7](#_Toc456798329)

[Travel Distribution 8](#_Toc456798330)

[Modal Choice 9](#_Toc456798331)

[Special Area Sub-Models 11](#_Toc456798332)

[Travel Assignment 12](#_Toc456798333)

[Longitudinal Tests 13](#_Toc456798334)

[Documentation of Validation Tests 15](#_Toc456798335)

[Next Steps 16](#_Toc456798336)

#### List of Tables

[Table 1 ZDATA Validation Metrics 6](#_Toc456798337)

[Table 2 Highway and Transit Network Visual/Mapping Reasonableness and Logic Checks 7](#_Toc456798338)

[Table 3 Aggregate Travel Trip Rate Metrics and Benchmarks (applied to each county) 8](#_Toc456798339)

[Table 4 Travel Distribution Metrics and Benchmarks 9](#_Toc456798340)

[Table 5 Additional Travel Distribution Metric and Benchmark for I-4 BtU 9](#_Toc456798341)

[Table 6 Modal Choice Metrics and Benchmarks 10](#_Toc456798342)

[Table 7 Special Area Sub-Model Metrics and Benchmarks 11](#_Toc456798343)

[Table 8 Roadway Validation Metrics and Standards 12](#_Toc456798344)

[Table 9 Additional Roadway Validation Metrics and Standards for I-4 BtU 13](#_Toc456798345)

[Table 10 Transit Validation Metrics and Benchmarks 13](#_Toc456798346)

[Table 11 Longitudinal Reasonableness and Logic Checks 14](#_Toc456798347)

# Introduction

The Florida Department of Transportation, District 5 will soon conduct a validation exercise of its newest version of the Central Florida Regional (Travel) Planning Model (CFRPM), version 6.2. The base year for CFRPM v6.2 will be 2015. It is expected that this version will assist the Department in analyzing environmental traffic related aspects of the I-4 “Beyond the Ultimate” Project (I-4 BtU). The I-4 BtU project is studying the addition of express lanes between approximately the Kirkman Road and US27 interchange. This document details the metrics and process to be used for this particular model validation.

Validation is “the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model”[[1]](#footnote-1). “The process of model calibration and validation is vital to producing defensible travel demand forecasts”[[2]](#footnote-2). Validation helps to ensure the CFRPM reasonably reflects existing transportation network, travel patterns, and the usage of the network, and that it will be a useful tool for upcoming I-4 BtU activities. The validation results inform planners, policy and decision-makers of the model’s strengths and weaknesses beyond its immediate intended purpose, and identify and prioritize future CFRPM adjustments to address those weaknesses or accentuate its strengths.

It is important that a model can be considered valid for certain purposes without meeting every benchmark and standard. If fact, models that achieve every benchmark and standard, assuming the benchmarks and standard reflect all model components, are commonly found later to be over-calibrated. Over-calibrating occurs when the model is adjusted in a way – usually to achieve a particular set of benchmarks and standards – that does not directly conform to a well-studied aspect of travel behavior. Over-calibration deprives the model of its ability to properly react to changed socio-demographic or transportation conditions. Consequently, models that do not meet every benchmark can be considered valid, and sometimes more valid than those achieving extensive lists of benchmarks and standards.

# Proposed Validation Tests and Metrics

There are four categories of tests commonly used in travel model validation. The descriptions of these tests are taken from FHWA’s Reasonableness Manual (insert footnote here).

**Comparisons of base year model results to observations or benchmarks** might be considered “traditional” validation. The comparisons might be of model results to disaggregate data such as data from a supplementary survey not used for model estimation or to aggregate data such as traffic counts or transit boardings. Comparing base year model results to different aggregations of the data used to estimate or calibrate a model is not as sound of a validation practice as comparing to independent data. However, for some validation tests, the data used for model estimation or calibration are the only data available.

**Longitudinal tests** are important aspects of model validation since, by definition, it implies comparing model results to data not used in model estimation. Both backcasts and forecasts may be used for model validation. For example, if a model is estimated using 2007 survey data, the model could be used to backcast to 2000 conditions, and compared to year 2000 traffic counts, transit boardings, CTPP data, or other historical data. Likewise, if a model was estimated or calibrated using 2005 survey data, a “forecast” validation could be performed against 2008 data.

**Reasonableness and logic checks** include the comparison of estimated (or calibrated) model parameters against those estimated in other regions with similar models. Reasonableness and logic checks may also include “components of change” analyses and an evaluation of whether or not the models “tell a coherent story” as recommended by the FTA for New Starts analysis.

**Model sensitivity testing** includes several important types of checks including both disaggregate and aggregate checks. Disaggregate checks, such as the determination of model elasticities, are performed during model estimation. Aggregate sensitivity testing results from temporal validation. Sensitivity testing can also include model application using alternative demographic, socioeconomic, transportation supply, or policy assumptions to determine the reasonableness of the resulting travel forecasts.

The process recommended for CFRPM v6.2 intends to use all four categories. The tests will be applied to all components of CFRPM: input data, travel generation, travel distribution, mode choice, special area sub-models and travel assignment.

The metrics used will consist of traditional travel model metrics and metrics related to performance-based planning. Recent guidance from FHWA (performance measures) and FDOT (managed lanes) have been included also.

Additional checks are included for the I-4 BtU project. The FDOT Standards recommend applying many of the same tests and metrics. Corridor-specific tests are included in separate tables.

The proposed process for each model component is:

1. Assemble the described observed data, if applicable.
2. Determine the extent the observed data can be used for validation testing. For example, the observed data could have systemic biases or variability that make it untenable for validation purposes.
3. Assemble the appropriate model outputs.
4. Compare the model outputs to the observed data and/or benchmarks and standards described in this report.
5. Assess the model’s performance given the quality of the observed data, identifying significant differences.
6. As schedule permits, locate the root cause of significant differences between model outputs and observed data. Develop a plan to correct the root cause, and enact the plan. The plan should be tied to a real-world behavior or observation and avoid over-calibrating the model. After enacting the plan, restart the validation process with Step 4. If the schedule does not permit enacting the plan, the issue should be recorded and addressed at the next opportunity.
7. Summarize the model’s performance, highlighting its strengths, weaknesses and unknowns.

The purpose of the plan in Step 6 is to improve the model’s ability to reflect travel characteristics. The plan should never over-calibrate the model. Over-calibrating occurs when the model is adjusted in a way that does not directly conform to a well-studied aspect of travel behavior with the purpose of exactly matching model output to observed values. Over-calibration causes the model to quickly lose its ability to properly react to changed socio-demographic or transportation conditions.

A travel model is designed by its nature to react and respond appropriately to reasonable changes in sociology-demographic variables and transportation systems. Over-calibrating, or more precisely tightly calibrating, the travel model restricts its ability to reasonably react to changed variables and systems. Unfortunately, over-calibrating is readily instinctive to modeling analysts because of the inherent desire to have the model match observed values as closely as mathematically possible. This desire is misplaced and therefore needs to be tempered with the realization that over-calibrating both restricts the model's ability to provide helpful information for project-level analysis and mistakenly disregards the natural variability of the observed data.

The following sections describe the validation tests, observed data and benchmarks or standards for each model component.

## Input Data

The checking of input data is a crucial step in the validation process. Significant errors in the input data typically leads to weak model performance in other validation tests. Three major types of input data need to be reviewed for validation purposes: the socio-demographic data (known as zonal data, or ZDATA), the highway network, and the transit network.

ZDATA is developed using standardized population figures by county, which are then distributed to TAZs within each county by the respective MPO. The benchmarks, taken directly from the Department’s Model Calibration and Validation Report, are largely based on historical demographic and socio-economic trends and are well-established in social science fields. The ZDATA validation tests in Table 1 can be found in the Olympus training model provided by the Department.

Table 1 ZDATA Validation Metrics

|  |  |  |
| --- | --- | --- |
| **Metric** | **Benchmark** | |
| **Low** | **High** |
| Visual inspection of population and employment and associated densities by TAZ and county | None (reasonable judgment) | |
| Regionwide persons/dwelling unit or persons/household | 2.0 | 2.7 |
| Regionwide employment/population ratio | 0.35 | 0.75 |
| Regionwide autos/dwelling unit or autos/household | 1.75 | 2.10 |
| Approximate population per TAZ | n/a | 3,000 |

Highway and transit networks require a detailed review by a professional familiar with the Central Florida transportation system. The highway network checks consist of characteristics that have an impact on estimated speeds and volumes. These characteristics are compared against readily-available observed information:

* HERE speed data observed from the night period to assess the reasonableness of free-flow roadway speeds,
* FDOT’s Roadway Characteristics Inventory to verify the facility types and number of lanes on roadway facilities, and
* Publicly available toll data will be used to evaluate the accuracy of network toll data.

No quantifiable observed data is available for area types and terminal times. These are based on a reasonable understanding of planning and modeling concepts. Terminal times, for example, are highly dependent on localized conditions but are typically higher in denser area types and special facilities like airports. It is important to review areas types and terminal times, however, as they factor into the travel distribution and modal choice computations.

Similarly, the transit network checks consist of characteristics that have an impact on estimated speeds, modal shares and route-level volumes. These characteristics are compared against readily-available observed information. Publicly available transit time tables are used to assess the accuracy of transit service frequencies modes, operators and travel times. General transit route and stop coding will be assessed by assigning transit trip tables from the 2010 transit onboard surveys to the transit network.

For both highway and transit networks, the review should be done more carefully within the I-4 BtU corridor.

Table 2 Highway and Transit Network Visual/Mapping Reasonableness and Logic Checks

|  |  |
| --- | --- |
| **#** | **Check** |
| 1 | Color-coded maps of area types |
| 2 | Color-coded maps of facility types |
| 3 | Color-coded maps of roadway lanes |
| 4 | Color-coded maps of traffic counts |
| 5 | Inspection of toll data |
| 6 | Color-coded maps of free-flow roadway speeds |
| 7 | Inspection of roadway turn penalties and prohibitions |
| 8 | Color-coded map of terminal times |
| 9 | Color-coded maps of transit service frequencies |
| 10 | Color-coded maps of transit modes |
| 11 | Color-coded maps of operators |
| 12 | Review of end-to-end travel time by transit route |
| 13 | Assignment of transit on-board survey to transit network and inspection of results |

## *Travel Generation*

Generation – estimating the magnitude of travel for each TAZ – is the product of the ZDATA and travel rates. Travel is computed in terms of (a) productions, the number of trips being produced within a TAZ, and (b) attractions, the number of trips attracted to a TAZ. The input data checks in the previous section provide critical detailed information for the validation relative to travel generation. The metrics in this section, therefore, focus on the productions and attractions computed in the generation step. Tests examine normalized trip rates across a variety of categories and the relative proportion of different trip purposes.

Production rates are typically computed from a recent, local household travel survey. Such a survey has not been conducted in Central Florida for decades. CFRPM’s trip rates are derived from a decades-old statewide report on average trip rates. Attraction rates are typically computed from location-based surveys. No such survey has ever been conducted in Central Florida.

The benchmarks are from the Department’s Model Calibration and Validation Standards Report produced in 2008. They are based on a variety of sources, including Census data, household travel surveys, NHTS tabulations, and Federal and state guidelines on modeling practice.

Table 3 Aggregate Travel Trip Rate Metrics and Benchmarks (applied to each county)

|  |  |  |
| --- | --- | --- |
| **Metric** | **Benchmark** | |
| **Low** | **High** |
| Relative comparison of trip rates by county | None (reasonableness and logic check) | |
| Person trips per TAZ | n/a | 15,000 |
| Person trips per person | 3.3 | 4.0 |
| Person trips per dwelling unit or household | 8.0 | 10.0 |
| HBW person trips/employee | 1.20 | 1.55 |
| Relative difference between unbalanced attractions to productions (all purposes) | 0-10% | 50% under certain conditions |
| Percent of HBW trips relative to all other trips | 12% | 24% |
| Percent of HBSH trips relative to all other trips | 10% | 20% |
| Percent of HBSR trips relative to all other trips | 9% | 12% |
| Percent of HBSC trips relative to all other trips | 5% | 8% |
| Percent of HBO trips relative to all other trips | 14% | 28% |
| Percent of HBNW trips relative to all other trips | 45% | 60% |
| Percent of NHB trips relative to all other trips | 20% | 33% |
| Percent of EE trips relative to all other trips | 4% | 21% |

## *Travel Distribution*

Travel distribution is the process of linking trip productions to attractions across the region. Distribution is evaluated in three aspects: district-to-district flows, average trip lengths by trip purpose, and the percentage of trips that occur within a single TAZ. The recommended metrics also include an inspection of sub-area constants, also called K-factors. These constants can be used to subtly alter district-to-district flows, but their use is discouraged. Constants of large magnitude are strongly discouraged, as they can render the model insensitive to roadway network or socio-economic changes over time.

Unfortunately, there is sparse local data to validate travel distribution. Two traditionally conducted tests, the relative difference between estimated and observed mean trip lengths and computing coincidence ratios, cannot be performed for CFRPM because of the absence of a regional household travel survey. AirSage data may be a viable source if the trip classifications are reliable and the districts are sufficiently sized.

ACS data is available to evaluate the district-to-district flows of HBW trips. A regional household travel survey is preferred to validate average trip lengths and intrazonal trips for the other trip purposes. There is no up-to-date survey for the Central Florida region. Therefore, benchmarks for average trip lengths and intrazonal trips will be used for validation.

The validation benchmarks are taken directly from the Department’s Model Calibration and Validation Report. They are based on Census data and household travel surveys.

Table 4 Travel Distribution Metrics and Benchmarks

|  |  |  |
| --- | --- | --- |
| **Metric** | **Benchmark** | |
| **Low** | **High** |
| Presence of sub-area constants or K-factors | Minimal application and all such values 0.80-1.20 (acceptable) | Absence of these constants/factors (preferable) |
| HBW average trip length (minutes) | 12 | 35 |
| HBSH average trip length (minutes) | 9 | 19 |
| HBSR average trip length (minutes) | 11 | 19 |
| HBSC average trip length (minutes) | 7 | 16 |
| HBO average trip length (minutes) | 8 | 20 |
| NHB average trip length (minutes) | 6 | 19 |
| IE average trip length (minutes) | 26 | 58 |
| County-to-county flows by time period and trip purpose | None (reasonable judgment) | |
| Percent of intrazonal HBW trips relative to all HBW trips | 1% | 4% |
| Percent of intrazonal HBSH trips relative to all HBSH trips | 3% | 9% |
| Percent of intrazonal HBSR trips relative to all HBSR trips | 4% | 10% |
| Percent of intrazonal HBSC trips relative to all HBSC trips | 10% | 12% |
| Percent of intrazonal HBO trips relative to all HBO trips | 3% | 7% |
| Percent of intrazonal NHB trips relative to all NHB trips | 5% | 9% |
| Percent of intrazonal total trips relative to all trips | 3% | 5% |
| Percent intrazonal trips relative to all observed trips | +/- 5% (acceptable) | +/- 3%  (preferable) |

An additional metric is included for the I-4 BtU. An inspection of the Intra-corridor and corridor-related trips is required to ensure that the trip movements are reasonable. AirSage data may be used for this comparison once it has been reviewed by the Department’s consultants.

Table 5 Additional Travel Distribution Metric and Benchmark for I-4 BtU

|  |  |
| --- | --- |
| **Metric** | **Benchmark** |
| I-4 BtU intra-corridor and corridor-related trips  by time period and trip purpose | None (reasonable judgement) |

## *Modal Choice*

Modal choice is the process of evaluating the available roadway and transit options and estimating the number of trips taking each option. Because modal choice models are calibrated directly from observed data, the validation efforts focus on the results of calibration: district to district flows by transit and access mode, the magnitude of the sub-modal, access and modal-specific constants, and the presence of non-logit decision rules. Review of the coefficients, which have asserted and based on nationwide experience, is also included.

The primary data source for evaluating the accuracy of trip flows is the 2010 transit surveys. Onboard travel surveys were conducted for LYNX, VOTRAN and Lake Xpress transit agencies. Equivalent data for SpaceCoast (Brevard County’s transit system) and the Polk transit agencies is not available. Consequently, the evaluation of trip flows will only be applicable to those counties with transit survey data. (Updated surveys for LYNX and VOTRAN are planned for late 2016, but it is anticipated that this data will not be finalized in time for this validation effort.) The 2010 surveys will be re-expanded to be consistent with the latest 2015 ridership data from the three agencies.

Some of the validation benchmarks are taken directly from the Department’s Model Calibration and Validation Report, while others are based on experiences with FTA’s New Starts (now called the Capital Investment Grant program) projects. Inspection of sub-modal, access and modal-specific constants is one such check. Constants that are large relative to their respective IVTT coefficient can over-ride the travel variables in mode choice equations, making the model directly produce pre-defined values. Thus the model becomes insensitive to socio-demographic or transportation changes and unhelpful for project-level evaluation.

Analyzing non-logit decision rules is another such check. Non-logit decision rules are inserted in the logit computations to temporarily improve calibration results. An example is a rule that automatically designates trips to auto modes if the total trip length is less than five minutes, rather than allowing the logit computations to estimate the auto and transit modal shares. Unfortunately, they sometimes cause undesirable (positive or negative) impacts on ridership and modal choice results. The benchmarks of constants and coefficients in the metrics are intended for evaluation at the top-most level of nested logit models.

Table 6 Modal Choice Metrics and Benchmarks

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Benchmark** | | |
| **Preferable** | **Acceptable** | |
| Regionwide transit trips versus observed targets | +/- 1% | +/- 2% | |
| Regionwide transit trips versus observed stratified by:  Trip purpose, time period, auto ownership, and sub-mode | +/- 1% | +/- 2% | |
| Modal shares versus observed targets | +/- 2  percentage points | +/- 2  percentage points | |
| Transit Trip flows within and between districts by mode (by access mode, purpose and auto ownership | Each cell with 30% (relative to cell) or  5% (relative to total transit trips) | | |
| Park-ride vehicle trips versus observed data by park-ride lot | +/- 20% | +/- 30% | |
| IVTT Coefficient – HBW | -0.05 | -0.01 | |
| IVTT Coefficient – HBNW | -0.033 | -0.007 | |
| IVTTT Coefficient – NHB | -0.05 | -0.01 | |
| Ratio of OVTT / IVTT Coefficients – HBW | 2.0-3.0 | | |
| Ratio of OVTT / IVTT Coefficients – HBNW |
| Ratio of OVTT / IVTT Coefficients – NHB |
| Implied value of time – percent of household income | 25%-33% | | |
| Implied value of time – HBW | $2.00-$5.00 | | |
| Implied value of time – HBNW | $0.50-$5.00 | | |
| Implied value of time – NHB | $0.20-$5.00 | | |
| Ratio of HBW constants to HBW IVTT coefficient | < 20 min | < 60 min | |
| Ratio of HBNW constants to HBNW IVTT coefficient | < 30 min | < 60 min | |
| Ratio of NHB constants to NHB IVTT coefficient | < 30 min | < 60 min | |
| Ratio of modal-specific constants to IVTT coefficient (each purpose) | < 10-12 min | < 15 min | |
| Ratio of access-mode constants to IVTT coefficient (each purpose) | < 20 min | < 40 min | |
| Absolute magnitude of sub-modal constants decreasing with increasing auto ownership (|0-car C| > |1-car C| > |2+car C|) | None (logic check) | | |
| LogSum coefficients | 0.3 – 0.7 | | 0.3 – 1.0 |
| Consistency of ratio of OVTT / IVTT parameters across modal choice and all path-building components | Exact match | Within 25% | |
| Consistency of ratio of implied VOT across modal choice and  all path-building components | Exact match | Within 25% | |
| Absence of non-logit decision rules | Total absence | Rules with proven avoidance of undesirable or illogical impacts to modal shares at a zonal level | |

## *Special Area Sub-Models*

Central Florida contains theme parks that are among the most attended in the world. These have been incorporated in CFRPM via special area sub-models, essentially microcosms of the regional model system. They have distinct travel generation, distribution and modal choice components. The special area sub-models to be validated are in the I-4 BtU corridor: Walt Disney World area, Universal Parks area, Sea World area and the Orange County Convention Center. The other three special area sub-models, the Orlando International Airport, Port Canaveral and Kennedy Space Center areas, will be validated in a subsequent validation effort.

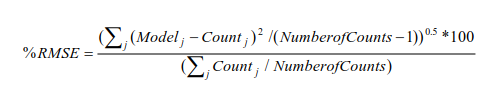
The metrics used to validate the special area sub-models represent a selection of metrics used to validate the main model’s travel generation, distribution and modal choice components. Travel generation will be compared to publicly released attendance figures. District-to-district trip flows will be compared to AirSage data, which is being collected for this validation effort. Before making these comparisons, the AirSage flow data will be reviewed for reasonableness. The modal choice metrics focus on the coefficients and non-logit decision rules since trip data by mode and auto occupancy is not available for any of the special areas.

Table 7 Special Area Sub-Model Metrics and Benchmarks

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Benchmark** | | |
| **Preferable** | **Acceptable** | |
| Person trip attractions compared to observed attendance | +/- 1% | +/- 2% | |
| Total district flows to/from the special area | Each cell with 30% (relative to cell) or  5% (relative to total transit trips) | | |
| IVTT Coefficient – HBW | -0.05 | -0.01 | |
| IVTT Coefficient – HBNW | -0.033 | -0.007 | |
| IVTTT Coefficient – NHB | -0.05 | -0.01 | |
| Ratio of OVTT / IVTT Coefficients – HBW | 2.0-3.0 | | |
| Ratio of OVTT / IVTT Coefficients – HBNW |
| Ratio of OVTT / IVTT Coefficients – NHB |
| Implied value of time – percent of household income | 25%-33% | | |
| Implied value of time – HBW | $2.00-$5.00 | | |
| Implied value of time – HBNW | $0.50-$5.00 | | |
| Implied value of time – NHB | $0.20-$5.00 | | |
| Ratio of HBW constants to HBW IVTT coefficient | < 20 min | < 60 min | |
| Ratio of HBNW constants to HBNW IVTT coefficient | < 30 min | < 60 min | |
| Ratio of NHB constants to NHB IVTT coefficient | < 30 min | < 60 min | |
| Ratio of modal-specific constants to IVTT coefficient (each purpose) | < 10-12 min | < 15 min | |
| Ratio of access-mode constants to IVTT coefficient (each purpose) | < 20 min | < 40 min | |
| LogSum coefficients | 0.3 – 0.7 | | 0.3 – 1.0 |
| Consistency of ratio of OVTT / IVTT parameters across modal choice and all path-building components | Exact match | Within 25% | |
| Consistency of ratio of implied VOT across modal choice and all path-building components | Exact match | Within 25% | |
| Absence of non-logit decision rules in modal choice model | Total absence | Rules with proven avoidance of undesirable or illogical impacts to modal shares at a zonal level | |

## *Travel Assignment*

Travel assignment is the loading of trips onto the roadway and transit facilities. Highway and transit assignment metrics have traditionally been the focus of past CFRPM validation efforts. Comparisons of volume to traffic counts are well established within the industry, and are based on FHWA guidance. Some volume-to-traffic tests require computing a %RMSE metric. The equation for this metric is:



Volume to traffic count comparisons will use 2015 FDOT traffic counts. New roadway metrics have been added to validate the congested travel times in selected corridors. These metrics will compare observed HERE data to the congested travel times.

Table 8 Roadway Validation Metrics and Standards

|  |  |  |
| --- | --- | --- |
| **Metric** | **Standards** | |
| **Acceptable** | **Preferable** |
| Freeway volume-over-count ratio (FT 10s, 80s, 90s) | +/- 7% | +/- 6% |
| Divided arterial volume-over-count ratio (FT 20s) | +/- 15% | +/- 10% |
| Undivided arterial volume-over-count ratio (FT 30s) | +/- 15% | +/- 10% |
| Collector volume-over-count ratio (FT 40s) | +/- 25% | +/- 20% |
| One-way/Frontage Road volume-over-count ratio (FT 60s) | +/- 25% | +/- 20% |
| Freeway AM peak volume-over-count ratio | 75% of links are within 20% | 50% of links are within 10% |
| Major arterial AM peak volume-over-count ratio | 75% of links are within 30% | 50% of links are within 15% |
| Assigned VMT-over-count ratio regionwide | +/- 5% | +/- 2% |
| Assigned VHT-over-count ratio regionwide | +/- 5% | +/- 2% |
| Assigned VMT-over-count ratio by FT / AT / NL | +/- 25% | +/- 15% |
| Assigned VHT-over-count ratio by FT / AT / NL | +/- 25% | +/- 15% |
| Volume-over-count ratios for external model cordon lines | +/- 1% | |
| Screenlines with greater than 70,000 AADT | +/- 10% | |
| Screenlines with 35,000 to 70,000 AADT | +/- 15% | |
| Screenlines with less than 35,000 AADT | +/- 20% | |
| Percent error for volume group < 10,000 AADT | 50% | 25% |
| Percent error for volume group 10,000-30,000 AADT | 30% | 20% |
| Percent error for volume group 30,000-50,000 AADT | 25% | 15% |
| Percent error for volume group 50,000-65,000 AADT | 20% | 10% |
| Percent error for volume group 65,000-75,000 AADT | 15% | 5% |
| Percent error for volume group 75,001+ AADT | 10% | 5% |
| RMSE for links with < 5,000 vehicles per day | 100% | 45% |
| RMSE for links with 5,000-9,999 vehicles per day | 45% | 35% |
| RMSE for links with 10,000-14,999 vehicles per day | 35% | 27% |
| RMSE for links with 15,000-19,999 vehicles per day | 30% | 25% |
| RMSE for links with 20,000-29,999 vehicles per day | 27% | 15% |
| RMSE for links with 30,000-49,999 vehicles per day | 25% | 15% |
| RMSE for links with 50,000-59,999 vehicles per day | 20% | 10% |
| RMSE for links with 60,000+ vehicles per day | 19% | 10% |
| RMSE regionwide | 45% | 35% |
| AM peak roadway travel times in selected travel corridors | 80% of corridors within 20% | 50% of corridors within 10% |
| Midday roadway travel times in selected travel corridors | 80% of corridors within 20% | 50% of corridors within 10% |
| PM peak roadway travel times in selected travel corridors | 80% of corridors within 20% | 50% of corridors within 10% |

Additional roadway validation metrics are included for the I-4 BtU project. They provide a more detailed assessment of the model within the I-4 BtU project corridor. The metrics and standards are from the Department’s Model Calibration and Validation Report.

Table 9 Additional Roadway Validation Metrics and Standards for I-4 BtU

|  |  |  |
| --- | --- | --- |
| **Metric** | **Standards** | |
| **Acceptable** | **Preferable** |
| Freeway volume-over-count ratio for I-4 BtU corridor | +/- 6% | +/- 5% |
| Divided arterial volume-over-count ratio for I-4 BtU corridor | +/- 10% | +/- 7% |
| Undivided arterial volume-over-count ratio for I-4 BtU corridor | +/- 10% | +/- 7% |
| Collector arterial volume-over-count ratio for I-4 BtU corridor | +/- 15% | +/- 10% |
| One-way/frontage road volume-over-count ratio  for I-4 BtU corridor | +/- 20% | +/- 15% |
| AM peak roadway travel times in I-4 BtU corridor | +/- 25% | +/- 15% |

Many of the transit assignment metrics are taken from the Department’s Model Calibration and Validation Report. Comparisons of estimated to observed boardings comprise the majority of the metrics. Ridership information from 2015 will be used as the basis of comparison. Additional metrics have been added for boardings at LYNX Central Station, the region’s largest transit station, and SunRail, the region’s first rail system.

Table 10 Transit Validation Metrics and Benchmarks

|  |  |  |
| --- | --- | --- |
| **Metric** | **Benchmarks** | |
| **Acceptable** | **Preferable** |
| Ratio of estimated-to-observed transit boardings  (each agency and mode) | +/- 9% | +/- 3% |
| Ratio of estimated-to-observed transit boardings  (Lynx Central Station) | +/- 25% | +/- 15% |
| Average transfer rate (each agency and mode) | Within 10 percentage points | Within 5 percentage points |
| SunRail ridership (aggregate) | +/- 20% | +/- 10% |
| SunRail ridership (for each station) | +/- 20% or within 300 boardings/day | +/- 10% or within 100 boardings/day |
| Route-level ridership: <1,000 boardings/weekday | +/- 150% | +/- 100% |
| Route-level ridership: 1,000-2,000 boardings/weekday | +/- 100% | +/- 65% |
| Route-level ridership: 2,000-5,000 boardings/weekday | +/- 65% | +/- 35% |
| Route-level ridership: 5,000-10,000 boardings/weekday | +/- 35% | +/- 25% |
| Route-level ridership: 10,000-20,000 boardings/weekday | +/- 25% | +/- 20% |
| Route-level ridership: 20,000+ boardings/weekday | +/- 20% | +/- 15% |
| Route-level end-to-end travel times (each route) | +/- 20 minutes | +/- 5 minutes |

## *Longitudinal Tests*

A longitudinal test is the review of model outputs to changed socio-economic and transportation conditions over time. These tests, not regularly performed in past CFRPM validations, assess CFRPM’s ability to reflect reasonably the impact of projected socio-demographic growth on the model system and transportation network. Longitudinal tests should only be performed after the model has been validated with base year data.

Two common longitudinal tests are backcasts and forecasts. Backcasts analyze the model’s reaction to socio-economic and transportation conditions from the recent past, usually within the past 10 years. Forecasts analyze the model’s reaction to conditions in the future, up to 30 years in the future is common.

Two longitudinal tests are suggested for CFRPM validation. One is a backcast to 2010 conditions. The 2010 scenario does not include SunRail (began operations in 2014) and the higher economic conditions that exist in 2015. The 2010 backcast will help determine CFRPM’s ability to evaluate large transit infrastructure investments and the changed socio-economic conditions. The other proposed longitudinal test is a forecast to 2040, the LRTP analysis year. Substantial development and transportation infrastructure investments are expected to occur by 2040. The 2040 forecast will help determine CFRPM’s ability to evaluate changes from the substantial changes in socio-economic conditions and transportation conditions.

The metrics for each longitudinal test are based on FHWA guidance and the Department’s Model Calibration and Validation Report.

Table 11 Longitudinal Reasonableness and Logic Checks

|  |  |
| --- | --- |
| **#** | **Check** |
| 1 | Changes in socio-demographic variables for each county |
| 2 | Changes in total trips generated within each county |
| 3 | Changes in county-to-county travel patterns |
| 4 | Changes in transit shares relative to changes in socio-demographic variables |
| 5 | Changes in auto occupancy relative to changes in socio-demographic variables |
| 6 | Changes in travel speeds (all major facilities, with special attention to I-4 BtU facilities) |
| 7 | Changes in VMT and VHT per capita by county |
| 8 | Evaluation of these changes and impacts on regional travel |
| 9 | Comparison to 2010 peak and off-peak travel speeds (backcast only) |
| 10 | Comparisons to daily traffic volumes (backcast only) |
| 11 | Changes in transit ridership by operator and mode |

# Documentation of Validation Tests

The model validation results should be documented in a way that provides a complete and straightforward analysis to the users of CFRPM results. A new format to the validation is recommended because many of the validation metrics, tests and checks in this document are new to CFRPM validation procedures. The outline presented here is an extension of the CFRPM Model Validation Report outline developed during the AIP in January 2016.

The proposed outline and format for the CFRPM v6.2 Model Validation Report borrows heavily from FHWA’s Travel Model Validation and Reasonableness Checking Manual. It is critical that this document provide analysis, rather than solely enumerating validation results. This document will be incorporated to CFRPM’s new wiki site.

1. **Executive Summary**. The objective for the executive summary is to summarize the report findings so that both experienced model users and non-modelers understand the usefulness and the limitations of the model. The summary should contain:
   1. Summary information about the model, including a map and description of the model area and major transportation facilities;
   2. A summary of the CFRPM process;
   3. A statement of the purpose and need for the validation;
   4. An overview of the validation process, information on the validation data;
   5. A summary of the validation results;
   6. A summary of model strengths and weaknesses; and
   7. Information regarding the types of studies for which the model is valid and for which it should not be used.
2. **Component-by-Component Review and Analysis.** Model components are input data, travel generation, travel distribution, modal choice, special area sub-models, and travel assignment. Each component should have its own chapter and include at least the following material:
   1. Summary description of the technical process of the component;
   2. Enumeration of the validation results and how they compare to benchmarks and standards;
   3. Documentation on model adjustments and corrections made during the validation process;
   4. An assessment of the validation results, including a summary of the component’s strengths and weaknesses; and
   5. Information regarding the types of studies for which the component is valid and for which it should not be used without further enhancement.
3. **Longitudinal Tests**. This chapter documents the results of each longitudinal test. It should include an assessment of the plausibility of the model’s reactions to the changes in socio-economic and transportation.
4. **Model System Validation**. This chapter distills the information in the preceding chapters and provides a summary of the model’s strengths and weaknesses given the cumulative validation results of the validation components.
5. **Next Steps**. This chapter will include suggestions for future efforts on model component updates and future data collection efforts and surveys.

# Next Steps

Preparation of the observational datasets should begin immediately for the forthcoming validation effort. The observed datasets mentioned in this report are:

* FDOT’s Roadway Characteristics Inventory – prepare number of lanes and facility type information for Input Data reasonableness and logic checks;
* FDOT’s Traffic Count Inventory – prepare 2010 and 2015 AADT counts for Input Data reasonableness and logic checks;
* Florida’s Turnpike and CFX publicly available toll data – gather existing toll rate data for Input Data reasonableness and logic checks;
* HERE data – develop average observed speeds for AM, midday, PM peak and night for all major facilities in CFRPM and all facilities in the I-4 BtU corridor;
* Publicly available transit schedules and maps – gather 2010 and 2015 travel time, service frequency and routing information from LYNX, VOTRAN, Lake Xpress, and SpaceCoast transit agencies. Develop an observed table of AM peak and midday end-to-end run times for all routes and operators;
* Transit ridership information – gather transit ridership information by bus route and by SunRail station for 2010 and 2015 from LYNX, VOTRAN, Lake Xpress, and SpaceCoast;
* 2010 LYNX, VOTRAN and Lake Xpress transit survey dataset – gather the 2010 dataset, re-expand the 2010 dataset to 2015 boarding data and develop a 2015 transit trip table for LYNX, VOTRAN and Lake Xpress;
* AirSage data – assemble AirSage data for district-to-district trip flow comparisons;
* Theme park attendance – download and assemble a table of theme park attendance from the 2015 AECOM and TEA Theme and Museum Index Report (<http://www.aecom.com/theme-index/>);
* ACS data – develop district-to-district flows using the most recent data available at a TAZ level (2006-2010); and
* I-4 BtU traffic counts – code these project counts, collected April-June 2016, on the 2015 highway network.

#### References

Federal Highway Administration. Travel Model Validation and Reasonableness Checking Manual (2nd Edition). 2010.

Florida Department of Transportation. Model Calibration and Validation Standards. 2008.

1. DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A), ), DoD Instruction 5000.61, December 9, 2009. <http://www.dtic.mil/whs/directives/corres/pdf/500061p.pdf> [↑](#footnote-ref-1)
2. Florida Department of Transportation. FSUTMS-Cube Framework Phase II, Model Calibration and Validation Standards. 2008. <http://www.fsutmsonline.net/images/uploads/reports/FR2_FDOT_Model_CalVal_Standards_Final_Report_10.2.08.pdf> [↑](#footnote-ref-2)