National Capital Region
Transportation Planning Board

MEMORANDUM

TO: Dusan Vuksan and Mark Moran, TPB staff

FROM: Feng Xie, TPB staff

SUBJECT: Year-2014 Validation of TPB's Version 2.3 Travel Demand Model

DATE: March 12, 2019

INTRODUCTION

TPB's current travel forecasting process is known as the Version 2.3 Travel Demand Model. Since TPB adopted the Version 2.3 model in November 2011, it has evolved into the current build #75 (also known as Version 2.3.75) with periodic updates. Prior to 2019, the Version 2.3 model's ability to produce credible future-year forecasts had been assessed in two major model validation efforts. The first assessment was carried out during 2009-2011 when build #36 of the model was calibrated and validated to year-2007 conditions using the 2007/08 Household Travel Survey and on-board transit surveys data [1]. The second assessment took place in 2013 with a focus on validating build #39 of the model to year-2010 conditions [2].

Entering 2019, a federal requirement associated with air quality conformity (AQC) determination prompted TPB staff to conduct a third validation of the Version 2.3 model. The air quality conformity analysis, which determines if long-range plans conform to air quality improvement goals embodied in the Clean Air Act, is an important application of the Version 2.3 model. The procedures for AQC determination must comply with federal transportation conformity rules set by Environmental Protection Agency [3]. One of those rules implies that TPB's Version 2.3 model with a validated base year of 2010 can only be used for conformity determination up to and including 2020.[1] As the deadline approached, TPB staff validated the Version 2.3.75 model to year-2014 conditions during January-February 2019, thereby enabling the Version 2.3 model to be used for AQC analyses in the next five years.[2]

In addition to the re-validation of TPB's Version 2.3 production model, TPB staff is undertaking two model development activities that involve model validation. TPB's Version 2.5 model is currently under development, and TPB staff is in the process of evaluating an initial round of model validation results generated with consultant assistance. TPB staff is also preparing the scope of work (SOW) for the TPB's next-generation travel model (known as Generation-3 or Gen3), which will include the task of developing a model validation plan.

---

[1] Specifically, the federal rule established in §93.122(b)(1)(i) requires "network-based travel models must be validated against observed counts (peak and off-peak, if possible) for a base year that is not more than 10 years prior to the date of the conformity determination". That implies a regional travel demand model with a validated based year of 2010 can only be used for conformity determination up to and including 2020.

[2] It is important to note that although the Version 2.3.75 model was validated to more recent conditions, this model, consistent with the rest of the Version 2.3 model family, was calibrated to the 2007/08 Household Travel Survey and on-board transit surveys data.

This memorandum documents the results from the recent re-validation of the Version 2.3 model and provides insights for TPB's ongoing model validation efforts. Starting with a review of the literature on model validation, the next section clarifies the definition of key terms and identifies areas where TPB's model validation routines could be enhanced or improved. A validation plan for this re-validation of the Version 2.3 model is then presented in the "Validation Plan" section, which is followed by a summary of the validation results in the "Validation Results" section. The last section provides a summary of the main findings and offers suggestions on the validation of TPB's two developmental models.

OVERVIEW

In the past, a wide body of literature has been produced on model validation because of the critical role it plays in travel demand forecasting. Drawing on a selective set of national and state guidelines [4]-[7], TPB technical reports or memoranda [1]-[2], consultant reports [8]-[10], and reports from fellow MPOs [11]-[16], this overview is not intended to duplicate the information that is already presented in the reference material. Instead, it focuses on 1) clarifying the definition of key terms on model validation and 2) identifying areas where TPB's routine model validation practice could be enhanced or improved.

Key Terminology and Definitions

Model validation is a loosely defined term and its definition varies in different references with different focuses. Federal Highway Administration (FHWA)'s Travel Model Validation and Reasonableness Checking Manual, one of the most definitive reference sources for model validation in the U.S., views model validation as an integral component of the model development and application process that follows the steps of model estimation/assertation, model calibration, model validation/sensitivity testing, and model application, and defines validation as "the application of the calibrated models and comparison of the results against observed data" [4]. With a few exceptions,[3] the definitions in other references generally align with FHWA's.

Model validation should be distinguished from **model calibration**. By FHWA's definition, model calibration is the "adjustment of constants and other model parameters in estimated or asserted models in an effort to make the models replicate observed data for a base (calibration) year or otherwise produce more reasonable results". In contrast, model validation doesn't involve adjusting model constants or parameters and it should use additional data that are not used for model estimation or calibration (although this is not always feasible in practice).

Depending on whether the validation data come from the same year as model estimation/calibration or from an alternative year, model validation can be categorized as **traditional validation** or **temporal**

[3] For instance, Florida Department of Transportation (FDOT) views "calibration" and "validation" as sometimes interchangeable terms except that "calibration" is focused on model steps before assignment (matching travel behavior) while "validation" on highway and transit assignment (matching ground counts) [7]. The Metropolitan Transportation Commission (MTC) in San Francisco Bay Area, also focusing on validation for assignment steps only, defines model validation as a stage to verify "the ability of the models to reproduce specific observations of travel patterns seen on the actual network that were not used in model calibration" [14].

validation.[4] While traditional validation, as part of model development, usually entails validating each model component in a travel demand forecasting process, many temporal validations are less rigorous and "focus on matching traffic counts and transit boardings" from the assignment steps "due to the general availability of those data" [4]. Temporal validation can either be conducted along with traditional validation when a travel model is developed or serve as periodic check to "ensure that established travel models continue to reasonably reproduce observed traffic and transit ridership" over time [5].

Model validation should include **sensitivity testing** which evaluates "the application of the models and the model set using alternative input data or assumptions" [4]. Model validation should also include **reasonableness and logic checks** which examine if model results are reasonable and logically coherent. In stricter terms, however, sensitivity testing or reasonableness/logic checks are different from validation tests as they don't involve a comparison against observed data, especially against locally collected data.

Metrics generated by validation tests are usually benchmarked against validation standards or guidelines. FDOT uses the term "**standards**" to describe desirable accuracy levels for comparing estimated and observed metrics and the term **"benchmarks"** to represent typical values or ranges documented in national publications, FDOT technical reports, or model validation studies [7].[5] FHWA and some other state DOTs (such as VDOT), on the other hand, use the less rigid term "**guidelines**" to represent all standards, threshold values or benchmarking ranges, emphasizing that those are considered useful targets, but not pass/fail tests [4][6]. In fact, the use of rigid standards is not recommended in both FDOT's and FHWA's guidance, as the attempt to meet a hard standard might make a model worse in other ways, e.g., overfitting the model, which can worsen its predictive capability.

Areas for Improvement

The survey of the state of the practice indicates room for improving TPB's routine model validation practices. Particularly, **six areas are identified below for potential improvement.** While some of these improvements are implemented in this re-validation of the Version 2.3 model, others remain to be addressed in the future.

1. Validating every step of the model chain
Although many validation efforts are predominantly focused on assignment statistics due to limited data availability, it is still considered best practice to validate every step of the model chain, especially when a travel model is developed. For a more complex model system such as an activity-based model (ABM), this task may become especially challenging, as there could be 10 to 20 model components requiring validation.

2. Developing a validation plan

---

[4] The Virginia Department of Transportation (VDOT) uses the alternative terms of "**static validation**" and "**dynamic validation**" and further categorizes dynamic validation into "**backcasting**" **validation** or **forecasting validation** depending on whether the validation year is prior to or after the base year [6].

[5] Comprehensive checklists of validation metrics by model step, along with standards or benchmarking ranges published nationwide, can be found in the appendices of FDOT's Model Calibration and Validation Standards Final Report [7].

As FHWA's model validation handbook points out, "the development of a model validation plan will enhance the likelihood of a successful validation process" [4]. Although it is useful to specify a validation plan prior to the initiation of model validation, it is even better practice to develop a model validation plan at the outset of the model development process when important decisions such as those related to validation data, scope and costs are made.

3. Including sensitivity testing
Model validation should include sensitivity testing. There is a wide range of sensitivity tests available for inclusion: sensitivity tests conducted for individual model components or for the entire model set; sensitivity tests using alternative demographic, socioeconomic, or network inputs data, sensitivity analyses carried out for a base year or for a future year, etc. Selection of sensitivity tests depends on modeling needs as well as federal and state guidance.[6]

4. Conducting model validation in an iterative fashion
Model development and application is an iterative process by which issues uncovered during model validation or application can be led back to preceding model steps. Ideally, model validation can be improved with additional rounds of model estimation or calibration addressing those issues. In practice, however, it may not be feasible due to the limitations of time and resources.

5. Benchmarking validation metrics
Local validation metrics could be checked against available accuracy standards or benchmarking ranges. Although those checks should not be considered as rigid pass/fail tests, they can help modelers gauge and improve model performance.

6. Introducing complementary validation tests
While TPB's past validation efforts already covered a variety of validation tests, additional tests could be used to further improve model validation. For instance, a consultant to the TPB staff has suggested including transit volume validation on screenlines to enhance the transit validation of the Version 2.3 model. In recent years, the increasing availability of commercial crowdsourced travel time data (such as the INRIX data) makes it possible to validate traffic speeds on roadways. It is reported that several large MPOs have already conducted speed studies and included comparisons of observed versus estimated speeds [7]. It should be noted, however, that there is often a direct conflict between model adjustments used to match volumes versus those to match speeds, and that TPB currently prioritizes validating to traffic counts over observed speeds.

VALIDATION PLAN

This re-validation effort, which is planned to be carried out during January - February of 2019, validates the TPB's current officially adopted travel demand model (Version 2.3.75) to Year-2014 conditions. Staff selected Year-2014 for this re-validation as the observed data for this analysis year had already been developed for other parallel efforts, especially for Version 2.5 model development.

The main purpose of this re-validation is to validate the model to more recent traffic conditions in compliance with the federal requirement regarding the air quality conformity (AQC) determination. It also serves as a periodic check of the Version 2.3 model to ensure that the model continues to

[6] For instance, FDOT emphasizes the importance of future-year sensitivity analysis, indicating that "model validation should not be considered complete until forecast year sensitivity tests are completed" [7].

produce reasonable travel forecasts one decade after it was developed. This re-validation is thus intended to be a temporal validation focusing on matching ground counts rather than a full-blown traditional validation that is usually conducted when a travel model is developed.

Based on the observed and estimated 2014 data, this re-validation performs a range of validation tests focusing on highway and transit assignment steps. The resulting validation metrics are benchmarked against metrics obtained in TPB's prior validation efforts, as well as against applicable accuracy standards or benchmarks found in federal or state guidance. This validation also includes a discussion of sensitivity testing results from a stand-alone sensitivity analysis based on the Version 2.3 model.

The remainder of the validation plan describes **the step-by-step methodology** for implementing this re-validation.

1. Literature Review
2. Validation Data Preparation
   a. Assemble observed data from the following sources:
      i. Year-2014 Vehicle Miles of Travel (VMT) by jurisdiction from HPMS, which can be extracted from the Version 2.5 validation report [9],
      ii. Year-2014 highway count data assembled for Version 2.5 model validation [17],
      iii. Additional 2014 highway count data provided by TPB staff per internal request,
      iv. Year-2014 observed transit ridership data by transit sub-mode prepared by consultant for Version 2.5 model validation, mainly based on household travel survey and transit on-board survey data provided by MWCOG/TPB [9],
      v. Year-2014 Metrorail ridership by station published by Washington Metropolitan Area Transit Authority (WMATA) [18], and
      vi. Year-2014 Metrorail station-to-station Origin/Destination (O/D) volume data published by WMATA [19].
   b. Extract benchmarking data[7] from the following sources:
      i. TPB Version 2.3 Travel Forecasting Model Calibration Report [1],
      ii. TPB technical memorandum on the 2010 validation of the Version 2.3 model [2],
      iii. TPB Travel Forecasting Subcommittee status report on Version 2.5 travel model development and evaluation [20],
      iv. TPB technical memorandum on WMATA Silver Line Ridership Forecast [21],
      v. FHWA Travel Model Validation and Reasonableness Checking Manual [4],
      vi. VDOT Manual on Travel Demand Modeling Policies and Procedures [6], and
      vii. FDOT Report on Model Calibration and Validation Standards [7].
3. Execute Travel Demand Model
   a. Assemble model inputs for Year 2014; [8]

[7] Validation metrics are benchmarked against two sets of reference data: one contains the validation metrics generated in TPB's past validation efforts, and the other consists of accuracy standards or guidelines specified in federal or state model validation guidance. Specifically, model validation manuals developed by FHWA, VDOT and FDOT are used. Virginia is chosen because Northern Virginia is part of TPB's modeling area; Florida is chosen because FDOT's model validation report, which contains a comprehensive survey of the state-of-the-practice validation metrics, standards and benchmarks, is frequently cited nationwide.
[8] Year-2014 model inputs data include Round 9.1 Cooperative Land Use Forecasts and the "Visualize 2045" Long-Range Plan (LRP) network database. Since Year-2014 was not an analysis year for the Visualize 2045 LRP evaluation, the 2014 network needs to be developed specifically for this re-validation.

   b. Conduct 2014 model runs;
   c. Create standard summaries for quality assurance;
4. Conduct Highway Validation
   a. Verify the quality of the additional highway count data; [9]
   b. Update the 2014 highway count data with the additional counts;
   c. Based on the updated count data, update input files to a model performance summary program TPB staff developed for Version 2.5 highway validation;
   d. Update the model performance summary program to include additional validation tests;
   e. Execute the model performance summary program to generate highway validation metrics;
5. Conduct Transit Validation
   a. Update the 2014 observed Metrorail ridership with a fix related to the Silver Line;
   b. Develop 2014 observed Metrorail ridership by station group;
   c. Develop observed Metrorail average weekday volumes crossing selected screenline/cordon line using the Metrorail station-to-station O-D volume data;
   d. Extract 2014 simulated transit ridership by transit sub-mode using LineSum; [10]
   e. Extract 2014 simulated Metrorail ridership by station and calculate simulated Metrorail ridership by station group;
   f. Calculate simulated Metrorail volumes crossing selected screenline/cordon line in ArcGIS;
   g. Develop validation metrics by comparing the simulated to observed data;
6. Benchmarking Validation Metrics
   a. Compare resulting validation metrics to metrics developed in TPB's prior model validations;
   b. Compare resulting validation metrics to available accuracy standards or benchmarks in federal or state guidance;
7. Documentation
   a. Summarize preliminary validation results in tables and graphics;
   b. Discuss preliminary validation results internally;
   c. Incorporate sensitivity testing from a stand-alone planning study;
   d. Document validation results and findings in a technical memorandum.

VALIDATION RESULTS

Highway Validation

For the Year-2014 validation of the Version 2.3 model, TPB staff performed an array of highway validation tests based on observed VMT or highway count data vs. estimated traffic volumes from highway assignment. Most of these tests generated validation metrics by comparing the estimated to the observed data. TPB staff then determined whether a specific validation metric meets the accuracy standard by comparing validation metrics against the corresponding standard where applicable.

Based on type of observed data and type of metrics being generated, these highway validation tests can be loosely divided into four categories: 1) daily VMT by jurisdiction tests based on the HPMS

---

[9] The verification of additional count data follows the same six-step procedure for cleaning highway hourly count data documented in a TPB technical memorandum. [19]

[10] Specifically, the "ACCESS_Report" method in LineSum is used to extract the estimated Metrorail ridership as the Metrorail observed data is faregate counts that don't include in-system transfers; the "ON_OFF_Report" method is used for other transit sub-modes as their observed counts are derived from transit on-board surveys which may include transfers.

data, 2) daily or time-of-day VMT by facility type tests and daily VMT by area type tests based on daily or hourly link counts, 3) link-based daily volume correlation tests (R-Squared or % RMSE), and 4) daily volume screenline test. Appendix A summarizes detailed testing results[11] in tables and graphics, while Table 1 below outlines key findings from the appendix with reference to corresponding appendix tables or graphics.

As shown in Table 1, daily VMT estimated to observed (E/O) ratio is 1.02 areawide (HPMS-based), which meets the standard for preferable accuracy level. Table A1 in Appendix A provides a detailed breakdown of estimated vs. observed VMT comparison by county. As shown in Table A1, county-level E/O ratios range from 0.87 to 1.38 [12] for the entire modeled area, and from 0.92 to 1.13 for the TPB Planning Area, which is most relevant to transportation planning studies in this region. Eight (8) out of 10 TPB member jurisdictions are within 10% and 9 out of 12 non-member jurisdictions are within 15%, indicating that VMT validates reasonably well at the county level.

Table 1. Summary of Year-2014 Highway Validation Metrics for the TPB Modeled Area

Validation Test (Estimated vs. Observed)	Metrics	Accuracy Standard (Acceptable/Preferable)	Benchmarking Results	Reference
Daily VMT Areawide (HPMS based)	1.02	±5% / ±2%	Preferable	Table A1
Daily VMT by County (HPMS based)	0.87-1.38	N/A	N/A	Table A1
Highway Links Daily Count Coverage	20.2%	N/A	N/A	Table A2
Highway Links Hourly Count Coverage	5.1%	N/A	N/A	Table A2
VMT by Facility Type (Daily Count-based)	1.06	±5% / ±2%	Marginally acceptable	Table A3-1
Freeway	1.07	±7% / ±6%	Acceptable	
Major Arterial	1.07	±15% / ±10%	Preferable	
Minor Arterial	1.13	±15% / ±10%	Acceptable	
Collector	0.74	±25% / ±20%	Marginally acceptable	
Expressway	0.95	±15% / ±10%	Preferable	
VMT by Area Type (Daily Count-based)	0.95-1.22	±25% / ±15%	Mostly preferable	Table A3-2
Time-of-Day VMT (Hourly Count-based)	0.92-1.12	N/A	N/A	Table A4
Daily Volumes R-Squared	0.90	0.9	Met standard	Figure A1
Daily Volumes % RMSE Areawide	42.6%	40%	Marginally acceptable	Table A5-1
Daily Volumes % RMSE by Facility Type	13.4%-76.0%	N/A	N/A	Table A5-1
Daily Volumes % RMSE by Volume Group	19.4%-110.1%	19%-100%	Marginally acceptable	Table A5-2
Daily Volumes on Regional screenlines	0.70-2.21	±10%(vol>50k); ±20% (vol<=50k)	14 out of 34 screenlines met standard	Map A1 Table A6

As shown in Table A3-1, E/O ratios resulting from the daily VMT by Facility Type test range from 0.95 to 1.13. In this test, the simulated and observed VMT are calculated by aggregating traffic volumes on 6,693 directional highway links with daily counts (AAWDT) available, which account for 20.2% of all links coded in the 2014 planning network (Table A2). As expected, freeways (1.07), major arterial (1.07) and expressways (0.95) validate better than minor arterials (1.13) and collectors (0.74). Collectively, the E/O ratio is 1.06, which is marginally acceptable against the standard.

[11] Please note that the estimates included in the appendix are directly extracted or computed from model outputs. Number of significant figures in those estimates doesn’t necessarily indicate level of accuracy.
[12] As noted in Table A1, Spotsylvania County has the observed VMT for the entire county but the estimated VMT for northern portion of county only, so its outlier statistics (0.65) is excluded from this analysis.

Daily VMT by Area Type, on the other hand, validates well against the standard (Table A3-2). The resulting E/O ratios vary between 0.95 and 1.22. With only one exception (Area Type 6; E/O = 1.22), the benchmarking results for all area types are "preferable."

As shown in Table A4, VMT by facility type and by time-of-day are validated using hourly traffic counts on 1,676 highway links. Such a table is a new addition that has not been part of recent validation efforts at COG. Resulting E/O ratios by time of day can be found in the bottom row of the table. These E/O ratios indicate that PM Peak VMT validates the best (0.92) while Mid-Day VMT the worst (1.20), and that PM Peak is the only time period for which VMT is under-estimated. As a quality assurance check, this test also generates daily VMT by facility type (using the 1,676 links with hourly counts) and the resulting E/O ratios (0.72-1.22) track well with those shown in Table A3-1 (using the 6,693 links with daily counts). When interpreting the results of this time-of-day VMT test, though, one needs to be mindful that these validation results, developed based on hourly counts, may not be as reliable as those developed based on daily counts, since the 1,676 links with hourly cover only 5.1% of the planning network (Table A2).

Figure A1 displays a scatter plot between simulated daily volumes and AAWDT counts. A regression model between the simulated and the observed indicates the two variables are close to perfect agreement (Slope=0.99). The resulting R-Squared is 0.90, which met the R-Squared standard (0.90) set by VDOT for large model regions.

Percent Root Mean Squared Error (% RMSE) is another important metric to gauge the correlation between simulated and observed daily volumes.[13] Table A5-1 and Table A5-2 contain daily volume % RMSE by facility type and by volume group, respectively. As expected, the resulting % RMSE is better for link groups on higher road hierarchies or with larger traffic volumes. Overall % RMSE (42.6%) and % RMSE for volume groups with daily volume over 15,000 vehicles marginally met accuracy standards, while the two volume groups below 15,000 didn't meet the standards.

Validation of daily volumes on regional screenlines indicates room for improvement - only 14 out of the 34 screenlines with count met the accuracy standards (Table A6). Map A1 displays the geographical locations of these screenlines and colors them according to their daily volume E/O ratios. While most screenlines located in regional core and inner suburb validate well (shown in green), many screenlines near external count stations, especially those in Maryland, are over-estimated (shown in blue or purple).[14] Another interesting observation is that, as shown in the inset map, although Screenline #20 (Potomac River Screenline) validates fairly well (0.93), the two Virginia screenlines intersecting with Screenline #20 (#1 and #3) are both under-estimated while the two DC screenlines (#2 and #4) are both over-estimated, indicating a possible "two-wrongs-make-a-right" situation regarding the validation on the river screenline.

Lastly, the above Year-2014 highway validation results for Version 2.3.75 were compared to TPB's prior validation results, specifically, to Year-2010 validation results for Version 2.3.39 [2] and preliminary Year-2014 validation results for Version 2.3.66_SIP and Version 2.5.9 [20]. In general, the results are comparable. Although there are ups and downs, the resulting highway metrics largely follow the same patterns.

---

[13] Formulae for calculating RMSE and % RMSE can be found in the FHWA manual [4].

[14] TPB staff has recently been testing the adjustment of modeled external trips based on AirSage data. Such an adjustment may be incorporated into a future model version and may improve the validation on those screenlines.

Transit Validation Results

For the transit validation of the Version 2.3.75 model, three validation tests were performed: transit daily ridership by transit sub-mode, Metrorail daily ridership by station group and Metrorail daily volumes on selected cordon lines/screenlines. Table 2 below summarizes key findings and Appendix B provides detailed testing results in tables and graphics.

Table 2. Summary of Year-2014 Transit Validation Metrics for the TPB Modeled Area

Validation Test (Estimated vs. Observed)	Metrics	Accuracy Standard (Acceptable/Preferable)	Benchmarking Results	Reference
Transit Daily Ridership by Sub-Mode	1.04	±2% / ±1%	Marginally acceptable	Table B1
Metrorail	1.01	N/A	N/A	
Commuter Rail	0.60	N/A	N/A	
All Bus	1.10	N/A	N/A	
Metrorail Ridership by Station Group	0.71-1.56	±20% / ±15% (riders>20k) ±25% / ±20% (riders=10k-20k)	17 out of 21 station groups met standard	Table B2, Map B1
Metrorail Screenline Volume				
I-495 Cordon Line	1.18	±20% / ±10%	Acceptable	Table B3,
Potomac River Screenline	0.99	±20% / ±10%	Preferable	Map B2

As shown in Table B1, overall estimated to observed transit ridership is 1.04, which marginally met the standard. Metrorail ridership validates well (1.01). Commuter rail ridership (0.60), on the other hand, is significantly under-estimated[15] and bus ridership (1.10) over-estimated. The validation results clearly indicate a need to adjust the calibration/validation of the mode choice model and to improve transit assignment validation in a cyclic fashion. This is not done in this Version 2.3 re-validation due to its purpose as a quick periodic check of model performance. For the models that are currently under development, however, the validation for transit sub-modes other than Metrorail will need to be improved. Because of the almost perfect agreement between the estimated and observed total Metrorail ridership, the subsequent station-group and screenline validation tests are focused on Metrorail.

As Table B2 indicates, Metrorail ridership by station group validates well, with 17 out of 21 station groups meeting the standards.[16] The results are consistent with those documented in a prior TPB memorandum which examined 2015 observed vs. 2016 simulated Metrorail ridership by station group [21]. Map B1 displays the geographical locations of the station groups. Among the four station groups that didn't meet the standards (#6, #16, #18 and #21), Orange Line – VA Arlington non-Core (1.56) and Silver Line Phase 1 (1.49) are the two being most over-estimated. When Silver Line Phase 2 opens in 2020, however, both the simulated and the observed ridership for the Silver Line Phase 1 group are expected to change significantly.

[15] Although it is extremely important to better validate the ridership for commuter rail, the current under-estimation doesn't have a significant regional impact, due to the relatively small market share of commuter rail in this region (commuter rail served about 4% of all transit trips in 2014).
[16] As a regional planning model, the Version 2.3. model was calibrated and validated for Metrorail station groups but not for individual stations. Thus, the Metrorail ridership validation was not examined at the station level in this re-validation.

Lastly, TPB staff examined the estimated to observed transit volumes by Metrorail on two important regional screenlines, namely, the Capital Beltway (I-495) Cordon Line and the Potomac River Screenline (Table B3). To our knowledge, this summary has not been conducted by TPB staff in past validation efforts. Estimated Metrorail volumes on the selected screenlines were derived from the loaded transit network by summing up the simulated volumes on Metrorail links that intersect with each screenline. Observed Metrorail volumes, on the other hand, were developed on Capital Beltway (I-495) Cordon Line and Potomac River Screenline using the methodologies illustrated in Maps B2-1 and B2-2, respectively.[17] Both screenlines validate well – while the cordon line validation (1.18) met the standard for acceptable accuracy, the river screen (0.99) validates almost perfectly.

Sensitivity Testing

In 2018, TPB developed and evaluated "Visualize 2045", a long-range transportation plan for this region. As part of the performance analysis of Visualize 2045, a sensitivity analysis was conducted based on the Version 2.3.75 model [22]. This analysis, described below, can serve as an independent sensitivity test of the model. TPB staff didn't conduct additional sensitivity testing in this re-validation effort due to its quick turnaround time.

As with past forecasts, the performance analysis of Visualize 2045 primarily compared conditions today with a future, labelled "2045 Build," which includes the transportation projects anticipated under the plan's financially constrained element. For the first time, however, the analysis also looked at a "2045 No-Build" scenario, which represents a future in which no new transportation projects would be constructed, but anticipated population and job growth would still occur. Model runs for both 2045 "No-Build" and "Build" scenarios were executed using the Version 2.3.75 model, and the differences in model outputs were subsequently analyzed. The analysis found that changes to travel demand and system performance measures (such as mobility, accessibility and emissions) going from No-Build to Build were consistent with changes to network inputs, suggesting that the Version 2.3.75 model responded to changes in network inputs in a reasonable way.

CONCLUDING REMARKS

The Year-2014 validation of TPB's Version 2.3 model documented in this memorandum is the third validation TPB staff conducted for the model. Due to its purpose as a periodic check of model performance, this re-validation of the Version 2.3 model was carried out as a temporal validation, with the focus on matching traffic counts and transit ridership on the ground. In addition, this re-validation updated the validation year from 2010 to 2014, thereby fulfilling a federal requirement in

[17] As shown in Map B2-1, Metro stations are colored according to their geographical locations relative to Capital Beltway: those inside the Beltway are display in white while those outside the Beltway are displayed in different colors in cluster. It's assumed that a rider will cross the Beltway cordon line only when the origin and destination stations of his/her trip belong to different color clusters. The rider will cross the cordon line twice if both end stations are located outside the Beltway and will cross once if one of them is located inside the Beltway. For the Potomac River Screenline, it's assumed that a rider will cross the line once when he/she travels between a station on the West side of the River and a station on the East side of the river (Map B2-2). Using the average weekday Metrorail volume (number of riders) information for each pair of origin and destination stations published by WMATA [19], observed Metrorail volumes on Capital Beltway Cordon Line and Potomac River Screenline can then be developed by aggregating the screenline crossings for all O-D pairs.

relation to Air Quality Conformity Determination by 2020 (i.e., the gap between the validation year of a travel model and the AQC analysis year should not exceed ten years).

A survey of the state of the practice identified a number of areas where TPB's routine model validation practice could be improved. This validation effort addressed some of them by including a validation plan, a discussion on sensitivity testing and a comparison of validation metrics against accuracy standards. This validation also introduced several complementary validation tests (e.g., time-of-day estimated to observed daily VMT, R-Squared for estimated to observed daily highway link volumes, daily volumes % RMSE performance by volume group, and transit volumes on screenlines). Other potential improvements, such as validating each model step on the model chain and validating the model in an iterative fashion by feeding validation/application results back to model estimation/calibration, could be considered in a more rigorous, traditional model validation effort usually coupled with a model development process.

This validation carried out a range of validation tests for both highway and transit assignment steps. Benchmarking results indicate that the resulting validation metrics largely aligned with those from TPB's prior validation efforts and met the accuracy standards specified in federal and state modeling guidance. Results from a stand-alone sensitivity test indicate that model outputs responded to changes in network inputs in a consistent and reasonable way. Less satisfactory validation results (such as those from highway screenline validation and transit ridership by mode validation), on the other hand, indicate directions for future model improvement, especially for the Version 2.5 model that is currently under development and for the Gen3 model that will soon be.

In conclusion, the results from this Year-2014 validation of TPB's Version 2.3.75 model suggest that **the performance of TPB's Version 2.3 model remains to be reliable at an acceptable level for the purposes of regional planning.** TPB staff will continue to use the Version 2.3 model for air quality conformity analysis and regional planning studies, until the Version 2.5 model or the Gen3 model is ready for production.

REFERENCES

[1] MWCOG/TPB, "Calibration Report for the TPB Travel Forecasting Model, Version 2.3, on the 3,722 - Zone Area System", January 10, 2012.

[2] Ronald Milone, "2010 Validation of the Version 2.3 Travel Demand Model", MWCOG/TPB Memorandum, June 30, 2013.
https://www.mwcog.org/assets/1/28/2010_Validation_Memo_v3.pdf

[3] Environmental Protection Agency, "§ 93.122 Procedures for determining regional transportation-related emissions", https://www.govinfo.gov/content/pkg/CFR-2012-title40-vol21/pdf/CFR-2012-title40-vol21-sec93-122.pdf, last retrieved on Feb 19, 2019.

[4] Cambridge Systematics, Inc., Travel Model Validation and Reasonability Checking Manual, Second Edition. Washington, D.C.: Travel Model Improvement Program, Federal Highway Administration, September 24, 2010.
https://connect.ncdot.gov/projects/planning/tpb%20training%20presentations/fhwa%20model%20validation%20handbook.pdf, last retrieved on Feb 19, 2019.

[5] Cambridge Systematics, Inc., Vanasse Hangen Brustlin, Inc., Gallop Corporation, Chandra R. Bhat, Shapiro Transportation Consulting, LLC, and Martin/Alexiou/Bryson, PLLC. NCHRP Report 716: Travel Demand Forecasting: Parameters and Techniques. National Cooperative Highway Research Program. Washington, D.C.: Transportation Research Board of the National Academies, 2012.

[6] Virginia Department of Transportation (VDOT), "Travel Demand Modeling Policies and Procedures", June 2014.
http://www.virginiadot.org/projects/resources/vtm/VTM_Policy_Manual.pdf

[7] Cambridge Systematics, Inc. FSUTMS-Cube Framework Phase II: Model Calibration and Validation Standards: Final Report. Tallahassee, Florida: Florida Department of Transportation, Systems Planning Office, October 2, 2008.
http://www.fsutmsonline.net/images/uploads/reports/FR2_FDOT_Model_CalVal_Standards_Final_Report_10.2.08.pdf.

[8] Cambridge Systematics, Inc. Fiscal Year 2011 Task Reports. Final Report. Washington, D.C.: National Capital Region Transportation Planning Board, June 30, 2011.
http://www.mwcog.org/transportation/activities/models/review.asp.

[9] Cambridge Systematics, Inc., and Gallop Corporation. FY 17 Task Orders. Final Report. Metropolitan Washington Council of Governments, National Capital Region Transportation Planning Board, June 30, 2017.
https://www.mwcog.org/file.aspx?&A=YiUe54YhmPVA0q1lahkVpmf4CjB%2fkVfhr3mZDJJ1ACM%3d.

[10] AECOM. Model Validation. Technical Memorandum, Draft. NVTA's TransAction, Transportation Action Plan for Northern Virginia. Northern Virginia Transportation Authority (NVTA), October 3, 2016.
http://www.vdot.virginia.gov/projects/northernvirginia/evaluating_significant_projects.asp.

[11] Florida Department of Transportation and Gannett Fleming, Inc. "Tampa Bay Regional Planning Model v8.0 Validation Report", Technical Report, April 24, 2015,
http://www.tbrta.com/downloads/TBRPM%20v8.0%20Validation%20Report_TR1_20150427.pdf

[12] Southern California Association Of Governments (SCAG), "SCAG Regional Travel Demand Model And 2012 Model Validation", Technical Report, March 2016,
http://www.scag.ca.gov/Documents/SCAG_RTDM_2012ModelValidation.pdf

[13] Cambridge Systematics, Inc. "SERPM 8.0 Model Update: Model Validation Plan", Technical Report, February 15, 2017.
http://www.fsutmsonline.net/images/uploads/md_lctr_2006/SERPM_8_Model_Validation_Plan_02152017.pdf

[14] Metropolitan Transportation Commission (MTC) and Parsons Brinckerhoff, Inc, "Travel Model Development: Calibration and Validation", Draft Report, May 17, 2012.
https://mtcdrive.box.com/s/7crr7bwhromi2au42jnpp11fqe5l24xq

[15] San Diego Association of Governments (SANDAG), "Activity-Based Travel Model Calibration And Validation For Base Year 2012", Technical Report, November 2016.
https://www.sandag.org/uploads/publicationid/publicationid_2097_21613.pdf

[16] New York Metropolitan Transportation Council, Systems Analysis Group, and Parsons Brinckerhoff, “2010 Base Year Update and Validation of the NYMTC New York Best Practice Model (NYBPM)”, Final Report, October 29, 2014. https://www.nymtc.org/LinkClick.aspx?fileticket=8WgNz6e-6dY%3D&portalid=0.

[17] Meseret Seifu, "2014 Daily and Hourly Traffic Counts", MWCOG/TPB Memorandum, May 22, 2017.

[18] Washington Metropolitan Area Transit Authority (WMATA), “Metrorail Average Weekday Passenger Boardings”. https://www.wmata.com/initiatives/plans/upload/2016_historical_rail_ridership.pdf, last retrieved on Feb 19, 2019.

[19] Washington Metropolitan Area Transit Authority (WMATA), “Metrorail Data Download, October 2014”, https://planitmetro.com/2015/01/26/metrorail-data-download-october-2014/, last retrieved on Feb 19, 2019.

[20] Ron Milone, “Ver 2.5 Travel Model Development and Evaluation Status Report”, presented at TPB Travel Forecasting Subcommittee, September 21, 2018.

[21] Ron Milone, Dusan Vuksan, Feng Xie, “WMATA Silver Line Ridership Forecast based on the 2016 Constrained Long Range Plan Estimates”, MWCOG/TPB Memorandum, February 8, 2017.

[22] MWCOG/TPB, “Visualize 2045 Plan Document”, October 17, 2018. https://www.mwcog.org/file.aspx?D=QzUYXpRjQptC2BOTuR1n7q620BSucaa4Bv1bpinuMr8%3d&A=GltaWDYK%2bTpCfsjquDZj4NlLjGX094QB8FR1tET5C9Y%3d

APPENDIX A

Table A1. Total Estimated and Observed * Year-2014 Daily VMT by Jurisdiction

	Observed ("O")	Estimated ("E")	Ratio (E/O)
District of Columbia	7,922,357	8,187,123	1.03
Montgomery County	19,757,260	21,596,642	1.09
Prince George's County	23,646,575	23,113,129	0.98
Arlington County	4,046,638	3,866,042	0.96
City of Alexandria	2,016,133	2,019,850	1.00
Fairfax County	26,663,007	26,631,226	1.00
Loudoun County	6,623,699	7,343,767	1.11
Prince William County	9,425,332	9,521,281	1.01
Frederick County	7,798,767	8,785,986	1.13
Charles County	3,276,575	3,020,140	0.92
TPB Planning Area	*111,176,343*	*114,085,186*	*1.03*
Stafford County	4,006,798	4,501,478	1.12
Calvert County	1,987,808	1,729,059	0.87
Howard County	10,546,027	11,317,730	1.07
Anne Arundel County	15,493,973	15,431,752	1.00
Carrol County	3,290,959	4,097,305	1.25
St. Mary's County	2,246,712	2,176,268	0.97
King George County	871,306	789,154	0.91
City of Fredericksburg	929,927	864,641	0.93
Spotsylvania County ł	3,442,058	2,246,698	0.65
Fauquier County ‡	3,439,861	3,520,312	1.02
Clarke County	810,485	1,114,449	1.38
Jefferson County	1,177,470	1,340,054	1.14
Non-TPB Member Area	*48,243,384*	*49,128,900*	*1.02*
Modeled Area Total:	***159,419,727***	***163,214,086***	***1.02 §***

Notes:

* The observed VMT data is from HPMS.

ł Observed VMT is for the entire Spotsylvania County while Estimated is for northern portion of county only

‡ Fauquier County urbanized area is part of TPB Planning Area. Fauquier is not included as a TPB member in this summary as the HPMS VMT data is only available for the whole county.

§ **FDOT standard** for estimated-over-observed VMT Areawide is ±5% (acceptable) and ±2% (preferable)

Table A2. Year-2014 Link Count Coverage by Facility Type

	Highway Links	Daily Counts		Hourly Counts	
		Links w/ Counts	Coverge (%)	Links w/ Counts	Coverge (%)
Freeway	2,792	517	18.5%	125	4.5%
Major Arterial	6,843	1,867	27.3%	543	7.9%
Minor Arterial	11,529	2,939	25.5%	596	5.2%
Collector	10,498	1,144	10.9%	319	3.0%
Expressway	691	224	32.4%	93	13.5%
Ramp	771	2	0.3%	0	0.0%
Total:	***33,124***	***6,693***	***20.2%***	***1,676***	***5.1%***

Table A3-1. Estimated and Observed 2014 Daily VMT by Facility Type*

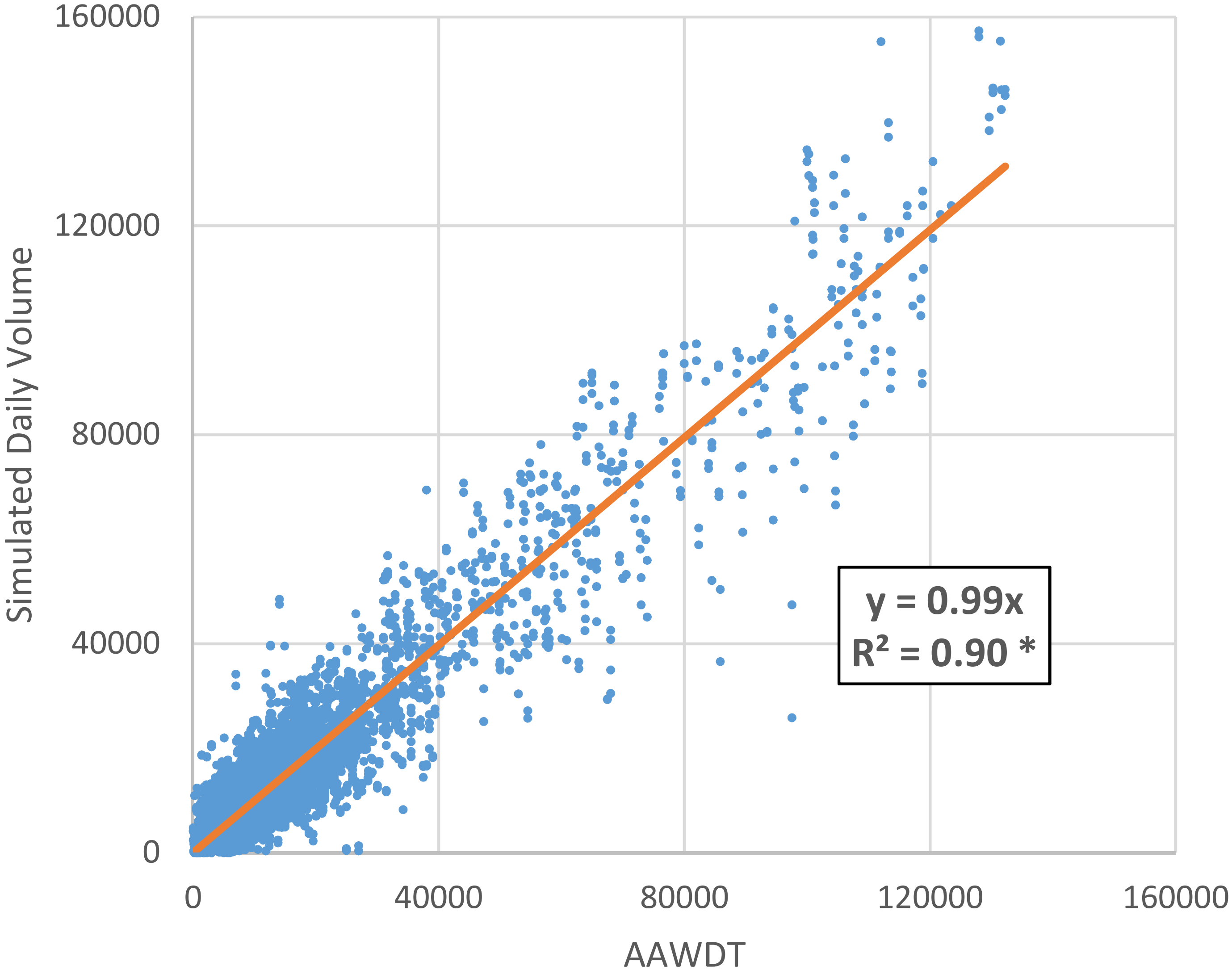
Facility Type	Links w/ Counts	Observed ("O")	Estimated ("E")	Ratio (E/O)	Standard ǂ	
					Acceptable	Preferable
Freeway	517	29,419,832	31,618,131	1.07	±7%	±6%
Major Arterial	1,867	14,795,795	15,845,341	1.07	±15%	±10%
Minor Arterial	2,939	10,897,071	12,343,027	1.13	±15%	±10%
Collector	1,144	2,311,056	1,718,105	0.74	±25%	±20%
Expressway	224	5,063,294	4,826,940	0.95	±15%	±10%
Ramp	2	30,176	26,161	0.87	N/A	N/A
Total:	***6,693***	***62,517,224***	***66,377,704***	***1.06***	***±5%***	***±2%***

Table A3-2. Estimated and Observed 2014 Daily VMT by Area Type (AT)*

Facility Type	Links w/ Counts	Observed ("O")	Estimated ("E")	Ratio (E/O)	Standard ǂ	
					Acceptable	Preferable
AT 1 (CBD)	634	1,543,036	1,585,012	1.03	±25%	±15%
AT 2	1,574	9,668,110	9,185,433	0.95	±25%	±15%
AT 3	1,145	14,648,461	14,113,874	0.96	±25%	±15%
AT 4	965	8,597,479	8,793,571	1.02	±25%	±15%
AT 5	1,066	14,854,152	16,548,738	1.11	±25%	±15%
AT 6 (Exurban)	1,309	13,205,985	16,151,076	1.22	±25%	±15%
Total:	***6,693***	***62,517,224***	***66,377,704***	***1.06***	***N/A***	***N/A***

Notes:
* Based on 6,693 directional links with daily traffic counts
ǂ **FDOT standards** for VMT by facility type, which are also cited in the FHWA and VDOT manuals
ǂ **FDOT standards** for VMT by area type, which are also cited in the FHWA and VDOT manuals

Figure A1. 2014 Simulated vs. Observed Daily Volumes



Note: ***VDOT standard** for R^2 in large model regions is 0.90 or higher.

Table A4. 2014 VMT Estimated to Observed Ratio (E/O) by Time Period and Facility Type*

	Links w/ Counts	AM Peak	Mid-day	PM Peak	Night	*Daily*
Freeway	125	1.12	1.40	0.93	1.12	***1.13***
Major Arterial	543	1.05	1.08	0.87	1.15	***1.02***
Minor Arterial	596	1.33	1.12	1.17	1.37	***1.22***
Collector	319	0.81	0.68	0.71	0.72	***0.72***
Expressway	93	0.91	1.07	0.82	0.98	***0.94***
Total:	***1,676***	***1.09***	***1.20***	***0.92***	***1.12***	***1.07***

Note: * Based on 1,676 directional links with hourly traffic counts (none of them are ramps)

Table A5-1. Daily Directional Volume % RMSE by Facility Type*

Facility Type	Links w/ Counts	% RMSE
Freeway	517	21.9%
Major Arterial	1,867	38.4%
Minor Arterial	2,939	51.5%
Collector	1,144	76.0%
Expressway	224	34.0%
Ramp	2	13.4%
Total:	***6,693***	***42.6% ł***

Table A5-2. Daily Directional Volume % RMSE by Volume Group*

Volume Range	Links w/ Counts	% RMSE	Standard ǂ
Less than 5,000	2,050	110.1%	100%
5,000-9,999	1,699	56.4%	45%
10,000-14,999	1,049	43.8%	35%
15,000-19,999	583	35.2%	30%
20,000-29,999	622	29.4%	27%
30,000-49,999	329	26.4%	25%
50,000-59,999	94	22.2%	20%
Greater than 60,000	267	19.4%	19%
Total:	***6,693***	***42.6% ł***	***40%***

Notes:

* Based on 6,693 directional links with daily traffic counts

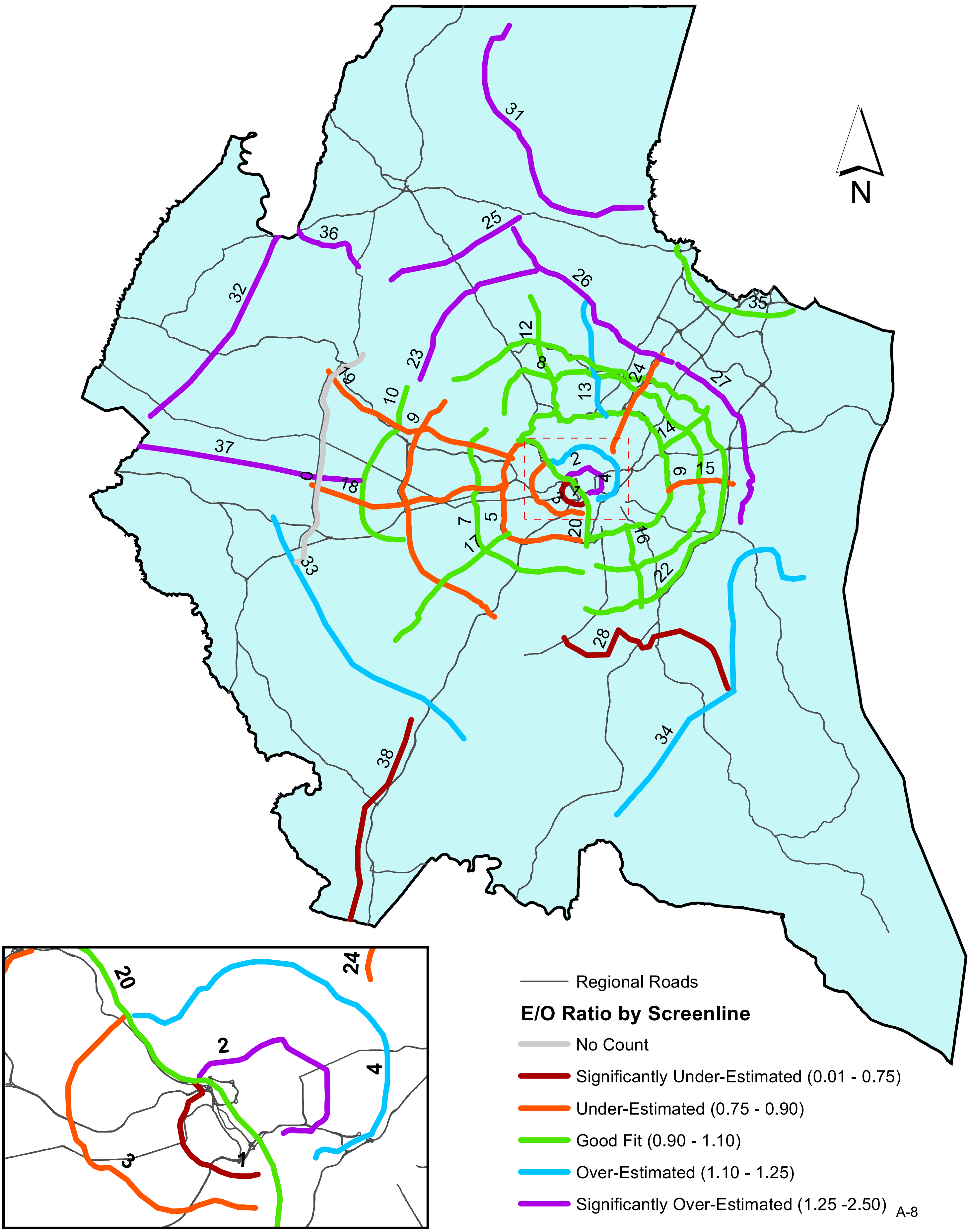
ł **VDOT standard** for percent RMSE areawide; **FDOT areawide standard** is 45% (acceptable) and 35% (preferable)

Table 6. Estimated and Observed 2014 Daily Vehicular Screenline Crossings

Screenline	Observed ("O")	Estimated ("E")	Ratio (E/O)	Standard *
1	189,600	140,495	0.74	±10%
2	363,864	457,183	1.26	±10%
3	242,200	214,927	0.89	±10%
4	562,162	684,763	1.22	±10%
5	454,700	384,217	0.84	±10%
6	1,207,388	1,240,057	1.03	±10%
7	561,400	542,844	0.97	±10%
8	1,053,952	1,158,729	1.10	±10%
9	328,000	253,049	0.77	±10%
10	125,000	118,229	0.95	±10%
12	399,264	399,605	1.00	±10%
13	271,530	329,609	1.21	±10%
14	242,602	258,354	1.06	±10%
15	323,004	290,837	0.90	±10%
16	157,428	147,215	0.94	±10%
17	133,300	121,969	0.91	±10%
18	438,500	386,601	0.88	±10%
19	346,150	278,359	0.80	±10%
20	905,074	837,437	0.93	±10%
22	826,658	869,792	1.05	±10%
23	38,446	61,658	1.60	±20%
24	359,688	323,225	0.90	±10%
25	100,842	132,846	1.32	±10%
26	38,998	81,919	2.10	±20%
27	137,466	204,726	1.49	±10%
28	214,260	161,233	0.75	±10%
31	64,798	143,386	2.21	±10%
32	37,000	65,260	1.76	±20%
33	47,000	52,152	1.11	±20%
34	101,990	119,545	1.17	±10%
35	725,446	677,333	0.93	±10%
36	25,412	53,699	2.11	±20%
37	23,500	47,084	2.00	±20%
38	163,600	113,862	0.70	±10%
Total:	***11,210,222***	***11,352,198***	***1.01***	***N/A***

Note:
* **FDOT standard** for screenline volumes is used (±10% for screenline volumes larger than 50k and ±20% for screenline volumes smaller than 50k). **VDOT standard** is much more stringent.

Map A1. Estimated to Observed (E/O) Ratios of 2014 Screenline Volumes



APPENDIX B

Table B1. 2014 Average Weekday Transit Ridership by Mode in Modeled Area

	Observed ("O")	Estimated ("E") ƚ	Ratio (E/O)
Metrorail	737,679 *	746,541	1.01
Commuter Rail	54,217	32,275	0.60
MARC	36,051	28,200	0.78
VRE	18,166	4,075	0.22
All Bus	648,083	715,273	1.10
Total:	***1,439,979***	***1,494,089***	***1.04* ǂ**

Notes:

* Observed 2014 Metrorail ridership data is extracted from WMATA Crystal ReportsSystem. Since Silver Line opened in July 2014 and its 2014 ridership data is not available, 2015 Silver Line station counts are used instead.

ƚ The estimated ridership data is computed by averaging ("balancing") simulated boardings and alightings in Production/Attraction format from the travel demand model.

ǂ **FDOT standard** for total area transit trips from Mode Choice is±2% (acceptable) and ±1% (preferable).

Map B1. MWCOG / TPB Definition of Metrorail Station Groups

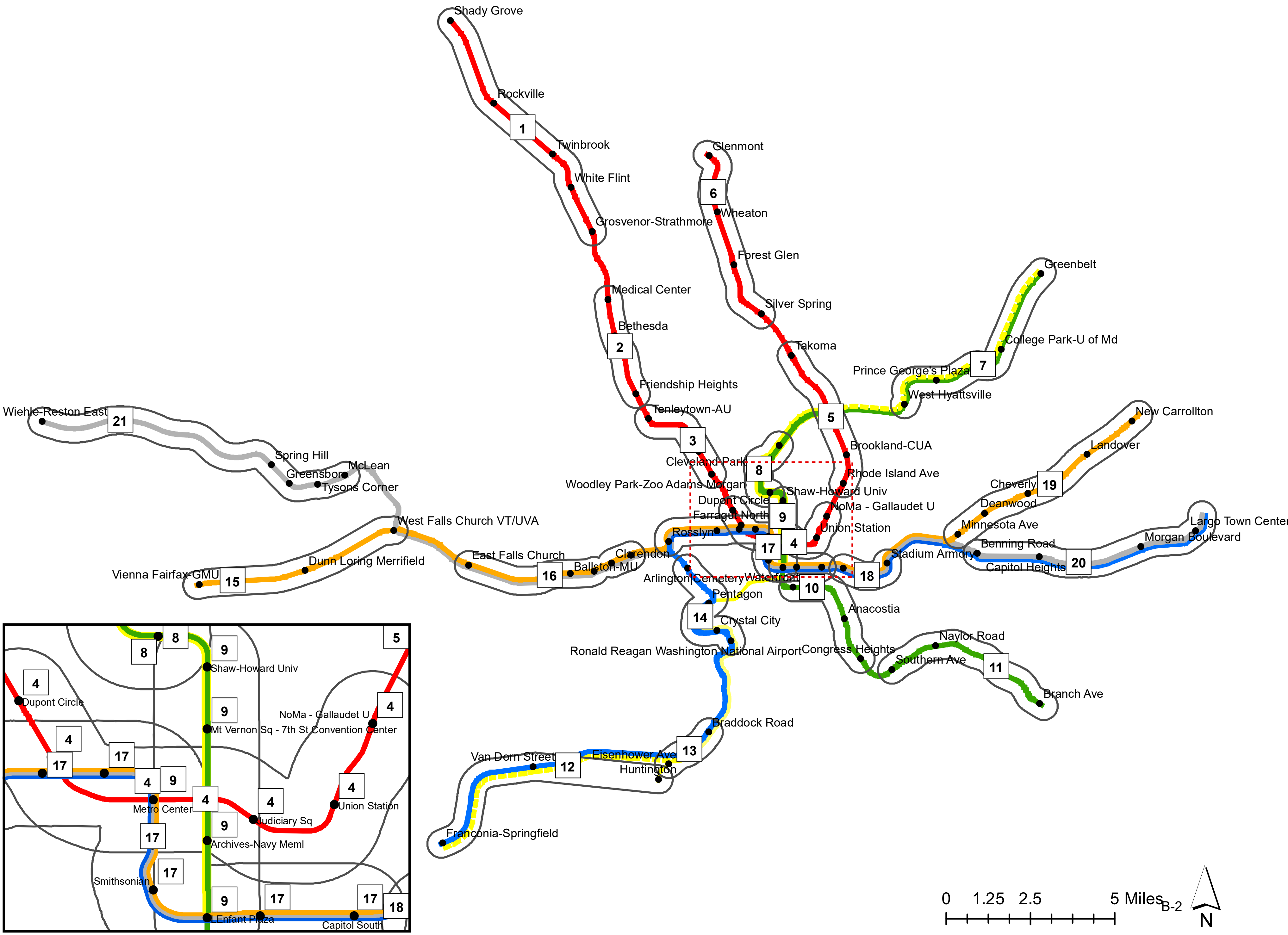
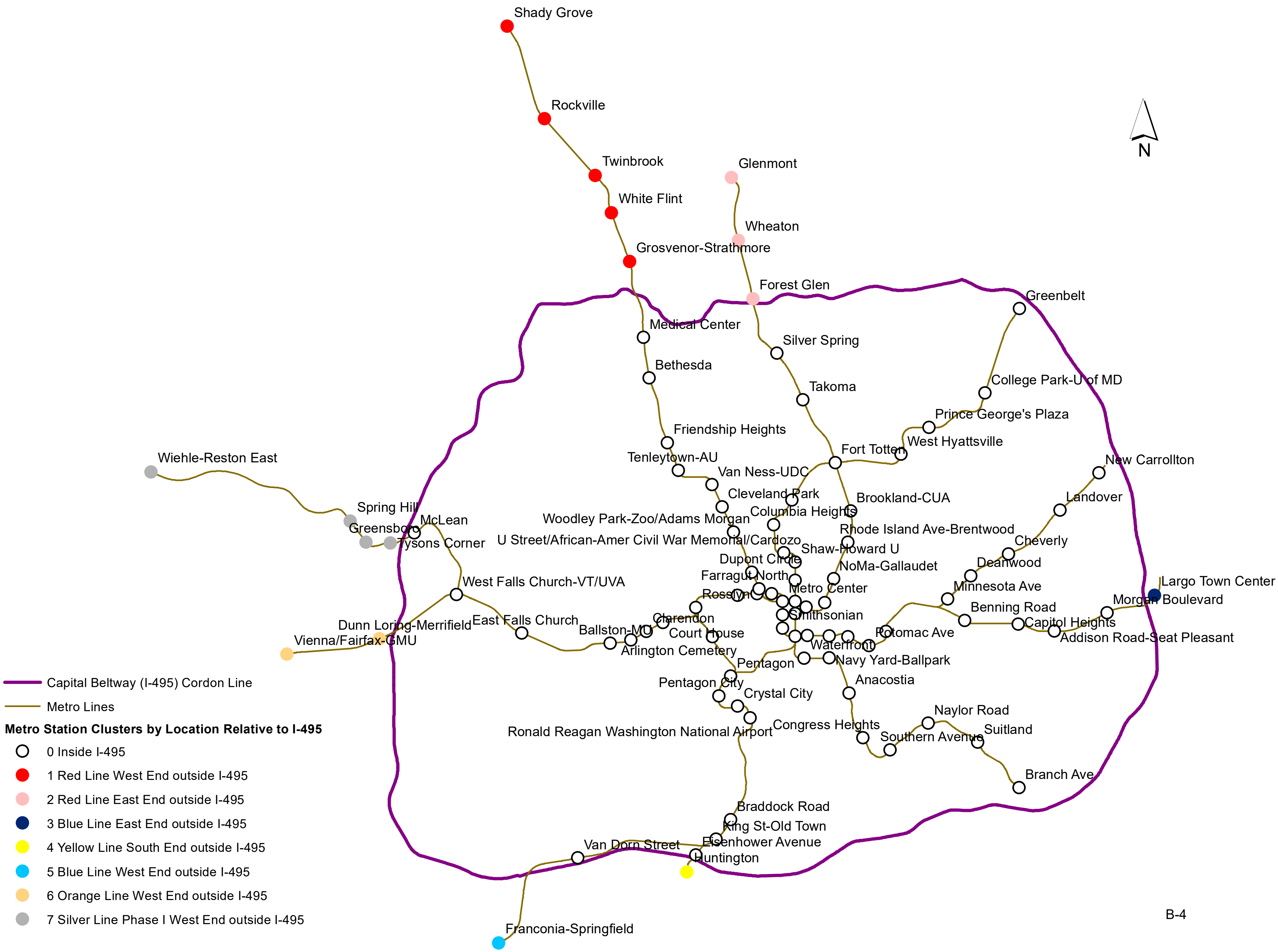


Table B2. 2014 Metrorail Average Weekday Ridership by Station Group: Observed vs. Estimated

	Station Group	Observed ("O")	Estimated ("E")	Ratio (E/O)	Standard *
1	Red Line - "A" route MD outside Beltway	32,231	35,876	1.11	±20%
2	Red Line - "A" route MD inside Beltway	26,483	31,724	1.20	±20%
3	Red Line - "A" route DC non-core	24,995	23,343	0.93	±20%
4	Red Line - DC core	149,787	120,392	0.80	±20%
5	Red Line - "B" route DC non-core	26,532	30,829	1.16	±20%
6	Red Line - "B" route MD	26,134	34,166	1.31	±20%
7	Green Line - "E" route MD	20,273	17,864	0.88	±20%
8	Green Line - "E" route DC non-core	27,131	24,333	0.90	±20%
9	Green Line - DC core	38,906	42,227	1.09	±20%
10	Green Line - "F" route DC non-core	24,526	24,048	0.98	±20%
11	Green Line - "F" route MD	20,518	18,927	0.92	±20%
12	Blue/Yellow Line - VA Fairfax	19,863	21,573	1.09	±25%
13	Blue/Yellow Line - VA Alexandria	15,720	15,330	0.98	±25%
14	Blue/Yellow Line - VA Core	51,911	50,034	0.96	±20%
15	Orange Line - VA Fairfax	28,891	20,551	0.71	±20%
16	Orange Line - VA Arlington non-core	31,877	49,600	1.56	±20%
17	Orange/Blue Line - VA/DC core	109,967	115,363	1.05	±20%
18	Orange/Blue Line - DC non-core	13,117	14,079	1.07	±25%
19	Orange Line - DC/MD	17,347	15,024	0.87	±25%
20	Blue Line - DC/MD	15,595	17,565	1.13	±25%
21	Silver Line (Ph. 1)	15,875	23,698	1.49	±25%
	Total:	***737,679***	***746,541***	***1.01***	***N/A***

Note: * **FDOT standard** for transit ridership >20,000 passengers per day is ±20% (acceptable) and ±15% (preferable), and is ±25% (acceptable) and ±20% (preferable) for 10k-20k passengers per day.

Map B2-1. Transit Volume by Metrorail Validation on Capital Beltway (I-495) Cordon Line



Map B2-2. Transit Volume by Metrorail Validation on Potomac River Screenline

Table B3. 2014 Average Weekday Metrorail Cordon Line / Screenline Volumes: Observed vs. Estimated

	Observed ("O")	Estimated ("E")	Ratio (E/O)	Standard *
I-495 Cordon Line	175,339	207,699	1.18	±20%
Potomac River Screenline	222,484	221,097	0.99	±20%

Note: * **FDOT standard** for transit screenline is ±20% (acceptable) and ±10% (preferable) .