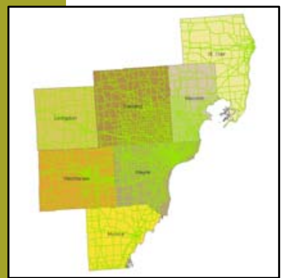
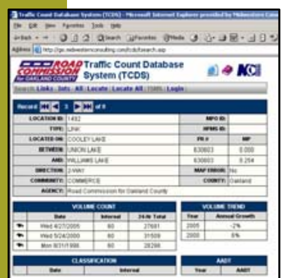
SEMCOG

Southeast Michigan Council of Governments





TRAFFIC COUNT FILE DEVELOPMENT FOR TRANSCAD MODEL CALIBRATION AND VALIDATION

Technical Report

Prepared by:

MIDWESTERN CONSULTING, LLC
3815 Plaza Drive
Ann Arbor, MI 48108
http://GIS.MidwesternConsulting.com

PARSONS
26777 Central Park Boulevard
Southfield, MI 48076
http://www.Parsons.com

October 2005

Table of Contents

Executive Summary

Part 1. Data Review and Model Calibration/Validation State-of-Practice Assessment

Part 2. Data Specification for Travel Demand Model Integration

Part 3. Validation File Development

Tables

Figures

Appendices

Executive Summary

Southeast Michigan Council of Governments (SEMCOG) has built a travel demand forecast model using TransCAD. The model calibration and validation require traffic count data input from numerous sources including the Michigan Department of Transportation (MDOT), county road commissions, and local communities. While there is a wealth of data, its wide variety and mixed formats pose a great challenge to create a dataset for periodic model calibration and validation. Another major issue is the reconciliation of the link definition difference between the TransCAD model network and the current traffic count locations. Therefore, an efficient process is needed to automatically retrieve, adjust, and format count data from available sources to meet the requirements of model calibration and validation. A statistically sound spatial analysis process is also necessary to select and associate traffic count data to corresponding TransCAD network links. Furthermore, it is necessary to identify the exact needs of additional traffic counts for future model validation and calibration.

To accomplish these objectives, this project includes three main parts:

1. Data Review and Model Calibration/Validation State-of-the-Practice Assessment
2. Data Specifications for Travel Demand Model Integration
3. Validation File Development

The key findings and recommendations from the project are as follows:

Traffic Count Data

SEMCOG has been a model of best practice in regional traffic count data management. All major road agencies in the region, except MDOT, currently use the Traffic Count Database System (TCDS) software. The standardization allows easy transfer of traffic count data into the SEMCOG's Regional Traffic Count Database (RTCD), which is the region's central data repository. The database provides a wealth of data that can be used for model calibration and validation. The RTCD is also linked to a GIS map (Michigan Geographic Framework) for easy display of count locations and count data. This linkage can be used as a basis for matching to the TransCAD-based travel model developed for SEMCOG, because its model highway network is derived from the same geographic map.

Building upon its success in the regional traffic count data management, SEMCOG has a great potential in moving up to the next level with the focus on the following areas:

- Data Detail: Significant amount of the counts in the region have been collected using a 1-hour time interval. The interval should be reduced to 15-minute, as it will provide sufficient details for most planning and engineering applications. The existing counts also lack of vehicle classifications and speed data. The modern "raw data" counting technology should be considered for acquiring all data types in one setup.
- Data Quality: Even though there are some QA/QC routines built in the TCDS and RTCD, additional procedures need to be implemented to further improve the data quality.

- Other Data Sources: There are vast amount of traffic count data available that can be gathered to supplement the regular counting programs. The data are being collected in various planning and engineering projects, such as traffic signal timing optimizations.

SEMCOG should consider setting up a regional data portal that allows more agencies and consultants to upload the data to the central data depository.

Ideals for Validation Dataset

A number of ideals should be considered for validation datasets created from the RTCD:

- Coverage: The validation dataset should provide adequate coverage in terms of number of links for the primary validation strata for network links. These strata include geographic strata such as county; area type; and functional class.
- Randomness: Randomness in link selection for the validation dataset would reduce the likelihood of inappropriate model adjustments.
- Screenlines: Screenline validation datasets should be separate from the "general" validation dataset. By nature, screenline validation data violate the randomness ideal.
- Consistency with the modeling network: Since the RTCD does not necessarily use the same link definition as the modeling network, care must be exercised in link selection to ensure that the counted location reflects the same location in the modeling network.
- Time-of-day counts: To the maximum extent possible, the validation dataset should include time-of-day counts in addition to the daily counts. The RTCD maintains counts on links for periods as small as 15 minutes and the model might use up to 10 time periods for assignment. The short time periods used for traffic assignment should not be considered for model validation, however, due to the normal variation in counts. At the most, validation to three or four time periods (which are aggregations of the periods used for traffic assignment) should be performed.
- Classification counts: To the maximum extent possible, the validation dataset should include vehicle classification counts in addition to the vehicle counts.
- Consistency with forecasting period: Links selected for the validation dataset should have counts that were collected on a weekday, not a weekend day. While factors can be developed to adjust weekend day counts to Average Annual Daily Traffic (AADT) or Average Weekday Daily Traffic (AWDT), weekend travel patterns and characteristics are so different from those on weekdays to make the selection of links with weekend counts a questionable practice for validation datasets. This ideal becomes even more important if time-of-day or classification counts are included on the validation dataset.

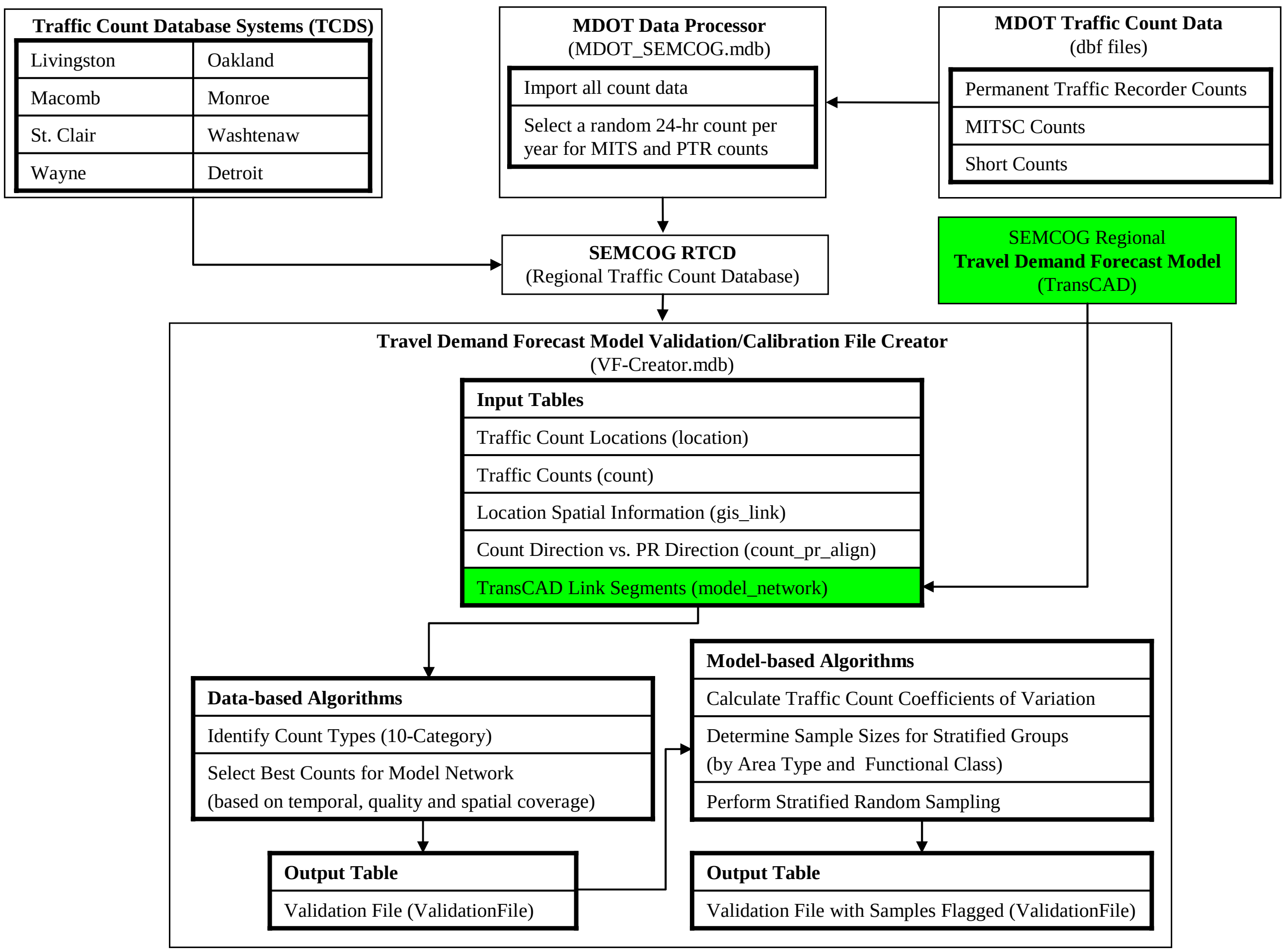
Sampling Methods

The cost of collecting traffic count data can be substantial. Statistical sampling methods can be used to create a validation dataset that uses as few data points as possible. There are two desirable sampling plans:

- Data-based Sampling Plan: The plan would only sample the model links with existing count data. It will create a feasible dataset based on the available traffic count data in the RTCD. This plan can be used for past year validation. While this plan would not require additional traffic counts, it may produce a biased dataset as the region is noticeably lacking count data for certain geographic areas and functional classes. This plan is comparable to procedures traditionally used throughout the country.
- Model-based Sampling Plan: Based on the required confidence level and precision rate, the plan would sample links from the whole universe of the TransCAD model network. It is the preferred procedure for developing a validation file. However, the model-based sampling plan can identify TransCAD links without associated counts in the RTCD and, for this reason, is primarily suited for the development of validation datasets for anticipated future year validations. Once the sample has been specified, appropriate actions should be taken by SEMCOG to ensure that the selected links are included in the regional counting program.

Validation File Development Process

As a final product of this project, a suite of MS Access database programs has been developed to automatically process traffic count data from MDOT and RTCD and generates traffic count input file for TransCAD. The system is based on the data specifications, business rules, and sampling methodology developed throughout the project. The system flow is illustrated in the figure below.



For further information about the project, please contact the following:

Tom Bruff
SEMCOG
535 Griswold Street, Suite 300
Detroit, MI 48226
Tel: (313) 324-3340
Fax: (313) 961-4869
Email: bruff@semcog.org
http://www.semcog.org

Ben Chen, PE, PTOE
Midwestern Consulting, LLC
3815 Plaza Drive
Ann Arbor, MI 48108
Tel: (734) 995-0200
Fax: (734) 995-0599
Email: bc@midwesternconsulting.com
http://GIS.MidwesternConsulting.com

Part 1. Data Review and Model Calibration/Validation State-of-the-Practice Assessment

INTRODUCTION

The purposes of Part 1, Data Review and Model Calibration / Validation State-of-the-Practice Assessment are:

- Review all available candidate data sources for travel model validation including the data sources, databases, and data collection methods.
- Assess current validation guidelines used by SEMCOG against state-of-the-practice procedures used in other regions.
- Provide recommendations regarding the quantity and quality of data available for model calibration, the data needs and issues that will likely be facing SEMCOG, and agency resources for future data acquisition.

It must be stressed that Part 1 is intended to “foster mutual understanding of SEMCOG’s needs (including a prioritization of those needs given the data available) and promote a clear vision of the data specification to be prepared under Part 2.”[1] As such, this Part has been accomplished based on the kick-off meeting and a brief review of the data and state-of-the-art practices. This report provides an overview intended to guide Part 2, Data Specification for Travel Demand Model Integration.

CANDIDATE DATA SOURCE REVIEW

DATA SOURCES / DATABASES

The preliminary review of the data sources and databases was focused on the types of validation data available, the consistency of the available data, and adequacy of the data. The review did not specifically focus on the data quality or accessibility through actual access and statistical review of the data, nor was the actual coverage (i.e., mapping of the data) reviewed. These items will be evaluated more fully in Part 2.

Information regarding the following data sources was reviewed:

- Regional Traffic Count Database (RTCD)
- Highway Performance Monitoring System Database (HPMS)
- MDOT Traffic Counts (short counts, permanent traffic recorder counts, and MITSC counts)
- Border Crossing Data
- SCATS Data

REGIONAL TRAFFIC COUNT DATABASE

The RTCD is intended to be SEMCOG’s central repository for all traffic count data. This Oracle database is a work in progress, but should provide a wealth of data that can be used for model calibration and validation. The RTCD is linked to a GIS map (Michigan Geographic Framework)

[1] Traffic Count File Development for TransCAD Model Calibration and Validation, Project Work Plan, Midwestern Consulting, November 2003.

for easy display of count locations and count data. This linkage should also provide for easy matching to the new TransCAD-based travel models developed for SEMCOG, because its model highway network is derived from the same geographic map.

A preliminary Access version of the dataset containing count data through 2003 for Oakland, Livingston, Macomb, Monroe, St. Clair, Washtenaw, and Wayne Counties was reviewed. Table 1 summarizes the number of records and unique links included in the preliminary database. The preliminary database contains an impressive number of links and counts and will simply continue to grow as 2003 data are added (anticipated in early 2004). Counts from the City of Detroit and the Michigan Department of Transportation (MDOT) will also be added in the future.

Table 1. Preliminary RTCD Database Summary

County	Count Records	Unique Links	Average Counts / Link	First Count Date	Last Count Date
Livingston	5,824	2,815	2.1	3/18/1982	11/21/2002
Macomb	5,607	2,292	2.4	8/9/1993	8/1/2003
Monroe	1,917	892	2.1	2/1/1978	3/18/2003
Oakland[1]	4,099	1,803	2.3	7/1/1996	12/8/2003
St. Clair	911	744	1.2	6/19/2000	8/15/2002
Washtenaw	2,043	471	4.3	1986	2002
Wayne	1,756	1,245	1.4	7/13/1999	12/12/2002
Total	22,157	10,262	2.2		

[1] An additional 1,249 counts at intersections (986 unique links) are available in the count database.

The primary source of data included in the RTCD is the Traffic Count Database System (TCDS) developed by Midwestern Consulting. Based on our working knowledge with the TCDS users in the SEMCOG region, the preliminary assessment of the availability of the traffic count data is as follows:

Table 2. TCDS Data Availability Summary

County	Location Definition	Daily Counts	Hourly Counts	Remarks
Livingston	✓	✓	✓	
Macomb	✓	✓	✓	
Monroe	✓	✓	Historical data import required	Historical hourly count files are available
Oakland	✓	✓	✓	
St. Clair	✓	✓	Historical data import required	Historical hourly count files are available
Washtenaw	✓	✓	Historical data import required	Historical hourly count files (in Peek and Jamar format) are available
Wayne	✓	✓	Data cleanup required	
City of Detroit	✓	✓	Historical data import required	Historical hourly count files (in Peek format) are available

While the simple summary performed for Table 1 shows an impressive amount of data, it also suggests at least two issues with the existing count data:

- The earliest year for which counts can be summarized for the entire region is 2000 unless St. Clair County is excluded. Even when St. Clair County is excluded, the earliest year is 1999 based on the earliest date for Wayne County count data. Due to its size, central location, and importance to the region, exclusion of Wayne County from a validation dataset would be unreasonable[2].
- All dates for counts in Washtenaw County were listed as January 1 of the year of the count. As a result, factoring of specific counts to average annual daily traffic (AADT) or average annual weekday traffic (AWDT) for Washtenaw County will be impossible.

The above two issues highlight an issue raised in the kick-off meeting for the project, specifically, quality assurance / quality control (QA/QC) of the data entered into the RTCD. Since the count data are provided by the seven counties in the region, the City of Detroit, and MDOT, there can be variations in the data collection methods, data recorded, or naming conventions. Some standardization has been achieved through the use of TCDS by the seven counties and City of Detroit. The QA/QC has been further enhanced by SEMCOG efforts to run error checks and standardizations of location descriptions.

The primary focus of this project is the development of methods to create count data sets for the validation of the regional travel model, not the development of QA/QC procedures for inputting and maintaining the RTCD. Nevertheless, due to the impact on the model validation, a consistent set of QA/QC criteria and standards will be developed as part of Part 2. The QA/QC procedures developed in Part 2 will use current SEMCOG and MDOT procedures along with FHWA's *Traffic Monitoring Guide* as bases.

An important QA/QC criterion will be "coverage." Three types of coverage are of primary concern:

- Temporal coverage
- Facility type coverage
- Geographic coverage

As noted above, there are temporal coverage issues with the preliminary data from the standpoint that the earliest date that a full validation dataset could be extracted from the RTCD is 2000. This is probably a minor issue from the standpoint that the current TransCAD version of the model has been calibrated / validated to 2000. However, as will be discussed more fully in a later section, "stronger" validations can be achieved when a model is validated to multiple years (e.g., 1997 and 2000) especially if there have been significant changes to the transportation system or regional demographics over the time period. To this end, general procedures will be developed in Part 2 so that SEMCOG can extract validation datasets for various years. This will allow, for example, the creation of 2000 and 2005 validation datasets in late 2005 or early 2006 after the new household surveys have been administered and revised models have been calibrated.

Since the preliminary dataset was known to be incomplete, detailed checking of facility type and geographic coverage issues was not performed. Concerns were raised in the kick-off meeting regarding the lack of counts on freeways. Based on comments from MDOT staff, this difficulty

[2] SEMCOG could import historical counts that are part of sample locations for the validation process identified in Part 2. Even though such an effort may be laborious, it should cost less than $25,000.

cannot be addressed until the MDOT traffic count data are added to the RTCD. When the data are entered, the type of count (raw or AADT) must be clearly identified. Both raw counts and AADT can be stored in the RTCD. Even if both types of counts are stored, a decision regarding the preferred method for providing and storing counts must be made. This issue is discussed more fully in a subsequent section, *Recommendation for Validation File Seasonal Adjustment Factors*.

For Part 2, a summary of the number of unique counts by facility type and area type (and/or county) will be prepared and reviewed. In addition, a GIS map of count locations will be prepared for visual checking of the distribution of counts.

HPMS DATABASE

The HPMS database includes AADT volumes prepared from counts provided by the counties and the City of Detroit along with traffic count data from MDOT. Since the county data are included in the RTCD and MDOT data will be added to the RTCD, the HPMS database is superfluous for the development of a validation dataset.

While the HPMS database is superfluous for the development of a validation dataset, it cannot be entirely ignored in the model validation process since it is the basis for an estimate of the vehicle miles of travel (VMT) for the region that will be used as a standard for measuring the overall reasonability of the travel model. However, the regional travel model does not attempt to estimate traffic for all roadways in the region (most local roads are excluded). This suggests that the validated regional travel model should produce a regional VMT estimate that is below the HPMS estimate.

In the kick-off meeting, questions regarding the veracity of the data on non-sampled segments of the network were raised. Some questions were raised about the stability of AADT adjustment factors used to develop the HPMS database; 1997-1999 AADT adjustment factors had been noted to vary substantially from 2000-2002 factors, especially for rural areas. Finally, it was noted that MDOT AADT adjustment factors have been based on statewide count data while SEMCOG has developed region specific AADT adjustment factors.

The HPMS database will not be used for the development of the travel model validation file even though the estimate of regional VMT based on the HPMS will be used as a general guide for modeled VMT in the region. Since the link counts collected for the HPMS are included in the RTCD, some or all of the counts might be used for the validation file, but they will be supplemented with counts from other sources. Detailed, VMT-based validation statistics will be based only upon links with traffic counts.

MDOT TRAFFIC COUNTS

MDOT collects traffic counts and vehicle classification counts through the use of short counts (greater than 24 hours), machine derived classified traffic counts, permanent traffic recorder (PTR) counts, and counts from imbedded loops and detectors from the MITSC system. SEMCOG has requested and received five years' worth of raw MDOT traffic count data and will incorporate the data into the RTCD once an efficient procedure can be developed. MDOT will continue to provide traffic count data to SEMCOG on an annual basis.

In order to incorporate MDOT's data into the RTCD and then match to model highway network, all traffic count locations need to be referenced with Framework's Physical Reference Number

(PR) and mile point (MP). SEMCOG needs to coordinate with MDOT to determine whether the geographic reference is readily available.

In some cases, the MDOT' traffic count data can be extensive:

- Examples of the PTR count data included hourly counts for each day of the year
- MITSC counts can be provided by lane by minute

Since the MDOT data will be included in the RTCD, it should not form an independent source for the development of the model validation data file.

BORDER CROSSING AND SCATS DATA

SEMCOG has information regarding the structure of border crossing data and data from the Sydney Coordinated Adaptive Traffic System (SCATS) used in a portion of Oakland County. However, SEMCOG is not planning to transfer those data into the RTCD in the near future. The data structures have not been provided for review. The border crossing data do not constitute independent validation data. External stations are established at or near border crossings. Since external station trips ends are set to match observed traffic volumes at external stations, the inclusion of border crossing data in the validation data file would tend to cause the validation statistics to be overstated.

SCATS data, like MITSC data, are very detailed. The SCATS data could provide substantial additional traffic count data and data useful for the development of "intersection-based" traffic assignment procedures. However, the amount of data available along with potential difficulties associated with processing the data obviates the value of the data for a validation data file. The SCATS data should be maintained for special analyses or to supplement validation data files when key information such as a count on a screenline is missing, provided efficient methods for processing the data can be developed. This need might become more important if the outcome of the sampling plan developed in Part 2 shows that certain SCATS data could enhance the sampling dataset. SEMCOG needs to solicit local agency support to archive the SCATS data and develop automatic data transfer and import processes.

SAMPLING ISSUES

The RTCD is extensive in its coverage and the number of links with counts included in the database. Figure 1 shows a plot of links with traffic counts based on the preliminary Access dataset provided by SEMCOG. As can be seen in the figure, there seems to be extensive coverage of some counties while counts in other areas (e.g., Washtenaw County and City of Detroit) are sparse. In addition, Figure 1 suggests that there might be some county ID coding errors in the RTCD. For example, a number of links have no county ID coded and some links in counties appear to have county IDs from a neighboring county.

Figure 2 shows a plot of the links included in the RTCD superimposed on the TransCAD highway network. In effect, Figure 2 plots the RTCD links as GIS "layer" on top of a GIS layer plotting the TransCAD links. No merging of the RTCD link data with the TransCAD network was performed. Thus, some of the links in the RTCD might not be included in the TransCAD network and, as a result, might be superfluous, at least as they relate to providing counts for a TransCAD validation database.

Figure 1. Links with Counts by County

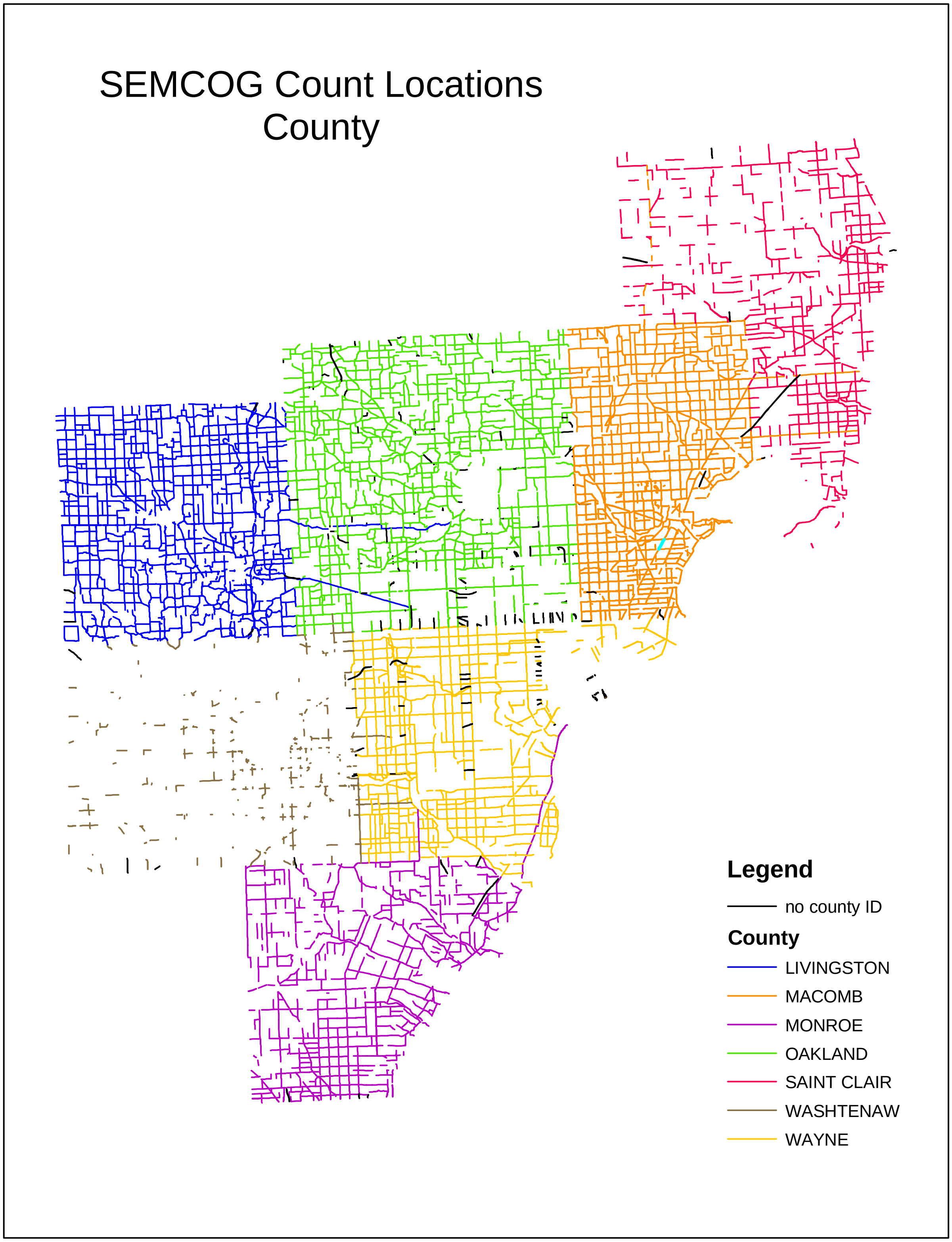


Figure 2. Links with Counts Superimposed on TransCAD Highway Network

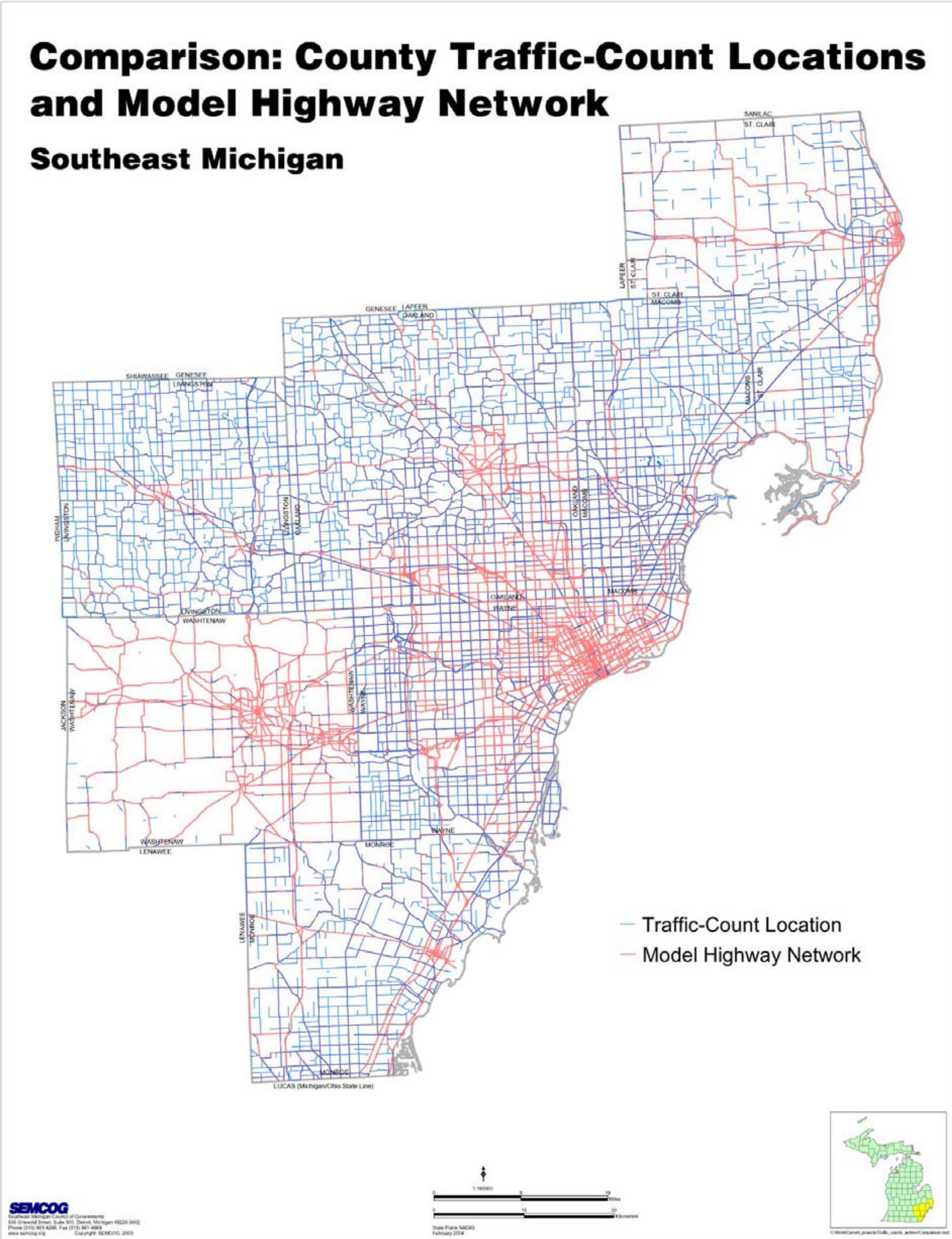


Figure 2 shows that RTCD count coverage is substantial for much of the region. Nevertheless, there are noticeable areas where traffic counts are lacking, particularly in the City of Detroit, Washtenaw County and, to lesser extents, in the Cities of Pontiac, Monroe, and Port Huron. The lack of counts on freeway facilities is also notable. A portion of Part 2 will be devoted to the review of the types of facilities with counts throughout the region with an emphasis on identifying types and locations of facilities where additional counts might be required. This effort relates to the statistical reliability of the validation dataset.

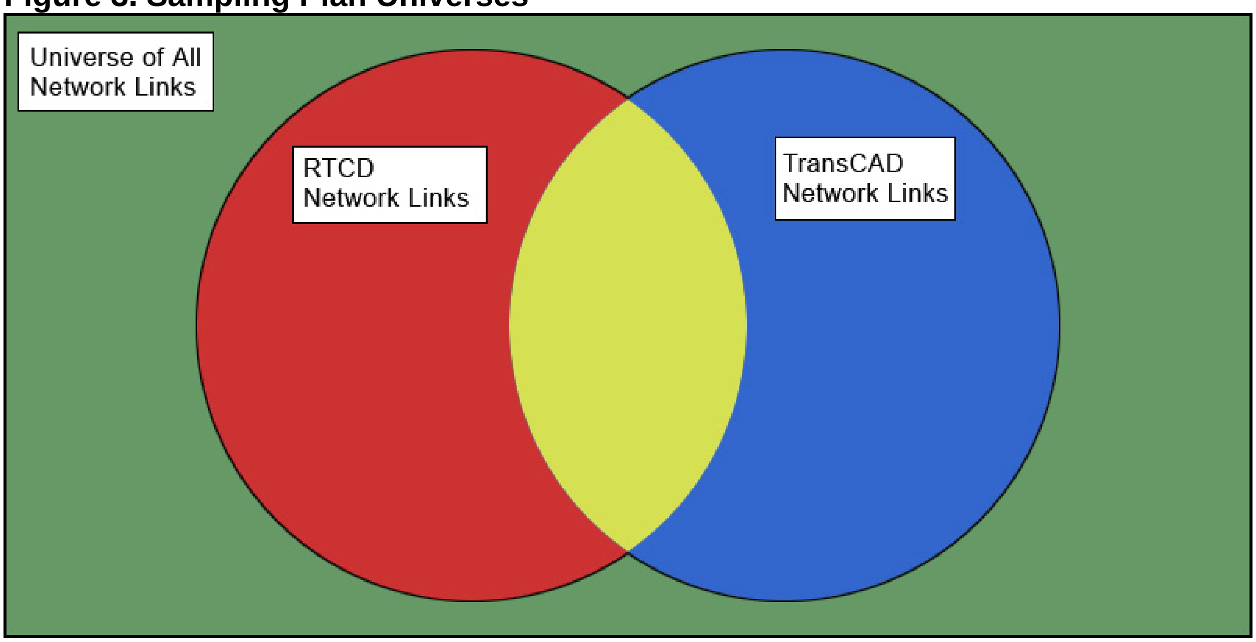
A second issue might occur in areas with substantial numbers of links with counts. Specifically, it might not be necessary to use all links with counts for a validation dataset. Thus, it might be worthwhile to develop a sampling plan for a subset of counts existing in the RTCD that are corresponding to the model highway network for the development of the validation dataset.

There are five potential sampling plans:

a) Sampling all roadway links in the region to produce an “ideal” validation dataset for a “perfect” travel demand model.
b) Sampling roadway links in the existing RTCD to produce the “most cost-effective” validation dataset.
c) Sampling roadway links in the existing TransCAD network to produce the “best model-based” validation dataset.
d) Sampling roadway links contained in either the existing RTCD or the TransCAD datasets.
e) Sampling roadway links contained in both the existing RTCD and the TransCAD to produce the “best data-based” validation dataset.

The concepts summarized above are shown in Venn Diagram format in Figure 3. The circles shown in the Venn Diagram represent the various universes from which model validation datasets can be selected.

Figure 3. Sampling Plan Universes



Sampling plan "a" is the selection of a sample from the universe of all network links in the region (represented by the green-shaded rectangle in Figure 3). The HPMS data might be considered an example of the sampling plan "a." This sample might be considered the most ideal sample of links for model validation since it, in theory, includes all roadway network links in the region. The problem with this sampling procedure is that not all links selected would necessarily be included in the existing RTCD nor would they necessarily be included in the TransCAD network. Such a sampling process could be used to identify locations for future traffic counts and, possibly, network additions for modeling purposes. The regional travel model should, by design, include all "important" network links from a traffic carrying / mobility standpoint. Thus, the primary reason for including additional links in the TransCAD network based on this sampling plan would be solely for less biased estimates of regional VMT.

Sampling plan "b," represented by the red circle, is the most cost-effective from the standpoint that additional count locations do not have to be specified. However, as shown in Figures 1 and 2, the procedure is flawed since there are areas and facilities in the region where counts are noticeably lacking. Thus, sampling from just RTCD links is likely to produce a biased validation dataset. In addition, this sampling plan might include counts on links not included in the regional model. As stated above, the regional travel model should, by design, include all "important" network links from a traffic carrying / mobility standpoint. Thus, the identification of additional links to include in the TransCAD network based solely on the fact that they have traffic counts would be unproductive.

Sampling plan "c," represented by the blue circle, would produce the most cost-effective validation dataset from a modeling validation standpoint. Some additional locations for traffic counts to be included in the RTCD might be identified.

Sampling plan "d," represented by the union of the red and blue circles, probably is closest to sampling plan "a" for producing an ideal validation dataset. The inclusion of all links in the RTCD would probably require the addition of some links in the TransCAD network that would, at least theoretically, provide for better estimates of regional VMT. The inclusion of all links in the TransCAD network should help fill in the areas and facilities where the RTCD is lacking counts. Nevertheless, there are some dangers and inefficiencies with this sampling scheme. First, since not all links are included in the RTCD and links included in the RTCD were not necessarily selected at random, there is the danger of introducing sampling biases into the validation dataset. Second, the regional travel model is not really designed to reproduce travel on very minor facilities in the region. Travel forecasts for minor facilities such as local streets are typically looked upon with little confidence. Thus, requiring their addition to the TransCAD network for travel model validation would be unproductive.

Sampling plan "e," represented by the intersection of the circles, the area shaded in yellow, is the most effective sampling plan in terms of producing links with counts that can be used for validation since the selected links would be known to have counts and would be known to be in the TransCAD network. However, the plan has the same flaws associated with sampling plan "b" since it is has been shown that the RTCD is lacking in counts for certain areas and facilities.

The primary goal of this project is to develop a traffic count dataset to validate and calibrate a given TransCAD model. Therefore, based on the characteristics of each sampling plan, Part 2 should focus on the following among the five possible sampling plans:

1. Model-Based Sampling Plan (Plan "c") – It will identify the best sampling dataset and provides a basis for count data deficiency analysis.

2. Data-Based Sampling Plan (Plan "e") – It will create a feasible dataset based on the available traffic count data.

SEASONAL ADJUSTMENT FACTORS

SEMCOG has developed seasonal adjustment factors for use in the RTCD. The seasonal adjustment factors provide a mechanism to standardize reporting of counts collected on different days throughout the year or even in different years. In Part 1, a quick review of the procedures used for the development of the factors has been made; a more thorough review of the data used to develop the factors and actual factors will be made in Part 2. While this section offers some critiques of the factors, it should not distract from the acclamation that SEMCOG should receive for developing the factors. Many regions collect traffic counts from any source possible, often assuming that the counts have been reasonably factored to AADT. The RTCD should have a level of consistency provided from the development and application of locally developed factors that should be among the best in the country. In addition, the factors provide a mechanism to summarize traffic counts for the validation database that are more consistent with the modeling process, specifically AWDT.

COMMENTS ON CURRENT PROCEDURES

It was somewhat discouraging to find that the seasonal adjustment factors were derived from only 29 "good" PTR stations in the SEMCOG area and that even if data from all PTR stations in the SEMCOG area were considered to be "good," only 41 total stations were available.

The lack of stations limited the ability of SEMCOG to stratify the adjustment factors by groups that could provide better estimates of traffic for model validation. Several different grouping schemes were investigated:

- MDOT's grouping procedure based on methods recommended in FHWA's *Traffic Monitoring Guide*. The grouping procedures were based on statistical analyses to identify patterns and groupings. Note that the geographic placement of the counter was not a factor in determining the grouping patterns.
- A grouping scheme based on an analysis of the coefficient of variation of the count data that was used to identify three primary groups: inner, mid, and outer region groups.
- Single group or, in other words, no grouping.

Due to some difficulty with developing factors based on the three group scheme, SEMCOG chose the single group scheme for the seasonal adjustment factors currently used in the RTCD. However, a future re-evaluation of this decision was recommended. It would be quite informative to stratify the groups by facility type as well as the three regions.

Growth factors to estimate volumes on facilities for years when no counts were taken were also developed by SEMCOG. A three year limit is typically used for growth factoring. SEMCOG made the decision to use a one percent per year growth factor for all counts and count locations currently. SEMCOG recommended that this procedure be re-evaluated "as soon as the new model network has been developed."

When the growth factors are revised, they should be stratified. Options include stratification by area type and facility type. If areas can be easily classified, rate of growth (of traffic) areas such as high growth, medium growth, low or no growth, and negative growth could be substituted for area type.

Storage of Raw Counts versus AADT Counts

Counts for most links in the region are available as raw counts and include information on the date the count was taken. Region specific seasonal adjustment factors can be developed for the links with raw counts and dates and the adjustment factors can be tailored to estimate AADT or AWDT values. In general, the storage of raw counts in the RTCD should be preferred to provide for flexibility in factoring data to produce consistent "average daily" counts as required for different analyses. As stated in the following section, it might be desirable to use AWDT weighted toward the spring or fall for model validation. However, other analyses might require AWDT (based on an annual average not weighted to spring or fall) or AADT.

Only AADT is available for a number of links in the region including freeway counts from MDOT. If possible, raw counts should be obtained for all new counts entered into the RTCD for those facilities that currently have only AADT recorded. For historical counts for the subject links, it might be possible to estimate annual and spring / fall AWDT from the AADT. Care will need to be used to ensure that types of counts are not simply mixed in the validation dataset. In other words, links with only AADT recorded (and there are no estimates of commensurate AWDT) can be included in a validation dataset only when AADT is used for all other links.

Recommendation for Validation File Seasonal Adjustment Factors

The day-of-week factor analysis performed by SEMCOG using the PTR data produced an interesting result. Specifically, Monday through Thursday adjustment factors were very similar to each other but the Monday through Thursday adjustment factors were quite different from Friday adjustment factors. Separate analyses of MITS loop count data for freeways suggests that Tuesday through Friday adjustment factors are more similar to each other than they are to Monday factors[3]. While contradictory, these findings underscore the need to develop a procedure that should lead to better estimates of "observed" counts (i.e., better than AADT) for the validation file.

A quick review of the model calibration process is helpful to understanding the needs for the model validation adjustment factors. Travel models are estimated from household travel surveys. The travel surveys are typically taken in the spring and fall to avoid "abnormal" travel caused by winter weather or summer vacations. In some areas, surveys on Mondays and Fridays have not been permitted in order to avoid inclusion of "abnormal" travel patterns that might occur on those days in the household survey data. In short, travel models tend to reflect average weekdays in the spring or fall.

The model validation file should reflect a comparable time period. To accomplish this, monthly adjustment factors could be developed to adjust traffic counts to reflect the average March through May and September through November traffic volumes rather than the average annual traffic volumes. Second, day of week factors should be developed to estimate adjustments necessary to produce average weekday traffic. This process will be especially important if seasonal adjustment factors are stratified by facility type and area type or some other geography identifier.

[3] The results of the MITS loop count data analyses were reported in a telephone conversation with SEMCOG staff. Actual detail of the analyses were unavailable.

The resulting validation period average weekday traffic should be compared to AADT using the same validation criteria as typically used for model validation. High levels of correlation and low root mean squared errors should be expected even when the validation measures are stratified by area type and facility type. However, if there are some unexpectedly large differences, they would suggest that some model adjustments made to produce a better match to AADT could have been caused by seasonal variations in traffic rather than errors in the model.

Since most planning for roadways is based on AADT or AWDT, it will still be important to for the models to match those volumes as closely as possible. If comparisons of AADT and AWDT show high levels of correlation and low root mean squared errors across all area types and facility types, model adjustments to better match AADT are appropriate. However, if there are unexpectedly large differences between AADT and AWDT for some area types or facility types, it might be more appropriate to make model adjustments to reproduce *validation period* average weekday traffic. This procedure might prevent introducing incorrect sensitivities into specific model components (e.g., trip generation or trip distribution). Since planning would still require AADT estimates, post-assignment adjustments would be applied to convert validation period average weekday traffic to AADT.

IDEALS FOR VALIDATION DATASETS

A number of ideals can be stated for validation datasets that will be created from the RTCD:

- Coverage: The validation dataset should provide adequate coverage in terms of number of links for the primary validation strata for network links. These strata include geographic strata such as districts or counties; joint facility type / area type strata; and, volume class strata.
- Randomness: Greater randomness in link selection for the validation dataset should reduce the likelihood of inappropriate model adjustments. As an extreme example, suppose the validation dataset included every I-94 network link, but no links from other freeways. With such a validation dataset, changes could be made to the models that would cause a close match between modeled and observed volumes on I-94, but poor matches between modeled and observed volumes on other freeways. This ideal suggests that validation datasets do not have to include every possible link in the RTCD.
- Screenlines: Screenline validation datasets should probably be separate from the "general" validation dataset. By nature, screenline validation data violate the randomness ideal.
- Consistency with the modeling network: Since the RTCD does not necessarily use the same network structure as the modeling network, care must be exercised in link selection to ensure that the counted location reflects the same location in the modeling network. This is especially true around ramps and in areas where the modeling network is somewhat abstract.
- Time-of-day counts: To the maximum extent possible, the validation dataset should include time-of-day counts in addition to the daily counts. The RTCD maintains counts on links for periods as small as 15 minutes and the model might use up to 10 time periods for assignment. The short time periods used for traffic assignment should not be considered for model validation, however, due to the normal variation in counts. At the most, validation to three or four time periods (which are aggregations of the periods used for traffic assignment) should be performed. For example, validation to three time periods – morning peak, afternoon peak, and off-peak, is reasonable. The key in the

development of the validation file is the posting time-of-day counts that match the aggregations of the time periods used for traffic assignment. This might require posting counts for periods that start or end on the "half-hour."

- Classification counts: To the maximum extent possible, the validation dataset should include vehicle classification counts in addition to the daily counts.
- Consistency with forecasting period: Links selected for the validation dataset should have counts that were collected on a weekday, not a weekend day. While factors can be developed to adjust weekend day counts to AADT or AWDT, weekend travel patterns and characteristics are so different from those on weekdays to make the selection of links with weekend counts a questionable practice for validation datasets. This ideal becomes even more important if time-of-day or classification counts are included on the validation dataset.

STATE-OF-THE-PRACTICE GUIDELINES FOR MODEL VALIDATION

PURPOSE OF VALIDATION

The FHWA's *Model Validation and Reasonableness Checking Manual*[4] (hereafter referred to as the "FHWA Validation Manual") defines the purpose of model validation as follows:

> In order to test the ability of the model to predict future behavior, validation requires comparing the model predictions with information other than that used in estimating the model. This step is typically an iterative process linked to model calibration. It involves checking the model results against observed data and adjusting parameters until model results fall within an acceptable range of error. If the only way that a model will replicate observed data is through the use of unusual parameters and procedures or localized "quick-fixes", then it is unlikely that the model can reliably forecast future conditions.

In effect, model validation is the process that demonstrates that the travel models are working reasonably. The FHWA Validation Manual further points out that model validation should be performed throughout the model estimation process, not just at the end. This project, however, is focusing on the final process validation that "proves" that a model is working reasonably by reproducing observed traffic volumes. It can be argued that the validation against traffic counts is a "super-point." If traffic volumes cannot be reasonably reproduced, the travel models are suspect regardless of the quality of the validation of individual model components.

The FHWA Validation Manual proposes that there are two types of validation checks: reasonableness checks and sensitivity checks. Reasonableness checks are intended to show that a travel model reasonably reproduces travel behavior for a specific point in time; sensitivity checks are intended to show that a travel model responds properly to changes in the transportation system or socioeconomic characteristics of the region.

[4] Barton-Aschman Associate, Inc. and Cambridge Systematics, Inc., *Validation and Reasonableness Checking Manual*, Federal Highway Administration Travel Model Improvement Program, February 1997.

SUMMARY OF PRACTICES

The following sections provide a summary of model validation practices used in various regions throughout the country. The examples are based on reports and consultant experience. The validation practices focus on the final, process validation.

FHWA VALIDATION MANUAL

Excerpts from the FHWA Validation Manual have been included in Appendix A to this report. The FHWA Validation Manual provides a good set of reasonableness checks along with guidelines for acceptable results. This manual provides guidelines, not summaries of actual practices.

MDOT

In 1993, MDOT produced a brief document of *Urban Model Calibration Targets*. A number of the calibration targets were included in the FHWA Validation Manual included in Appendix A. This manual provides guidelines, not summaries of actual practices.

OHIO DEPARTMENT OF TRANSPORTATION

The Ohio Department of Transportation (ODOT) published a manual on traffic assignment procedures[5] that included a section on assignment checks / validation (see Appendix B). This manual provides guidelines, not summaries of actual practices.

DENVER REGIONAL COUNCIL OF GOVERNMENTS

The Denver Regional Council of Governments (DRCOG) has recently completed a model refresh project. This project updated existing travel models based on travel survey data collected in 1997 and completed the transfer of the models from MINUTP to TransCAD modeling software. The highway validation focused on standard validation measures: VMT, correlation between modeled and observed counts, and percent RMSE for the region and by area type / facility type strata. The models were deemed valid when validation measures met or exceeded those achieved using the previous MINUTP version of the travel model.

The model validation included several rarely performed tests:

- Model validation was performed for two separate years 1997 and 2001. A major light rail transit line was opened between the calibration year (1997) and 2001.
- The model validation included a validation of assignment speeds by time-of-day for 2001 (based on a speed study performed in 2002). Volume-delay functions were adjusted so that both traffic volumes and estimated speeds were validated simultaneously.
- A detailed transit assignment validation was performed. The transit validation focused on total system boardings, boardings by transit service type, and parking demand at major park-and-Ride lots. In addition, transit screenlines were defined (for comparison of modeled and observed ridership on transit vehicles crossing the screenlines). As with the highway side validation, the transit validation was performed for 1997 and 2001.

---

[5] Giaimo, Gregory, *Travel Demand Forecasting Manual 1 – Traffic Assignment Procedures*, Ohio Department of Transportation, Division of Planning, Office of Technical Services, August 2001.

The performance of the validation for two separate years provided a good sensitivity check increasing the credibility of the model.

LAS VEGAS

The models maintained by the Regional Transportation Commission were recently updated for a study of US 95 through Las Vegas. The highway validation focused on standard validation measures: VMT, correlation between modeled and observed counts, and percent RMSE for the region and by area type / facility type strata. Validation measures were also summarized by geographic area. A substantial number of screenlines and cutlines were defined for the region.

SOUTHERN CALIFORNIA

In the late 1990s, the travel model maintained by the Orange County Transportation Authority (OCTA) was updated in concert with a model update performed by the Southern California Association of Governments (SCAG). The model validation focused on reproducing regional VMT from HPMS summaries and reproducing traffic crossing a standard set of screenlines. In addition, correlation between modeled and observed counts, and percent RMSE for the region and by area type / facility type strata were summarized. Surprisingly, the primary source of traffic count data was from the screenlines. Few, if any, traffic counts were available for links not on a screenline.

SOUTH EAST FLORIDA

The South East Florida Regional Planning Model combines county planning models for Miami-Dade, Broward, and Palm Beach Counties into one unified model, SERPM. In the early 2000s, the fifth version of the model, SERPM-V, has been calibrated and validated. Extensive model validation was performed throughout the model development process. The final traffic assignment validation was performed using 1999 as a model year. The validation focused on regional statistics and statistics specifically to each of the three counties included in the model. Validation statistics included the ratio of total modeled to count VMT on links with counts by county and for the region. VMT summaries were not stratified by facility type and area type. Percent RMSE was calculated by volume class for all links with counts by county and for the region. In addition to the VMT and %RMSE summaries 2025 forecasts were summarized and compared to 1999 results for reasonability.

SEMCOG VALIDATION PROCEDURES

The recently updated SEMCOG travel models were to have been validated using standard measures: comparisons of modeled and observed VMT by facility type and area type, screenline volumes, and RMSE. However, an issue arose with the available validation data:

> There were problems with the traffic count data available for model validation in 2002, as documented in a Cambridge Systematics memorandum to SEMCOG. The issue is that the new TransCAD model produces about the same total VMT as the original TRANPLAN model, but the traffic counts are significantly higher (about 15 to 20 percent) than the modeled volumes. This meant that there was no reliable data source for the highway assignment validation checks described above. While it is important that validation be performed in the future as soon as

count data become available, the best that could be done at the time was to compare VMT by functional class and area type to the original TRANPLAN model.[6]

SEMCOG has continued to perform validation testing as new data from the RTCD become available and as the TransCAD model continues to be improved and tested. The following modifications to the TransCAD model have resulted in improved validation statistics:

- Time-of-day factors were adjusted to better reflect travel for the different times-of-day.
- The entire TransCAD network was recoded using GIS.
- External station related travel was corrected.
- Volume-delay functions for traffic assignment were adjusted to better reflect the impact of congestion on travel time.

The changes listed above have resulted in the model producing substantially better distributions of VMT by functional class throughout the region, although some questions regarding freeway VMT still exist due to the lack of freeway counts in the RTCD. No formal validation documentation has been produced.

SUMMARY OF VALIDATION RECOMMENDATIONS AND PRACTICES

Table 3 summarizes the primary validation recommendations from FHWA along with the actual practices used in the non-SEMCOG regions summarized above. Table 3 also recommends validation criteria for SEMCOG based on the FHWA and Ohio DOT recommendations, validations performed elsewhere, and desires and needs for the region. Data for most of the validation statistics should be readily available in the RTCD. Table 3 does not specify calculation procedures for the statistics or standards. Existing FHWA, Michigan DOT, and Ohio DOT manuals provide calculation procedures and reasonable standards for most of the criteria.

RECOMMENDATIONS

The development of the RTCD provides SEMCOG with the potential for setting an example of best practices for the maintenance of traffic count data for model validation. However, several issues, as recommended below, should be addressed as part of this project or in future projects.

RECOMMENDATIONS FOR PART 2

Based on the Part 1 effort, the following recommendations should be considered as high priority for Part 2.

- <u>Develop Two Sampling Plans</u>: The model-based sampling plan will provide a basis for traffic count data deficiency analysis and identify data requirements to achieve the best model calibration/validation result. The data-based sampling plan will generate a feasible validation dataset given the current data availability.

[6] Cambridge Systematics, Inc., SEMCOG Travel Model Documentation, prepared for SEMCOG, 2003, p. 10-6.

- Develop QA/QC Procedures and Standards: Much work is being performed in the acquisition and entry of count data into the RTCD. Some data checking is being performed, perhaps even against a standard set of rules. However, as noted, some data that might be unusable for a validation dataset (due to lack of day-of-week of count information) were discovered in a cursory check of the example RTCD data. Also, in the kick-off meeting, MDOT staff and SEMCOG staff seemed to be unaware of each others' QA/QC procedures. A documented set of QA/QC standards should be developed and, if feasible, automated into TCDS or other software.
- Review AADT Factors and Develop Procedures for Estimating Validation Period Weekday Traffic: The reasonability of developing the alternative seasonal adjustment factors should be more fully investigated along with a test of the significance of the alternative factors.

Table 4 lists the specific Part 2 work items along with a relative prioritization considering the findings of Part 1 along with the budget for Part 2.

Table 3. Comparison of Validation Practices and Procedures

Validation Practice / Procedure	**FHWA**	**Ohio DOT**	**DRCOG**	**Las Vegas**	**OCTA / SCAG**	**SE Florida**	**Recommended SEMCOG**
Comparisons of Daily Assigned and Count VMT							
Regional	✓	✓	✓	✓	✓	✓	✓
By Facility Type / Area Type	✓	✓	✓	✓	✓		✓
By District, Ring, or Sector		✓	✓	✓		✓	✓
Comparison to HPMS VMT	✓						✓
R^2 of Daily Assigned and Counted Vehicles							
Regional	✓	✓					
By Facility Type / Area Type	✓	✓	✓	✓	✓		✓
By Volume Class					✓		✓
Percent Root Mean Squared Error of Daily Assigned and Counted Vehicles							
Regional	✓	✓	✓	✓	✓	✓	✓
By Facility Type / Area Type	✓	✓	✓	✓	✓		✓
By Volume Class						✓	✓
Visual check of Daily Assigned vs. Count Differences (from plots)	✓	✓	✓	✓	✓		✓
Time-of-Day Comparisons of VMT, R^2, & %RMSE							<1>
Time-of-Day Average Speeds by Facility Type / Area Type	✓		✓				<2>
Screenline Crossings	✓	✓	✓	✓	✓		✓
Vehicle Class							<3>
Multi-year Validation			✓				✓

Notes:

1 Due to variation in the time-of-day count data, these validation criteria should focus on the match between modeled and observed VMT by time period, possibly stratified by facility type and area type, and by district. Attempts to validate time-of-day assignments using R^2 and %RMSE are probably of little value.

2 Speed data will need to be collected.

3 Vehicle classification data will need to be collected.

Table 4. Part 2 Work Items and Priorities

Work Item	Clarification	Priority
Develop a statistically sound sampling plan for count collection/compilation, considering balances on spatial distribution of samples, such as county, functional class, area types, etc.	Issues are of particular concern: • Development of counts for a validation dataset from existing counts • Identification of locations in the region requiring new counts	**High**
Review screenline adequacy (number and geographic location).		**Medium**
Identify additional data needs, based on the quality and quantity of available existing data. The data needed may comprise: new types of data, alternative geographic locations (to supplement locations of existing data), and/or alternative stratifications of current data. If more data is required, the assessment will be made about how sound the validation file will be, given existing data.	Vs. sampling needs Stress the importance of having sufficient data to meet the needs of developing a sound validation file (if data needs are not met).	**High**
Determine categories and the appropriate number of counts required, including specifications of a set of weekday traffic counts, statistically sound, that will facilitate testing the desired validation criteria		**High**
Recommend count collection strategies on vehicle classification counts, hourly counts, and weekend counts	And sampling needs.	**Medium**
Recommend an equivalence/conversion process for vehicle classification data used in the FHWA Scheme F, the travel model (light, medium, and heavy), and the Mobile6 model.	SEMCOG has been provided a process from air quality consultants. SEMCOG will provide the process for review and comment.	**Low (Focus on review of recommended process)**
Recommend a method to derive hourly counts considering SEMCOG's 4-period model and its future 10-period model.	Validation should focus on matching overall statistics for specified time periods rather than matching detailed hourly counts by link.	**Low**
Identify a quality control/quality assurance procedure to apply to count data that measures the consistency and reasonableness of the data being collected, prior to inclusion in the validation file.	The development of these procedures is crucial to the development / maintenance of the RTCD.	**High**

Work Item	Clarification	Priority
Provide a recommendation on the methods and frequency of data collection for future updates to the validation file.		**Medium**
Review current monthly and daily adjustment factors and procedures for AADT and AWDT. Make recommendations for change, if warranted. The recommendation will include the methodology on computation of adjustment factors		**High**
Assess current SEMCOG methodology used to “grow” traffic count data from previous years to the present year when current counts are not available. Recommendations for change will be made as appropriate.		**High**
Recommend sample size and stratification of a weekend traffic count database. The database should be able to permit the analysis of differences in weekday vs. weekend travel by facility type, vehicle type, and time of day.	This should be a future task. A weekend travel model must be developed prior to a validation dataset being needed.	**Low**

RECOMMENDATIONS FOR FUTURE PROJECTS

- Import Existing Hourly Counts into RTCD: A wealth of hourly data is contained in files stored in various formats by the City of Detroit and several counties. SEMCOG should consider acquiring and importing these files into the RTCD. The import effort, however, can be laborious because a count location identifier may not have been consistently saved in these files.
- Collect Speed Data for Model Validation: The work performed by DRCOG on the collection and use of speed data for model calibration and validation constitutes a best practice that should be emulated by SEMCOG.
- Specify Transit Validation Data: SEMCOG will continue to enhance transit modeling capabilities and must collect the transit data necessary to validate the results.
- Specify Requirements for Weekend Travel Model Validation Data: Sample size and stratification of a weekend traffic count database will be required to support the validation of a weekend travel model. The database should be able to permit the analysis of differences between weekday and weekend travel by facility type, vehicle type, and time of day.

Part 2. Data Specification for Travel Demand Model Integration

INTRODUCTION

The purpose of Part 2, *Data Specification for Travel Demand Model Integration*, is to prepare a specification for data that will subsequently be integrated into the travel demand model (through the process developed in Part 3, *Traffic Count File Development*). Based on Part 1, *Data Review and Model Calibration / Validation State-of-the-Practice Assessment*, additional detail regarding the accomplishment of this Part was developed. Table 5 shows the recommended Part 2 work items and priorities as developed in Part 1. In addition, Table 5 lists the sections of this report where the work items are discussed.

Table 5. Part 2 Work Items and Priorities

Work Item	**Clarification**	**Priority**	**Discussed in Section**
Develop a statistically sound sampling plan for count collection/compilation, considering balances on spatial distribution of samples, such as county, functional class, area types, etc.	Issues are of particular concern: • Development of counts for a validation dataset from existing counts • Identification of locations in the region requiring new counts	High	Validation Procedures
Review screenline adequacy (number and geographic location).		Medium	Assessment of SEMCOG RTCD
Identify additional data needs, based on the quality and quantity of available existing data. The data needed may comprise: new types of data, alternative geographic locations (to supplement locations of existing data), and/or alternative stratifications of current data. If more data is required, the assessment will be made about how sound the validation file will be, given existing data.	Vs. sampling needs Stress the importance of having sufficient data to meet the needs of developing a sound validation file (if data needs are not met).	High	Assessment of SEMCOG RTCD
Determine categories and the appropriate number of counts required, including specifications of a set of weekday traffic counts, statistically sound, that will facilitate testing the desired validation criteria		High	Validation Procedures
Recommend count collection strategies on vehicle classification counts, hourly counts, and weekend counts	And sampling needs.	Medium	Validation Procedures

Work Item	Clarification	Priority	Discussed in Section
Recommend an equivalence/conversion process for vehicle classification data used in the FHWA Scheme F, the travel model (light, medium, and heavy), and the Mobile6 model.	SEMCOG has been provided a process from air quality consultants. SEMCOG will provide the process for review and comment.	Low (Focus on review of recomm ended process)	Other Validation Issues
Recommend a method to derive hourly counts considering SEMCOG's 4-period model and its future 10-period model.	Validation should focus on matching overall statistics for specified time periods rather than matching detailed hourly counts by link.	Low	Validation Procedures
Identify a quality control/quality assurance procedure to apply to count data that measures the consistency and reasonableness of the data being collected, prior to inclusion in the validation file.	The development of these procedures is crucial to the development / maintenance of the RTCD.	High	Quality Assurance / Quality Control
Provide a recommendation on the methods and frequency of data collection for future updates to the validation file.		Medium	Quality Assurance / Quality Control
Review current monthly and daily adjustment factors and procedures for AADT and AWDT. Make recommendations for change, if warranted. The recommendation will include the methodology on computation of adjustment factors		High	Quality Assurance / Quality Control
Assess current SEMCOG methodology used to "grow" traffic count data from previous years to the present year when current counts are not available. Recommendations for change will be made as appropriate.		High	Quality Assurance / Quality Control
Recommend sample size and stratification of a weekend traffic count database. The database should be able to permit the analysis of differences in weekday vs. weekend travel by functional class, vehicle type, and time of day.	This should be a future task. A weekend travel model must be developed prior to a validation dataset being needed.	Low	Validation Procedures

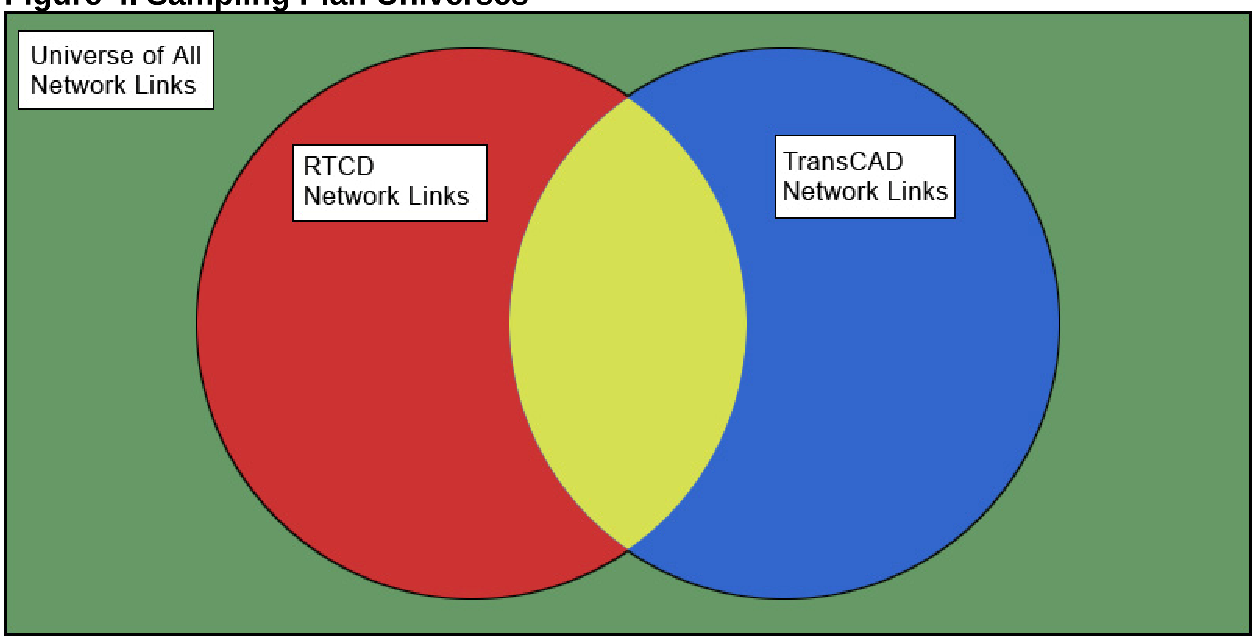
VALIDATION PROCEDURES

COUNT FILE SELECTION

Figure 4 is a Venn Diagram describing the universes of all roadway links in the SEMCOG region, the network links with counts included in the RTCD and the network links included in the network used for the TransCAD model. The selection of the count file for model validation could be based on any of the following five sampling plans:

f) Sampling all roadway links in the region to produce an "ideal" validation dataset for a "perfect" travel demand model.
g) Sampling roadway links in the existing RTCD to produce the "most cost-effective" validation dataset.
h) Sampling roadway links in the existing TransCAD network to produce the "best model-based" validation dataset.
i) Sampling roadway links contained in either the existing RTCD or the TransCAD datasets.
j) Sampling roadway links contained in both the existing RTCD and the TransCAD to produce the "best data-based" validation dataset.

Figure 4. Sampling Plan Universes



As described in the Part 1 technical memo, the model-based (Plan "c") and the data-based (Plan "e") sampling plans are the best candidate sampling plans for the development of a validation count file.

MODEL-BASED SAMPLING PLAN

The model-based sampling plan would sample links from the universe described by the blue and yellow circle in Figure 4. All network links included in the TransCAD model network would be candidate links for the validation dataset. Procedures outlined in subsequent parts of this report would be used to identify a random sample of links for the validation and traffic count data from the RTCD would be posted on the sample links for validation purposes. The sample produced using this procedure would be the least biased sample since it would be based on all links included in the TransCAD database. Thus, the model-based sampling plan would be the most cost-effective procedure for producing a validation dataset from a model validation standpoint.

One drawback to the model-based sampling plan is that it could identify links without traffic counts as validation links, making the use of this sampling plan for years prior to 2004 infeasible. In fact, it is probably impractical to acquire the additional traffic counts that could be required for a 2004 validation dataset in a timely manner. Thus, the model-based sampling plan should be considered only for the future development of validation datasets for 2005 and beyond to allow time to schedule and collect traffic counts for the selected links.

Data-Based Sampling Plan

The data-based sampling plan would sample links from the universe described by the yellow area in Figure 4. All network links included in both the TransCAD model network and the RTCD would be candidate links for the validation dataset. As with the model-based sampling plan, procedures outlined in subsequent parts of this report would be used to identify a random sample of links for the validation. The sample selection might be further refined to limit link selection to a reasonable range of years around the years specified for the validation. Traffic count data from the RTCD would be posted on the sample links for validation purposes.

The data-based sampling plan should be very cost-effective since all links included in the validation dataset would be guaranteed to have traffic counts. The data-based sampling plan is the only plan that can be used to develop validation datasets for past years. However, some bias in the validation dataset will be introduced since it is limited to links with existing traffic counts. The RTCD has been shown to be lacking counts for certain areas (e.g. Washtenaw County and the City of Detroit) and functional classes (e.g. freeways and interstate highways). Even though some bias will be introduced through the use of this sampling plan, it describes the de-facto state-of-the-practice for the development of validation datasets.

Stratified Random Sampling

A sampling methodology is required regardless of whether the model-based or the data-base sampling plan is used. For the model-based sampling plan, it would be infeasible to select all network links in the TransCAD network since the collection of traffic count data for all of the links would be infeasible. For the data-based sampling plan, a sampling methodology should be used since some geographic areas or functional classes are over-sampled in comparison to others.

Two sampling methodologies can be considered: simple random sampling and stratified random sampling. With the simple random sampling approach, any link in the universe of links should have an equal probability of being selected for the sample. This approach could work under the model-based sampling plan but would not be useful for removing any of the identified biases inherent in the data-based sampling plan.

Stratified random sampling is a variation of the simple random sampling methodology. Under this approach, the universe of links is stratified into different groups and simple random samples are selected for each group. Within each stratum, descriptive statistics based on the assumption of simple random sampling (e.g. mean, standard deviation, statistical confidence intervals, etc.) can be calculated and relatively simple procedures can be used to estimate the weighted statistics for the overall universe.

The stratified random sampling methodology has the added benefit of providing a means to remove some of the known bias associated with the data-based sampling plan. This can be accomplished through the use of disproportionate random sampling where specified numbers of samples are selected for each stratum. Examples can be used to demonstrate the procedure. Table 6 shows the numbers of directional roadway links by area type and functional class included in the TransCAD network for 2002 and Table 7 shows the number of directional links with counts. For definitional purposes, one two-way link counts as two directional links; a one-way link counts as one directional link. Tables 8 and 9 show similar information stratified by county and functional class.

Table 6. Directional Links in 2002 TransCAD Network Stratified by Area Type and Functional Class

	Functional Class				
Area Type	Freeway[1]	Principal Arterial	Minor Arterial	Collector/ Local	Total
Urban Business	229	1,255	1,063	982	3,529
Urban	602	3,420	3,416	1,862	9,300
Suburban / Rural Place	474	1,896	3,025	1,725	7,120
Rural	266	165	664	3,151	4,246
Total Roadway Links	1,571	6,736	8,168	7,720	24,195

[1] Excluding system-to-system ramps and ramps.

Table 7. Directional Links with Counts in 2002 TransCAD Network Stratified by Area Type and Functional Class

	Functional Class				
Area Type	Freeway[1]	Principal Arterial	Minor Arterial	Collector/ Local	Total
Urban Business	41	395	189	75	700
Urban	115	1,000	702	139	1,956
Suburban / Rural Place	116	986	1,657	917	3,676
Rural	108	91	453	2,232	2,884
Total Roadway Links[1]	380	2,472	3,001	3,363	9,216

[1] Excluding system-to-system ramps and ramps.

Table 8. Directional Links in 2002 TransCAD Network Stratified by County and Functional Class

County	Functional Class				Total
	Freeway[1]	Principal Arterial	Minor Arterial	Collector/ Local	
Livingston	102	40	197	744	1,083
Macomb	140	858	1,136	751	2,885
Monroe	101	90	391	781	1,363
Oakland	315	1,584	1,835	1,435	5,169
St. Clair	80	131	315	953	1,479
Washtenaw	158	440	654	1,102	2,354
Wayne	299	1,714	1,823	716	4,552
City of Detroit	376	1,879	1,817	1,238	5,310
Total Roadway Links[1]	1,571	6,736	8,168	7,720	24,195

[1] Excluding system-to-system ramps and ramps.

Table 9. Directional Links with Counts in 2002 TransCAD Network Stratified by County and Functional Class

County	Functional Class				Total
	Freeway[1]	Principal Arterial	Minor Arterial	Collector/ Local	
Livingston	37	24	118	578	757
Macomb	39	428	590	412	1,469
Monroe	28	48	208	506	790
Oakland	64	715	803	706	2,288
St. Clair	35	82	128	524	769
Washtenaw	52	146	220	356	774
Wayne	62	694	770	263	1,789
City of Detroit	63	335	164	18	580
Total Roadway Links[1]	380	2,472	3,001	3,363	9,216

[1] Excluding system-to-system ramps and ramps.

As an example, suppose that 30 links with counts are desired in each stratum for statistical purposes. Tables 10 through 13 show the proportions of total links and links with counts that are included in the samples. As is obvious, the proportions of links sampled vary by stratum. As shown in Table 10, if the sample is stratified by area type and functional class, and all links are considered, the proportion of sampled links by stratum varies from one percent to 18 percent, with the overall sample being two percent of the total directional links. However, when the sample is stratified by area type and functional class and only links with counts are considered, the sample proportions vary from about one to 73 percent, with the overall sample being five percent of the links with counts.

Table 10. Example Proportion of Directional Links in 2002 TransCAD Network Stratified by Area Type and Functional Class (Assuming 30 Samples per Stratum)

Area Type	Functional Class				Total
	Freeway[1]	Principal Arterial	Minor Arterial	Collector/ Local	
Urban Business	13% (30/229)	2% (30/1,255)	3% (30/1,063)	3% (30/982)	3% (120/3,529)
Urban	5% (30/602)	1% (30/3,420)	1% (30/3,416)	2% (30/1,862)	1% (120/9,300)
Suburban / Rural Place	6% (30/474)	2% (30/1,896)	1% (30/3,025)	2% (30/1,725)	2% (120/7,120)
Rural	11% (30/266)	18% (30/165)	5% (30/664)	1% (30/3,151)	3% (120/4,246)
Total Roadway Links[1]	8% (120/1,571)	2% (120/6,736)	1% (120/8,168)	2% (120/7,720)	2% (480/24,195)

[1] Excluding system-to-system ramps and ramps.

Table 11. Example Proportion of Directional Links with Counts in 2002 TransCAD Network Stratified by Area Type and Functional Class (Assuming 30 Samples per Stratum)

Area Type	Functional Class				Total
	Freeway[1]	Principal Arterial	Minor Arterial	Collector/ Local	
Urban Business	73% (30/41)	8% (30/395)	16% (30/189)	40% (30/75)	17% (120/700)
Urban	26% (30/115)	3% (30/1,000)	4% (30/702)	22% (30/139)	6% (120/1,956)
Suburban / Rural Place	26% (30/116)	3% (30/986)	2% (30/1,657)	3% (30/917)	3% (120/3,676)
Rural	28% (30/108)	33% (30/91)	7% (30/453)	1% (30/2,232)	4% (120/2,884)
Total Roadway Links[1]	32% (120/380)	5% (120/2,472)	4% (120/3,001)	4% (120/3,363)	5% (480/9,216)

[1] Excluding system-to-system ramps and ramps.

Table 12. Example Proportion of Directional Links in 2002 TransCAD Network Stratified by County and Functional Class (Assuming 30 Samples per Stratum)

County	Functional Class				Total
	Freeway[1]	Principal Arterial	Minor Arterial	Collector/ Local	
Livingston	29% (30/102)	75% (30/40)	15% (30/197)	4% (30/744)	11% (120/1,083)
Macomb	21% (30/140)	3% (30/858)	3% (30/1,136)	4% (30/751)	4% (120/2,885)
Monroe	30% (30/101)	33% (30/90)	8% (30/391)	4% (30/781)	9% (120/1,363)
Oakland	10% (30/315)	2% (30/1,584)	2% (30/1,835)	2% (30/1,435)	2% (120/5,169)
St. Clair	38% (30/80)	23% (30/131)	10% (30/315)	3% (30/953)	8% (120/1,479)
Washtenaw	19% (30/158)	7% (30/440)	5% (30/654)	3% (30/1,102)	5% (120/2,354)
Wayne	10% (30/299)	2% (30/1,714)	2% (30/1,823)	4% (30/716)	3% (120/4,552)
City of Detroit	8% (30/376)	2% (30/1,879)	2% (30/1,817)	2% (30/1,238)	2% (120/5,310)
Total Roadway Links[1]	15% (240/1,571)	4% (240/6,736)	3% (240/8,168)	3% (240/7,720)	4% (960/24,195)

1 Excluding system-to-system ramps and ramps.

Table 13. Example Proportion of Directional Links with Counts in 2002 TransCAD Network Stratified by County and Functional Class (Assuming 30 Samples per Stratum)

County	Functional Class				Total
	Freeway[1]	Principal Arterial	Minor Arterial	Collector/ Local	
Livingston	81% (30/37)	100% (24/24)[2]	25% (30/118)	5% (30/578)	16% (114/757)
Macomb	77% (30/39)	7% (30/428)	5% (30/590)	7% (30/412)	8% (120/1,469)
Monroe	100% (28/28)[2]	63% (30/48)	14% (30/208)	6% (30/506)	15% (118/790)
Oakland	47% (30/64)	4% (30/715)	4% (30/803)	4% (30/706)	5% (120/2,288)
St. Clair	86% (30/35)	37% (30/82)	23% (30/128)	6% (30/524)	16% (120/769)
Washtenaw	58% (30/52)	21% (30/146)	14% (30/220)	8% (30/356)	16% (120/774)
Wayne	48% (30/62)	4% (30/694)	4% (30/770)	11% (30/263)	7% (120/1,789)
City of Detroit	48% (30/63)	9% (30/335)	18% (30/164)	100% (18/18)[2]	21% (108/580)
Total Roadway Links[1]	63% (238/380)	9% (234/2,472)	8% (240/3,001)	7% (228/3,363)	10% (940/9,216)

1 Excluding system-to-system ramps and ramps.
2 Fewer than 30 links with counts.

Tables 12 and 13 show many of the same characteristics as Tables 10 and 11. Note, however, since the number of strata are greater when the sample is stratified by county and functional class then when the sample is stratified by area type and functional class, more total samples would be selected. Using this example stratification, the overall sample size would be doubled to four percent for the model-based sampling approach (Table 12 versus Table 10). For a data-based sample, the overall sample size would be almost doubled to 940 samples (Table 13 versus Table 11). However, three of the 32 sample strata would have insufficient links with counts to satisfy the example sample requirements of 30 links per stratum.

Within each stratum shown in Tables 10 through 13, descriptive statistics and confidence intervals can be estimated directly based upon the samples. However, if regional statistics and confidence intervals are desired, the summary statistics for each stratum will need to be weighted based upon the sample proportions (see Equations 2 and 3, below).

HIGHWAY PERFORMANCE MONITORING SYSTEM SAMPLING PROCEDURES

The Highway Performance Monitoring System (HPMS) Field Manual recommends a stratified random sampling approach for collecting data required for HPMS reporting[7]. The stratified sampling procedure recommended in the manual is based on thirteen average annual daily traffic (AADT) volume groups and five functional classes. While the HPMS is targeted at specific reporting requirements that are not necessarily the focus of a model validation, the sample selection procedures and methods for determining statistical significance are useful for the validation file development process. In particular, Appendix C of the HPMS provides a formula for the estimation of the number of samples required for estimation of information at a chosen level of statistical significance. The formula includes a “finite universe” correction factor:

$$n = \frac{Z^2 C^2 / d^2}{1 + (1/N)\left[\left(Z^2 C^2 / d^2\right) - 1\right]} \qquad (1)$$

where: n = required sample size
- Z = value of the standard normal statistic for an alpha confidence level (two sided):
  - = 1.96 for 95 percent confidence
  - = 1.645 for 90 percent confidence
  - = 1.282 for 80 percent confidence
  - = 1.040 for 70 percent confidence
- C = coefficient of variation
  - = population standard deviation / population mean
- d = desired precision rate
- N = Universe or population of the stratum being measured

For example, there are a total of 229 directional urban business freeway links in the TransCAD network. If the coefficient of variation is 0.75 and ±10 percent accuracy at the 90 percent confidence level is desired, traffic counts for 92 randomly selected links would be required.

The above formula is based on the assumption that links are selected at random. The random sample selection process can be easily performed, even for the stratified random sampling procedures since each stratum can be treated separately. One procedure that can be used to

[7] Highway Performance Monitoring System Field Manual, US Department of Transportation, Federal Highway Administration, December 2002, Chapter 7 and Appendices C and D.

select the links for the sample is to assign a random number to each TransCAD network and then sort the links by random number within each stratum. After this work is performed, the first "n" links required can be selected as the sample. This strategy works reasonably for future year calibration datasets (i.e. using the model-based sampling plan) when the sample can be specified prior to the counts being taken. For past years (i.e. based on the data-based sampling plan), this procedure will be impossible and either an alternative sampling strategy will need to be used (depending on the counts available) or sample statistics that account for non-random sampling will need to be estimated.

The HPMS stratified random sampling approach described above provides a good link selection procedure for the validation files. The approach is, however, based upon knowing the coefficient of variation of the variable being reviewed. The coefficients of variation can be different for different variables. For the validation datasets, the coefficient of variation should be based on AADT. This is the same variable that drives the development of the HPMS, so coefficient of variation information might be "borrowed" from the HPMS Field Manual.

TRADITIONAL VALIDATION STRATA REVIEW

Part 1 included a review of validation procedures used by a number of agencies and Metropolitan Planning Organizations throughout the US. Most regions use functional class and area type as the primary strata for model validation although a number also use district, ring, or sector of the region in addition to the functional class and area type strata. Of course, regional summaries were also important in the validation process.

RECOMMENDED SEMCOG VALIDATION STRATA

The Part 1 of this report also recommended validation measures and strata for the SEMCOG region. Specifically, it was recommended that functional class and area type be used as validation strata along with district (or county) summaries and summaries for the region. The stratified random sampling approach will provide the necessary data for estimation of region-wide measures, although formulae for estimating weighted statistics will need to be applied for regional summaries (or any summaries that are aggregations of the strata used for the sample selection such as summing over area types to get an estimated weighted population mean for a specific functional class for the region). The estimated weighted population mean can be estimated as:

$$\overline{X} = \sum_{i=1}^{k} W_i \overline{X}_i \qquad (2)$$

where: $\overline{X}$ = estimated weighted population mean

W_i = weight (sample proportion) of the ith stratum in the population ($\sum_{i=1}^{k} W_i = 1$)

$\overline{X}_i$ = estimated sample mean for stratum i

The estimated weighted population variance using stratified random sampling can be estimated using the following formula:

$$\hat{\sigma}_{\overline{X}}^{\,2} = \sum_{i=1}^{k} W_i^2 \,\hat{\sigma}_{\overline{X}_i}^{\,2} \qquad (3)$$

where: $\hat{\sigma}_{\overline{X}}^{\,2}$ = estimated weighted population variance of the mean

$\hat{\sigma}_{\overline{X}_i}^{\,2}$ = estimated variance of the mean within the i^{th} stratum

There are two different options that can be considered for the stratification of the validation measures by functional class, area type, and county. One option would be to consider the stratification to be three dimensional (e.g. five functional class strata by four area type strata by seven county strata would result in 140 different strata). While such a stratification scheme would provide a substantial amount of information about the model validation, it is likely to be prohibitively expensive to collect the traffic counts necessary for the model-based sampling plan. In addition, as can be inferred from the example shown in Table 13, insufficient counts exist in the existing RTCD for the data-based sampling plan.

An alternative sampling scheme would focus on the functional class by area type stratification, but sample links in such a way that good coverage of the region is obtained. Based on this sampling scheme, it would be possible to summarize regional statistics by functional class and area type and also summarize certain statistics by county. The procedure would require fewer traffic counts and, thus, should be more feasible, especially for the data-based sampling plan. The sampling procedure outlined in Exhibit 1 can be used. The procedure will not guarantee that a specific number of links are selected for each county, but it will ensure that selected links are spread throughout the region.

Exhibit 1
Recommended Sampling Procedure for Validation Files

1) Assign a random number to each candidate TransCAD directional link.
2) For each functional class / area type, sort the candidate links by the random number.
3) Determine the total number of links in the functional class / area type stratum (i.e. the population of links) and the number of samples required. Divide the population of links by the number of required samples to obtain a sampling interval, "m."
4) Determine a random number, "r," between one and "m." Select sample "r" from the list of links developed in Step 2 and then select every "m^{th}" link thereafter.
5) Repeat the procedure for each functional class / area type.

The Part 1 of this report suggested that statistics should be estimated by volume classes in addition to functional class, area type, and county. As with the county statistics, it is suggested that volume classes simply be considered over the region, not within each functional class, area type, or county stratum. Since traffic volumes will vary by functional class, area type, and county, the selection procedure outlined above should produce a sufficient number of samples for each volume class.

The recommended sample size(s) must be determined prior to the sample selection. As described in the section on HPMS Sampling Procedures, sample size will depend, in part, on the variation of the variable being analyzed. Analyzing different variables could result in different required sample sizes for a specific stratum. Since the selection of multiple samples would add substantial complication to the validation process, the selection of a single "primary" variable for sample selection is suggested. A number of reasonable candidate variables could be considered, but the two primary variables are probably AADT and VMT. As noted above, it is recommended that AADT be used for the SEMCOG region.

Table 14 summarizes coefficients of variation for observed counts by functional class and area type based on the 2002 analysis count file. Table 15 shows recommended coefficients of variation to use for estimation of sample sizes. The recommended coefficients of variation have generally been increased over those shown in Table 14 to provide a safety margin. Table 16 shows required samples sizes for ±10 percent accuracy at the 90 percent confidence level for the model-based sampling plan using the recommended coefficients of variation. Table 17 shows the same information for the data-based sampling plan. Both tables were estimated using the formula shown in Equation 1. Tables 18 and 19 show the proportions of directional TransCAD links and the proportion of directional links with counts that must be included in the samples for ±10 percent accuracy at the 90 percent confidence level.

Table 14. Coefficients of Variation in Observed Counts based on 2002 Count File

	Functional Class			
Area Type	Freeway[1]	Principal Arterial	Minor Arterial	Collector / Local
Urban Business	0.37	0.60	0.50	0.65
Urban	0.47	0.57	0.51	0.83
Suburban / Rural Place	0.43	0.50	0.56	0.72
Rural	0.42	0.42	0.64	0.91

[1] Excludes freeway ramps and freeway-to-freeway connectors

Table 15. Recommended Coefficients of Variation for Sample Size Determination

	Functional Class			
Area Type	Freeway[1]	Principal Arterial	Minor Arterial	Collector / Local
Urban Business	0.40	0.60	0.50	0.65
Urban	0.50	0.60	0.55	0.85
Suburban / Rural Place	0.45	0.50	0.60	0.75
Rural	0.45	0.45	0.65	0.95

[1] Excludes freeway ramps and freeway-to-freeway connectors

Table 16. Model-Based Sampling Plan Sample Sizes (±10 Percent Accuracy at the 90 Percent Confidence Level)

Area Type	Functional Class				
	Freeway[1]	Principal Arterial	Minor Arterial	Collector / Local	Total Links to Sample
Urban Business	37	90	64	103	294
Urban	61	95	80	177	413
Suburban / Rural Place	49	65	94	140	348
Rural	46	41	98	227	412
Total Links to Sample	193	291	336	647	1,467

1 Excludes freeway ramps and freeway-to-freeway connectors

Table 17. Data-Based Sampling Plan Sample Sizes (±10 Percent Accuracy at the 90 Percent Confidence Level)

Area Type	Functional Class				
	Freeway[1]	Principal Arterial	Minor Arterial	Collector / Local	Total Links to Sample
Urban Business	21	78	50	46	195
Urban	43	89	73	81	286
Suburban / Rural Place	37	63	92	131	323
Rural	37	34	91	220	382
Total Links to Sample	138	264	306	478	1,186

1 Excludes freeway ramps and freeway-to-freeway connectors

Table 18. Proportions of TransCAD Directional Links Sampled for Model-Based Sampling Plan (±10 Percent Accuracy at the 90 Percent Confidence Level)[1]

Area Type	Functional Class				
	Freeway[1]	Principal Arterial	Minor Arterial	Collector / Local	Total Links to Sample
Urban Business	16%	7%	6%	10%	8%
Urban	10%	3%	2%	10%	4%
Suburban / Rural Place	10%	3%	3%	8%	5%
Rural	17%	25%	15%	7%	10%
Total Links to Sample	12%	4%	4%	8%	6%

1 Total population of directional links is 24,195
2 Excludes freeway ramps and freeway-to-freeway connectors

Table 19. Proportions of Directional Links with Counts Sampled for Data-Based Sampling Plan (±10 Percent Accuracy at the 90 Percent Confidence Level)[1]

Area Type	Functional Class				
	Freeway[1]	Principal Arterial	Minor Arterial	Collector / Local	Total Links to Sample
Urban Business	51%	20%	26%	61%	28%
Urban	37%	9%	10%	58%	15%
Suburban / Rural Place	32%	6%	6%	14%	9%
Rural	34%	37%	20%	10%	13%
Total Links to Sample	36%	11%	10%	14%	13%

1 Total population of directional links with counts is 9,216
2 Excludes freeway ramps and freeway-to-freeway connectors

PAST-YEAR AND FUTURE-YEAR VALIDATION CONSIDERATIONS

The model-based sampling plan is the preferred procedure for developing a validation file. However, as mentioned previously, the model-based sampling plan can identify TransCAD links without associated counts in the RTCD and, for this reason, is primarily suited for the development of validation datasets for anticipated future year validations, especially those anticipated for 2005 and beyond. For future year validations, all TransCAD links, including those anticipated to be constructed by the future year should be considered as candidate links in the sampling methodology outlined in Exhibit 1. Once the sample has been specified, appropriate actions should be taken by SEMCOG to ensure that the selected links are included in the count program for the future year.

The model-based sampling plan runs the risk of producing a sample that substantially increases the efforts required for the regional count program since only a portion of the links might have counts. There are several ways that might mitigate the problem. First, counts scheduled for links for other sampling purposes can be used for the validation sample (if they happen to be selected) with no loss of generality. For example, if the sample for the validation file selects a link that is also counted for the HPMS, the count normally taken for the HPMS can be used for the validation file. Second, counts do not necessarily have to be taken in the validation year. Counts from up to about three[8] years prior to the validation year can be factored up for the future year validation file. For example, if a selected link for a 2006 validation file had a count taken in 2004, the 2004 count could be factored up to 2006 based on average annual growth factors. This procedure will introduce some bias. However, the use of factored counts is standard practice for the creation of validation count files created for cities throughout the US.

A third procedure to mitigate the cost of the development of a future year validation file can be considered. Specifically, some counts could be “replaced” if they are not scheduled with counts that are scheduled. This procedure will introduce bias and should be used sparingly. It can be implemented two ways. The first option is to create an over-sample of count links for the validation file. For example, if 60 links are required for a stratum, select a sample of, say, 120 links. Then, if a link is not scheduled for counting or has not been counted two years, it can be skipped and the next randomly sampled link can be selected. Obviously, taken to the extreme, this procedure would produce a data-based sample rather than a model-based sample. Thus, the number of replaced links should be limited to, say, 25 percent of the total links.

The second option for replacement of traffic counts introduces bias since it is not a random procedure. For this reason, it should be used only as a last resort for completing a sample. Basically, the procedure is to use an existing count from a nearby link for the same facility. Care would be required to ensure that there were no major intersecting facilities, no major changes in the facility (such as a change in the number of lanes), and no major changes in adjacent land uses between the selected link and the proposed replacement link. This

[8] The three year limit was selected based on an estimated average growth in traffic volume on facilities of about a two percent per year. A two percent per year growth rate would result in a 6.1 percent adjustment to the count volume over three years. In rapidly growing areas where the growth rate exceeds two percent per year, adjustment periods might be limited to two years. For example, a three percent per year growth rate would result in about a 9.3 percent adjustment to the traffic volume over a three year period but only a 6.1 percent adjustment over two years. In areas of low growth, the three year limit might be increased. A one percent per year growth rate would result in only a 5.1 percent adjustment in traffic volumes over a five year period. For reasonability, traffic count adjustments greater than about six percent should be avoided.

replacement procedure would need to be implemented manually by an analyst familiar with the roadway network and area surrounding the selected and replacement links.

Selection procedures for validations up to and including 2004 will, by necessity, be based on the data-based sampling plan. Mechanically, the only change to the sampling procedure outlined in Exhibit 1 will be that only links with counts for the validation year should be considered as candidate links. The population of candidate links can be expanded to use links with counts if they were taken within a specified number of years of the validation year, provided the counts can be factored up or down as necessary to account for annual growth in traffic.

For past year validations, the sample frame will not be a random sample TransCAD model links since only links with existing counts can be included. The sampling procedure should, however, produce a random sample of links with counts. The bias introduced by including only links with counts in the sample frame will have the impact of increasing the confidence intervals associated with the various statistics (e.g. VMT) estimated for the validation. In other words, if the model-based sample selection procedure is used, observed VMT for modeled urban principal arterials should be reported as “x” miles ±10 percent at the 90 percent confidence level. For the data-based sampling procedure, the error at the 90 percent confidence level would be greater than ±10 percent. There is, unfortunately, no way to determine the actual impact of the sample bias on the error level.

Given that the impact of the sample bias on the sample error cannot be fully determined for the data-based sampling frame, an alternative strategy is recommended. Specifically, there is no cost, or almost no cost, for including all links with counts in the RTCD in the validation data file. If all links with counts are selected for the validation data file, the coefficients of variation listed in Table 15 can be used along with the number of links with count by stratum to estimate the sample error using the following transformation of Equation 1:

$$d = \frac{ZC}{\sqrt{\frac{n(N-1)}{(N-n)}}} \tag{4}$$

where d, Z, C, n, and N are as defined for Equation 1.

Note that the sample error estimated using the above formula is simply a lower limit of the sample error that would be achieved if the counted links had been from a random sample. The actual sample error will be greater since the counted links were not from a random sample. Nevertheless, this procedure, using all links with counts from the RTCD for the validation data file, is comparable to procedures traditionally used throughout the country.

Exhibit 2
Summary of Recommendations for Validation File Sampling Plans

Model-Based Sampling Plans

1. Use for validation files starting in 2005.
2. Stratify by area type and functional class with sample sizes as shown in Table 16.
3. Use the sampling procedures outlined in Exhibit 1.
4. Draw an over-sample of links for each stratum to allow for limited substitution of counts to reduce overall traffic counting costs.

Data-Based Sampling Plans

1. Use for validation files for 2004 or prior years.
2. Select all links with counts and estimate minimum sample errors using Equation 4.

COUNT FILE DEVELOPMENT ISSUES

REVIEW OF RECOMMENDED SEMCOG VALIDATION MEASURES

The Part 1 of this report recommended validation practices and procedures be used by SEMCOG for the model validation process. The procedures and practices directly affected by the validation file creation procedures are as follows:

- *Comparison of daily assigned VMT and count VMT for the region, by functional class and area type, and by district (or county)*. These comparisons will be based on the sample of counts selected for the validation file.
- *R^2 for daily assigned volumes and traffic counts by functional class and area type, and by volume class*[9]. These comparisons will be based on the sample of counts selected for the validation file.
- *Percent root mean squared error of daily assigned and counted vehicles for the region, by functional class and area type, and by volume class*. These comparisons will be based on the sample of counts selected for the validation file.
- *Visual check of daily assigned and count differences*. These comparisons will be based on the sample of counts selected for the validation file. It is suggested that the errors between assigned and observed counts be sorted by decreasing magnitude of the error and that a set number (such as the top 100 errors) be reviewed by an analyst. As well as listing the pertinent link information, the locations of the links should be mapped and reviewed to identify geographic biases associated with the model.
- *Time-of-day comparisons of VMT, R^2, and percent RMSE*. These comparisons will be based on the sample of counts selected for the validation file. Although the goal should be to collect time-of-day counts for all count locations, not all links selected for the

[9] Unlike facility type and area type, volume class is not a specific stratum considered in the model-based sample selection procedures for validation files. Nevertheless, due to the natural variation in volumes by facility type and area type and the random procedures used to pick model-based sample files, there should not be any problems with stratifying validation measures by volume class.

validation file will have time-of-day counts, especially for early generations of the RTCD. R^2 measures are likely to be lower and percent RMSE measures are likely to be higher for time-of-day validations than for daily validations. This will be mainly due to natural variation in the time-of-day traffic counts but could easily be incorrectly attributed to inaccurate model results. Thus, until substantial experience has been gained regarding the accuracy and natural variation of time-of-day counts, validation measures should focus on the reproduction of VMT by time period.

- *Screenline crossings*. These comparisons will be independent of the sample of counts selected for the validation file. In effect, the screenlines are a supplementary list of counts that should be maintained in the RTCD for validation purposes. Note that screenline links should not be excluded from either the data-based or the model-based sampling procedures.
- *Multi-year validation*. Whenever models are recalibrated, validations should be performed for several different years to demonstrate model stability over time. Multi-year validations might also be considered whenever the travel model is used to support the development of the Regional Transportation Plan. The multi-year validation can be performed by simply re-sampling the RTCD or using the same samples of links for each year of the validation (with appropriate count data for each specific validation year).

Several validation measures were recommended in the Part 1 of this report that are independent of the RTCD or cannot yet be fully supported by the RTCD. Thus, the development of a validation file from the RTCD will not be discussed for the following measures:

- *Comparison of daily assigned VMT to HPMS-derived VMT for the region*. This comparison is independent of the validation file since the HPMS counts constitute an independent set of traffic counts. The HPMS counts are related to the RTCD only from the standpoint that they are used to supplement the RTCD. It is hoped that regional VMT estimates from the validation file are reasonably close to the regional VMT estimates derived from the HPMS. However, if this is not the case, the validation should hinge on comparisons to the regional validation file rather than the HPMS file.
- *Vehicle-class*. The incorporation of vehicle classification counts from the RTCD has not been investigated. Due to the impacts on traffic flow, roadway maintenance, and air quality, the development and validation of detailed truck travel models for the SEMCOG region is important. Nevertheless, this is a future task to be considered for the region. At the current time, emphasis should be placed on acquiring and recording traffic counts by vehicle class whenever possible.
- *Time-of-day average speeds by functional class and area type*. Speed data are not currently collected and stored in the RTCD. Speed collection and storage procedures should be developed in the future. Two types of speed data can be collected: spot speed data and travel time data. Spot speed data can be collected on a link-by-link basis and could reasonably be stored in the RTCD. However, spot speed data are not as useful for model validation as travel time data (over a collection of links). Travel time data are necessary to average the localized variation of speeds resulting from signal control, localized congestion, and other interruptions to traffic flow. The RTCD might not be the correct database for the storage of travel time data.
- *Weekend counts*. A weekend travel model has not been developed for the region. Weekend count data are available in the RTCD and, as a result, a weekend count validation file could be developed using the sampling procedures outlined above. However, at the current time, the only issue with weekend count data should be the collection and recording of weekend traffic counts whenever they are available.

COUNT DATA TO BE USED FOR VALIDATION

The Part 1 of this report suggested the use of "spring or fall" AWDT rather than AADT or annual AWDT for the model validation. This suggestion was based on the fact that models are generally developed based on data collected during the spring or fall rather than summer or winter. At the same time, the Part 1 report acknowledged that most planning for roadways was based on AADT or AWDT and that a validation to AADT or AWDT should also be performed. If large discrepancies were found in the validation to spring or fall AWDT and the validation to AADT, the Part 1 of this report suggested that the validation to spring or fall AWDT be the primary validation and that post-model adjustment factors of modeled volumes to AADT be developed.

An emphasis has been placed on the storage of raw counts in the RTCD along with the information necessary for factoring the raw counts to spring or fall AWDT, AADT, or annual AWDT as necessary. Table 20 summarizes the type of count data stored in the RTCD that were available for the 2002 validation dataset. Raw traffic count data were stored for all of the 9,216 available links with counts and 8,766 or 95 percent, of the links had the information necessary to factor the raw count data to AWDT or AADT.

Table 20. Traffic Count Storage Method for Links Included in the 2002 Analysis Count File

Storage Method	Links With Counts[1]		Unique Locations with Counts[1]	
	Number	Percent	Number	Percent
Two-way AADT counts	0	0%	0	0%
Directional AADT Counts	0	0%	0	0%
Raw daily two-way count w/o info for AADT calculation	358	4%	179	3%
Raw daily two-way count w/ info for AADT calculation	1,154	13%	577	11%
Raw time-of-day two-way count w/o info for AADT calculation	0	0%	0	0%
Raw time-of-day two-way count w/ info for AADT calculation	3,192	35%	1,596	31%
Raw daily directional count w/o info for AADT calculation	66	1%	34	1%
Raw daily directional count w/ info for AADT calculation	124	1%	64	1%
Raw time-of-day directional count w/o info for AADT calculation	26	0%	14	0%
Raw time-of-day directional count w/ info for AADT calculation	4,296	47%	2,762	53%
Total	9,216	100%	5,226	100%

[1] For links with counts, counts on one-way links each contributed one to the total and counts on two-way links each contributed two to the total. For unique locations, counts on one-way links and counts on two-way links contributed only one to the count total.

SEMCOG has developed monthly and day-of-week seasonal adjustment factors to estimate AADT from raw data counts. Day-of-week adjustment factors were developed for four groupings of day-of-week: Monday through Thursday, Friday, Saturday, and Sunday. Since the focus of this study is the development of a validation file for the regional travel model, only counts taken during the week should be used for the validation file; weekend travel is different by nature and should not be considered in the validation of a travel model focused on weekday travel.

VALIDATION UNCERTAINTY CAUSED BY VARIATION IN TRAFFIC COUNTS

As pointed out in a previous section, the RTCD comprises a sample of counts for the region. Sampling error will exist in the counts selected for the validation file. Thus, any estimates based on the count data should be stated with confidence intervals as determined using Equation 4, shown previously.

When estimates of observed data are produced for the validation, the model data should fall within the confidence interval to be considered valid. For example, if the urban business principal arterial VMT was estimated to be 30,000 ±3,000 (10 percent) at the 90 percent confidence level, the model should be considered "valid" for that measure if the modeled VMT is in the range 27,000 to 33,000.

A NEW VALIDATION PARADIGM

As discussed in previous sections, traffic counts used for travel model validations are samples and, as such, are subject to sampling error. Resources precluded the development of a new validation paradigm that explicitly considers and accounts for the sampling error inherent in the traffic count data. Nevertheless, the concepts of the new paradigm are discussed below for consideration when time and resources permit. The procedures outlined in the new paradigm are a departure from traditional validation procedures such as those described in other sections of this report. SEMCOG should develop the procedures only if the potential benefits of using the alternative procedures outweigh the costs of developing the procedures. The primary benefit would be a possible relaxation of validation criteria through the explicit consideration of sampling error in the traffic count data. If acceptable model validations are achieved using the traditional procedures outlined elsewhere in this report, little benefit would be realized from developing the new validation procedures outlined below.

The new validation paradigm is based on a desire to use statistics in validation procedures. The new paradigm focuses on the statistical significance of differences in the mean values for the various validation strata. Two different types of statistical tests are considered:

- The first type of test determines the significance of the differences in the mean values from two different samples. This test would be implemented by drawing a sample of traffic counts for various strata such as functional class and area type. A second sample of assigned traffic volumes would also be selected for the same strata (but not necessarily the same links). The means and standard deviations of the samples of traffic counts and assigned volumes would then be used to determine the statistical significance of the differences of the mean sample values for the various strata. If the means for the two samples were not significantly different, model validation would be assumed.
- The second type of test employs paired t-tests of traffic counts and assigned traffic volumes to determine the statistical significance of the differences. It is implemented by drawing a sample of traffic counts for different strata such as functional class and area type. Assigned traffic volumes for the same set of links are also summarized so that the paired t-tests can be performed. This test, in effect, considers the observed traffic volumes to represent a "pre-treatment" measure and the assigned traffic volumes to represent a "post-treatment" measure.

The new validation paradigm is most appropriate for use with the model-based sampling plan for future validations since that plan is more solidly based on random sampling techniques.

The first type of test, the comparison of the differences in mean values, should be used to compare measures such as VMT by different strata. The results of the statistical test would lead to validation statements such as, "the total count VMT and assigned VMT are not statistically significantly different by functional class and area type." For example, the average daily count per lane for urban freeways might be determined to be 20,100 ±1,000 at the 95 percent confidence level while the average assigned volume per lane might be 18,700 ±700 at the 95 percent confidence level. If the samples represented links with an average of, say, 4.3 lanes and an average length of 1.2 miles, the count sample would represent 103,716 VMT while the sample of assigned volumes would represent 96,492 VMT. Traditional validations might deem the apparent seven percent under-estimation of VMT by the model as unacceptable. Yet, based on sample statistics, the difference between daily volumes per lane (1,400 vehicles) might not be statistically significant and the seven percent under-estimation of VMT could be deemed to be not statistically significant.

The second type of test, the paired t-tests of counted traffic volumes and assigned traffic volumes, would be used for a more direct comparison of assignment results with traffic counts. The results of the second type of statistical test would lead to validation statements such as, "the differences between the assigned and observed traffic by functional class and area type are not statistically significant." For example, suppose a sample of five links were selected with observed counts and assigned traffic volumes as shown below:

Link	Count	Assigned Volume	Difference
1	10,500	10,000	-500
2	9,300	9,800	500
3	12,200	11,700	-500
4	14,100	14,600	500
5	13,000	12,500	-500

The paired t-test for the above results in a t-score about 0.4; the differences in traffic counts and assigned traffic volumes above are significant at only the 0.7 level of significance. Typically, a t-score of 1.96 or higher, or a level of significance of 0.05 or less is considered statistically significant. Note that if all of the assigned volumes were 10 vehicles greater than the counts, the differences would be highly significant since they would all be relatively uniformly greater than the counts. The paired t-test can be easily applied using Microsoft Excel.

ASSESSMENT OF SEMCOG RTCD FOR VALIDATION FILE CREATION

DEVELOPMENT OF VALIDATION DATABASE

Three of the high priority work items for Part 2 for the Traffic Count File Development For TransCAD Model Calibration and Validation project are:

- Develop a statistically sound sampling plan for count collection/compilation, considering balances on spatial distribution of samples, such as county, functional class, area types, etc.

- Identify additional data needs, based on the quality and quantity of available existing data.
- Determine categories and the appropriate number of counts required, including specifications of a set of weekday traffic counts, statistically sound, that will facilitate testing the desired validation criteria.

The above three tasks have been discussed elsewhere in this report. The performance of the above tasks required a dataset that combined information from the TransCAD network dataset and the Regional Traffic Count Dataset (RTCD). This dataset was developed using 2002 TransCAD network data. The creation of comparable validation file datasets for future years will be required. This section describes the creation of the existing validation dataset with the expectation that the procedures will be used for the creation of future validation datasets.

Two issues are important regarding the specification of the validation dataset:

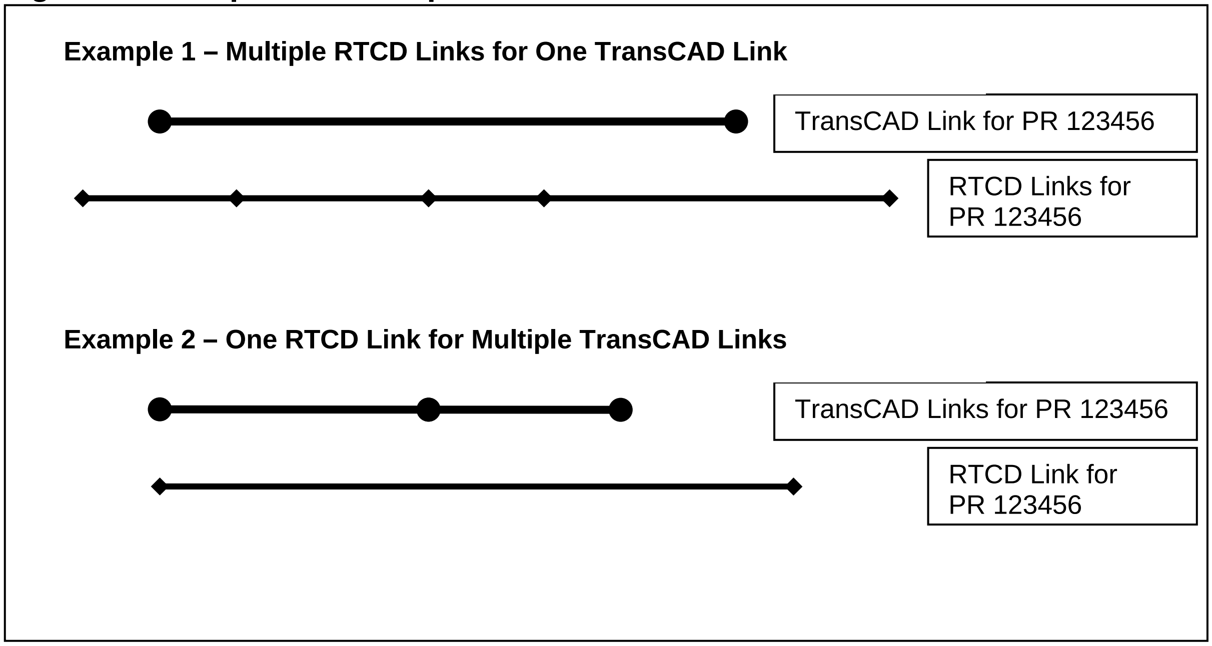
- The records to be included in the combined datasets
- The data items to be included in the combined datasets.

The data records to be included in the combined dataset should be the *union* of the TransCAD dataset and the RTCD (see the Venn diagram shown in Figure 4). One record from each dataset should be included. In cases when the network record exists in both datasets, data from each dataset should be included. If the record exists in only one of the datasets, the appropriate data from the dataset will be include and all other data for the record (that would have been obtained from the other dataset) should be filled with dummy or null values.

While the above concept of one record for each link is easy to define, it is not as easy to accomplish due to the often competing goals for the RTCD and the specific needs of the TransCAD network dataset. While both datasets are maintained using GIS principles and were originally derived from the same GIS base, the networks are, in fact, different. Issues affecting the merging of the RTCD data onto the TransCAD network dataset include:

- The only common identifier for both the RTCD and the TransCAD network is the data item, PR, or Physical Route. Unfortunately, PR does not provide a unique identifier for specific links. Beginning and ending mile post (BMP and EMP) must also be known to uniquely identify a link in the RTCD. While a corresponding section of roadway might be included in the TransCAD network, there is no requirement that it be defined by the same BMP and EMP. Figure 5 illustrates the lack of correspondence of between RTCD links and TransCAD network links for a common section of roadway. As can be seen in Figure 5, one TransCAD link might cover multiple RTCD links or portions of multiple RTCD links. Likewise, one RTCD link might cover multiple TransCAD links.
- While traffic counts are, by definition, point data, current RTCD standards are to post count data to links. Thus, in cases where one TransCAD link covers multiple RTCD links and, possibly, multiple counts, it can be difficult to determine which count is most appropriate for the TransCAD link. Conversely, if one RTCD link covered multiple TransCAD links, it would be difficult to determine which TransCAD link should be assigned the count.
- Each RTCD link might have several associated counts. These counts might include directional counts, counts for various years, or multiple counts taken the same year using different count procedures (i.e. link counts or intersection counts).

Figure 5. Example of Correspondence between RTCD and TransCAD Networks



GROUND RULES FOR POSTING RTCD COUNTS ON TRANSCAD LINKS

Coding of Counts to Points

Perhaps the most important change to the RTCD to improve the posting of traffic counts on TransCAD network links would be the posting of locations of the counts for each count. In effect, each count would become a point count. Within the existing RTCD, this can be accomplished by posting the coordinates of the count or the count milepost on the RTCD count record rather than modifying the structure of the RTCD to make count information "point data."

The posting of exact count locations on existing counts in the RTCD is impossible. Exact count location information was probably not recorded for most counts and would be impossible to locate the exact locations of the counts retroactively. In these cases, it is suggested that the count location be estimated as the mid-point of the RTCD link on which the count is posted. For future counts, provisions should be made and standards established for the recording the count location (e.g. by milepost) on the actual count record.

Identification of Directional Counts

The RTCD is a "two-way" network file. In other words, a link for a specific PR between mileposts "a" and "b" is not matched by a companion link between mileposts "b" and "a." Several possibilities exist for the coding of directional data:

- Directional counts by time-of-day are available for the link with directional information being posted as -1 for "b" to "a," 0 for two-way, or 1 for "a" to "b." This is the most general situation; data could be used for the generation of time-of-day specific TransCAD validation files.
- Daily directional counts are available for the link with directional information being posted as noted above.

- Only daily two-directional counts are available. Directional counts for TransCAD validation files must be estimated as one-half of the daily count.

A simple code for the identification of the type of directional count has been created and posted in the validation file (the code is the number identifying each of the ten general categories of links listed in the next section). This code allows for an analysis of the directional and time-of-day counts available and, in the future, should simplify the creation of time-of-day TransCAD validation files. Due to concerns about the direction codes in the current RTCD, only daily TransCAD validation files will be created.

Identification of Type of Count Factoring

Ten general categories of count data are available in the RTCD:

1. Two-way AADT counts
2. Directional AADT counts
3. Raw daily two-way count data without the necessary information to factor to AADT counts
4. Raw, daily two-way count data with all necessary information for factoring to AADT counts
5. Raw time-of-day, two-way count data without the necessary information to factor to AADT counts
6. Raw, time-of-day, two-way count data with all necessary information for factoring to AADT counts
7. Raw daily directional count data without the necessary information to factor to AADT counts
8. Raw, daily directional count data with all necessary information for factoring to AADT counts
9. Raw time-of-day, directional count data without the necessary information to factor to AADT counts
10. Raw, time-of-day, directional count data with all necessary information for factoring to AADT counts

The numbering above provides a hierarchy of detail from lowest level of information to highest level. The numbering can be used as a code to identify the type of count information provided. Typically, the highest level of count information represented by category 10, above, is desired. The detailed data provides the means necessary to use SEMCOG specific data for factoring from raw counts to the “average” counts required for validation such as AADT counts or, if desired, AWDT counts representative of spring or fall for the validation year.

Multiple Counts from Permanent Traffic Recorders and MITS Sites

During the course of the project, an issue regarding the handling of multiple counts from PTR and MITS sites. Specifically, traffic counts are available for every day of the year for those sites. To make matters more complicated, the data are available for each entire day by very small time intervals – in some cases, to the level of minute-by-minute count data for each lane counted. In addition to severely complicating the decision of which count to select for the validation dataset, the amount of data available could overwhelm the RTCD. To address this problem, the following actions have been suggested:

- Data from all PTR and MITS sites should be acquired maintained separately from the RTCD. In this way, data will be available for the calculation of factors to convert raw traffic count data to AADT or AWDT.
- One count for each PTR and MITS site should be posted in the RTCD for each year. A random weekday that is not a holiday or a day prior to or after a holiday should be selected for each site. The random day should be selected independently for each site. To simplify the selection, the day can be picked from the March through June or September through October time frames.
- Count data from the PTR and MITS sites should be aggregated to a minimum of 15-minute time periods prior to posting in the RTCD.

BUSINESS RULES FOR DEVELOPING A TRANSCAD COUNT FILE

The following provides a suggested set of decision rules for developing the validation file. The rules are based on the assumption that the ground rules outlined above are in place. The following business rules should be followed sequentially:

1) Determine RTCD links / records covered fully or partially by a TransCAD link
2) Select all RTCD links / records with counts for each TransCAD link
3) From the set of RTCD links / records with counts, select the RTCD link(s) / record(s) with a count for the year closest to the desired year
4) Initially break ties (to selectively include links / records) resulting from Rule 3 by selecting the link(s) with the highest type of count codes (as listed in the previous section)
5) If multiple links remain, select the RTCD link(s) / record(s) with a mid-point closest to the mid-point of the TransCAD link.
6) If ties still remain, select one of the remaining candidate RTCD links / records at random.

COMBINED COUNT FILE ANALYSIS DATASET

Table 21 summarizes the data items contained in the combined count file analysis dataset. The dataset is stored in Access format for easy statistical analysis. It includes all links from the RTCD and the TransCAD network and is, in effect, the union of those two datasets. The file can be used as the sample-based validation dataset. Sampling procedures outlined previously in this report should be used with TransCAD networks for future years (e.g. 2006) to determine links to be included in the validation process. Then, after the links are counted and added to the RTCD, the model-based validation datasets.

COUNTS BY AREA TYPE BY FUNCTIONAL CLASS BY COUNTY

A substantial number of counts by area type by functional class by county exist in the RTCD. Previous sections of this report summarized the numbers of the independent counts available for a validation file and the number of counts that would be required for reasonable levels of statistical significance for various levels of stratification. Those analyses showed that insufficient count data existed for stratifying the validation count file by area type by functional class by county. Thus, the lowest level stratification suggested is area type by functional class. However, the procedures designed to ensure that reasonable samples of counts are selected from each county should be used whenever samples of links by area type and functional class.

Table 21. Validation Dataset Specification

Data Item	Format	Source	Notes
TRANSCAD_ID	Number	TransCAD	SEMCOG's TransCAD ID Number
PR_NUM	Number	TransCAD	Physical Route
BEGIN_MP	Number	TransCAD	Beginning Mile Post
END_MP	Number	TransCAD	Ending Mile Post
MID_MP	Number		Mid-point Mile Post Calculated from Above Two Fields
FENAME	Text	TransCAD	Facility Name
FETYPE	Text	TransCAD	
FUNCLASS	Text	TransCAD	Functional Class
AREATYPE	Text	TransCAD	Area Type
COUNTY	Text	TransCAD	County Name (or City of Detroit)
LENGTH	Number	TransCAD	Link Length in Miles
LANES_AB	Number	TransCAD	Number of Lanes from TransCAD A-node to B-node
LANES_BA	Number	TransCAD	Number of Lanes from TransCAD B-node to A-node
ESTIMATED_AB	Number	TransCAD	Assigned Traffic Volume from A-node to B-node
ESTIMATED_BA	Number	TransCAD	Assigned Traffic Volume from B-node to A-node
AB_DIRECTION	Number	TransCAD	1 = AB Follows Traffic Direction -1 = AB Opposes Traffic Direction 0 = Two-way Link
MGFV	Number	TransCAD	
SEM_ID	Number	RTCD	SEMCOG's RTCD ID Number
CLMP	Number	RTCD	Count Location Mile Post
COUNT_ID	Text	RTCD	SEMCOG's RTCD Count ID Number
ST_DATE	Date	RTCD	Start Date for Count
ST_TIME	Date	RTCD	Start Time for Count
Type_of_Count	Number	VFC[1]	Type of Count [2,3]
OBSERVED_AB	Number	RTCD	24-hour Count from A-node to B-node
OBSERVED_BA	Number	RTCD	24-hour Count from B-node to A-node
SAMPLED	Yes/No	VFC[1]	Whether the links are sampled for validation

1 VFC is the Validation File Creator program
2 Data item calculated from RTCD data.
3 Type of count is a 1-10 code that identifies whether the count is based on AADT or raw data, directional counts or two-way counts, and daily or time-of-day counts.

SCREENLINES

Figure 6 shows the screenlines currently in use for model validation in the SEMCOG region. As noted in the Model Validation and Reasonableness Checking Manual[10]:

> *Screenlines* typically extend completely across the modeled area and go from boundary cordon to boundary cordon. For example, a river that passes completely through the area makes an excellent screenline. Travel demand that goes from one side of the river to the other must cross this river screenline within the study area boundary. Screenlines are often associated with physical barriers such as rivers or

10 Model Validation and Reasonableness Checking Manual, prepared for the Federal Highway Administration, Travel Model Improvement Program by Barton-Aschman, Inc. and Cambridge Systematics, Inc., February 1997.

> railroads, however jurisdictional boundaries such as county lines that extend through the study area make excellent screenlines.

Unfortunately, there are no major rivers or other natural boundaries in the SEMCOG region that can be used to create screenlines as noted above. However, the validation manual also states:

> If multiple counties are included in the modeled area, then each county boundary can form either a cordon or screenline, dependent upon its location within the area.

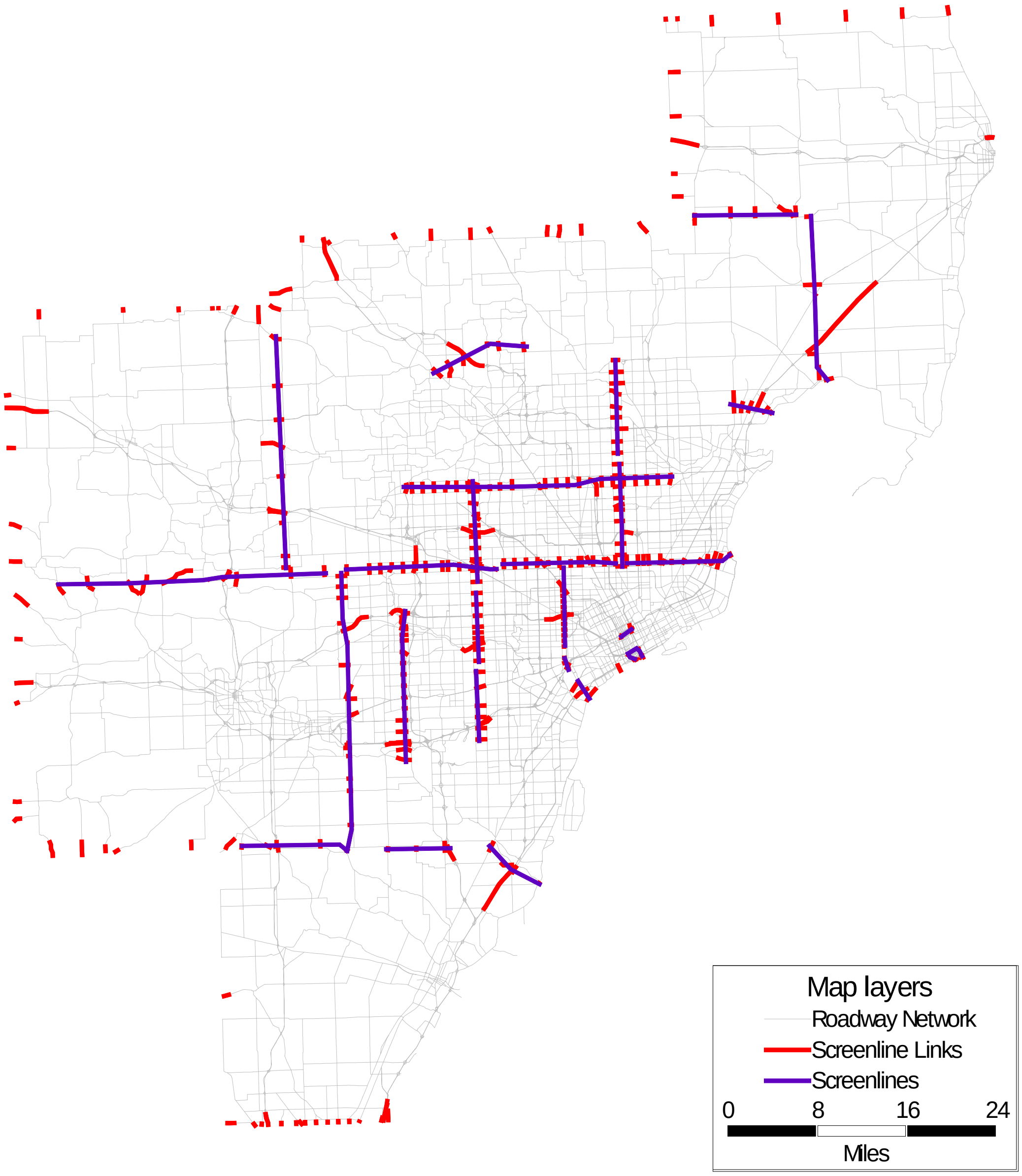
The screenlines and cutlines[11] defined for the SEMCOG region nicely define county boundaries. Most of the county boundary cutlines have good coverage of the crossing links with only several appearing to require additional links. For example, the cutlines between Washtenaw and Wayne counties and between Wayne County and the City of Detroit appear to be missing several crossing links and the cutline between Oakland and Macomb counties should be extended to the regional boundary on the north.

While the county boundaries are well defined, additional cutlines within counties would be useful for model validation. The following provides some examples of cutlines and sets of cutlines that would be useful for model validation and diagnosis, especially for subarea and corridor level planning:

- In Washtenaw County, define a set of three or four cutlines around Ann Arbor. The freeways ringing the city provide a good manmade boundary.
- In Wayne County, add a cutline from the southern end of the central north-south screenline to Lake Erie on the southeast.
- In Macomb County, extend the east-west cutline from just north of State Road 59 west to the county boundary with Oakland County.
- In Oakland County, extend the east-west cutline north of Pontiac to both the east and west county boundaries; extend the southernmost east-west cutline west to the border with Livingston County; add a north-south west of Pontiac cutline from the boundary with Wayne County to the regional boundary on the north.
- In St. Clair County, create a cordon cutline around Port Huron.
- In Monroe County, create an east-west cutline along the Raisin River from the Lenawee County boundary to Lake Erie.

[11] According to the Model Validation and Reasonableness Checking Manual, *cutlines* extend across a corridor containing multiple facilities. They should be used to intercept travel along only one axis.

Figure 6. Screenlines and Cordon Lines for the SEMCOG Region



- In the City of Detroit, add two or three short cutlines to form a complete north-south cutline paralleling Interstates 75 and 375; add two or three short cutlines to form a complete cutline paralleling the Ambassador Bridge and Interstates 75 and 96 from the Detroit River to City of Detroit border on the west.

The addition of the cutlines should provide a good basis for the analysis of intra-regional flows. Such analysis is useful for determining the reasonability of the trip distribution models and the need for K-factors in the trip distribution (if the gravity model continues to be used for distribution).

As noted in the Part 1 of this report, the screenline validation file should be separate from the main validation file. Note, however, that links included in the RTCD can be used to populate the main validation file.

QUALITY ASSURANCE / QUALITY CONTROL

REVIEW OF QA/QC PROCEDURES USED IN OTHER AREAS

QA/QC procedures for traffic count programs have been suggested for count programs by the Federal Government and in a number of studies. QA/QC procedures specified in four different sources were reviewed:

- *Traffic Monitoring Guide* prepared by the US Department of Transportation, Federal Highway Administration, Office of Highway Policy Information, May 1, 2001;
- Wright, Hu, and Young, *Variability in Traffic Monitoring Data Final Summary Report*, prepared by Oak Ridge National Laboratory for the US Department of Energy, August 1997;
- Flinner and Horsey, *Traffic Data Edit Procedures Pooled Fund Study*, prepared for the US Department of Transportation, Federal Highway Administration, Office of Highway Policy Information, undated; and
- Correspondence from Larry Whiteside describing QA/QC procedures used by the Michigan Department of Transportation for the traffic count program.

TRAFFIC MONITORING GUIDE QA/QC PROCEDURES

The Traffic Monitoring Guide (TMG) is available in Adobe Acrobat® PDF format on the FHWA, Office of Policy Information website. The QA/QC procedures in this document focus on portable short duration and permanent continuous count programs. The procedures focus more heavily on the actual performance of the traffic counts rather than the verification of the resulting counts. The TMG makes the following recommendations for count programs:

- Vehicle classification counts should be performed for about 25 to 30 percent of all traffic counts performed using portable, short duration traffic counters.
- Data from permanent traffic counters are crucial for the development of day-of-week and month-of-year adjustment factors and for the development of growth factors for counts performed using portable, short duration traffic counters.
- Short duration traffic counts should be 48 hours in duration with hourly recording of traffic volumes.
- Short duration traffic counts should be adjusted for axles (if necessary), day-of-week, and month-of-year to produce AADT estimates.

Several recommendations are made in the TMG regarding the QA/QC procedures for count procedures and the analysis of the traffic counts:

- The accuracy of traffic count equipment should be verified by producing a simultaneous manual count.
- There should be no missing hours for counts collected using the short duration, portable or permanent traffic counters.

- For 48-hour counts, total counts for one day should be compared to the second day and “unusual circumstances” should be reviewed to ensure that the counts were not affected by equipment failure, a special event, or an incident.
- Formal rules should be established for the exclusion of traffic counts from the count program to maintain “truth in data” standards. Traffic counts should not be excluded solely on the basis, “they look bad.”

VARIABILITY IN TRAFFIC MONITORING DATA

This study was primarily focused on the sources of variation in traffic counts that adversely impact estimates of AADT. The study attempted to identify the sources of variability in traffic count data, characterize the variability, and determine the impacts of the variability on the precision of AADT estimates produced using the traffic count data. The study summarized a number of findings, six of which are important to the SEMCOG RTCD and the development of TransCAD validation files:

- Weekend days, winter months, and holidays contribute substantially to the variability observed in traffic data. These results confirm the need to annualize short-term monitored data by day-of-week and month-of-year factors.
- Even after accounting for the impact of weekends and holidays, day-of-week variance in traffic volumes is significant. Adjustment of short-term traffic count data by weekday and weekend factors, at the minimum, is essential.
- Even after accounting for the impact of winter months and holidays, monthly variation in traffic volumes, especially by vehicle class, is significant. Adjustment of short-term traffic count data to account for monthly variation is essential.
- More commonality results in less variability in traffic volumes. This can be seen both in total volumes and in volumes by vehicle classes. High volume traffic estimates have low associated coefficients of variation while low volume traffic estimates have associated high coefficients of variation. This fact is borne out somewhat by the data summarized in Table 14. This finding is especially important for classification data, suggesting that less common vehicle classes be combined for reliable AADT estimates.
- Traffic differs significantly by direction. Traffic monitoring needs to be conducted in both directions.
- Adjustment factors based on data from continuous traffic counters for similar functional classes work reasonably well for expanding short-term traffic counts to AADT estimates.

TRAFFIC DATA EDIT PROCEDURES POOLED FUND STUDY

The Traffic Data Edit Procedures (TDEP) pooled fund study SPR-2 (182) began in 1995 when fourteen states and the Federal Highway Administration (FHWA) expressed the desire to work together to:

- Learn from other experts in the field who are in charge of the operation and data quality control of traffic data collection equipment
- Document and qualify existing and potential traffic data screening methods
- Refine the list of traffic data screening methods into a logically integrated "rule base"
- Test the rule base in the context of actual data and expert analysis
- Develop software to assist in the evaluation of the rule base and to put revised software into production

The first three objectives of the study were fully met. A prototype version of software intended to test the effectiveness of the coded portions of the rule base was developed. The resulting prototype was successfully run in September 1999 on a data set provided by the Minnesota Department of Transportation. However, lack of funding prevented the full development and testing of the software. While the rule base developed in this effort represented the consensus expertise of traffic analysis experts from around the country, it was not validated by rigorously testing it against actual data from a spectrum of traffic monitoring sites. Prior experience with knowledge-based systems suggests that such testing would have resulted in further modification of the rule base and a better understanding of the types of traffic monitoring sites where the rule base would be most applicable.

While the study was not fully completed and the rule base was not fully tested, the TDEP study provides a number of useful rules for the checking of traffic counts included in the RTCD. The prototype rule base was copied from the FHWA, Office of Highway Policy website and has been included in Appendix C. Suggested QA/QC rules for the RTCD will be listed in a subsequent part of this report.

QA/QC Procedures for the MDOT Traffic Count Program

Larry Whiteside (MDOT) provided a brief description of the QA/QC procedures used by the MDOT for their traffic count program (see Appendix D). The MDOT procedures are based on the TMG and, for this reason, focus mainly on count procedures. However, manual QA/QC procedures are used for the evaluation of traffic counts:

> The counts and classifications are manually evaluated for quality. This is done by creating a report which compares the AADT as generated from the data against the previous year's AADT. These reports use the monthly and daily seasonal adjustment factors to generate a preliminary AADT based on the data. These reports are manually reviewed for both distribution of the traffic, abnormalities, and change against previous years AADT. With classification data, a similar procedure is done in which the axle adjustments are compared to last year factors as well as a computed AADT.

Review of SEMCOG RTCD QA/QC Procedures

Count Program and Procedures

The RTCD is a central repository for traffic count data collected in the SEMCOG region. Count data are provided to SEMCOG by the seven counties in the region, the City of Detroit, and MDOT. Since the data are provided by different agencies, there can be variations in collection methods, collection frequency, data recorded, local QA/QC procedures, and naming conventions. As documented above, MDOT uses QA/QC procedures based on the TMG. Some standardization in count procedures, data recorded, naming conventions, and QA/QC procedures have been achieve through the relatively widespread use of the TCDS by the seven counties and the City of Detroit.

Frequency of Count Collection

Since SEMCOG does not control or administer the count programs in the region, standards for frequency of count collection and other procedures must be obtained through voluntary compliance with SEMCOG standards. While there is a desire by the counties and local governments to perform counts on a regular basis, local needs and budgets can affect the frequency and duration of traffic counts.

ASSESSMENT OF SEMCOG AADT / AWDT FACTORING PROCESS

Table 22 shows the monthly adjustment factors, the Monday through Thursday adjustment factors by month, and the Friday adjustment factors by month estimated for the SEMCOG region using data from the 29 permanent traffic recording (PTR) stations in the region. Table 22 also shows composite adjustment factors for the Monday through Thursday by month and Friday by month. The composite factors are simply the products of the monthly adjustment factors and the day-of-week adjustment factors by month. The AADT adjustment factors shown in Table 22 provide the best adjustment factors available for the region based on currently available data.

Table 22. Seasonal Adjustment Factors for AADT Estimates

Month	Monthly Factor	Monday – Thursday Factor	Friday Factor	Composite Monthly & Day-of-Week Factors	
				Monday – Thursday Factor	Friday Factor
January	1.22	1.13	1.28	1.38	1.56
February	1.14	1.07	1.21	1.22	1.38
March	1.08	1.03	1.02	1.11	1.10
April	1.05	1.00	0.97	1.05	1.02
May	1.02	0.98	0.97	1.00	0.99
June	0.97	0.95	1.19	0.92	1.15
July	1.01	0.97	1.22	0.98	1.23
August	0.96	0.93	1.05	0.89	1.01
September	1.00	0.98	0.99	0.98	0.99
October	1.02	0.97	1.04	0.99	1.06
November	1.04	0.99	1.02	1.03	1.06
December	1.16	1.10	1.11	1.28	1.29

The PTR data were not re-analyzed to verify the adjustment factors in Table 22. However, the monthly factors sum to 12.67 which suggests that the average monthly AADT adjustment factor is about 1.056. Procedures listed in the *Traffic Monitoring Guide*[12] should be used to verify the factors. Since the average monthly factor should, logically, be 1.0, SEMCOG should verify the calculations of the monthly adjustment factors.

SEMCOG should develop comparable sets of adjustment factors to estimate annual AWDT and spring / fall (March through May and September through November) AWDT from the 29 PTR stations in the region. As long as only 29 PTR stations are available in the region, there should not be an attempt to subdivide the adjustment factors by area type or functional class since too few data points would exist for each stratum. The three sets of factors should be used with the

[12] *Traffic Monitoring Guide*, Section 2 – Introduction to Traffic Monitoring, U.S. Department of Transportation, Federal Highway Administration, Office of Highway Policy Information, May 1, 2001. This document can be downloaded for free in pdf file format from the U.S. Department of Transportation website.

2002 validation file to develop estimates of AADT, annual AWDT, and spring / fall AWDT for the region.

TCDS COUNT DATA ENTRY PROCEDURES

The current version of the TCDS includes four QA/QC tests:

1. Have complete count date and time data been recorded?
2. Has location identification information (i.e. the count location description) been recorded?
3. Have count data for a full 24 hours been recorded?
4. Do any of the count intervals have a null value?

The TCDS prohibits automatic saving of the count if any of the above criteria are violated. More recent versions of TCDS do not allow saving of incomplete data. However, the above tests were not included in older versions of the software. In addition, while it is not recommended, some users have imported data directly into the TCDS database, completely bypassing the interface that performs the above checks. Therefore, incomplete count data can be imported to the RTCD even though the data were processed using the TCDS.

A range of procedures for reducing or eliminating the import of incomplete data can be considered by SEMCOG. The procedures range from simple “arm twisting” to the development of more formal and costly procedures that reproduce the TCDS data checks when TCDS data are imported into the RTCD. Since the extent of the problem is not yet known, the communication of the need for clean, complete data to member agency personnel collecting and inputting the TCDS data is the recommended approach for addressing the issue.

If incomplete TCDS data is identified as a major problem in the future, the TCDS QA/QC tests should be reproduced in software used to import the TCDS data into the RTCD (see Other RTCD QA/QC Procedures, below). The software could be designed to report frequency of rejected data for each agency providing TCDS data. The need for remedial training of agencies and personnel with abnormally high rates of rejects or missing data can be determined on an ad hoc basis. This approach is based on the philosophy, “if it isn’t broken, don’t fix it.”

OTHER RTCD QA/QC PROCEDURES

SEMCOG has performed a number of ad hoc data checks in an effort to correct and clean data included in the RTCD. Examples of these checks include:

- Standardization of location descriptions.
- Ad hoc tests on TCDS data imported to the RTCD

The ad hoc TCDS checks are performed by SEMCOG are independent from the TCDS QA/QC procedures programmed into the TCDS software. The checks are performed on all TCDS imported into the RTCD. When TCDS files are provided to SEMCOG, they are converted into an independent MS Access database. SEMCOG personnel code the locations with the appropriate PR numbers and milepoints, and runs the data validation checks described in Table 23. After the checks are complete and any errors are corrected, the data are imported into the RTCD.

Table 23. QA/QC Checks Performed on TCDS Data Prior to Entry into the RTCD

1. Is each field in each of the three tables (location, count, and gis_link) of the appropriate type and length?
2. Is LOCAL_ID a valid primary key for the location table?
3. For each record in the location table, does SOURC_ID contain the correct value?
4. For each record in the location table, does ON_ROAD contain a non-null value?
5. For each record in the location table, does DIR contain a value of either 2-WAY, EB, NB, SB, or WB, or RAMP?
6. For each record in the location table, does TYPE contain a value of either I-SECTION or LINK?
7. For each I-SECTION location record, does AT_ROAD contain a non-null value?
8. For each I-SECTION location record, do FROM and TO contain null values?
9. For each I-SECTION location record, does APPROACH contain a value of either EAST OF, NORTH OF, SOUTH OF, or WEST OF?
10. For each LINK location record, do FROM and TO contain non-null values?
11. For each LINK location record, do APPROACH and AT_ROAD contain null values?
12. Is each record unique when grouped by ON_ROAD, DIR, FROM, TO, APPROACH, and AT_ROAD?
13. For each I-SECTION location record, does the direction in APPROACH exactly match or directly oppose the direction in count?
14. Does each record in the count table have a corresponding record in the location table (when linked via LOCAL_ID)?
15. For each record in the count table, do ST_DATE and ST_TIME contain non-null values?
16. For each record in the count table, is END_DATE exactly one date after ST_DATE?
17. For each record in the count table, is END_TIME exactly equal to ST_TIME?
18. For each record in the count table, does INTERVAL contain a value of 15 or 60?
19. For each record in the count table, does each hourly field (H01, H02, etc.) contain non-null values?
20. For each record in the count table, does the value in TOTAL exactly equal the sum of the hourly values?
21. For each count-table record with INTERVAL equal to 60, do the 15-minute fields (H01_1, H01_2, H01_3, H01_4, H02_1, etc.) all contain null values?
22. For each count-table record with INTERVAL equal to 15, do the 15-minute fields all contain non-null values?
23. For each count-table record with INTERVAL equal to 15, does each hourly field exactly equal the sum of its four 15-minute fields?
24. Does each count location description conform to SEMCOG's standards?
25. Does each record in the gis_link table have a corresponding record in the location table (when linked via LOCAL_ID)?
26. For each record in the gis_link table, does MGFV contain the correct value?
27. For each record in the gis_link table, do PR and BMP together identify a specific node in the MGF (not always necessary, though)?
28. For each record in the gis_link table, do PR and EMP together identify a specific node in the MGF (not always necessary, though)?
29. For each record in the gis_link table, is BMP less than EMP?
30. For each record in the gis_link table, does the linear referencing information match the location description (is the count coded to the correct location)?

RECOMMENDED RTCD QA/QC PROCEDURES

Table 24 recommends a set of formal QA/QC procedures to be incorporated into the maintenance of the RTCD. The recommendations are based on the reviews of QA/QC procedures documented above and focus primarily on the veracity of the data. The tests can be coded into software used to input data into the RTCD or to check existing RTCD counts. While the effort is somewhat redundant, some checks might also be coded into the TCDS software. However, there would be no guarantee that local agencies would use the updated software and would not bypass the checks; thus, the checks should not be coded solely in the TCDS software.

Table 24. Rules for Inclusion of RTCD Data in TransCAD Validation Data Set

QA/QC Procedure	Error Level[1]	Database for Check[2]	Comment
Date is not correct or unique	Error	TCDS & RTCD	If the date is missing, the traffic count cannot be adjusted.
Date is a major holiday	Error	Validation	Counts taken on New Year's day, Memorial Day, July 4th, Labor Day, Thanksgiving, Christmas and, possibly, other holidays are by nature abnormal and should not be factored to AADT
Extreme hourly volume per lane	Warning	RTCD & Validation	The hourly volume in any lane will be reported as anomalous if exceeds the estimated capacity per lane used in the TransCAD modeling process (as determined from a look-up table stratified by area type and functional class).
1:00 AM to 2:00 AM Volume vs. 1:00 PM to 2:00 PM Volume	Warning	TCDS & RTCD	1:00 AM to 2:00 AM volumes greater than the 1:00 PM to 2:00 PM volumes of the same day are unlikely occurrences.
Consecutive hourly zero volumes between 5 AM and Midnight	Warning	TCDS & RTCD	Consecutive zero volumes on links outside the hours midnight to 5 AM might suggest counter failure.
Consecutive hours with the same non-zero volume	Warning	TCDS & RTCD	Consecutive hours with the same non-zero volume might suggest counter failure.
Daily directional split for links with directional counts outside the 60/40 range.	Warning	TCDS & RTCD	For most two-way links with directional counts, daily volumes should be reasonably symmetrical.
For 48-hour counts, percent difference in hourly volume between day 1 and day 2 is greater than 10 percent.	Warning	TCDS & RTCD	Substantial differences in hourly volumes on consecutive days might suggest counter failure or roadway incident (e.g., upstream or downstream accident).
For 48-hour counts, percent difference in daily volumes adjusted to AADT is greater than 10 percent.	Warning	RTCD	Substantial differences in AADTs for consecutive days estimated from counts might suggest counter failure or roadway incident (e.g., upstream or downstream accident) or error in AADT factoring.
Full day of data does not exist.	Error	TCDS & RTCD	For 24 or 48 hour counts, count information must be provided for either 24 or 48 hours if hourly data are provided.

"Best" daily AADT more than 10 percent different than "second best" daily AADT for same link.	Warning	Validation	For links where AADT can be estimated from more than one count, the best and second best counts should be determined. If second best counts are for a different year, they should be "grown" to the same year as the "best" count. Differences of more than 10 percent between AADTs determined from the counts suggest possible anomalous counts.
"Best" daily AADT more than 10 percent different than "second best" daily AADT for adjacent links.	Warning	RTCD & Validation	If AADT can be determined from two different count types for the same PR number, differences of more than 10 percent between AADTs determined from the counts suggest possible anomalous counts.
For 48-hour counts, percent difference in daily volumes adjusted to AADT is greater than 10 percent.	Warning	RTCD	Substantial differences in AADTs for consecutive days estimated from counts might suggest counter failure or roadway incident (e.g., upstream or downstream accident) or error in AADT factoring.
Full day of data does not exist.	Error	TCDS & RTCD	For 24 or 48 hour counts, count information must be provided for either 24 or 48 hours if hourly data are provided.
"Best" daily AADT more than 10 percent different than "second best" daily AADT for same link.	Warning	Validation	For links where AADT can be estimated from more than one count, the best and second best counts should be determined. If second best counts are for a different year, they should be "grown" to the same year as the "best" count. Differences of more than 10 percent between AADTs determined from the counts suggest possible anomalous counts.
"Best" daily AADT more than 10 percent different than "second best" daily AADT for adjacent links.	Warning	RTCD & Validation	If AADT can be determined from two different count types for the same PR number, differences of more than 10 percent between AADTs determined from the counts suggest possible anomalous counts.
Observed AB and BA directional counts inconsistent with TransCAD link directional coding.	Warning	Validation	TransCAD links are coded with -1, 0, or 1 directional code to designate one-way B-to-A, two-way, or one-way A-to-B links. Observed AB and BA directional counts must be consistent with TransCAD directional codes.
Print out links 100 highest absolute differences between count & TransCAD assigned volume	Information	Validation	While this type of analysis is typically part of the validation process with an emphasis on determining errors in the modeling process, it can also be used to uncover anomalous traffic counts.

[1] Error levels are as follows:
Information is provided for optional analysis review.
Warning is issued when an analyst should review information for possible error.
Error is issued when there is a definite error with the count data.

[2] Database for check means that QA/QC procedure should be coded into software creating or maintaining the named database.

Finally, Table 24 suggests that several of the checks be coded into the software used to build validation files. This has been suggested when the QA/QC procedure might be easier to implement when the validation dataset is constructed, when the QA/QC procedure requires information from TransCAD, or when the QA/QC check is designed to address issues that affect only model validation. For example, it might be worthwhile to maintain holiday traffic counts as “special” counts, but those counts should never be used in a model validation.

The Part 1 of this report also suggested that temporal, functional class and geographic coverage were important QA/QC procedures. QA/QC procedures for coverage criteria cannot be as formal as the criteria listed in Table 24. Rather, the coverage QA/QC procedures should focus on visual inspections of count coverage as shown in Figure 7 and summaries of links with counts as shown in Tables 7 and 25.

Table 25. Percent of TransCAD Directional Links with Traffic Counts

	Functional Class				
Area Type	Freeway[1]	Principal Arterial	Minor Arterial	Collector/ Local	Total
Urban Business	18%	31%	18%	8%	20%
Urban	19%	29%	21%	7%	21%
Suburban / Rural Place	24%	52%	55%	53%	52%
Rural	41%	55%	68%	71%	68%
Total Roadway Links[1]	24%	37%	37%	44%	38%

[1] Excluding system-to-system ramps and ramps.

Figure 7. TransCAD Links from 2002 Validation Effort with at Least One Count

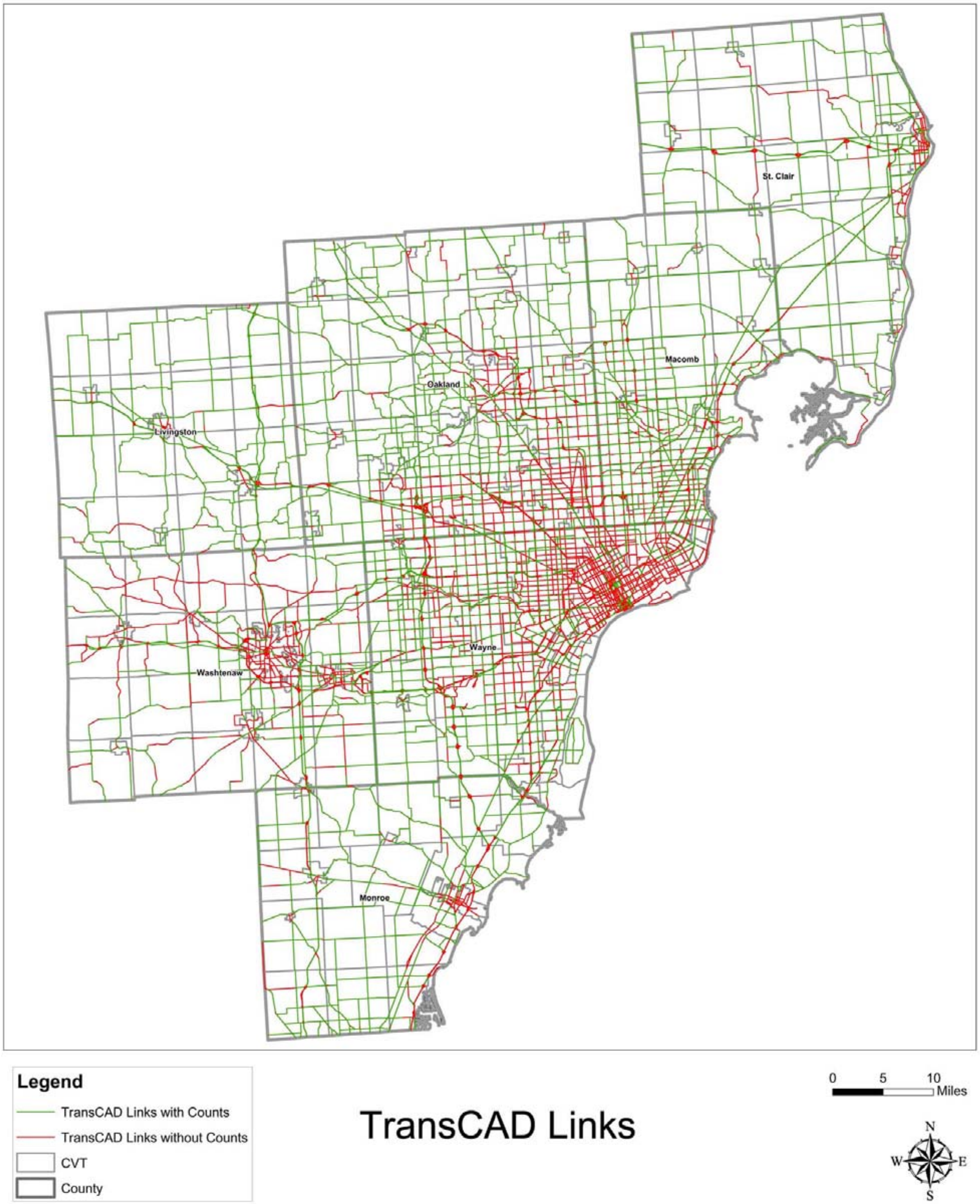


Figure 7 shows the TransCAD network links for 2002 that have at least one traffic count. As can be seen, there is generally good coverage of the network, except in selected areas such as the cities of Detroit, Ann Arbor, Pontiac, and Port Huron, and areas of Oakland and Wayne Counties. Table 7 (shown at the beginning of this report) shows the number of TransCAD links with traffic counts. Table 25 shows percentages of TransCAD links with traffic counts. The only strata that look at all deficient in the number of counts are urban freeways, rural principal arterials, and urban collector / local streets.

OTHER VALIDATION ISSUES

REVIEW OF RECOMMENDED VEHICLE CLASSIFICATION SCHEME

SEMCOG has recommended using data from 11 Automatic Vehicle Classification (AVC) sites[13] maintained by MDOT to estimate the proportions of traffic by hour by vehicle class by day of week. While this proposal provides a reasonable starting point, additional work and modified procedures are strongly recommended:

- The data from the 11 AVC sites should be supplemented with data from other sites as available. Potential sources include the MITS data and data from short-term traffic counts performed using mechanical counters. Ideally, some validation of the accuracy of the mechanical counters would be performed prior to accepting data from the counters for use.
- Vehicle classification data are required at a sufficient number of sites to account for the variation in vehicle mix by area type and functional class. It might be possible to collect sufficient classification counts to summarize the data by only one of the stratifications. If this is the case, the focus should be on the summarizing the mix of vehicles by functional class.
- It is likely that estimation of traffic by hour by vehicle class by day of week is stretching or exceeding the accuracy limits of the available classification data, especially when the data are summarized by functional class and/or area type. SEMCOG should focus on first producing reasonable classification estimates by functional class and area type for an average day (based on either AADT or AWDT). If sufficient data exist, it might be reasonable to stratify the proportions of vehicles by type by time-of-day (morning peak period, midday, afternoon peak period, and night).

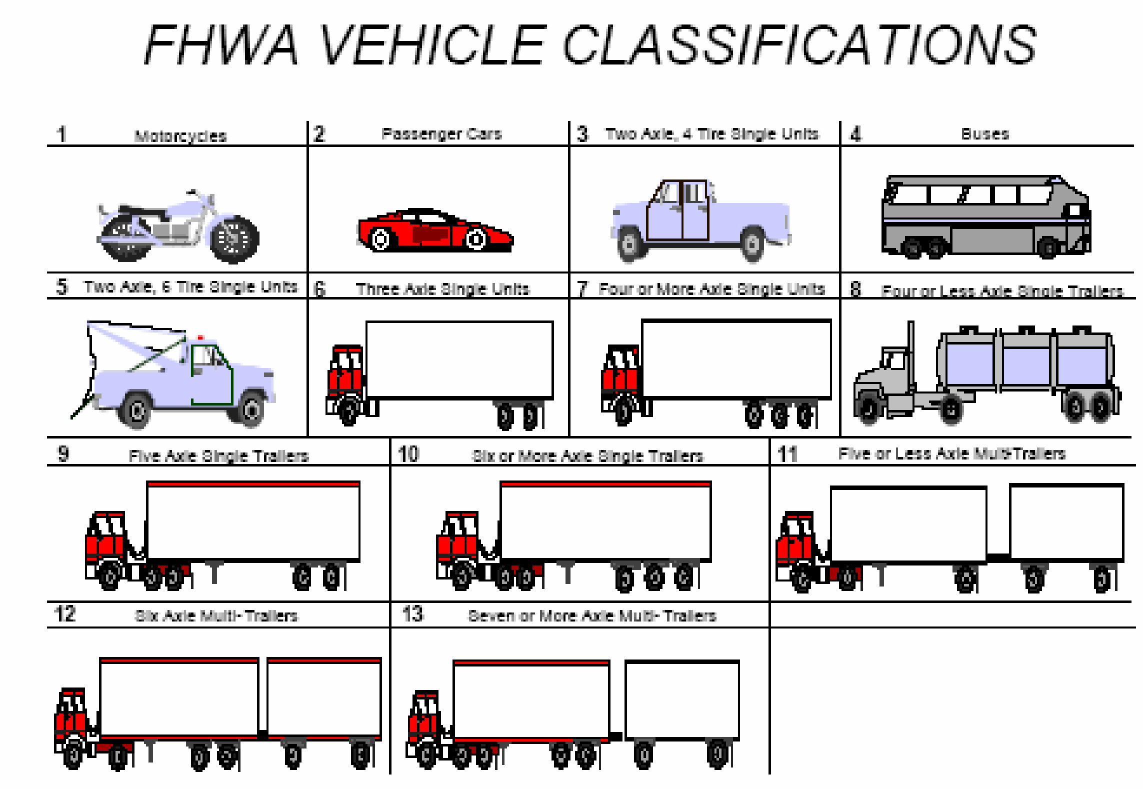
SEMCOG has further recommended aggregating vehicle classes in a manner consistent with classes used for MOBILE 6 air quality modeling (see Table 26 and Figure 8). This is a logical approach to take since the stratification will implicitly be made within MOBILE 6 if the data are not explicitly input in the various classes[14]. In effect, vehicles must be classified by four aggregate strata for the MOBILE 6 groupings suggested by SEMCOG as shown in Table 26:

[13] The AVC sites are actually PTRs with vehicle classification capabilities.
[14] A review of the actual required MOBILE 6 inputs used in the SEMCOG region should be made with the staff responsible for applying MOBILE 6 for the region.

Table 26. Vehicle Classification Aggregations for MOBILE 6 Air Quality Modeling Recommended by SEMCOG

MOBILE 6		FHWA Vehicle Classification		
Classification	**Description**	**Proportion**	**Bin**	**Description**
MC	Motorcycles	1.000	Bin 1	Motorcycles
LDGV	Light Duty Gas Vehicles	0.676	Bin 2	Passenger Cars
LDDV	Light Duty Diesel Vehicles	0.005	Bin 2	Passenger Cars
LDGT1	Light Duty Gas Trucks Less than 8,500 Pounds GVW	0.234 0.732	Bin 2 Bin 3	Passenger Cars Two Axle, Four Tire Single Units
LDGT2	Light Duty Gas Trucks 8,500 Pounds or More GVW	0.083 0.260	Bin 2 Bin 3	Passenger Cars Two Axle, Four Tire Single Units
LDDT	Light Duty Diesel Trucks	0.002 0.008	Bin 2 Bin 3	Passenger Cars Two Axle, Four Tire Single Units
HDGV	Heavy Duty Gas Vehicles	0.305	Sum of Bins 4-13	Buses; Two Axle, Six Tire Single Units; Three or more axle single units; combination units
HDDV	Heavy Duty Diesel Vehicles	0.695	Sum of Bins 4-13	Buses; Two Axle, Six Tire Single Units; Three or more axle single units; combination units

Figure 8. FHWA Vehicle Classification Bins



- FHWA bin 1, motorcycles
- FHWA bin 2, passenger cars
- FHWA bin 3, two axle, four tire single units
- FHWA bins 4-13, all other vehicle types

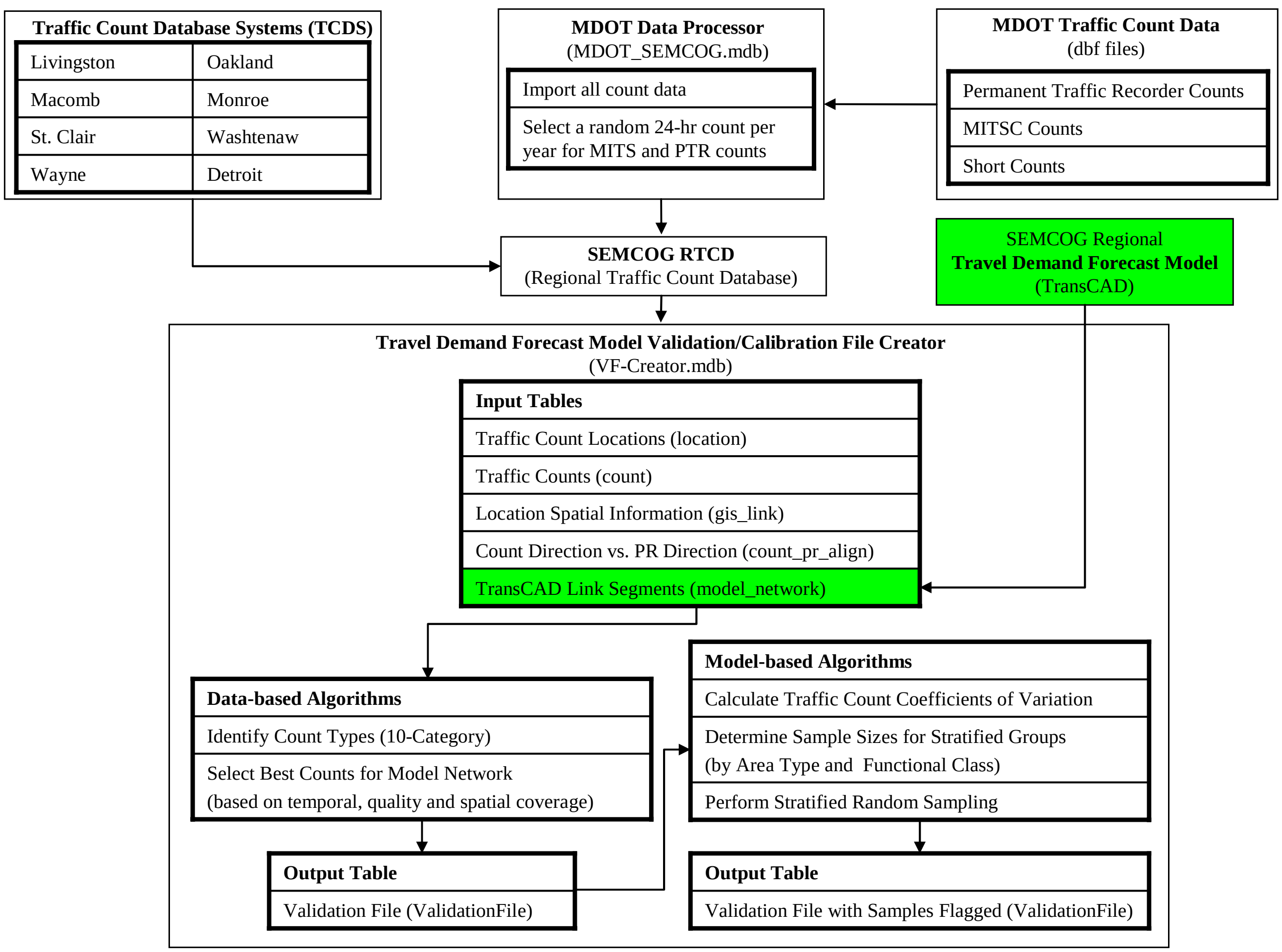
The proportions shown in Table 26 for estimating the eight vehicle classification groups used for MOBILE 6 suggest that the proportions of the above four aggregate groups of vehicles is constant over the region and over time. Without detailed vehicle classification studies by area type and functional class, this is probably the best assumption that can be made. However, the proportions should be checked periodically on a regional level through analyses of the AVC data and, possibly, regional vehicle registration data.

Part 3. Validation File Development

INTRODUCTION

The purposes of Part 3, Validation File Development is to develop a database program that automatically processes traffic count data from various sources and generates traffic count input file for TransCAD. Based on the data specifications and business rules developed in the previous Parts, the system flow is setup as shown in Figure 9.

Figure 9. Validation File Development System Flow Chart



MDOT DATA PROCESSOR

SYSTEM DESCRIPTION

To migrate MDOT MITS, PTR and Short Counts to the SEMCOG database format, an MS Access database (mdot_semcog.mdb) was developed. The MITS and PTR are continuous all year, while a validation file needs only a 24-hr count. Therefore, within the database, options for randomly selecting a 24-hr MITS or PTR count per year were also developed.

<u>SYSTEM PROCESS TABLES</u>

COUNT_MDOT

SEMCOG count format table for housing migrated data. There are additional fields used for determining count blocks and count source (MITS, PTR, Short Count)

COUNT_MDOT_RANDOM_MIT

Holds the random selection of MIT counts after running the "Generate Random MITS Count" process

COUNT_MDOT_RANDOM_PTR

Holds the random selection of PTR counts after running the "Generate Random PTR Count" process

COUNT_MDOT_DELETE

Temporary table used in import process

<u>SYSTEM ALGORITHMS</u>

See <u>Appendix F – MDOT Count Import Algorithms</u> for detail of the algorithms behind each process.

<u>SYSTEM OUTPUT TABLE</u>

COUNT_MDOT

MITS, PTR and Short Counts are distinguished by the MDOT_SRC field. This table is also used as a source for generating random MITS and PTR counts.

<u>SYSTEM INSTRUCTIONS</u>

1. Browse to the mdot_semcog.mdb Microsoft Access database file on the computer hard drive with Windows Explorer.
2. Double-click ("run") the database file to display the import process main menu.
3. Follow the "Preliminary Measures" and review the "MDOT Import Menu - Overview" and each desired process detailed later in this document.
    a. Single Import - Short Count
    b. Single Import - MITS
    c. Single Import - PTR
    d. Multiple Import
    e. Random Count Selection

    NOTE: Each import process and random count selection process can take 5-10 minutes to run depending on the number of counts and computer hardware running the process. If multiple imports are processed, a warning pops up before each (described in steps below), so the range of time is still 5-10 minutes for user interaction.

PRELIMINARY MEASURES

1. Make sure the database DBF files received from MDOT conform to the MDOT data format specifications for each count type:
    a. Short Count Format (see Appendix E - Short Count Format)
    b. MITS Data Format (see Appendix E - MITS Data Format)
    c. PTR Format (see Appendix E - PTR Data Format)
2. Verify the MDOT DBF filenames are at most 8 characters with a "dbf" extension (an error will occur when importing the tables if the filename does not follow this)
    a. Browse to the DBF files and Windows Explorer
    b. Rename the files to follow this file format (e.g. from mits_2004.dbf to mits2004.dbf)
3. Import the corresponding MDOT DBF tables into the database used in this process
    a. Open the MS Access database for this process (e.g. mdot_semcog.mdb)
    b. If the "mdot_semcog : Database" view is not available
        i. Select "Unhide..." from the "Window" menu
        ii. Select "mdot_semcog : Database" from the list and click "OK"
    c. Select "Get External Data" -> "Import" from the "File" menu
    d. Change the "Files of type" dropdown to "dBASE IV (*.dbf)"
    e. Browse to the directory of one of the DBF files provided by MDOT
    f. Select the DBF file and click "Import"
    g. The new table will be named by the first part of the filename (e.g. mits2004.dbf would yield the table named mits2004)
4. The SEMCOG count format specification for this process follows the data layout presented in Appendix E - SEMCOG Count Format

MDOT IMPORT MENU - OVERVIEW

1. Exit button - Exit the main menu (i.e. to import a table).
2. Process Data button - When an imported table is selected, click this button to process each of the selected tables for each count type (MITS, PTR, Short Count).
3. Reset Values button - Click this button to reset the values in the table selection dropdown menus.
4. Select Short Count Table dropdown - The list of tables in the database that are available for import. Be sure the selected table format follows that of the Short Count specification in Appendix E - Short Count Format.
5. Select MITS Count Table dropdown - The list of tables in the database that are available for import. Be sure the selected table format follows that of the MITS specification in Appendix E - MITS Data Format.
6. Select PTR Count Table dropdown - The list of tables in the database that are available for import. Be sure the selected table format follows that of the PTR specification in Appendix E - PTR Format.
7. Generate Random MITS Count button - Click this button after importing a MITS Count table to generate a random count selection of MITS data.
8. Generate Random PTR Count button - Click this button after importing a PTR Count table to generate a random count selection of PTR data.

SINGLE IMPORT - SHORT COUNT DATA

1. Select the imported Short Count table from the table list
2. Click the Process Data button

3. A warning will pop up notifying that all data in the COUNT_MDOT table will be cleared before running the process. Click OK to start the import or Cancel to return to the main menu.
4. After clicking OK to import the table, another warning will pop up confirming the table being processed and that the process could take several minutes to run.
5. The status area in the lower left portion of MS Access will show the progress of each step in the process, will display "Ready" when complete, and a pop up message will notify "Process Complete!". For large data sets, a warning may pop up at least once (possibly 2-3 times) notifying "There isn't enough disk space or memory to undo the data changes this action query is about to make." Click "Yes" to continue each time this warning is displayed.
6. The processed data is now stored in the COUNT_MDOT table with the value of "CNT" in the MDOT_SRC field.
7. Click the Exit button to leave the main menu.
8. Select "Unhide..." from the "Window" menu
9. Select "mdot_semcog : Database" from the list and click "OK"
10. The table is located in the "Tables" area of MS Access under "Objects".
11. This table can be queried or exported at the user disgression and must NOT be altered in structure or deleted

SINGLE IMPORT – MITS DATA

1. Select the imported MITS table from the table list
2. Click the Process Data button
3. A warning will pop up notifying that all data in the COUNT_MDOT table will be cleared before running the process. Click OK to start the import or Cancel to return to the main menu.
4. After clicking OK to import the table, another warning will pop up confirming the table being processed and that the process could take several minutes to run.
5. The status area in the lower left portion of MS Access will show the progress of each step in the process, will display "Ready" when complete, and a pop up message will notify "Process Complete!". For large data sets, a warning may pop up at least once (possibly 2-3 times) notifying "There isn't enough disk space or memory to undo the data changes this action query is about to make." Click "Yes" to continue each time this warning is displayed.
6. The processed data is now stored in the COUNT_MDOT table with the value of "MIT" in the MDOT_SRC field.
7. Click the Exit button to leave the main menu.
8. Select "Unhide..." from the "Window" menu
9. Select "mdot_semcog: Database" from the list and click "OK"
10. The table is located in the "Tables" area of MS Access under "Objects".
11. This table can be queried or exported at the user disgression and must NOT be altered in structure or deleted

SINGLE IMPORT – PTR DATA

1. Select the imported PTR table from the table list
2. Click the Process Data button

3. A warning will pop up notifying that all data in the COUNT_MDOT table will be cleared before running the process. Click OK to start the import or Cancel to return to the main menu.
4. After clicking OK to import the table, another warning will pop up confirming the table being processed and that the process could take several minutes to run.
5. The status area in the lower left portion of MS Access will show the progress of each step in the process, will display "Ready" when complete, and a pop up message will notify "Process Complete!". For large data sets, a warning may pop up at least once (possibly 2-3 times) notifying "There isn't enough disk space or memory to undo the data changes this action query is about to make." Click "Yes" to continue each time this warning is displayed.
6. The processed data is now stored in the COUNT_MDOT table with the value of " PTR" in the MDOT_SRC field.
7. Click the Exit button to leave the main menu.
8. Select "Unhide..." from the "Window" menu
9. Select "mdot_semcog : Database" from the list and click "OK"
10. The table is located in the "Tables" area of MS Access under "Objects".
11. This table can be queried or exported at the user disgression and must NOT be altered in structure or deleted

MULTIPLE IMPORT

1. Select the imported table for each type (Short Count, MITS, PTR) from the table lists
2. Click the Process Data button
3. A warning will pop up notifying that all data in the COUNT_MDOT table will be cleared before running the process. Click OK to start the import or Cancel to return to the main menu.
4. After clicking OK to import the table, another warning will pop up confirming the table being processed and that the process could take several minutes to run. This warning pops up for each table being processed.
5. The status area in the lower left portion of MS Access will show the progress of each step in the process, will display "Ready" when complete, and a pop up message will notify "Process Complete!". For large data sets, a warning may pop up at least once (possibly 2-3 times) notifying "There isn't enough disk space or memory to undo the data changes this action query is about to make." Click "Yes" to continue each time this warning is displayed.
6. The processed data is now stored in the COUNT_MDOT table and distinguished by the MDOT_SRC field ("CNT", "MIT", "PTR").
7. Click the Exit button to leave the main menu.
8. Select "Unhide..." from the "Window" menu
9. Select "mdot_semcog : Database" from the list and click "OK"
10. The table is located in the "Tables" area of MS Access under "Objects".
11. This table can be queried or exported at the user disgression and must NOT be altered in structure or deleted

RANDOM COUNT SELECTION

1. Click the button corresponding to the random count selection
   a. Generate Random MITS Count button for random MITS data
   b. Generate Random PTR Count button for random PTR data

2. The processed data is now stored in the corresponding random selection table
    a. COUNT_MDOT_RANDOM_MIT for random MITS selection
    b. COUNT_MDOT_RANDOM_PTR for random PTR selection
3. Click the Exit button to leave the main menu.
4. Select "Unhide..." from the "Window" menu
5. Select "mdot_semcog : Database" from the list and click "OK"
6. The table is located in the "Tables" area of MS Access under "Objects".
7. This table can be queried or exported at the user disgression and must NOT be altered in structure or deleted

VALIDATION FILE CREATOR

SYSTEM DESCRIPTION

This MS Access database (VF-Creator.mdb) was developed to generate a traffic count validation dataset for a TransCAD model based on actual traffic counts observed on a road network.

SYSTEM INPUT TABLES

Location –	traffic count locations (exported from RTCD[15])
Count –	traffic counts (exported from RTCD)
GIS_link –	PR and mile point information for traffic count locations (exported from RTCD)
Count_pr_align –	the relationship between traffic flow direction and mile point direction
Model_network –	travel demand model network (generated from TransCAD model)

SYSTEM ALGORITHMS

1. Count Quality Assessment – The algorithm processes the traffic count data and determines the data quality based on the 10-level system described in the Part 2 of this report (page 42)
2. Count Assignment – The algorithm selects the best count for each model network link based on the business rules described in the Part 2 of this report (page 43)

SYSTEM OUTPUT TABLE

Validationfile – traffic count validation file with field specification described in the Part 2 of this report (page 44)

SYSTEM INSTRUCTIONS

1. Update the system input tables
   (location, count, gis_link, count_pr_align, and model_network)
2. Open Access Forms “Validation”
3. Select model validation target year (e.g. 2002)
4. Click on “Start” button

The algorithms will take about 25 minutes to run, depending on the computer CPU power and dataset size. The output table “Validationfile” will contain all data relevant to model validation analysis.

[15] SEMCOG Regional Traffic Count Database

Appendix A

FHWA Validation Manual Validation Guidelines

Coefficient						
	70 mph	60 mph	50 mph	70 mph	60 mph	50 mph
✂	0.88	0.83	0.56	1.00	0.83	0.71
👓	9.80	5.50	3.60	5.40	2.70	2.10

The speeds shown in the above table are design speeds of the facility, not the free flow speeds. Capacities used in the v/c ratio are ultimate capacity, not a design capacity as used in the standard BPR curve. The curves based on the HCM exhibit a speed of about 35 mph at a v/c ratio of 1.0. This is consistent with standard capacity rules that the denser traffic flows occur at this speed. Note that the BPR curve has a much higher speed at a v/c equal to 1.0 than does the HCM curves.

The ultimate capacity used for these curves was 1800 vehicles per hour per lane for a one mile section. This value is the ultimate capacity for typical prevailing conditions, not those under ideal conditions which would have a capacity of 2000 vehicles per hour per lane (and even higher based on recent changes to the Highway Capacity Manual). The curves extend beyond the point where the v/c ratio is 1.0, or where the flow has reached capacity. In capacity analysis, this portion of the curve is considered unstable. However, for travel demand modeling, the curve must extend beyond 1.0 to account for the theoretical assignment of the traffic.

The calibration and validation of the assignment model includes both the systematic adjustment of any lookup speed and capacity tables as well as the adjustment of the coefficients of the volume-delay function, by facility type.

7.1.3 Validation Tests

The validation tests for highway assignment are presented at three levels; systemwide, corridor, and link specific. This increasing detail of validation tests is correlated to the step(s) in the model chain that could be the cause of the possible error(s).

There are several systemwide or aggregate validation checks of the auto assignment process. The checks are generally made on daily volumes, but it is prudent to make the checks on volumes by time-of-day as well. Systemwide checks include Vehicle Miles of Travel (VMT), Vehicle Hours of Travel (VHT), cordon volume summaries and screenline summaries. In addition to checking summations of VMT, VHT, and volumes, the average VMT and VHT per household and person should be checked.

Vehicle Miles of Travel (VMT)

Validation of the model using VMT addresses all major steps in the travel demand models including trip generation (the number of trips), trip distribution (the trip lengths), and assignment (the paths taken).

VMT validation is particularly important in urban areas that are designated by the Environmental Protection Agency (EPA) as non-attainment for moderate and serious carbon monoxide (CO). The EPA has published guidance for the forecasting and tracking of VMT as required by Section 187(a) of the Clean Air Act Amendments of 1990 (CAAA). This guidance should be read and understood by those developing travel demand models for these urban areas. The document can be found on the Internet at http://www.bts.gov/smart/cat/vmt.html. The Bureau of Transportation Statistics has an Internet home page at www.bts.gov and this is an excellent resource for all information relating to transportation statistics.

The first check is observed versus modeled Vehicle Miles of Travel (VMT). VMT is simply the product of the link volume and the link distance, summed over the desired geographic area and facility types. The observed VMT is a product of a comprehensive traffic count program. Since not every link in the network will be counted for the validation year, estimates of observed VMT must be developed.

The primary source of observed VMT is the Highway Performance Monitoring System (HPMS) data. The VMT tracking and forecasting guidance issued by the EPA requires that the HPMS be used for tracking VMT in urban areas that are in violation of the air quality standards. The HPMS estimate for VMT is calculated from samples of observed traffic counts in a region and updated regularly. It is part of the reporting requirements to the Federal Highway Administration. The FHWA publishes a report, *Highway Performance Monitor System (HPMS) Field Manual* that should be referred to when comparing HPMS VMT with modeled VMT.

When using the HPMS estimate of VMT, is it important to account for the basic differences in the highway system covered by HPMS and that included in the typical highway network for the travel demand model. The HPMS data includes VMT estimates for all functional classifications of roadways within the Federal Aid Urbanized Area (FAUA), including local streets. Most regional model networks do not include local streets. The lowest level of roadway in most models is the collectors. The local streets are typically represented by the centroid connectors. Recognizing this difference, the direct estimates of VMT from the model should be lower than the HPMS estimate of VMT.

In addition to the differences in the functional classification of the highway system, the different geographic areas covered by each estimate of VMT must be recognized. The HPMS is designed primarily for the area within the FHWA's designated Federal Aid Urbanized Area (FAUA). On the other hand, when the EPA designates an area as being in non-attainment, the area usually includes all counties within the nonconforming area. This non-attainment area is typically larger than the FAUA. The EPA's guidance for VMT forecasting and tracking allows for non-HPMS methods to be used in the non-attainment areas that are outside of the FAUA. Therefore, it is important to reconcile the various geographic areas of the modeled area, the HPMS area, and the non-attainment area.

While the EPA requires the HPMS method be used for *tracking* VMT, the network based travel demand model is the preferred method for *forecasting* VMT in non-attainment areas. In order to simplify the forecasting of VMT for air quality purposes, many urban areas have elected to include the entire non-attainment area in the travel demand model. This has the added advantage of not only covering the entire FAUA as required by the FHWA, but also allows for forecasting travel demand in areas that are likely to become urbanized in the future, as required by the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991.

- *Check VMT values for the region, per household, and per person.* There are many useful statistics that can be calculated for the systemwide-level validation of VMT. These include both the absolute and relative

A

(percent) difference. Compare current estimates of regionwide VMT with the historical trend and rate of growth from HPMS.

The absolute difference is the simple difference between observed and modeled VMT. The difference is typically large for high-volume links and low for low-volume links, so the size of the numerical difference does not reliably reflect the true significance of error.

Percent difference is often preferred to absolute difference since its magnitude indicates the relative significance of error. Modeled regional VMT should generally be within five percent of observed regional VMT. This five percent difference is particularly important in light of the accepted error that EPA allows for VMT tracking using the HPMS data. The EPA has allowed margins of error in VMT estimates as high as five percent in 1994 to a new margin of three percent in 1996 and afterwards.

Table 7-3 is an example of a VMT validation summary.

Table 7-3
Example VMT Validation Summary

	VMT		Error		VMT Distribution	
Facility Type	Estimated[1]	Observed[2]	Difference	Percent	Estimated	Observed
Freeways						
Principal Arterials						
Minor Arterials						
Collectors						
Total					1	1

Notes: 1 - Estimated is the VMT produced by the model
2 - Observed is based on either traffic counts or the HPMS estimates of VMT

Typical distributions of VMT by facility type are presented in Table 7-4.

Table 7-4
Urban Area VMT by Facility Type

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

Minor Arterials	25-28%	18-22%	18-22%
Collectors	12-15%	8-12%	8-12%

Source: Christopher Fleet and Patrick De Corla-Souza, *Increasing the Capacity of Urban Highways - The Role of Freeways*, presented at the 69th Annual Meeting of the TRB, January 1990

As noted, VMT per household and VMT per person are useful measures to determine if the modelled estimates of VMT are within reasonable limits. These unit measures of VMT are also useful in determining the source of modelling error. A model that underestimates regional VMT, yet has reasonable VMT per household may have errors in the household data (underestimation of the number of households). All of these pieces of data assist the analyst in determining the cause of the modelling error and the associated adjustment or correction.

Reasonable ranges of VMT per household are 40-60 miles per day for large urban areas and 30-40 miles per day for small urban areas. The 1990 NPTS reported an average of 41.37 vehicle miles traveled per household daily. Reasonable ranges of VMT per person are 17-24 miles per day for large urban areas and 10-16 miles per day for small urban areas.

When models are originally calibrated from survey data (or transferred from other regions), the modeled regional VMT will frequently be substantially lower than the observed regional VMT. An initial response to this occurrence is often to increase trip generation rates, especially for home-based non-work and non-home-based trips, under the justification that these trips are the most commonly under-reported trips in a household travel survey. Frequently, increases in modeled trip rates of 10 to 20 percent produce modeled results that reasonably match the observed regional VMT. However, some regions have increased trip rates by as much as 60 to 70 percent.

Traffic Volumes

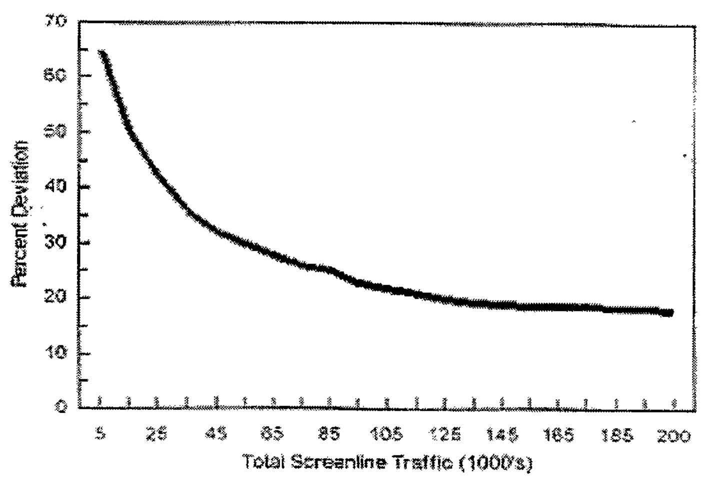
After validation of the VMT, the next level of validation of the highway assignment is the comparison of observed versus estimated traffic volume on the highway network. The observed count data are derived from the ongoing traffic counting and monitoring program in the urban area as described in section 2.3. This data may be developed primarily for the HPMS requirements and supplemented as required. Traffic volumes are validated at the systemwide level by comparing summations of volumes at both cordons and screenlines. While the comparison of volumes on cutlines can be used as a systemwide measure, it will be treated as a localized measure in this document.

- *Compare observed versus estimated volumes by screenline.* The Michigan Department of Transportation (MDOT) has targets of 5% and 10% for screenlines and cutlines, respectively, for percent differences in observed and estimated volumes by screenline. Figure 7-2 shows maximum desirable deviation in total screenline volumes according to the observed screenline volume.

- *Compare observed versus estimated volumes for all links with counts.* With the use of the on-screen network editors and plots of network attributes, the checking of link level counts visually is relatively simple. In addition to visually checking the correlation of the counts to volumes, (Figure 7.3) it is also useful to compute aggregate statistics on the validity of the traffic assignment. Two measures can be computed; the correlation coefficient and the Percent Root Mean Square of the Error. Each is discussed below.

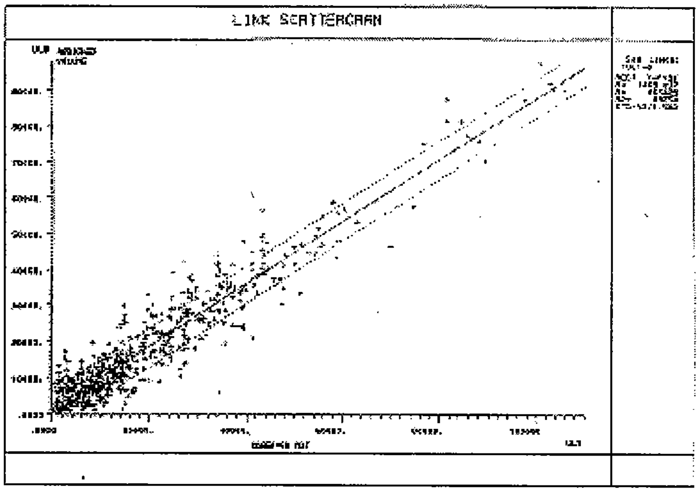
Calculate R^2 (Coefficient of Determination) comparing regionwide observed traffic counts versus estimated volumes. R^2 regionwide should be greater than 0.88. Another useful validation tool is to plot a scattergram of the counts versus the assigned volumes. Any data points (links) that lie outside of a reasonable boundary of the 45° line should be reviewed.

Figure 7-2
Maximum Desirable Deviation in Total Screenline Volumes



Source: NCHRP 255 p.41 (cited in FHWA, Calibration and Adjustment of System Planning Models, Dec. 1990)

Figure 7-3
Assigned versus Observed Average Daily Traffic Volumes



#	*Calculate percent RMSE as follows:*

$$\%RMSE = \frac{(\Sigma_j(\,Model_j - Count_j\,)^2 \,/\, (Number\ of\ Counts - 1)\,)^{0.5} * 100}{(\Sigma_j\, Count_j \;\; / \; Number\ of\ Counts)}$$

The Montana Department of Transportation (MDT) suggests that an appropriate aggregate %RMSE is less than 30%. The %RMSE can be calculated for all links with counts or by facility type and area type as shown in Tables 7-5 and 7-6.

Table 7-5
Percent Root Mean Square Error Comparisons

	Reno		*Phoe nix*	*Concord*	
Facility	*PM*	*ADT*	*ADT*	*PM*	*AM*
Freeway	*18.3*	*18.6*	*25.4*	*NA*	*NA*
Arterial	*39.2*	*36.8*	*38.5*	*NA*	*NA*
Collector	*76.1*	*77.5*	*62.7*	*NA*	*NA*
Total	*39.9*	*36.8*	*40.6*	*31.1*	*36.8*

Table 7-6
Percent Root Mean Square Error - 24-Hour Assignment (Reno)

	Area Type					
Facility Type	*1*	*2*	*3*	*4*	*5*	*All*
Freeway	*11.649*	*18.092*	*21.891*	*0.000*	*11.271*	*18.334*
Major Art	*22.547*	*37.778*	*42.209*	*43.162*	*43.283*	*36.768*

Collector	*0.000*	*52.953*	*88.920*	*115.326*	*70.148*	*77.482*
Minor Art	*25.874*	*44.072*	*52.353*	*28.367*	*60.121*	*43.895*
Ramp	*24.237*	*63.524*	*47.574*	*80.649*	*131.009*	*74.846*
Total	*21.303*	*37.210*	*37.793*	*43.742*	*38.694*	*36.767*

Assigned Speeds

If actual observed speed data are available, this should be summarized in highway segments consisting of a number of links and intersections so that intersection-based delay is averaged into highway travel time. Speed observations should be classified by facility type and area type to compare with modeled speeds for the same categories. Checks of highway skims include the following:

- Summarize link speeds by facility type and area type, showing the minimum, maximum, and average speed for each category. Compare assigned speeds with speeds used for distribution and mode choice.
- Compare observed and estimated speeds by highway segments, if available.

Model Parameters

Once the cordon lines and screenlines are validated and the trip distribution model is judged to be producing acceptable results, the assignment volume-delay functions can be modified systematically to produce the desired assignments. It has been the practice in some urban areas to adjust individual link attributes to get an assignment that matches the link counts. In many cases, these adjustments have produced unrealistic values of link speeds and capacities (free-flow speeds of 5mph for example) that only worked to get the desired assignment results. The adjustment of link attributes should be limited to minor systematic adjustments to speeds and capacities for groups of links that have the same facility and area type.

There are a number of parameters in highway assignment that are potential sources of error. While the actual parameters and calculation options involved depend on the modeling software and assignment methodology being used, possibilities include:

- Assignment procedures including number of iterations, expansion of incremental loads, and damping factors,
- Volume-delay parameters such as the BPR coefficient ✂ and exponent 👓.
- Peak-hour conversion factors used to adjust hourly capacity and/or daily volumes in volume-delay function.
- Scaling or conversion factors to change units of time, distance, or speed (mi/hr or km/hr).
- Maximum/minimum speed constraints.
- Preload purposes (HOV, through trips, trucks, long/short trips).
- Toll queuing parameters (diversion, shift constant, etc.)

Other validation tests include:

- Path trees based on assigned travel times.

- Select Link Analysis

- Assign through trip table separately to check routing of external-external trips. Should use higher-level facilities.

7.2 Transit Assignment

The primary validation check of the transit assignment process is of observed versus modeled boardings. These should be checked for the region, by mode and possibly sub-mode, and by trip length. In addition, a check of observed versus modeled boardings per trip (transfer rates) is a more detailed check that tests reasonability of the number of transfers made per trip.

Model Calibration

The first step of the validation of a transit assignment occurs during the mode-choice model calibration. In the calibration step, the mode-specific constants for a region are derived so that the mode-choice model produces the appropriate share of transit trips for the region. The structure of the mode-choice model will affect the order in which the bias constants are derived. In a multinomial logit model, the bias constants for all transit modes can be derived simultaneously. If a nested logit model is employed, the bias constants for the lower levels of the nest should be derived first, then the next higher level, until the top level of the nest is reached. Several iterations of this process are normally required before an acceptable set of bias constants are derived. Note: care should be used to avoid bias constants that have an absolute value greater than 2.0 or 3.0 at the top level of the nest. If the constants are too large, the model will lose its sensitivity to level of service changes.

Validation

The amount of time and effort required to validate a transit assignment is directly correlated with the level of precision demanded. For highway planning purposes, it is generally sufficient to validate to the regional number of boardings, so that the appropriate number of person trips are removed from the highway network. For transit planning purposes, however, it may be necessary to validate to the mode, corridor, route, segment, or even station level of detail. Such precision is very difficult to attain with a fully synthetic model. (One option available when a finer level of detail is required is to utilize a pivot-point model.)

A few of the common problems that occur when validating a transit assignment are discussed in the following paragraphs, along with suggested solutions.

Number of Transfers - It is very common for a transit assignment to produce more transfers than are occurring in the actual transit systems. This problem can sometimes be solved by adjusting the transfer penalty.

However, the problem of assigning too many transfers may also result from having a shortage of walk access links to serve the transit system. The walk access links should be checked to make sure that each transit route has walk access to each TAZ within the accepted walking distance, especially when an automated access coding routine is employed. This can be difficult in CBD areas, where numerous transit routes often serve even more numerous TAZs. In order to avoid the problem of having to code too many walk access links, a CBD walk network should be employed.

Trip Length Frequency Distribution - If the average trip length for the assigned transit trips is not right, check the trip length frequency distribution for the person trip table used to create it. If the person trip distribution reflects the same pattern as the transit trip, the problem may be attributed to the trip distribution model.

Otherwise, the district-to-district transit trip summaries should be examined. The problem of an erroneous trip length frequency distribution may result from trips associated with a specific zone or district in the region. If the comparison of an observed transit trip table vs. an estimated trip table shows a large imbalance for a specific area, the route and access coding for that area should be checked first. If that network coding is reasonable and consistent with the rest of the model, you may wish to derive and apply a bias constant specific to that district.

Express or Limited Service - During the transit validation process it is often helpful to examine the relative assignments of different types of transit service. For example, it may be helpful to compare the assignments of local bus service and express bus service to determine whether or not a pattern can be found.

If the express service is being under-assigned the cause could be insufficient drive access, since express bus riders are more likely than local bus riders to drive to either a formal or an informal park-and-ride lot along the route. Alternately, the under-assignment could be due to an excess of wait time, since express bus riders who know the schedule of their service would not need to wait as long as the infrequent level of service would tend to indicate.

On the other hand, if express bus ridership is overestimated in comparison to local service, you may wish to check the transit route coding to make sure that the route is not allowed to collect passengers on the limited- or non-stop portions of the journeys.

Corridor Analysis - Most transit systems have corridors, of varying lengths, that are served by more than one transit route. These corridors have the benefit of improving the perceived, or composite, frequency of service for some of the potential transit riders in that corridor. However, with most transportation planning software, care must be taken when coding the transit lines in these corridors to ensure that the stop sequence is identical, or else a composite headway will not be calculated for that trip.

Another aspect of corridors served by multiple routes is the assignment of trips to competing transit routes. The most common practice is to have the software distribute the trips to the competing routes based on the relative frequency of service. However, this practice is only valid if certain assumptions are true: 1) the potential riders must be aware of all routes that serve their particular trip; and 2) the transit service must be spaced evenly between the competing routes.

Since these assumptions are usually *not* true in real life, it is unlikely that the assignment of transit trips to competing routes in a corridor will be consistent with reality. Therefore, it is appropriate in the validation phase to analyze competing routes as a group, and to ignore the assignments to the individual routes.

Summary

In summary, the transit validation can include analysis of the following comparisons:

- Observed vs. estimated boardings for region, by mode, by time of day, and by trip length;
- Observed vs. estimated transfers per trip;
- Observed vs. estimated screenline volumes;
- Observed vs. estimated boardings by route or group of routes;
- Observed vs. estimated district-to-district transit trips.

Most modeling software platforms can generate a number of reports useful in the validation process, both at the regional and local levels. Typical reports provide information relating to:

- Passenger loadings by line, company, and mode;
- Access modes;
- Station-to-station/transfer nodes;
- Specified/calculated headways;
- Passenger- and vehicle-hours or miles of service;
- Peak loads.

Data Sources

The primary data source for transit ridership data is from the transit operator(s) within the region. Transit ridership data that can be obtained from transit operators include:

- System-wide linked trips, unlinked trips, and transfer rates;
- Route-specific boardings and fare collection data;
- Boardings and alightings at transit stations;
- Passenger-hours and passenger miles of service.

Additional ridership data can be obtained with the use of field surveys. The most common forms of transit survey include on-board surveys, ride-check surveys, and load-check surveys. These transit surveys can be conducted separately or in concert with each other.

Ride-check surveys are conducted by placing an observer on a transit vehicle to collect on/off count data at each stop. The observer is trained to record the stop location, time, the number of passengers boarding and alighting at each stop, and the passenger load following the stop. The observer can also be trained to collect other information about the passengers, such as gender, age, or the method of fare payment. The ride-check data can be used to calculate the peak load-point along a route.

On-board transit surveys involve the use of questionnaires which ask transit riders to provide information such as the origin and destination of their trip, modes of access and egress, trip purpose, and personal information such as gender, age, income level, and automobile availability. When conducted in conjunction with a ride-check survey, information from an on-board survey can be geo-coded and expanded to build trip tables describing the zone-to-zone trips made by the riders on a specific route.

Load-check transit surveys are used to count the number of passengers boarding and alighting at a transit stop, and the number of passengers on the transit vehicles travelling through that stop. Load-check surveys are used for two main purposes, to count the transit traffic at a major terminal or transfer location, and to count the number of passengers passing through a peak load point.

7.3 Validation Targets

Although absolute criteria for assessing the validity of all model systems cannot be precisely defined, a number of target values have been developed. These commonly-used values provide excellent guidance for evaluating the relative performance of particular models.

As noted earlier, observed versus estimated volumes should be checked by facility type and geographic area. The Federal Highway Administration (FHWA) and Michigan Department of Transportation (MDOT) define targets for daily volumes by facility type as shown in Table 7-7.

Table 7-7
Percent Difference Targets for Daily Traffic Volumes by Facility Type

Facility Type	FHWA Targets	MDOT Targets
Freeway	+/- 7%	+/- 6%
Major Arterial	10%	7%
Minor Arterial	15%	10%
Collector	25%	20%

Sources: FHWA, *Calibration and Adjustment of System Planning Models*, 1990; Michigan Department of Transportation (MDOT), *Urban Model Calibration Targets*, June 10, 1993

The Contra Costa Transit Authority (CCTA) in the San Francisco Bay Area has developed the following targets for peak-hour model validation:

- 75% of all freeway links must be within 20% of traffic counts.
- 50% of all freeway links must be within 10% of traffic counts.
- 75% of all major arterial links must be within 30% of traffic counts.
- 50% of all major arterial links must be within 15% of traffic counts.
- 50% of all intersection major turning movements must be within 20% of traffic counts.
- 30% of all intersection secondary turning movements must be within 20% of traffic counts.

For the CCTA, a major arterial is defined as one that carries over 10,000 vehicles per day, a major turning movement is defined as over 1,000 vehicles per hour, and a secondary turning movement is defined as 500-1,000 vehicles per hour.

R^2 and %RMSE values for VMT can be calculated for subsets of links, such as by facility type, volume range, or district.

Standards also exist for comparing observed versus modeled volumes for individual links. Table 7-8 shows percent difference targets for individual links as defined by FHWA and MDOT.

Table 7-8
Percent Difference Targets for Daily Volumes for Individual Links

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

Source: MDOT, *Urban Model Calibration Targets*, June 10, 1993

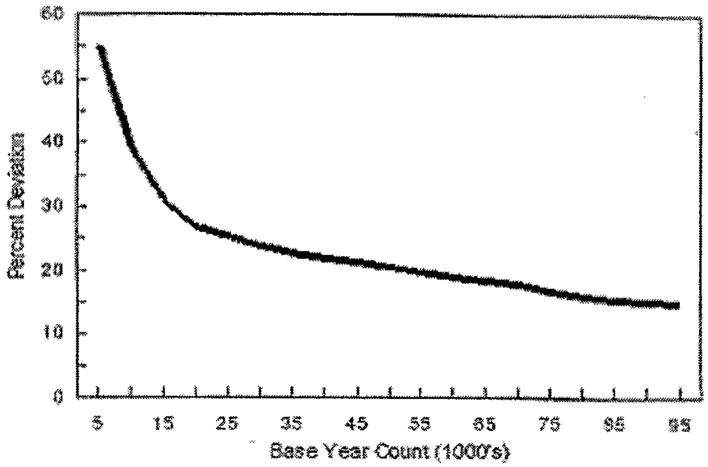
The FHWA targets are displayed graphically in Figure 7-3.

Additional checks should be made of observed versus modeled VHT and observed versus average speeds by facility

A

type, area type, and district.

Figure 7-4
Maximum Desirable Error for Link Volumes



Appendix B

Ohio Department of Transportation Validation Guidelines

- For most areas it is recommended that EPS be set to 0.0002 and EQUILIBRIUM ITERATIONS be set to 11. If 11 iterations are not obtained with this EPS, a lower value should be used.

- If the error reported by Tranplan at 11 iterations is above 0.05, then the EQUILIBRIUM ITERATIONS parameter should be increased until the error is reduced below 0.05.

Feedback Loops

- Implementation of feedback loops to trip distribution and beyond is recommended when an area has implemented a new model.

- For the initial run of the demand model prior to assignment the Model Calibration Manual (25) recommends that 87% of the free flow speed be used as a first approximation. Another possible method is to retain the old LOS C based speeds for this purpose, therefore, it is recommended that the old LOS C speeds on the network be maintained in the PEAKSPD field. It is recommended that program CSPD2000 be used to maintain this speed.

Count Restraint and Count Based Trip Table Estimation

- It is recommended that neither count restraining or count based matrix estimation techniques be used as part of the standard model run.

IV. Assignment Limitations, Checks and Refinements

Assignment Checks/Validation

- Model validation is the process of comparing model results to an **independent** data source.

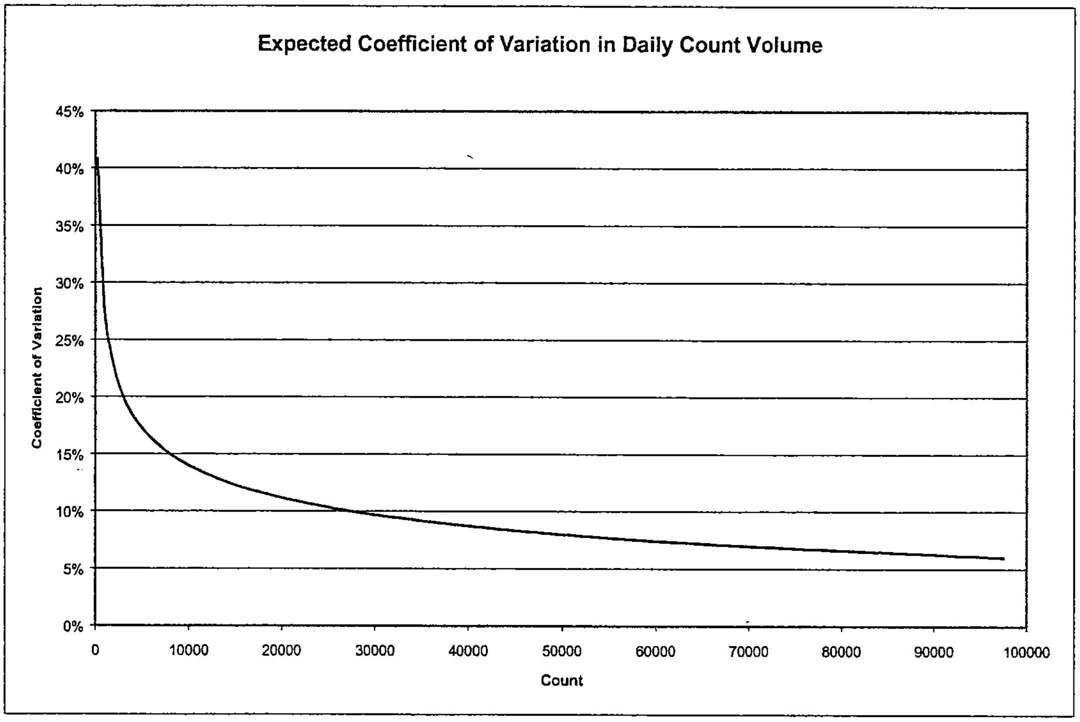
- Model validation involves checks of every step of the travel demand model process. This document only seeks to describe validation checks of the final traffic assignment volumes, not those needed for intermediate model steps.

- Because a readily available source of speed data is not available, assignment speeds are generally not validated and therefore are not used in Ohio for post processing such as air quality analysis.

- When making assignment volume checks, it is important to realize that there is error inherent in traffic counts.

- When gross discrepancies between a model volume and a count exist the counts should be verified first to ensure that the count is at least representative of the volume on the given roadway.

- In Ohio, all 24 hour model validation work shall report the proportion of links whose percent error with respect to counts exceeds the "Expected Coefficient of Variation in Daily Count Volume" curve shown in the following figure. This is not required of hourly validations. At this time there is no specific standard as to the maximum value, however, values less than 33% will be regarded with caution by ODOT.



- This curve can be represented by the equation CV=(3.706633/ln(COUNT))-0.264598 and in tabular form as follows:

count	cv	count	cv
250	41%	11250	13%
1000	27%	13750	13%
2000	23%	16250	12%
3000	20%	18750	11%
4000	18%	22500	11%
5000	17%	30000	10%
6250	16%	45000	8%
7750	15%	65000	7%
9250	14%	97500	6%

<u>There are four basic checks that should be made on a base year assignment. These are the root mean square error check, vehicle miles traveled (VMT) check, screen line check and plot check of individual link volumes. In addition, a fifth check, the select link check, is made when necessary to diagnose assignment problems.</u>

Assignment checks are not conducted on base year assignments that benefited from the aid of count based trip table adjustments, rather they are based upon model results from the model as it will function in forecast years. Count based trip table estimation is best used to produce a validation trip table against which the demand model trip table can be compared.

Plot Check

- A plot is made of the network with assignment volume and ground count annotate on each link. Each link is inspected for assignment accuracy and routes are analyzed for assignment consistency.

% Root Mean Square Error Check

- The volume group ranges to be used for %RMSE validation of traffic assignment models in Ohio is as follows:

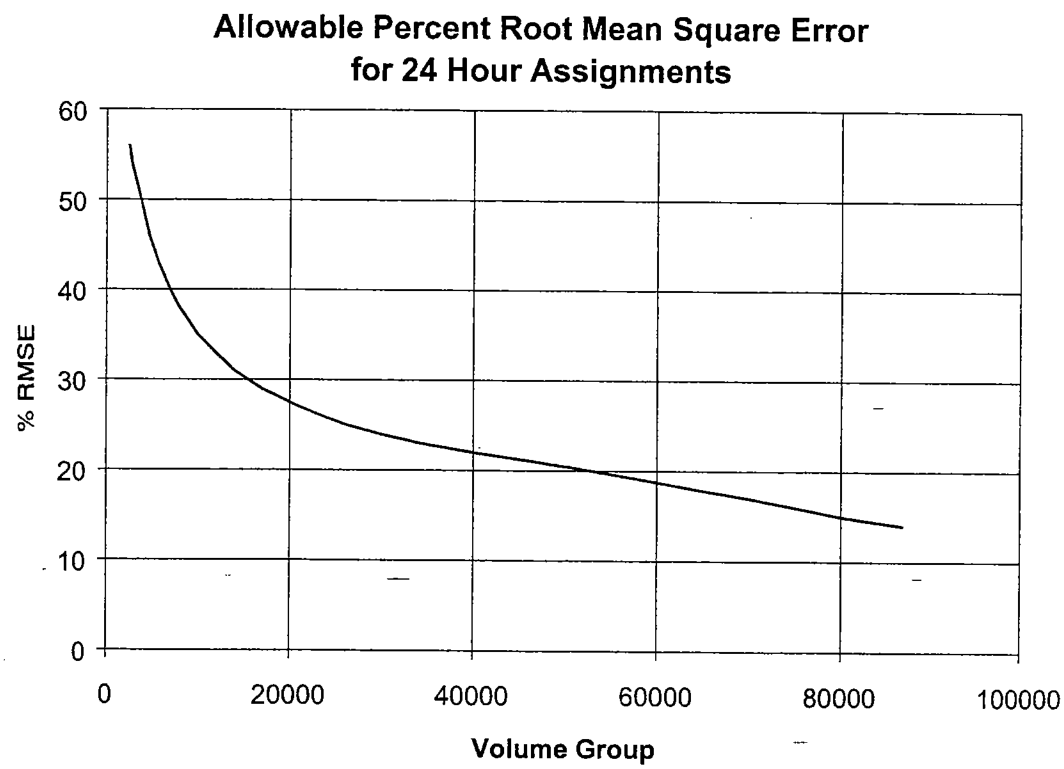
Required Group #	Daily Assignment	Hourly Assignment
1	0-499	0-49
2	500-1499	50-149
3	1500-2499	150-249
4	2500-3499	250-349
5	3500-4499	350-449
6	4500-5499	450-549
7	5500-6999	550-699
8	7000-8499	700-849
9	8500-9999	850-999
10	10000-12499	1000-1249
11	12500-14999	1250-1499
12	15000-17499	1500-1749
13	17500-19999	1750-1999

Optional Group #	Daily Assignment	Hourly Assignment
14	20000-24999	2000-2499
15	25000-34999	2500-3499
16	35000-54999	3500-5499
17	55000-74999	5500-7499
18	75000-120000	7500-12000

- In some cases a study area will have very few if any counts in the higher volume groups. In these cases the highest volume groups from the optional group may be dropped if they contain no data. In addition, if the highest optional volume groups contain very few counts (less than 10) they may be grouped together into a larger volume group. When combining groups, the group number should be listed as the lowest and highest group number separated by a hyphen. Examples of an areas top two groups could be 13 and 14-15 or 14 and 15-17 or 12 and 13 etc.

- Summaries of %RMSE for links without ground counts should not be analyzed.

- The add a lane, drop a lane criteria results in a curve of maximum %RMSE (4) for a 24 hour assignment as follows:



B

- This curve is shown in tabular format below for both daily and hourly assignments:

Group	Group Midpoint (Daily)	Group Midpoint (Hourly)	Desired %RMSE
1	250	25	200%
2	1000	100	100%
3	2000	200	62%
4	3000	300	54%
5	4000	400	48%
6	5000	500	45%
7	6250	625	42%
8	7750	775	39%
9	9250	925	36%
10	11250	1125	34%
11	13750	1375	31%
12	16250	1625	30%
13	18750	1875	28%
14	22500	2250	26%
15	30000	3000	24%
16	45000	4500	21%
17	65000	6500	18%
18	97500	9750	12%

- The %RMSE from each of the defined volume groups should be plotted vs. this curve.

- There is no specific criteria for the over-all %RMSE. A good rule of thumb, however, is that over-all %RMSE should be about 40% or less.

Vehicle Miles Traveled Check

- Regionwide VMT must be within 3% of the count total.

- With respect to functional class, this check should result in VMT's within 7% of the counts on freeways, 10% on arterials and 15% on lower classes. With respect to administrative class, the VMT should be within 7% of counts on interstates, 10% for state and US routes and 15% for all others. Certain classes may have only a few links with counts (less than 10) such as local roads and township roads. These classes may have higher errors than those specified.

- Rings and sector are aggregates of zones which have been coded in most networks and should be maintained. Generally the individual rings and sectors should be within about 10% of the ground count VMT. This is often impossible to achieve, however, for the CBD ring and sector.

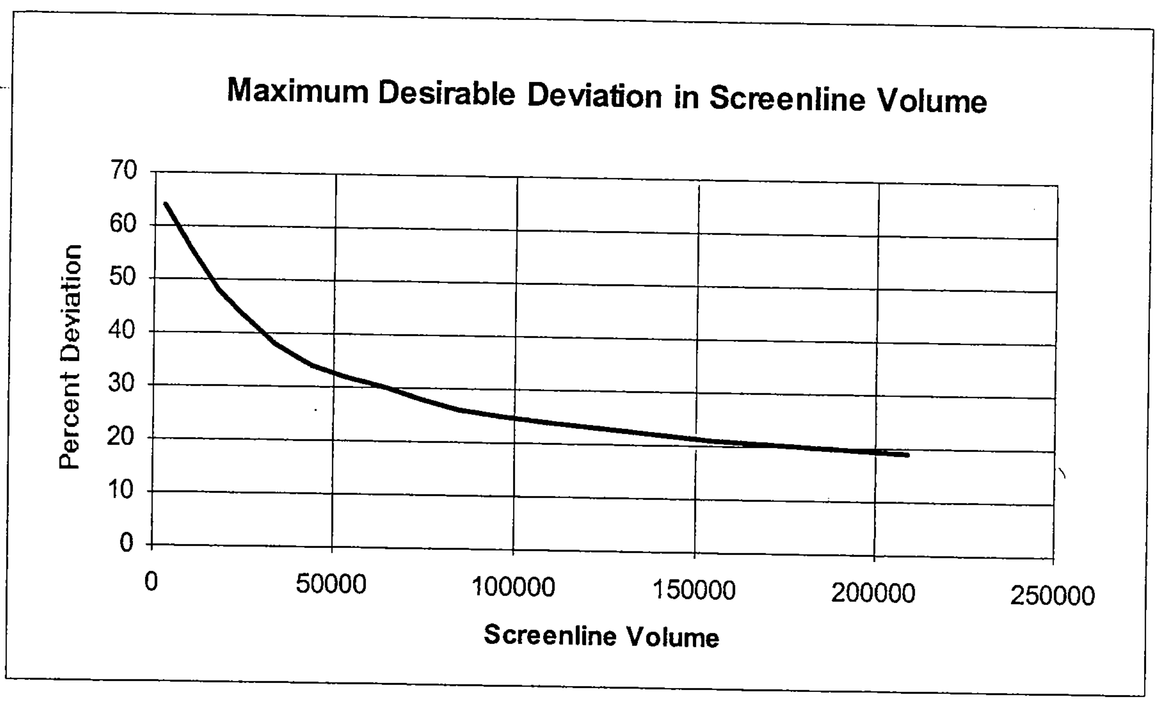
Screen Line Check

- Each area has a number of screen lines and cordon lines which have been traditionally used. These lines are not sacred, however, and may be changed to suit the analysis at hand.

- Screen lines should be drawn taking into account the following factors:

- They should intersect only parallel routes which form a corridor. There should be no diagonal streets. This only applies to auxiliary screen lines. The main screen line and cordon lines intersect any link regardless of direction.
- They should not intersect centroid connectors. If they must, the centroid connector volume is not included in the screen line analysis.
- They should include from 3 to 7 roadways. Again this only applies to auxiliary screen lines, the main screen line intersects as many roads as necessary to bisect the area and the cordons intersect as many as are necessary to delineate a cordon.
- They should be drawn to intersect the least number of roads as possible to delineate a cordon or corridor. Thus they should be drawn along natural choke points such as streams and railroads if possible.
- They should be drawn such that all links intersected have ground counts coded. This criteria may sometimes cause existing screen lines to be adjusted slightly so that all links have counts.

- The following curve given in NCHRP 255 (23) is used to check screen line differences for 24 hour assignments:



B

- This curve is shown in tabular format below for both 24 hour and hourly assignments:

Volume **	3	10	14	18	24	29	34	44	53	65	74	85	96	108	154	209
Allowable %Error	64	56	52	48	44	41	38	34	32	30	28	26	25	24	21	19

**The listed volumes are in 1000's for daily assignments and in 10's for hourly assignments

Select Link Check

- Select link analysis is not a required validation check and their are no standards for comparison.

Assignment Refinements

Post Refinement of Assignment Output

- Refining of assignment results is necessary for most project level analysis, particularly those meant to produce design traffic. Traffic assignment results should only be used directly for early planning analysis and aggregate system wide analysis such as generation of system wide VMT, congestion and air pollution burden. At the project level, the relationship between final base year assigned volume and ground counts should be consulted when using traffic assignment forecasts.

Refinement of the Assignment Process

- When using AON assignment there are basically only 2 things that can be changed, first, turn penalties can be used or discarded from the network, however it is recommended that they be used. Second, instead of doing 24 hour assignment, hourly assignments as discussed previously might be used.

- When using capacity restraint assignments, in addition to the modifications listed with AON assignments, the BPR curves can be modified (if local data has been collected which supports this change).

- If using equilibrium assignment, use of hourly assignments is recommended. In addition, it is possible to adjust the number of iterations of equilibrium.

- As mentioned previously, the VMT check may point out problems with trip generation or auto occupancy rates which may make it necessary to factor up or down certain trip generation/auto occupancy rates or special generators may need to be added if the problem is more localized.

- The screen line check might also point out trip generation/auto occupancy problems, but might also point out problems with trip distribution. When the distribution of trips between areas is the problem it is sometimes necessary to penalize or expedite travel across the screen line by changing the link impedances of all screen line links by a certain amount for trip

APPENDIX C

TRAFFIC DATA EDIT PROCEDURES - TDQ PROTOTYPE SOFTWARE RULE LIST

(Source: FHWA, Office of Highway Policy Website)

Traffic Data Edit Procedures - TDQ prototype software Rule List

Rule_ID A.3 Rule #	Rule Name	Rule Description
0 V42	Date is Correct and Unique	If the date of the input data is not correct or unique, the record will not be loaded into the database. An input error message will be reported.
1 V43	Lane and Direction are Correct	If the lane or direction fields in the input data do not match the station record, the input data will not be loaded into the database. An input error message will be reported.
2 C49	Number of Axles = Number of Axle Spaces + 1	Any vehicle record where the number of axles does not equal the number of axle spaces plus one will be flagged.
3 W70	Number of Axles = Number of Axle Weights	Any vehicle record where the number of axles does not equal the number of axle weights will be flagged.
4 W35	Sum of Axle Weights Does Not = GVW	Any vehicle record where the sum of the axle weights does not equal the recorded GVW will be flagged.
5 V1	Completeness of Data	If the input data is insufficient or invalid in any way, an error message will be reported.
6 V2	Zero Volume for an Hour	Any hourly volume of zero in any lane will be flagged.
7 V4	Extreme Hourly Volume per Lane	The hourly volume in any lane will be reported as anomalous if exceeds this global extreme maximum:
8 V32	1:00 AM to 2:00 AM Volume vs. 1:00 PM to 2:00 PM Volume	If the 1:00 AM to 2:00 AM volume is greater than the 1:00 PM to 2:00 PM volume of the same day, a warning will be reported.
9 C1	No Classification Data	If no volumes for any vehicle classes are present in the input data, an error message will be reported.
10 W51	Record Contains Valid Date	Any vehicle record containing an invalid or unexpected date will be flagged
11 W52	Record Contains Valid Lane Number	Any vehicle record containing a lane that does not match the station record will be flagged.
12 W53	Record Contains Valid Class Number	Any vehicle record containing an invalid class number will be flagged.
13 C24	Number of Axles Min/Max	Any vehicle having more or less than the number of axles in this range will be flagged:
14 W36	Wheelbase Exceeds Value for Class	Any vehicle of this class having a recorded wheelbase greater than this maximum will be flagged:
15 W39	GVW Exceeds Value for	Any vehicle of this class having a recorded

	Class	GVW greater than this maximum will be flagged:
16 W28	Front Overhang Out of Range	Any vehicle with a front overhang outside of this range will be flagged:
17 W26	Rear Overhang Out of Range	Any vehicle with a rear overhang outside of this range will be flagged:
18 W30	Sum of Axle Spaces > or = Recorded Vehicle Length	Any vehicle where the sum of the axle spaces is greater than the recorded vehicle length will be flagged.
19 W24	Record Contains Off-Scale Warning	Any vehicle record containing a vendor's off-scale warning code will be flagged.
20 W46	Wheelpath Imbalance Exceeds Threshold	Any vehicle with the total weight on one side exceeding the total weight on the other side by more than this maximum will be flagged:
21 C35	Vehicle Exceeding Speed Min/Max	Any vehicle with a recorded speed outside of this range will be flagged:
22 W25	Extreme Speed	Any vehicle with a recorded speed greater than this global extreme maximum will be flagged:
23 W43	Heavy Class 6 Vehicle With Close Follower	Any class 6 vehicle with an excessive GVW and followed within 2 seconds by another vehicle will be flagged.
24 C26	Extreme Axle Spacing	Any vehicle with any axle space greater than this maximum will be flagged:
25 C27	Minimum First Axle Space	Any vehicle with a first axle space (following the steering axle) less than this minimum will be flagged:
26 C28	Minimum Subsequent Axle Space	Any vehicle with any axle space less than this minimum will be flagged:
27 C29	Minimum Spacing Between Axle Groups	Any vehicle with a tandem or tridem axle space less than this minimum will be flagged:
28 W37	Axle Spacings vs. Min/Max Default Values for Class	Any vehicle of this class will be flagged if this particular axle space is greater or less than this range :
29 W40	Axle Weights vs. Min/Max Default Values for Class	Any vehicle of this class will be flagged if this particular axle weight is greater or less than this range :
30 C30	3S-2 Drive Tandem Spacing	Any 3S-2 tractor with a drive tandem spacing outside of this range will be flagged:
31 W50	Class 9 Front Axle Weight vs. Default Min/Max	This rule is implemented by rule W40 in the TDQ Prototype
32 W50	Class 11 Front Axle Weight vs. Default	This rule is implemented by rule W40 in the TDQ Prototype

	Min/Max	
33 V3	Consecutive Hourly Zero Volumes	The number of consecutive zero-volume hours in any one lane will be reported as anomalous if it exceeds this daily maximum:
34 V7	Consecutive Hours with Same Non-Zero Volume	The number of consecutive hours with the same non-zero volume in the same lane will be reported as anomalous if it exceeds this daily maximum:
35 V28	Sunday Hourly Directional Split	Sunday's hourly directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
36 V28	Monday Hourly Directional Split	Monday's hourly directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances
37 V28	Tuesday Hourly Directional Split	Tuesday's hourly directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
38 V28	Wednesday Hourly Directional Split	Wednesday's hourly directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
39 V28	Thursday Hourly Directional Split	Thursday's hourly directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
40 V28	Friday Hourly Directional Split	Friday's hourly directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
41 V28	Saturday Hourly Directional Split	Saturday's hourly directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
42 V9	Hourly Volume vs. Next/Prior Day	The total hourly volume will be reported as anomalous if it is greater than or less than the total volume for that hour of the previous or following day by these tolerances:

43 V17a	Daily Directional Volume vs. AADT	The daily directional volume will be reported as anomalous if it is greater or less than the previous year's adjusted directional AADT by this tolerance:
44 V33	Daily Combined Volume vs. AADT	The daily combined volume will be reported as anomalous if it is greater or less than the previous year's adjusted AADT by these tolerances:
45 V5	Sunday Daily Directional Split	Sunday's daily directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
46 V5	Monday Daily Directional Split	Monday's daily directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
47 V5	Tuesday Daily Directional Split	Tuesday's daily directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
48 V5	Wednesday Daily Directional Split	Wednesday's daily directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
49 V5	Thursday Daily Directional Split	Thursday's daily directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
50 V5	Friday Daily Directional Split	Friday's daily directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
51 V5	Saturday Daily Directional Split	Saturday's daily directional split will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
52 C48	Full Day of Data Exists	If less than 24hours of data is present, a warning will be reported as anomalous.
53 C4	Extreme Daily Percent in Any Class Except 2	The daily percent of vehicles binned to any class except 2 (cars) will be reported as anomalous if it exceeds this maximum:
54 C37	Excessive Daily Percent	The daily percent of vehicles binned to any

	by Class	class except 2 or 3 will be reported as anomalous if it exceeds this maximum:
55 C38	Excessive Daily Volume by Class	The daily volume of vehicles binned to any class except 2 or 3 will be reported as anomalous if it exceeds this maximum:
56 W16	Unloaded Class 9 GVW Distribution Peak	The majority of unloaded class 9 GVWs are expected to fall within this weight range:
57 W16	Unloaded Class 11 GVW Distribution Peak	The majority of unloaded class 11 GVWs are expected to fall within this weight range:
58 W17	Loaded Class 9 GVW Distribution Peak	The majority of loaded class 9 GVWs are expected to fall within this weight range:
59 W17	Loaded Class 11 GVW Distribution Peak	The majority of loaded class 11 GVWs are expected to fall within this weight range:
60 W68	Percent of Vehicles With GVW Out of Range for Class	The daily percent of vehicles flagged for excessive GVW will be reported as anomalous if it exceeds this maximum:
61 W67	Percent of Vehicles With Invalid Class	The daily percent of vehicles flagged for an invalid class designation will be reported as anomalous if it exceeds this maximum:
62 W21	Average Class 9 Steering Axle Weight	The daily average class 9 front axle weight will be reported as anomalous if it falls outside of this range:
63 W21	Average Class 11 Steering Axle Weight	The daily average class 11 front axle weight will be reported as anomalous if it falls outside of this range:
64 W65	Percent of Records With Invalid Dates	The daily percent of vehicle records flagged for an invalid date will be reported as anomalous if it exceeds this maximum:
65 W66	Percent of Records With Invalid Lane	The daily percent of vehicle records flagged for an invalid lane will be reported as anomalous if it exceeds this maximum:
66 W56	Average Steering Axle Weight for Light-GVW Class 9s	The average steering axle weight of all class 9 vehicles with a GVW of less than 32,000 lbs. will be reported as anomalous if it falls outside of this range:
67 W56	Average Steering Axle Weight for Mid-GVW Class 9s	The average steering axle weight of all class 9 vehicles with a GVW of between 32,000 lbs. And 70,000 lbs. will be reported as anomalous if it falls outside of this range:
68 W56	Average Steering Axle Weight for Heavy-GVW Class 9s	The average steering axle weight of all class 9 vehicles with a GVW of more than 70,000 lbs. will be reported as anomalous if it falls outside of this range:
69 W56	Average Steering Axle Weight for Light-GVW	The average steering axle weight of all class 11 vehicles with a GVW of less than 32,000

	Class 11s	lbs. will be reported as anomalous if it falls outside of this range:
70 W56	Average Steering Axle Weight for Mid-GVW Class 11s	The average steering axle weight of all class 11 vehicles with a GVW of between 32,000 lbs. And 70,000 lbs. will be reported as anomalous if it falls outside of this range:
71 W56	Average Steering Axle Weight for Heavy-GVW Class 11s	The average steering axle weight of all class 11 vehicles with a GVW of Class 11s more than 70,000 lbs. will be reported as anomalous if it falls outside of this range
72 W58	Percent of Class 9s With Front Axle Weight Flags	The daily percent of class 9 vehicles flagged for an out-of-range front axle weight will be reported as anomalous if it exceeds this maximum:
73 W58	Percent of Class 11s With Front Axle Weight Flags	The daily percent of class 11 vehicles flagged for an out-of-range front axle weight will be reported as anomalous if it exceeds this maximum:
74 C2	Percent of Records With Vendor Warning Codes	The daily percent of vehicle records containing a vendor's warning code will be reported as anomalous if it exceeds this maximum:
75 W62	Percent of Vehicles Where GVW Is Not = Sum of Axle Weights	The daily percent of vehicle records where the GVW is not equal (within rounding error) to the sum of the axle weights will be reported as anomalous if it exceeds this maximum:
76 W60	Percent of Vehicles With Overhang Flags	The daily percent of vehicles with overhang flags will be reported as anomalous if it exceeds this maximum:
77 W8	Percent of Vehicles Where Length < Wheelbase	The daily percentage of vehicles where the sum of the axle spaces is greater than the recorded vehicle length will be reported as anomalous if it exceeds this maximum:
78 W10	Class 9 Average Length Within Range + Average Wheelbase	The average class 9 vehicle length and wheelbase relationship will be reported as anomalous if the average length is not within the sum of the average wheelbase and this range:
79 W10	Class 11 Average Length Within Range + Average Wheelbase	The average class 11 vehicle length and wheelbase relationship will be reported as anomalous if the average length is not within the sum of the average wheelbase and this range:
80 W45	Percent of Records With Off-Scale Warnings	The daily percent of vehicle records containing a vendor's off-scale warning will be reported as anomalous if it exceeds this maximum:
81 W47	Pattern of Vehicles With	An otherwise anomalous percent of wheelpath

	Wheelpath Imbalance	imbalances will not be reported as anomalous if opposite wheelpath imbalances are detected in opposite directions (likely due to crosswinds).
82 W54	Percent of Vehicles With Wheelpath Imbalance	The daily percent of vehicles with wheelpath imbalance flags will be reported as anomalous if it exceeds this maximum:
83 W59	Percent of Vehicles that Exceed Extreme Max Speed	The daily percent of vehicles with globally extreme speed flags will be reported as anomalous if it exceeds this maximum:
84 C40	Percent of Vehicles Slower Than Speed Min	The daily percent of vehicles with speeds less than the station minimum will be reported as anomalous if it exceeds this maximum:
85 C40	Percent of Vehicles Faster Than Speed Max	The daily percent of vehicles with speeds greater than the station maximum will be reported as anomalous if it exceeds this maximum:
86 W61	Percent of Heavy Class 6 Vehicles With Close Follower	The percent of class 6 vehicles flagged for excessive GVW with a closely following vehicle will be reported as anomalous if it exceeds this maximum:
87 C15	Average 3S-2 Drive Tandem Spacing	The daily average drive tandem spacing for 3S-2 vehicles will be reported as anomalous if it falls outside of this range:
88 W63	Percent of Vehicles With Wheelbase or Axle Spacing Flags	The daily percent of vehicles with wheelbase or axle spacing flags set by the default values for their class will be reported as anomalous if it exceeds this maximum:
89 W64	Percent of Vehicles With an Axle Weight Flag	The daily percent of vehicles with an axle weight flag set by the default values for their class will be reported as anomalous if it exceeds this maximum:
90 W55	Average Left Axle Weight vs. Average Right Axle Weight	The average left and right axle weights for all vehicles will be reported as anomalous if they differ by more than this maximum percent:
91 V19	Hourly Directional Volume vs. History	An hourly directional volume will be reported as anomalous if it differs from its historical minimum or maximum for that hour by more than these tolerances:
92 V40	Hourly Combined Volume vs. Recent History	An hourly combined volume will be reported as anomalous if it differs from its historical minimum or maximum for that hour by more than these tolerances:
93 V39	Daily Combined Volume vs. Recent History	A daily combined volume will be reported as anomalous if it differs from its historical minimum or maximum by more than these

		tolerances:
94 V17b	Daily Directional Volume vs. History	A daily directional volume will be reported as anomalous if it differs from its historical minimum or maximum by more than these tolerances:
95 V29	Daily Percent Distribution by Lane vs. History	The daily lanal distribution will be reported as anomalous if any lane differs from its historical average percent by more than these tolerances:
96 C12	Daily Volume Binned to One Class vs. History	The daily volume binned to a single vehicle class except 2 or 3 will be reported as anomalous if it differs from its historical minimum or maximum volume by more than these tolerances:
97 C11	Daily Percent Binned to One Class vs. History	The daily percent binned to a single vehicle class will be reported as anomalous if it differs from the historical average percent for that class by more than these tolerances:
98 C23	Daily Volume of Both Class 6 and 1 Exceed History	The daily volumes of class 1 and class 6 vehicles will be reported as anomalous if both are greater than their average historical values.
99 C22	Daily Ratio of Class 2 to 3 vs. History	The daily ratio of class 2 vehicles to class 3 vehicles will be reported as anomalous if the number of class 2s per one class 3 varies by more than these tolerances:
100 C42	Daily Ratio of Class 9 to 8 by Lane vs. History	The daily ratio of class 9 vehicles to class 8 vehicles in a lane will be reported as anomalous if the number of class 9s per one class 8 differs from the historical minimum or maximum ratio by more than these tolerances:
101 C19	Daily Ratio of Class 9 to 8 by Direction vs. History	The daily ratio of class 9 vehicles to class 8 vehicles in each direction will be reported as anomalous if the number of class 9s per one class 8 differs from the historical minimum or maximum ratio by more than these tolerances:
102 C41	Daily Sum of Class 8 and 9 vs. History	The daily sum of class 8 and class 9 vehicles will be reported as anomalous if it differs from the historical minimum or maximum sum of these two classes by more that these tolerances:
103 C14	Daily Class 8 Directional Split vs. History	The daily directional split percentages for class 8 vehicles will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:

104 C13	Daily Class 9 Directional Split vs. History	The daily directional split percentages for class 9 vehicles will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
105 C43	Daily Sum of Class 8 and 9 Directional Split vs. History	The daily directional split percentages for the sum of class 8 and class 9 vehicles will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more than these tolerances:
106 C46	Daily Directional Split of Any Class (not 8 or 9) vs. History	The daily directional split percentages for any vehicle class will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more that these tolerances:
107 C17	Daily Directional Split of Sum of Class 4 thru 13 vs. History	The daily directional split percentages for the sum of all commercial vehicles will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more that these tolerances:
108 C47	Daily Directional Split of Class Groups vs. History	The daily directional split percentages for any class group (passenger, truck, semi-truck and multi-trailer) will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more that these tolerances:
109 C16	Monthly Directional Split of Sum of Class 4 thru 13 vs. History	The monthly directional split percentages for the sum of all commercial vehicles will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more that these tolerances:
110 C47	Monthly Directional Split of Class Groups vs. History	The monthly directional split percentages for any class group (passenger, truck, semi-truck and multi-trailer) will be reported as anomalous if the leading direction's percentage varies from its historical minimum or maximum by more that these tolerances
111 W18	Unloaded Class 9 GVW Distribution Peak Shift	A shift in the unloaded GVWs for class 9 vehicles will be reported if the central tendency of the input data is not within these percents of the historical central tendency
112 W19	Loaded Class 9 GVW Distribution Peak Shift	A shift in the loaded GVWs for class 9 vehicles will be reported if the central tendency of the input data is not within these percents of the historical central tendency

113 W23	Loaded vs. Unloaded Class 9 GVW Distribution Peaks	A parallel shift in Class 9 GVWs will be reported if the loaded central tendency's shift from its historical value minus the unloaded central tendency's shift from its historical value is not within these percent tolerances:
114 W20	Incidental Class 9 GVW Distribution Peak Shift	A shift in the major incidental GVW peak for class 9 vehicles (if there is one) will be reported if the central tendency of the input data is not within these percents of a matching historical central tendency
115 W18	Unloaded Class 11 GVW Distribution Peak Shift	A shift in the unloaded GVWs for class 11 vehicles will be reported if the central tendency of the input data is not within these percents of the historical central tendency
116 W19	Loaded Class 11 GVW Distribution Peak Shift	A shift in the loaded GVWs for class 11 vehicles will be reported if the central tendency of the input data is not within these percents of the historical central tendency
117 W23	Loaded vs. Unloaded Class 11 GVW Distribution Peaks	A parallel shift in Class 11 GVWs will be reported if the loaded central tendency's shift from its historical value minus the unloaded central tendency's shift from its historical value is not within these percent tolerances:
118 W20	Incidental Class 11 GVW Distribution Peak Shift	A shift in the major incidental GVW peak for class 11 vehicles (if there is one) will be reported if the central tendency of the input data is not within these percents of a matching historical central tendency
119 C6	Daily Average Speed per Lane vs. History	The average vehicle speed in each lane will be reported as anomalous if it differs from the historical average speed for that lane by more than these tolerances:

APPENDIX D

MDOT TRAFFIC COUNT METHODOLOGY

(Source: Unpublished summary of MDOT procedures provided by Larry Whiteside, MDOT)

Overview

In May of 2001, the Federal Highway Administration published the "Traffic Monitoring Guide" (TMG) which outlines procedures for the collection and analysis of traffic volume data, vehicle classification data and truck weight data. The current Guide updates the earlier publications. The philosophy of the revised Guide is unchanged. The Guide also outlines procedures for establishing statistical interpretations of the reliability of seasonal factorization and groupings and provides guidance crossing a broad spectrum traffic counting programs.

The methodology for the development of seasonal factors to expand short counts to Annual Average Daily Traffic (AADT) is addressed in Section 3 of the TMG. This approach is employed by the Michigan Department of Transportation (MDOT) in the development of the seasonal (monthly and day-of-week) factors used to expand 48 hour counts. This section also addresses the duration of counts, cycle, special needs counts, and adjustment factors.

MDOT has strived to match those presented in the guide’s section covering “Uses for Traffic Volume Data” Section 3-3. Our traffic monitoring program seeks to that collects hourly volumes by direction and lane to meet our user needs. They are used to address traffic count need with warrant studies (signals), turning movement studies, air quality analysis, noise studies, project specific planning studies, and travel demand modeling calibration needs. Attention is given the HPMS program requirements and compliance.

The traffic monitoring program development by the MDOT seeks to provide the extensive counts needed for seasonal adjustment factor development and to create the limited count coverage necessary for our users needs. This is based on Section 3-6 of the Guide regarding short duration volume counts. There are over 3988 segments of road in the current MDOT count program, with over 29,000 possible traffic count stations available.

The short count program collects at least 24 hours for special needs counts, and for the AADT program, at a minimum 48 hours. The recommended count length from the Guide for HPMS segments is a 48 hour monitoring period for both classifications and vehicle counts (TMG Section 3-10).

MDOT is on a two year cycle for traffic counts, and seek to count our segments every other year while allowing one year to catch-up on bad counts and where construction impaired collection efforts (TMG Section 3-12). There are ramp counts in metro areas that are collected every three years. We have over 140 permanent traffic recorders (PTRs) in operation on a variety of road types that are used to address our limited continuous counts needs.

The Guide provides a great deal of flexibility related to special needs traffic monitoring (TMG Section 3-16). Our methodology has been to collect at a minimum, 24 hour to 48 hours counts (in 1 hour or 15 minute intervals) for our various special studies (ramp counts, turning movement studies, and planning projects). This is done so that the results can be used in conjunction with the AADT monitoring program.

The adjustments factors that MDOT develops and use to modify our short counts for temporal bias into the annual average daily traffic volumes (AADT) are derived from our traffic recorders. They are in the form of a day-of-week and seasonal (monthly) factor. Our traffic monitoring program also uses axle adjustments and growth factors as needed to report AADT. (TMG Section 3-19)

Scheduling counts/classifications

The MDOT statewide crews collect counts at rest areas, airports, and trunkline roads. The Traffic Monitoring Unit submits a list of locations on the trunkline system to be counted to the MDOT Statewide staff to schedule. These locations are based on an even/odd year cycle with a catch up process if they weren't collected in the previous year. Scheduling is done in March of each year. The data collection effort starts in April and ends in November. The entire state is divided into six count areas. There are six crew members, one for each count area.

Office procedures

Traffic counts are collected by MDOT field staff and appended together into an ASCII file on disk. This is done programmatically. The raw data is provided in two format types; one format for vehicle counts and a second format for classification data. The raw data files are in an MDOT required format. This is a requirement that MDOT made to the traffic count manufacturers as a requirement for our purchase. The disks are then provided to the office staff for additional processing. Each count is recorded when they are received in the office, processed (converted into DBF file format), and stored.

Quality controls

The counts and classifications are manually evaluated for quality. This is done by creating a report which compares the AADT as generated from the data against the previous year's AADT. These reports use the monthly and daily seasonal adjustment factors to generate a preliminary AADT based on the data. These reports are manually reviewed for both distribution of the traffic, abnormalities, and change against previous years AADT. With classification data, a similar procedure is done in which the axle adjustments are compared to last year factors as well as a computed AADT.

Archive/Storage

After the counts and classifications have been reviewed and accepted, they are stored in two places; in our Oracle Corporate Database, and in our section FoxPro Database. Although the counts are stored, they can still be edited or deleted with tools in Oracle or in FoxPro.

Status

Once the counts have been stored in the corporate data bases, the counts/classifications are record in our status file. This status file (database) is used to report out back to the MDOT field staff several key monitoring features; the number of counts/classifications completed, whether a location needs to be reset because of bad data, and how many they have to set. This is done monthly.

APPENDIX E

MDOT AND SEMCOG TRAFFIC COUNT DATA FORMATS

MDOT Short Count Format

Record Length = Characters　　　　Blocking Factor = Records

Columns	Length	Description
1-2	2	CO (County)
3-5	3	CS_ (Rest of control section)
6-10	5	Milepoint (Mile Point, 99.999 implied)
11-14	4	Stat (Station Number)
15-21	7	PRNUM (Primary Route Number)
22-26	5	P.R. Mile Point 99.999 implied
27-30	4	Year
31-32	2	MO (Month)
33-34	2	DA (Day)
35	1	DW (Day of Week Code: Sunday = 1, Saturday = 7)
36-37	2	Dir (Direction of Travel: 01= N., 02 = N.E., 09 = N.S., 10 = E.W., etc.)
38	1	LN (Lane Code: 0= All lanes in a direction, 1 = drive lane, 2 = 2nd lane, etc.)
39	1	Cnttyp (Type of Count: (STWD, Operation, A.Q., etc.)
40-43	4	StartHR (Ending Time of the 1st count period)
44	1	Cnstrcode (Construction Code)
45-48	4	Machine (Machine number that did the counting)
49	1	HR_15 (Code: 1 = Hourly Counts, 2 = 15 Minute Counts)
50-53	4	HOA (First 15 minute recording, blank for hourly)
54-57	4	HOB (Second 15 minutes, blank for hourly)
58-61	4	HOC (Third 15 minutes, blank for hourly)
62-66	5	H1 (Fourth 15 minute period or First hourly recording)
67-457	4	Repeats columns 48-457 for a total of 96 15 minute periods or 24 hourly periods
458-459	2	MO0 (Start Month Block ID)
460-461	2	DA0 (Start Date Block ID)
462-464	3	HRS (Total Hours in Block)
465-467	3	Filler

Site ID = 15 characters

Character 1-2	County Number	(1st 2 characters of control section, col. 1-2 above)
Character 3-6	Station Number	(station number in col. 11-14 above)
Character 7-10	Machine Number	(machine number in col. 45-48 above)
Character 11-11	Day of Week	(Day of Week in col. 35 above)
Character 12-12	Type of count	(Type of Count in col. 39 above)
Character 13-13	Construction Code	(Construction code in col. 44 above)
Character 14-15	Direction	(Direction of Travel in col. 36-37 above)

Lane code (col. 38), ending hour of the first count period (col. 40-43), Date (col. 33-34) is taken from machine.

MDOT MITS Data Format

Columns	Length	Description
1-7	7	MITSID (MITS center ID
8-9	2	CONSEC (County code)
10-12	3	RESTOFSECT(Rest of the control section)
13-17	5	MILEPOINT (implied 3 decimal mile point)
18	1	CONSTCODE(not used)
19-20	2	DIRECTION (direction codes 1-N, 2-NE, 3-E, 4-SE, 5-S, 6-SW, 7-W, 8-NW)
21	1	LANE(0-ramp, 1-mainline)
22	1	RAMP(0-mainline, 1-off ramp, 2-on ramp)
23-26	4	YEAR(year of the data)
27-28	2	MONTH(month of the data)
29-30	2	DATE(day of the data)
31	1	DAYCODE(day of week code 1-Sun, 2-Mon, 3-Tue, 4-Wed,5-Thu, 6-Fri, 7-Sat)
32	1	CNTERROR (not used)
33-35	3	MACHINE (not used-default is 999)
36-39	4	STATION(MDOT count station number)
40	1	CNTTYPE (8-MITS)
41-45	5	HOUR1 (volume of vehicles for this hour of day)
46-50	5	HOUR2 (volume of vehicles for this hour of day)
51-55	5	HOUR3 (volume of vehicles for this hour of day)
56-60	5	HOUR4 (volume of vehicles for this hour of day)
61-65	5	HOUR5 (volume of vehicles for this hour of day)
66-70	5	HOUR6 (volume of vehicles for this hour of day)
71-75	5	HOUR7 (volume of vehicles for this hour of day)
76-80	5	HOUR8 (volume of vehicles for this hour of day)
81-85	5	HOUR9 (volume of vehicles for this hour of day)
86-90	5	HOUR10 (volume of vehicles for this hour of day)
91-95	5	HOUR11 (volume of vehicles for this hour of day)
96-100	5	HOUR12 (volume of vehicles for this hour of day)
101-105	5	HOUR13 (volume of vehicles for this hour of day)
106-110	5	HOUR14 (volume of vehicles for this hour of day)
111-115	5	HOUR15 (volume of vehicles for this hour of day)
116-120	5	HOUR16 (volume of vehicles for this hour of day)
121-125	5	HOUR17 (volume of vehicles for this hour of day)
126-130	5	HOUR18 (volume of vehicles for this hour of day)
131-135	5	HOUR19 (volume of vehicles for this hour of day)
136-140	5	HOUR20 (volume of vehicles for this hour of day)
141-145	5	HOUR21 (volume of vehicles for this hour of day)
145-150	5	HOUR22 (volume of vehicles for this hour of day)
151-155	5	HOUR23 (volume of vehicles for this hour of day)
156-160	5	HOUR24 (volume of vehicles for this hour of day)
161-170	10	TOTAL((total volume of vehicles for this day)
171-176	6	AVGTRAF (average of traffic - not used)
177-181	5	ROUTE (road name of the route, examples(M-39, US-23, I-75)
182-226	45	DESCRP (ON RAMP, OFF RAMP, LOCAL, OFF LOOP, ON LOOP, EXPRESS)
227-233	6	TRUEAVG (not used)

PTR Data Format

Record Length = Characters Blocking Factor = Records

Columns	Length	Description
1-4	4	Statnum (Station Number)
5	1	Direction (direction code 1-north ,2-ne,3-east,4-se,5-south ,6-sw,7-west ,8-nw , 9 is two way total)
6	1	Trav_lane (0,1,2)
7-10	4	All_year (year of the data 2002, 2003)
11-12	2	All_month (month of the data 1-12)
13-14	2	All_day (day 1-31)
15-19	5	All_1am (traffic volume)
20-24	5	All_2am(traffic volume)
25-29	5	All_3am (traffic volume)
30-34	5	All_4am (traffic volume)
35-39	5	All_5am (traffic volume)
40-44	5	All_6am (traffic volume)
45-49	5	All_7am (traffic volume)
50-54	5	All_8am (traffic volume)
55-59	5	All_9am (traffic volume)
60-64	5	All_10am (traffic volume)
65-69	5	All_11am (traffic volume)
70-74	5	All_12am (traffic volume)
75-79	5	All_1pm (traffic volume)
80-84	5	All_2pm (traffic volume)
85-89	5	All_3pm (traffic volume)
90-94	5	All_4pm (traffic volume)
95-99	5	All_5pm (traffic volume)
100-104	5	All_6pm (traffic volume)
105-109	5	All_7pm (traffic volume)
110-114	5	All_8pm (traffic volume)
115-119	5	All_9pm (traffic volume)
120-124	5	All_10pm (traffic volume)
125-129	5	All_11pm (traffic volume)
130-134	5	All_12pm (traffic volume)
135-141	7	Total_no
142	1	Del_flag (.T.=bad data, .F.=good data)
143-146	4	Reason_ln
147	1	Day_of_Wk (1-Sunday, 2-Monday ,3-Tuesday….7-Saturday)
148-149	2	County (1-82)

Note: PTR **direction** of "9"is the total volume for the PTR site for of day. There will be records for the directions of 1, 5, and 9 which equate to north bound, southbound and two way total by hours of the day.

SEMCOG Count Format

SEM_ID	(Long)	SEMCOG's ID number for the count location
LOCAL_ID	(Long/Text-10)	The counting agency's ID number for the count location
ST_DATE	(Short Date)	The date on which the count was begun
ST_TIME	(Short Time)	The time at which the count was begun
END_DATE	(Short Date)	The date on which the count was completed
END_TIME	(Short Time)	The time at which the count was completed
INTERVAL	(Long)	The count interval for the record, in minutes—15 or 60
TOTAL	(Long)	The total count for the 24-hour period
H01_1	(Long)	The count taken between 12:00 AM and 12:15 AM
H01_2	(Long)	The count taken between 12:15 AM and 12:30 AM
H01_3	(Long)	The count taken between 12:30 AM and 12:45 AM
H01_4	(Long)	The count taken between 12:45 AM and 1:00 AM
H01	(Long)	The total count taken between 12:00 AM and 1:00 AM
H02_1	(Long)	The count taken between 1:00 AM and 1:15 AM
H02_2	(Long)	The count taken between 1:15 AM and 1:30 AM
H02_3	(Long)	The count taken between 1:30 AM and 1:45 AM
H02_4	(Long)	The count taken between 1:45 AM and 2:00 AM
H02	(Long)	The total count taken between 1:00 AM and 2:00 AM
…………		
H24_1	(Long)	The count taken between 11:00 PM and 11:15 PM
H24_2	(Long)	The count taken between 11:15 PM and 11:30 PM
H24_3	(Long)	The count taken between 11:30 PM and 11:45 PM
H24_4	(Long)	The count taken between 11:45 PM and 12:00 AM
H24	(Long)	The total count taken between 11:00 PM and 12:00 AM

The field LOCAL_ID is the unique record identifier in the location table. Use LOCAL_ID to link the location table to the gis_link table and to link the location table to the count table.

APPENDIX F

MDOT COUNT IMPORT ALGORITHMS

Short Count

1. Insert into COUNT_MDOT table from source table
    a. Exclusions
        i. Block date fields for month (MO0) and day (DA0) not empty
    b. Transformations
        i. Count date fields MO, DA and YR are joined as ST_DATE
        ii. Count date fields are joined and incremented by one day for END_DATE
        iii. Count data fields are inserted as:
            1. H[x-1]A as H[x]_1
            2. H[x-1]B as H[x]_2
            3. H[x-1]C as H[x]_3
            4. H[x] as H[x]_4
            5. H[x] as H[x]
    c. Additional processing fields
        i. MO0 and DA0 are combined as BLOCK_DATE to help determine when a block starts and ends
        ii. HRS is inserted as NUM_HOURS to help determine the potential number of 24-hour counts in a block
        iii. LN is inserted as LANE for joining split counts
        iv. ST_TIME_TMP is the string version of ST_TIME
        v. END_TIME_TMP is the string version of END_TIME
2. Join lanes for split counts
    d. Insert duplicate of COUNT_MDOT into temporary table COUNT_MDOT_DELETE where lane is 6 or 9
    e. Open recordset of COUNT_MDOT_DELETE for scanning records
    f. Loop through the recordset
        i. Each location is determined by COUNTY, STATION, DIR
        ii. Each block is determined by BLOCK_DATE within a location
        iii. Each similar lane count is determined by the same ST_DATE within a block
        iv. Locate the first lane 6 and verify the next record is lane 9 of the same count
        v. Use the array arrH to hold the lane 6 data
        vi. Update the lane 6 record with the lane 9 data added and set the lane to 0
        vii. Skip records where the lane 6 and lane 9 identifiers do not match
    g. Delete all remaining lane 6 and lane 9 data (non-matching lanes/dates/locations/blocks)
    h. The temporary COUNT_MDOT_DELETE table is dropped
3. All H[x]_4 records are set to 0 where all H[x]_[y] fields are 0
4. All H[x] records are set to 0 where any H[x]_[y] fields are not 0
5. All records are set to INTERVAL 60
6. All records are set to INTERVAL 15 where any H[x]_[y] fields are not 0
7. Phase adjustments are made for each block start time
    i. Insert duplicate of COUNT_MDOT into temporary table COUNT_MDOT_DELETE
    j. Open recordset of COUNT_MDOT_DELETE for scanning records
    k. Loop through the recordset
        i. Each location is determined by COUNTY, STATION, DIR
        ii. Each block is determined by BLOCK_DATE within a location
        iii. ST_TIME for INTERVAL 15 is adjusted to beginning of hour (i.e. 0915 changes to 0900)

iv. ST_TIME for INTERVAL 60 is adjusted to beginning of previous hour (i.e. 0900 changes to 0800)
v. All start/end times for counts within a block are set to the ST_TIME of the first record in the block (eliminates 0100, 0115 for consecutive records)
vi. Potential number of 24-hour counts is determined by NUM_HOURS field as int_hours
vii. 24-hour count phase is determined by ST_TIME with int_hr
viii. Counts from current block are combined with counts from next block via the array arrH using int_hr for the start position and (int_hr - 1) for the end position
ix. A new block is determined if:
    1. Location is different
    2. BLOCK_DATE is different
    3. Number of 24-hour counts in a block ends
x. The COUNT_MDOT table is updated with the adjusted count data based on matching COUNTY, STATION, DIR, ST_DATE
xi. All remaining records have NULL for END_TIME and are deleted (partial counts)
xii. The temporary COUNT_MDOT_DELETE table is dropped

8. All INTERVAL 15 records are totaled as hourly totals as H[x]
9. All records are totaled from hourly totals (H[x]) as TOTAL
10. Update ST_TIME_TMP and END_TIME_TMP fields to 0000 where they are 2400
11. Update ST_TIME and END_TIME as the time format HH:MM of ST_TIME and END_TIME

<u>MITS Count</u>

1. Insert into COUNT_MDOT table from source table
    a. Exclusions
        i. Lane (LN) not equal 1 or 0
    b. Transformations
        i. Count date fields MONTH, DATE and YEAR are joined as ST_DATE
        ii. Count date fields are joined and incremented by one day for END_DATE
        iii. ST_TIME and END_TIME as 0000
        iv. INTERVAL as 60
        v. Count data fields are inserted as:
            1. 0 as H[x]_1
            2. 0 as H[x]_2
            3. 0 as H[x]_3
            4. 0 as H[x]_4
            5. HOUR[x] as H[x]
    c. Additional processing fields
        i. ST_TIME_TMP is the string version of ST_TIME
        ii. END_TIME_TMP is the string version of END_TIME
2. Delete partial counts
    a. All counts had valid starts, but some ended with 0 (96 of 51,072) and were ends of blocks (next record was new location or date greater than one day in future)
    b. Insert duplicate of COUNT_MDOT into temporary table COUNT_MDOT_DELETE
    c. Open recordset of COUNT_MDOT_DELETE for scanning records
    d. Loop through the recordset
        i. Each location is determined by COUNTY, STATION, DIR

        ii. A different block is determined when:
            1. Beginning/end of recordset
            2. Location is different
            3. The ST_DATE is more than one day in the future from previous record
        iii. Delete the last record in a block where H24 is 0 (partial count)
        iv. The temporary COUNT_MDOT_DELETE table is dropped
3. All INTERVAL 60 records are totaled
4. Update ST_TIME and END_TIME as the time format HH:MM of ST_TIME and END_TIME

<u>PTR Count</u>

1. Insert into COUNT_MDOT table from source table
    a. Exclusions
        i. MDOT delete flag (DEL_FLAG) not equal 0
        ii. Lane (TRAV_LANE) not equal 0
        iii. Direction (DIRECTION) equal 9
    b. Transformations
        i. Count date fields ALL_MONTH, ALL_DAY and ALL_YEAR are joined as ST_DATE
        ii. Count date fields are joined and incremented by one day for END_DATE
        iii. ST_TIME and END_TIME as 0000
        iv. INTERVAL as 60
        v. Count data fields are inserted as:
            1. 0 as H[x]_1
            2. 0 as H[x]_2
            3. 0 as H[x]_3
            4. 0 as H[x]_4
            5. All_[x]am as H[x]
            6. All_[x-12]pm as H[x]
    c. Additional processing fields
        i. ST_TIME_TMP is the string version of ST_TIME
        ii. END_TIME_TMP is the string version of END_TIME
2. Delete partial counts
    d. Many counts had partial count starts or ends (started with at least one 0 or ended with at least one 0) and were ends of blocks (next/previous record was different location or date greater than one day apart)
    e. Insert duplicate of COUNT_MDOT into temporary table COUNT_MDOT_DELETE (excluding COUNTY 77 and STATION 6109)
    f. Open recordset of COUNT_MDOT_DELETE for scanning records
    g. Loop through the recordset
        i. Each location is determined by COUNTY, STATION, DIR
        ii. A different block is determined when:
            1. Beginning/end of recordset
            2. Location changes
            3. The ST_DATE is more than one day in the future from previous record
        iii. Delete the last record in a block where H24 is 0 (partial count)
        iv. Delete the first record in a block where H01 is 0 (partial count)
        v. The temporary COUNT_MDOT_DELETE table is dropped
3. All INTERVAL 60 records are totaled

4. Update ST_TIME and END_TIME as the time format HH:MM of ST_TIME and END_TIME

Random Count Selection

1. This process is designed to select random 24-hour counts for MITS and PTR count types (noted by <count type> as "MIT" and "PTR") for each location and year.
2. After importing a MITS or PTR count data via the previously mentioned import process steps, copy and paste the COUNT_MDOT table (Structure and Data) as a new table named COUNT_MDOT_<count type>
3. Open the frmProcessCounts form and run one of the Random Count Selection processes via the button that matches that count type
4. The automated code behind the process runs as follows:
    a. Delete all records from the COUNT_MDOT_RANDOM_<count type> table
    b. Open a recordset (rst) using the _qry_rnd_<count type>_valid_days query
        i. This query selects the unique location (COUNTY, STATION, DIR) and number of 24-hour counts (NUM_COUNTS) at this location for each year
        ii. The following criteria are used for date range inclusions:
            1. Counts starting and ending on weekdays (Monday through Friday)
            2. Counts outside of the Holiday seasons:
                a. May 1 through May 21 (10 days padding for Memorial Day)
                b. June
                c. July 11 through July 31 (10 days padding for 4th of July)
                d. August
                e. September 11 through September 30 (10 days padding for Labor Day)
                f. October
        iii. The NUM_COUNTS field is used for random number generation
    c. Loop through the query recordset "rst"
        i. Generate a random number based off NUM_COUNTS
        ii. Re-select from the _qry_rnd_<count type>_valid_days query as a new recordset (rst2) where the current "rst" record matches on location and year
        iii. Move to the random record number (rnd_row) in "rst2"
        iv. Insert the random record data from "rst2" into the COUNT_MDOT_<count type> table as a new record