Traffic Volume Collection and Estimation on Non-Federal Aid System Roads

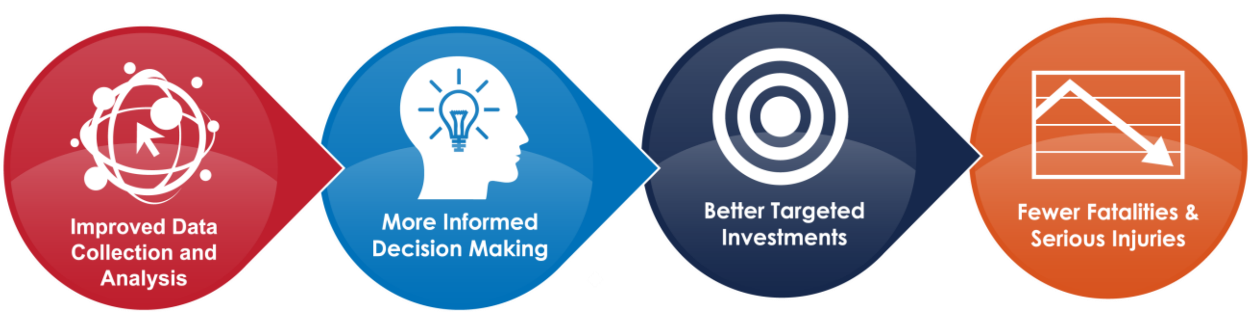
SOUTHEAST MICHIGAN COUNCIL OF GOVERNMENTS

INNOVATIVE TRAFFIC DATA QUALITY ASSURANCE/QUALITY CONTROL PROCEDURES AND AUTOMATING AADT ESTIMATION

CASE STUDY

FHWA-SA-17-035

Federal Highway Administration
Office of Safety
Roadway Safety Data Program
http://safety.fhwa.dot.gov/rsdp/



March 21, 2017

U.S. Department of Transportation
Federal Highway Administration

Safe Roads for a Safer Future
Investment in roadway safety saves lives
http://safety.fhwa.dot.gov

Notice

Quality Assurance Statement

TECHNICAL DOCUMENTATION PAGE

1. Report No. FHWA-SA-17-035	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Innovative Traffic Data QA/QC Procedures and Automating AADT Estimation		5. Report Date March 2017	
		6. Performing Organization Code	
7. Author(s) William Holik, Ioannis Tsapakis, Anita Vandervalk, Shawn Turner, John Habermann		8. Performing Organization Report No.	
9. Performing Organization Name and Address Texas A&M Transportation Institute (TTI) Texas A&M University System 3135 TAMU College Station, TX 77843-3135		10. Work Unit No.	
		11. Contract or Grant No. DTFH6116D00004	
12. Sponsoring Agency Name and Address Federal Highway Administration Office of Safety 1200 New Jersey Ave., SE Washington, DC 20590		13. Type of Report and Period Case Study, August 2016–February 2018	
		14. Sponsoring Agency Code FHWA	
15. Supplementary Notes The contract manager for this report was Stuart Thompson, Office of Safety.			
16. Abstract This case study highlights two noteworthy practices at the Southeast Michigan Council of Governments (SEMCOG) regarding short-duration traffic count validation procedures and an automated annual average daily traffic (AADT) estimation process. SEMCOG maintains a centralized traffic count database and receives traffic counts from the local agencies in southeast Michigan. SEMCOG conducts 46 validity checks on all traffic count data to identify invalid data available in the database but not adequate for analysis. After implementing the system, SEMCOG reduced labor costs associated with converting and entering data and was able to spend more time analyzing data. SEMCOG developed an algorithm that works inside its geographic information system (GIS) to improve AADT estimates by searching for uncounted segments with nearby counted segments. When a counted segment is identified, the algorithm calculates the weighted average of two nearby segments and assigns that AADT to the uncounted segment. This process was automated using Python scripts, which results in an increase in the number of AADT estimates without requiring additional field data collection.			
17. Key Words: short-duration traffic counts, AADT estimation, data quality control, data sharing, data viewer	18. Distribution Statement No restrictions.		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 24	22. Price

Form DOT F 1700.7 (8-72) Reproduction of Completed Pages authorized

ACRONYMS

Table 1. Acronyms.

Acronym	Description
AADT	Annual Average Daily Traffic
AWDT	Average Weekday Daily Traffic
DOT	Department of Transportation
GIS	Geographic Information System
LRS	Linear Referencing System
MPO	Metropolitan Planning Organization
NFAS	Non-Federal Aid System
QA/QC	Quality Assurance/Quality Control
SEMCOG	Southeast Michigan Council of Governments

EXECUTIVE SUMMARY

This case study highlights two noteworthy practices at the Southeast Michigan Council of Governments (SEMCOG) regarding short-duration traffic count validation procedures and an automated annual average daily traffic (AADT) estimation process. SEMCOG maintains a centralized traffic count database and receives traffic counts from local agencies in southeast Michigan. SEMCOG conducts 46 validity checks for all traffic count data to ensure quality data are used for analysis. Some of these data serve a purpose for a local agency and do not meet SEMCOG's requirements for short-duration traffic count data used for AADT estimation. Thus, SEMCOG identifies and excludes these invalid data from AADT analysis, but the data remain in the database. After implementing the system, SEMCOG reduced labor costs associated with converting and entering data and is able to spend more time analyzing data. SEMCOG developed an algorithm that works inside of its geographic information system to improve AADT estimates by searching for uncounted segments with nearby counted segments on either side. When such a segment is identified, the algorithm calculates the weighted average of the two nearby segments and assigns that AADT to the uncounted segment. This process was automated using Python scripts, which results in an increase in the number of AADT estimates without requiring additional field data collection.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The purpose of this case study is to highlight two noteworthy practices at the Southeast Michigan Council of Governments (SEMCOG) and demonstrate how they can be applied to non-Federal Aid System (NFAS) roads:

- Use of quality assurance and quality control (QA/QC) procedures to clean and filter data collected by local agencies.
- Use of automated annual average daily traffic (AADT) estimation procedures in a geographic information system (GIS) environment to reduce data collection requirements.

SEMCOG has a unique relationship with the public agencies in its region in that the local agencies and the Michigan Department of Transportation (DOT) collect all traffic data and SEMCOG stores, validates, and analyzes the data. Local agencies conduct two types of short-duration counts: those used to develop AADT values for road segments, and those used for special projects that may include a shorter period or only a single lane or direction of travel. Local agencies send both types of counts to SEMCOG, which then checks and validates the counts. To accomplish this task, SEMCOG developed 46 validity checks in a Microsoft Office Access database. The main benefit of these procedures is that other agencies can easily apply them to identify invalid or questionable data and increase the accuracy and reliability of traffic estimates. After implementing the system, SEMCOG reduced labor costs associated with converting and entering data and is able to spend more time analyzing data.

To improve AADT estimates on the 40,355 road segments in the SEMCOG area, SEMCOG developed a process to estimate AADT on segments meeting certain criteria instead of conducting short-duration counts on each segment. SEMCOG developed an algorithm that works inside of its GIS to search for uncounted segments with nearby counted segments on either side. When the algorithm identifies an uncounted segment, it calculates the weighted average of the two nearby counted segments and assigns that AADT to the uncounted segment. SEMCOG reports these values as interpolated segments in its traffic data viewer. This process would be time consuming to do manually, so SEMCOG automated the process using Python scripts. The benefit of the AADT estimation scripts are an increase in the number of AADT estimates that do not require additional data collection and an automated procedure to quickly estimate the AADT values.

BACKGROUND

SEMCOG is the metropolitan planning organization (MPO) for seven counties in southeast Michigan: Livingston, Macomb, Monroe, Oakland, Saint Clair, Washtenaw, and Wayne Counties. SEMCOG does not collect traffic data itself. Instead, it receives around 3,000 short-duration counts conducted at various locations every year by local agencies. The traffic data collectors apply axle correction factors to the short-duration counts. Additionally, SEMCOG develops and applies adjustment factors for day of week and month of year based on permanent stations grouped by county and functional class. Michigan DOT maintains the permanent count stations, and SEMCOG develops its own adjustment factors. SEMCOG has been archiving traffic data for about 15 years. During this period, data collectors have taken roughly 163,500 traffic counts on 40,355 individual links—or road segments delineated by functional class, county, and intersecting roads.

In the past, SEMCOG spent much time and effort converting data received from each agency into a commonly used format compatible with its database. This left little time to analyze the data. Now, SEMCOG hosts a common traffic count database system and provides software and support to all local agencies in the area to use the system. The system has some QA/QC procedures that data go through when uploaded, but SEMCOG validates all data using its own 46 validity checks when it downloads the data. This ensures all data used for analysis are valid.

SEMCOG has noteworthy practices for validating data received from other agencies that other states can use. Also, SEMCOG uses an automated process to estimate AADT values for uncounted segments. Agencies can expand this process to a statewide system to significantly reduce processing time and data collection requirements.

QA/QC PROCEDURES

DATA COLLECTION

Local agencies collect and report short-duration traffic count data to SEMCOG. Agencies conduct short-duration traffic counts on a four-year cycle and cover roughly 55 percent of the network in that four-year period. SEMCOG does not collect short-duration counts nor do they operate any permanent count stations. SEMCOG developed their own adjustment factors using data from the 35 Michigan DOT permanent count stations in the SEMCOG area. The permanent count station data are supplemented using local agency systems on traffic signals that collect continuous traffic data.

Local agencies have their own data collection plans and equipment. SEMCOG ensures that the traffic count database manufacturer supports all the data collection systems used by local agencies and helps set up traffic data collection programs as needed. SEMCOG and the local agencies have a user's group that meets to discuss data collection plans and locations. SEMCOG occasionally requests local agencies count specific locations. They will hire consultants to collect data when local agencies are not able to help. Most of the counts local agencies collect are short-duration traffic counts but may collect data for their purposes that cover a shorter period or fewer lanes or directions. SEMCOG's validity checks flag invalid data so they are not included in analysis.

Prior to 2001, SEMCOG received data from local agencies in over 10 different formats and spent too much time converting and uploading the data. As a result, SEMCOG now maintains a centralized traffic count database that is used by local agencies in the region so all data are compatible and easily transferable. SEMCOG provides the database software and login information to each local agency so users can deploy the software locally. SEMCOG works with the database system provider to ensure the traffic data collection systems that each local agency uses are compatible. This allows each agency to collect traffic data with its own equipment and processes, just as it did before, using the software. With the common software, each agency downloads data from its counters and uploads the data to the centralized database system. The software has built-in validity checks that require each agency to certify that the data are valid. However, agencies can bypass these validity checks. This typically only happens when an agency collects data for a specific project, such as peak-hour counts or one-directional counts. These data are not appropriate for AADT estimation but must be retained in the database for use by the agency that collected the data.

DATA PROCESSING

SEMCOG conducts validity checks for all traffic count data to ensure quality data are used for analysis. Since SEMCOG does not collect traffic data itself, local agencies upload their traffic count data to the central database. Some of these data serve a purpose for a local agency and do not meet SEMCOG's requirements for short-duration traffic count data used for AADT estimation. Thus, SEMCOG identifies and excludes these invalid data from AADT analysis, but the data remain in the database so that the local agency that collected the data can view and access them.

Local agencies in the SEMCOG area upload short-duration traffic count data directly into the traffic count database. SEMCOG works with the software vendor to ensure that the software can support all traffic data collection systems used by the local agencies. This eliminates the need to convert data and expedites the upload process by allowing local agencies to upload data themselves. When local agencies upload data, the software runs 15 validity checks that are built into the system. The system flags any errors that appear, and local agencies can remove the flagged data or submit all the data.

Of the 15 validity checks built into the database system, the software provider developed some, and SEMCOG specifically requested others. These procedures check for such things as missing count intervals, duplicate counts, tolerance compared to previous counts, directional split, and vehicle classification percentages.

In addition to the 15 built-in checks, SEMCOG developed 46 validity checks to clean and filter data downloaded from third-party software to SEMCOG's internal database. These procedures provide an additional layer of data cleaning and filtering to complement the data cleaning processes embedded in the third-party software database. The validity checks are conducted in Microsoft Access but implemented in MySQL. The switch is necessary because the database of count records is too large to handle efficiently with Microsoft Access.

SEMCOG stores short-duration count data in four tables in a relational database:

- **All Counts**—This table includes the location of the count, the start and end date of the count, the direction, the counts for 15-minute intervals, the total count, and the computed AADT and average weekday daily traffic (AWDT).
- **All GIS**—This table contains location information necessary to map short-duration traffic counts to roadway segments. This includes information such as segment ID, route number, milepoint, and distance. SEMCOG's and Michigan DOT use the same linear referencing system (LRS).

- **All Location**—This table describes the location where short-duration traffic counts are taken. This includes information such as road name, direction, milepoint, and cross streets.
- **All Source**—This table describes the agency that collected short-duration traffic counts.

All tables include primary and foreign keys to identify unique features and join tables. The validity checks are applied to one or more tables and include the following:

1. Is segment ID a valid primary key for the location table? Check for null and duplicate values.
2. For each record in the location table, does road name contain a non-null value?
3. For each record in the location table, does direction contain a value of either 2-Way, EB, NB, NEB, NWB, Ramp, SB, SEB, SWB, or WB?
4. For each record in the location table, does location type contain a value of either Intersection or Link?
5. For each Intersection location record, does At Road contain a non-null value?
6. For each Intersection location record, does From Road contain a null value?
7. For each Intersection location record, does To Road contain a null value?
8. For each Intersection location record, does Approach contain a value of either East Of, North Of, South Of, or West Of?
9. For each Link location record, does From Road contain a non-null value?
10. For each Link location record, does To Road contain a non-null value?
11. For each Link location record, does Approach contain a null value?
12. For each Link location record, does At Road contain a null value?
13. Is each location record unique when grouped by On Road, Direction, From Road, To Road, Approach, and At Road?
14. For each Intersection location record where Direction equals any valid value other than 2-Way or Ramp, and where Approach contains a valid value, do the directions in Approach in Direction run approximately parallel to each other?

15. Is Count ID a valid primary key for the count table?

16. Does each record in the count table match exactly one record in the location table when linked via Segment ID?

17. Does each record in the location table match at least one record in the count table?

18. For each record in the count table, does Local ID contain a non-null value?

19. For each record in the count table, does Start Date contain a non-null value corresponding to a date from between 01/01/1980 and the present?

20. For each record in the count table, does End Date contain a non-null value corresponding to a date from between 01/01/1980 and the present?

21. For each record in the count table, is End Date exactly one day after Start Date?

22. For each record in the count table, does Interval contain a value of either 15 or 60? If Interval is 24, delete all those records.

23. For each record in the count table, does each hourly field (HR01, HR02, HR03, etc.) contain a non-negative integer?

24. For each record in the count table, does the value in Total Count exactly equal the sum of the hourly values?

25. For each count record where Interval contains a value of 60, does each 15-minute field (HR01_1, HR01_2, HR01_3, HR01_4, HR02_1, etc.) contain a null value?

26. For each count record where Interval contains a value of 15, does each 15-minute field contain a non-negative integer?

27. For each count record where Interval contains a value of 15, does each hourly field exactly equal the sum of its four 15-minute fields?

28. For each record in the count table, if AWDT is not a positive integer, is there a value of X in Action?

29. For each record in the count table, if AADT is not a positive integer, is there a value of X in Action?

30. For each record in the count table, does Source ID contain a positive integer?

31. Is each count record unique when grouped by Source ID, Local ID, and Start Date?

32. Is each count record unique when grouped by Segment ID, Source ID, and Start Date?

33. Is GIS ID a valid primary key for the GIS table?

34. Does each record in the GIS table match exactly one record in the location table when linked via Segment ID? Note: All GIS records must match location records, but not all location records have corresponding GIS records.

35. For each record in the GIS table, does Primary Route Number contain a positive integer?

36. For each record in the GIS table, does Begin Milepoint contain a non-negative value?

37. For each record in the GIS table, does End Milepoint contain a positive value?

38. For each record in the GIS table, is End Milepoint ≥ Begin Milepoint?

39. For each record in the GIS table where End Milepoint = Begin Milepoint, does the record correspond with a location that is associated solely with Michigan DOT counts?

40. For each record in the GIS table where End Milepoint = Begin Milepoint, is that record the only GIS record associated with the corresponding record in the location table?

41. For each record in the GIS table, does the field Michigan Geographic Framework Value contain a reference to the appropriate Michigan Geographic Framework version number?

42. Is each GIS record unique when grouped by Segment ID and Primary Route Number?

43. Does each Primary Route Number/Begin Milepoint combination plot without error as a point?

44. Does each Primary Route Number/End Milepoint combination plot without error as a point?

45. For each location record without a corresponding GIS record, does the field GIS Comment contain an appropriate non-null comment?

46. For each location record with at least one corresponding GIS record, does the field GIS Comment contain either a null value or the value Ramp?

AADT ESTIMATION

OBJECTIVE

SEMCOG developed automated procedures to expedite high-quality AADT estimation without requiring additional data collection. SEMCOG estimates AADT in three ways:

- Observed—SEMCOG collects these counts in the field and annualizes them using adjustment factors.
- Interpolated—For segments that do not have an observed count but do have observed counts on either side of them, SEMCOG calculates the weighted average value of the adjoining segments. A Python script in ArcGIS interpolates the adjoining segment counts.
- Default—SEMCOG estimates an average value by grouping observed counts by county, number of lanes, and functional class, and then assigns this value to the remaining segments.

SEMCOG uses the interpolation and default procedures based on observed counts to provide additional coverage of AADT values without significantly increasing data collection costs.

DATA ANALYSIS

SEMCOG developed Python scripts in a GIS environment to automate the generation of AADT estimates for uncounted segments. The scripts interpolate traffic volumes from counted segments and produce AADT estimates for adjacent uncounted segments. SEMCOG counts roughly 48 percent of the network segments and uses an interpolated AADT values for about 40 percent of the network. The scripts assign a default AADT value to the remaining 12 percent of segments.

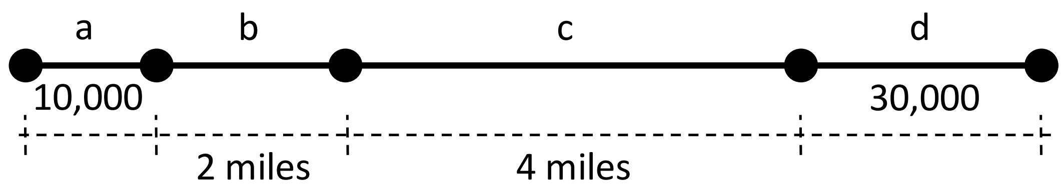
Python scripts optimize the interpolation method and quickly determine AADT values. These scripts systematically work through several steps. First, the scripts search for observed counts on links with the same primary route number. If no links of the same primary route have observed counts, then interpolation is not possible. If all links of the same primary route have observed counts, then interpolation is not required. For primary routes that have at least one link with an observed count, the scripts conduct interpolation as follows:

- The beginning and ending segments are assigned the nearest AADT value for a route of the same functional class.

- If only one observed count is found on the primary route, all links are given that AADT value.
- If multiple observed counts are found on the primary route, the inverse distance weighted average AADT is calculated for the links without an AADT value.

For example, the primary route shown in Figure 1 has known AADT values for roadway links (or roadway segments) a ($AADT_a$=10,000) and d ($AADT_d$=30,000). The AADT of links b and c is unknown.

Figure 1. Sample AADT Interpolation.



SEMCOG uses the inverse distance weighted average formula to determine the unknown AADT value. It takes the difference between the higher and lower AADT values and multiplies it by the distance from the endpoint of the link with the lower AADT value to the mid-point of the link with an unknown AADT. Then it divides it by the distance of the links with observed AADT values, all added to the lowest AADT. This can be expressed as Equation *(1)*.

$$AADT_{unc} = \frac{(AADT_{high} - AADT_{low}) * d_{End\ to\ MP}}{d_{tot}} + AADT_{low} \qquad (1)$$

Where:

$AADT_{unc}$ = the AADT being calculated for the uncounted link.

$AADT_{high}$ = the higher AADT value of the nearest link on either side of the uncounted link.

$AADT_{low}$ = the lower AADT value of the nearest link on either side of the uncounted link.

$d_{End\ to\ MP}$ = the distance from the end of the link with the lower AADT value to the mid-point of the uncounted link.

d_{tot} = the distance between links with observed AADT values on both sides of the uncounted link.

Solving Equation *(1)* for links b and c in Figure 1 results in the following:

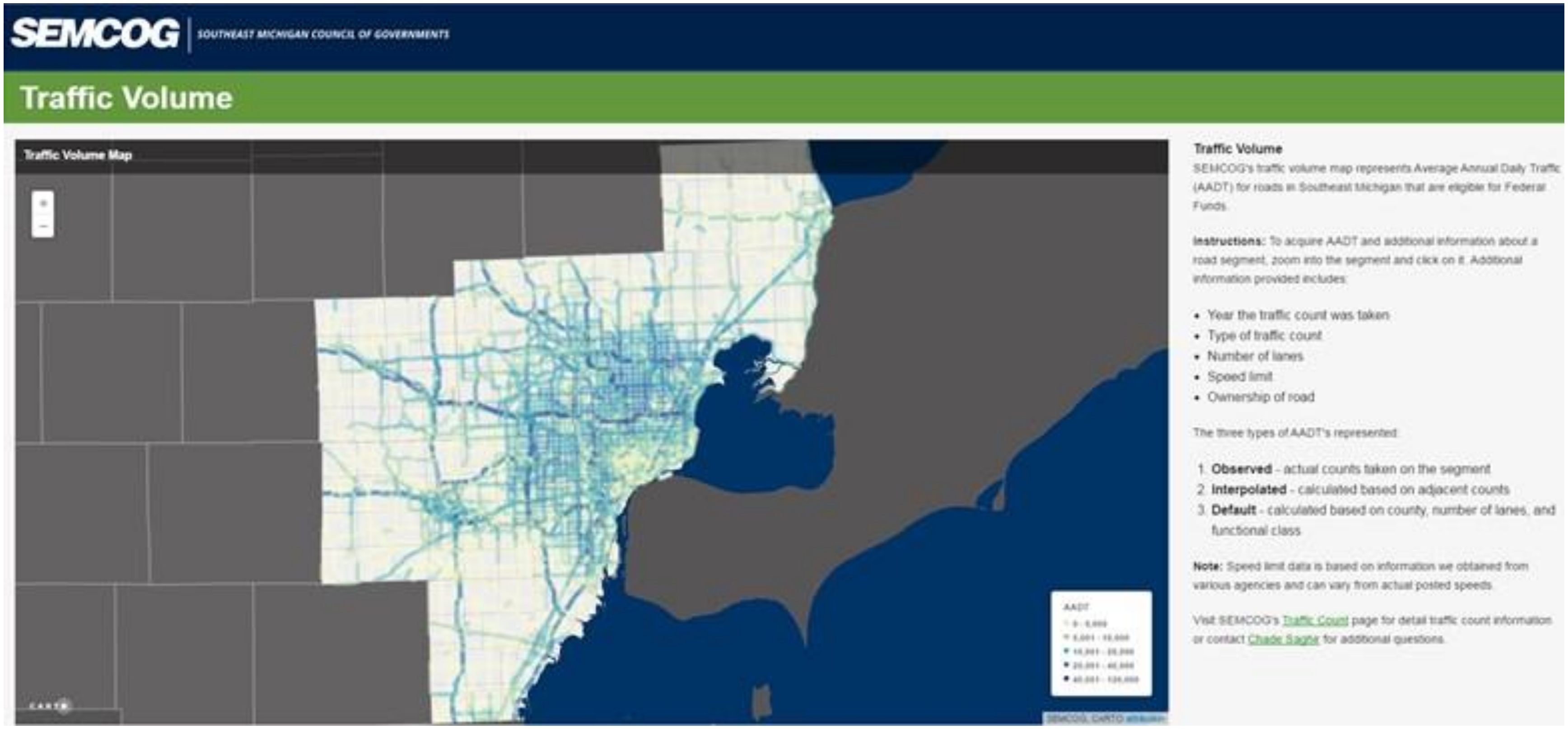
$$AADT_b = \frac{(30{,}000 - 10{,}000) * 1}{6} + 10{,}000 = 13{,}333$$

$$AADT_c = \frac{(30{,}000 - 10{,}000) * 4}{6} + 10{,}000 = 23{,}333$$

SEMCOG validated its interpolated and default data estimates by excluding observed counts and using the interpolation and default methods to estimate the AADT. The estimates compared closely with the observed value for the segment. As a result of the data validation, SEMCOG feels confident in its estimation procedures.

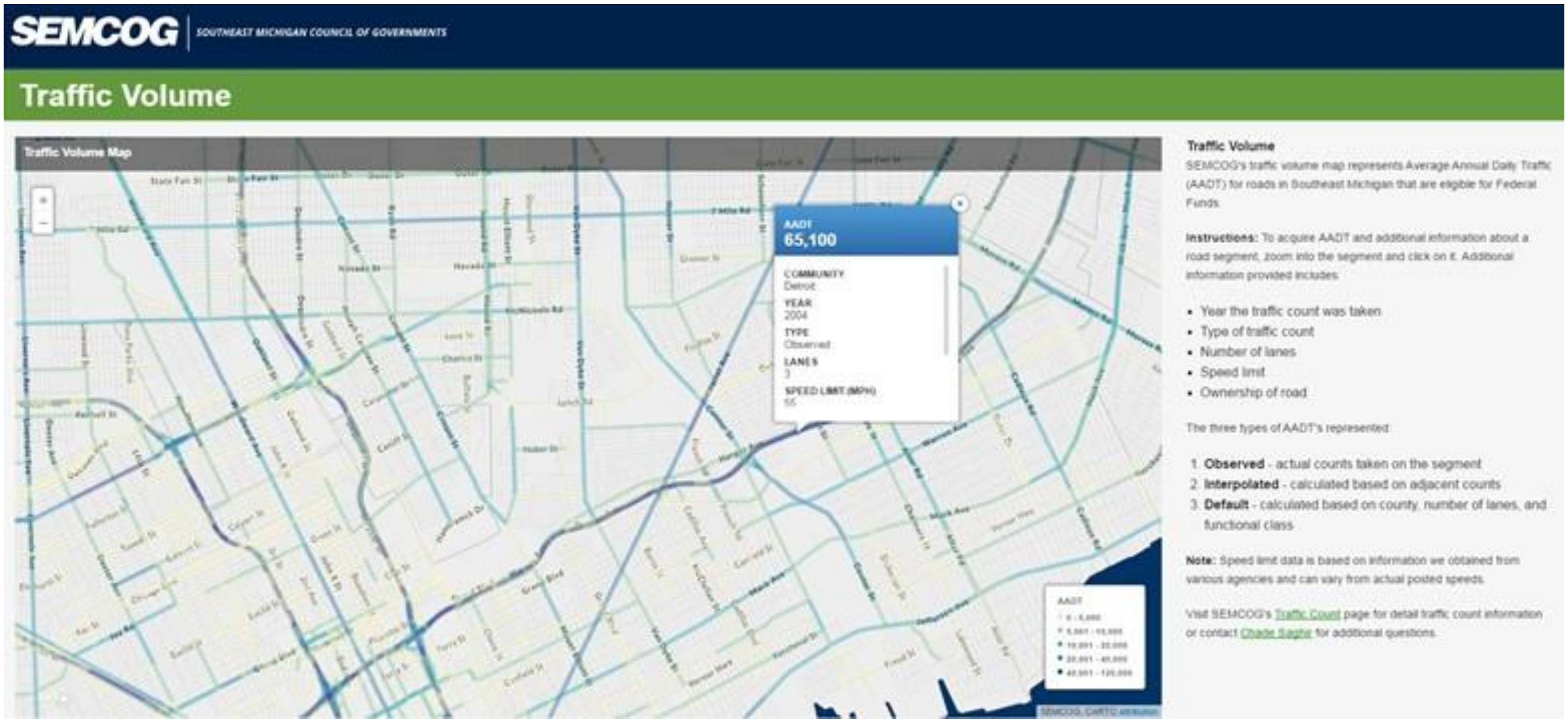
After SEMCOG staff receive data, conduct validity checks, and estimate AADT, they upload short-duration traffic counts to a web viewer. This viewer shows AADT values on roads in the SEMCOG area (Figure 2) and shows whether the AADT value is observed, interpolated, or default (Figure 3).

Figure 2. SEMCOG Traffic Volume Data Viewer.



Source: http://maps.semcog.org/TrafficVolume/

Figure 3. Traffic Volume Information for Individual Link.



Source: http://maps.semcog.org/TrafficVolume/

APPLICABILITY TO OTHER STATES

SEMCOG has implemented several innovative practices for its AADT data collection program, including 46 validity checks to ensure short-duration count accuracy and automated Python scripts to estimate and assign AADT values to uncounted roadway sections. Other agencies can adopt SEMCOG's validity checks and apply the Python scripts to interpolate AADT values to make certain elements of their AADT estimation practices more efficient. SEMCOG regularly helps other agencies develop similar validity checks and AADT interpolation code. To adopt similar validity check and AADT estimation systems, agencies would need personnel familiar with database management and programming. If an agency uses a third-party software vendor for its AADT data, the vendor may be able to implement additional specific validity checks.

QA/QC PROCEDURES

Lessons learned from SEMCOG regarding validity checks include the following:

- Vendor selection is critical. With the number of agencies and amount of customization required to fully implement the procedures, an off-the-shelf product would not have worked. SEMCOG vetted the vendor to ensure it would be able to customize its software to meet the specific needs of all agencies.
- Having in-house technical expertise helps ensure that the program accomplishes the intended purpose and is sustainable. This requires coordination and buy-in from IT personnel. Having a member of the team who is competent in SQL helps with implementing and maintaining the validity checks.
- Gathering and using data collected by local agencies can increase the number of short-duration traffic counts and eliminate duplicate data collection efforts. However, local agencies may collect data for purposes other than AADT estimation, thus creating the need for validity checks.
- Developing validity checks that exceed the procedures built into third-party software may be necessary. These checks will allow agencies to better customize procedures and handle erroneous data.
- Creating validity checks that flag and exclude erroneous data can greatly increase the accuracy and reliability of data used in different types of analyses.

AADT ESTIMATION

Lessons learned from SEMCOG regarding AADT estimation include the following:

- Keep the initial program simple. Start with a clear goal of what is important and what system can be implemented and maintained currently. Create the system to satisfy the immediate need and then expand upon it.
- Using interpolation can greatly reduce data collection needs for estimating AADT values. For example, SEMCOG estimates the AADT using interpolation on 40 percent of the roadway links in its area.
- Developing automated scripts to interpolate AADTs can make the interpolation process more efficient, resulting in time and money savings.
- By automating the interpolation process, agencies can easily validate the results by randomly selecting some counted segments and interpolating those segments as if they had not been counted. The interpolated AADT value can be compared to the actual observed AADT value. SEMCOG found the interpolated values to be accurate compared to observed values.

REFERENCES

Phone interviews and email conversations with Chade Saghir, Transportation Planner, SEMCOG.

SEMCOG, Traffic Volume Maps. http://maps.semcog.org/TrafficVolume/. Accessed December 19, 2016.

FHWA, Office of Safety

Stuart Thompson, P.E.

Federal Highway Administration

202-366-8090

Stuart.Thompson@dot.gov

ACKNOWLEDGEMENTS

The FHWA Office of Safety Data Management Systems and Processes Project Team would like to thank the following individuals, who graciously provided information needed to develop this case study.

Name	Title	Department	Email/Phone #
Chade Saghir	Transportation Planner	Southeast Michigan Council of Governments	saghir@semcog.org 313-324-3342
Delores Muller	Applications Developer	Southeast Michigan Council of Governments	muller@semcog.org 313-324-3412
Brian Mohr	Transportation Engineer	Southeast Michigan Council of Governments	mohr@semcog.org 313-324-3337