IDAHO TRANSPORTATION DEPARTMENT

RESEARCH REPORT

Off-System Public Roads Annual Average Daily Traffic (AADT) Estimation Study

RP 303

By

Peter J. Hylton, Ph.D., Bryce C. Miller

High Street Consulting

Prepared for

Idaho Transportation Department

ITD Research Program, Contracting Services

Highways Construction and Operations

28 February 2023

YOUR Safety • • • ▶ YOUR Mobility • • • ▶ YOUR Economic Opportunity

Disclaimer

Technical Report Documentation Page

1. Report No. FHWA-ID-23-303	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Off-System Public Roads Annual Average Daily Traffic (AADT) Estimation Study: Final Report		5. Report Date February 28, 2023 6. Performing Organization Code
7. Author(s) Peter J. Hylton, http://orcid.org/0000-0002-7214-6514 Bryce C. Miller		8. Performing Organization Report No. [Billing code]
9. Performing Organization Name and Address High Street Consulting 6937 Blenheim Ct Pittsburgh, PA 15208		10. Work Unit No. (TRAIS) 11. Contract or Grant No. T002785
12. Sponsoring Agency Name and Address Idaho Transportation Department (SPR) Highways Construction and Operations, Contracting Services, Research Program PO Box 7129 Boise, ID 83707-7129		13. Type of Report and Period Covered Final Report [12/01/2022 – 02/28/2023] 14. Sponsoring Agency Code RP-303
15. Supplementary Notes This project is performed in cooperation with the Idaho Transportation Department and Federal Highway Administration.		
16. Abstract This report describes a process to estimate AADT on non-Federal-Aid public roads (collectively called "off-system" roads) in Idaho. This will provide additional data for ITD and regional and local transportation agencies to conduct analysis, and it will meet new federal requirements for estimating AADT on all public roads. The report includes a review of AADT estimation approaches in research and in practice, documentation of ITD's current AADT estimation process, and summaries of federal guidance. It also includes an inventory of potential data sources. The project team proposed three methods (related to regression, travel demand models, and geospatial interpolation), and the technical advisory committee (TAC) selected the geospatial interpolation approach based on its flexibility, understandability, moderate technical requirements, and accuracy. Therefore, this report also details how to implement the geospatial interpolation method for estimating off-system AADT. Some methodological decisions remain to be made based on the data (such as the specific geospatial interpolation technique and variables to include), so the report describes the decisions to be made and how to approach them. Finally, the report provides an implementation and validation plan with next steps, implementation roles, schedule, and statistics for use in model validation. This process can be implemented using tools in Esri ArcGIS and Python.		
17. Key Words Annual average daily traffic, traffic volume, traffic estimation, non-Federal aid highways local roads, interpolation, geospatial data		18. Distribution Statement Copies available from the ITD Research Program

19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 119	22. Price None

Acknowledgments

This project has benefited from the feedback, experience, assistance, and insights of many organizations and people. These include but are not limited to—

- The Technical Advisory Committee (TAC) members, including those from the Idaho Transportation Department, the Local Highway Technical Assistance Council (LHTAC), and the Federal Highway Administration (FHWA).
- Members of the Jackalope users group who took the time to share their experience with estimating annual average daily traffic (AADT) on off-system roads, especially participants from state transportation agencies in Colorado, Iowa, Kansas, Minnesota, Missouri, Montana, Oregon, Utah, Wisconsin, and Wyoming.
- Representatives from Idaho's metropolitan planning organizations (MPOs) who informed the project team about their experiences with AADT estimation and count data availability, especially leaders and staff with the Bannock Transportation Planning Organization, Bonneville MPO, Community Planning Association of Southwest Idaho (COMPASS), Kootenai MPO, and Lewis-Clark Valley MPO (LCVMPO).

Technical Advisory Committee

Each research project is overseen by a Technical Advisory Committee (TAC), which is led by an ITD project sponsor and project manager. The TAC is responsible for monitoring project progress, reviewing deliverables, ensuring that study objectives are met, and facilitating implementation of research recommendations, as appropriate. ITD's Research Program Manager appreciates the work of the following TAC members in guiding this research study.

- Project Sponsor: Idaho Transportation Department
- Project Manager: Vicky Calderon
- TAC Members:
  - Idaho Transportation Department (ITD): Vicky Calderon, David Coladner, Nicole Hanson, Amanda Laib, Margaret Pridmore
  - Local Highway Technical Assistance Council (LHTAC): Kevin Kuther, Matthew Syphus
  - Federal Highway Administration (FHWA): Lance Johnson
- FHWA-Idaho Advisor: Lance Johnson

Table of Contents

List of Tables

List of Figures

Acronyms

AADT	Annual average daily traffic
AADTT	Annual average daily truck traffic
AASHTO	American Association of State Highway and Transportation Officials
AC	Asphalt-concrete
ACS	American Community Survey
ATR	Automatic traffic recorder
CAADT	Commercial AADT
CFR	Code of Federal Regulations
COMPASS	Community Planning Association of Southwest Idaho
CRCP	Continuously reinforced concrete pavement
CSV	Comma separated values
DMV	Department of Motor Vehicles
DOT	Department of transportation
FC	Functional classification
FHWA	Federal Highway Administration
GIS	Geographic information systems
GPS	Global positioning system
GWR	Geographically weighted regression
HPMS	Highway Performance Monitoring System
HQ	Headquarters
HSIP	Highway Safety Improvement Program
IAHD	Idaho Association of Highway Districts
ITD	Idaho Transportation Department
ITE	Institute of Transportation Engineers
JPCP	Jointed plain concrete pavement
JRCP	Jointed reinforced concrete pavement
IAHD	Idaho Association of Highway Districts
LCVMPO	Lewis-Clark Valley Metropolitan Planning Organization
LEHD	Longitudinal Employer Household Dynamics
LHTAC	Local Highway Technical Assistance Council
LODES	LEHD Origin-Destination Employment Statistics
LRI	Local road inventory
LRS	Linear referencing system
ME	Mean error
MAPE	Mean absolute percent error
MIRE	Model Inventory of Roadway Elements
ML	Machine learning
MPO	Metropolitan planning organization
MSE	Mean squared error
NAICS	North American Industry Classification System
NCHRP	National Cooperative Highway Research Program
NHS	National Highway System
NLCD	National Land Cover Database
NPMRDS	National Performance Management Research Data Set
PCC	Portland cement concrete

PMID	Pavement management ID
RMSE	Root-mean-square error
RPC	Regional Planning Council
SVM	Support vector machines
TAC	Technical advisory committee
TDM	Travel demand model
TMC	Traffic message channel
VMT	Vehicle miles traveled
WAC	Workplace area characteristics
WIM	Weigh-in-motion

Definitions

Annual average daily traffic (AADT): The AADT measure is an estimate of the average volume of vehicle traffic on a section of roadway over a full year. Discussion of how ITD currently calculates AADT begins on page 28, and examples of equations from federal guidance for calculating AADT based on different types and durations of traffic counts are on page 5.

Federal-Aid roads: Federal-Aid roads are those belonging to the National Highway System, the Eisenhower National System of Interstate and Defense Highways ("Interstate Highways"), and all other public roads not classified as local roads (functional classification 7) or rural minor collectors (functional classification 6) (23 CFR § 470.103).

Local road: As defined by the Idaho Transportation Department Systems Procedures, "local roads provide direct access to residential neighborhoods, local businesses, agricultural properties and timberlands. Volumes typically range from less than one-hundred to possibly thousands of vehicles per day. Roads not classified as arterials or collectors are considered local roads" (Idaho Transportation Department 2016).

Off-system roads: Off-system roads are public roads that are not a part of the Federal-Aid system. In Idaho, these are generally local roads (functional classification 7) and rural minor collectors (functional classification 6).

On-system roads: On-system roads are public roads that are part of the Federal-Aid System.

Public roads: Public roads are those "under the jurisdiction of and maintained by a public authority and open to public travel." A public authority is a "federal, state, county, town, or township, Indian tribe, municipal, or other local government or instrumentality with authority to finance, build, operate, or maintain toll or toll-free facilities" (23 USC § 101 (22-23)).

Urban roads: Urban roads are in an area where the population exceeds 5,000 (INSIDE Idaho 2020).

1. Introduction

This report describes a method to estimate traffic on every road in the state by supplementing the Idaho Transportation Department's (ITD) existing process for estimating annual average daily traffic (AADT) on all Federal-Aid roads in Idaho with a process for estimating AADT for remaining non-Federal-Aid public roads (collectively referred to as 'off-system' roads in this report). The AADT estimates will facilitate crash analytics by ITD's Traffic Safety group and respond to new federal recommendations for data collection described in the Model Inventory of Roadway Elements (MIRE). Additionally, the Highway Safety Improvement Program (HSIP) will require every paved road to have estimates of AADT made available by 2026 (Tsapakis, Holik, et al. 2020).

This report has five chapters in addition to this "Introduction," which is the first chapter. The second chapter ("Literature Review, Existing AADT Estimation Methodologies, and FHWA Requirements") reviews research for estimating AADT on public off-system or similar roads, describes ITD's existing AADT estimation methodology, and reviews FHWA requirements and recommendations around estimating AADT. The third chapter ("Data Inventory") details the data sets that are available for estimating off-system AADT. Many of these data sets are used in the AADT estimation methodology described in chapter 4 ("Methodology"). The geospatial interpolation method was selected from all the methods reviewed in the literature review for its alignment with ITD's goals and processes (including that it produce the most accurate estimates possible, be executable with in-house resources, and be relatively straightforward to explain and understand), and based on ITD's guidance and the technical advisory committee's (TAC) input. The final chapter ("Implementation and Validation Plans") provides additional guidance for executing and validating the process based on a 'dry run' of the process conducted by the consultant team, in addition to a draft implementation schedule and roles.

2. Literature Review, Existing AADT Estimation Methodologies, and FHWA Requirements

This chapter reviews federal requirements and guidance, current practices related by ITD staff, and literature from state departments of transportation (DOTs) and other research organizations that are studying AADT estimation to help ITD select a methodology to make these AADT estimates. These researchers and organizations have developed, improved, and /or applied several methodologies for estimating AADT on off-system roads in the last two decades with a broad range of expertise and data source requirements. The review reveals broad leeway for ITD to pursue AADT estimation techniques that make sense for its needs since there are few federal requirements around how to estimate AADT on off-system public roads. However, neither is there industry consensus around best practices.

This chapter has three sections after this introduction. The chapter is structured to bookend the review of methods in the federal regulatory framework and ITD's current practices. The first section summarizes federal regulations and rules contained in the *Traffic Monitoring Guide* related to AADT estimation. The second summarizes methods for estimating AADT on off-system public roads, including describing pros and cons of each method and locations where the method has been implemented. Academic research, a survey of states, and instances of practical application inform the methodological review. Finally, the third section describes ITD's current process for estimating segment-level AADT and statewide vehicle miles traveled (VMT).

Federal Regulations & Rules

The *Traffic Monitoring Guide* published by the Federal Highway Administration (FHWA) provides guidance to state highway agencies about the policies, standards, procedures, and equipment for traffic monitoring programs. The *Traffic Monitoring Guide* primarily consists of recommendations rather than requirements about how to conduct traffic monitoring, and the recommendations cover statewide traffic monitoring programs rather than being specifically tailored to AADT estimation off of the Federal-Aid System. However, it remains the most authoritative source for standards about collecting traffic monitoring. This section is based on the 2016 *Traffic Monitoring Guide* (Federal Highway Administration 2016) (cover page in Figure 2-1) and is supplemented by the newest version of the *Traffic Monitoring Guide* released in December 2022 (Federal Highway Administration 2022) (cover page in Figure 2-2). FHWA's *Traffic Data Computation Method* Pocket Guide provides additional guidance on calculating AADT (Federal Highway Administration 2018).

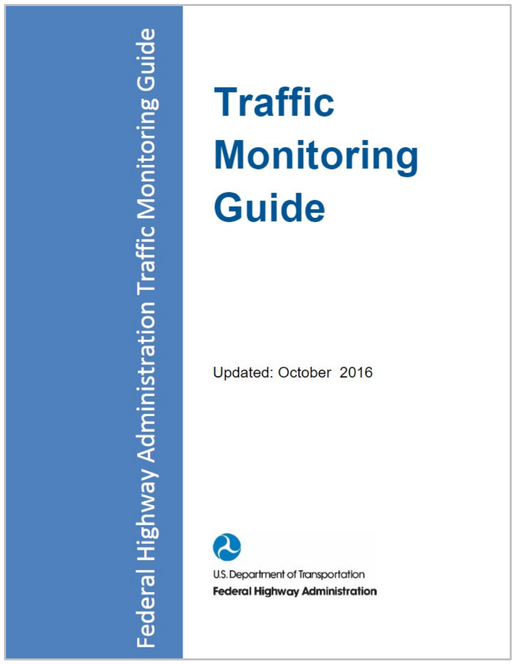


Figure 2-1. Cover Page of 2016 *Traffic Monitoring Guide*

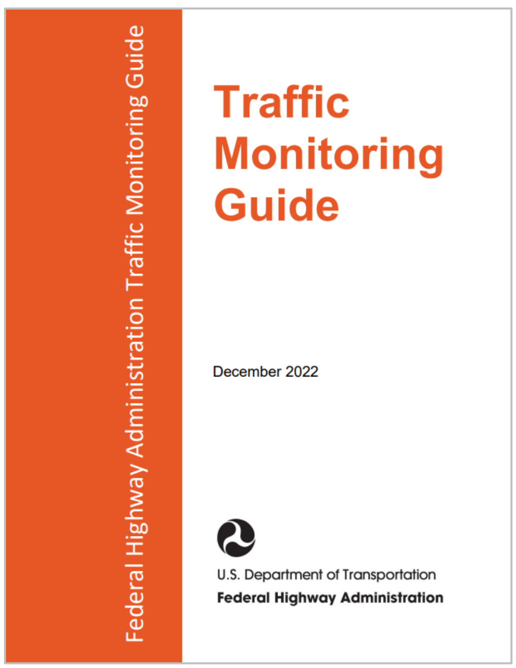


Figure 2-2. Cover Page of 2022 *Traffic Monitoring Guide*

This subsection addresses the recommendations that the *Traffic Monitoring Guide* provides related to AADT estimation, including traffic count programs, calculations to derive AADT from traffic counts, and related data submission requirements. It documents best practices related to traffic count programs, but few requirements and no best practices specifically geared toward AADT estimation for off-system roads.

The 2016 *Traffic Monitoring Guide* contains several relevant chapters to AADT estimation. Recommendations around the traffic count program are in chapter 2 ("Traffic Monitoring Program") and chapter 3 ("Traffic Monitoring Methodologies"). Recommendations related to calculations for deriving AADT are also in chapter 3 ("Traffic Monitoring Methodologies"). Recommendations related to data submission are in chapter 6 ("HPMS Requirements for Traffic Data") and in chapter 7 ("Traffic Monitoring Formats"). Other chapters address theory (chapter 1), the traffic monitoring program (chapter 2), traffic monitoring for non-motorized traffic (chapter 4), and transportation management and operations (chapter 5). The 2022 *Traffic Monitoring Guide* discusses the traffic count program in chapters 1, 2, and 3 ("Traffic Monitoring Program Introduction," "Traffic Data Collection Technology and Equipment," and "Methodologies for Traffic Data Collection and Processing" respectively). Recommendations related to calculations for deriving AADT are in chapter 3 ("Methodologies for Traffic Data Collection and Processing"). Other chapters address data formats (chapter 4), data reporting requirements (chapter 5), and third-party traffic data (chapter 6).

Traffic Count Programs

FHWA recommends for state DOTs to evaluate their traffic monitoring programs at least every five years, and that the evaluation should cover the entire program including equipment, selection of data collection sites, validation, and data analysis (2016 *Traffic Monitoring Guide* Section 2.3, 2022 *Traffic Monitoring Guide* Section 1.6). Most traffic monitoring programs will include continuous data programs for sites for which data is continuously collected and short-duration count sites. The *Traffic Monitoring Guide* provides methods for collecting and processing traffic data including data on traffic volume (source: 2016 *Traffic Monitoring Guide* Figure 3-2, 2022 *Traffic Monitoring Guide* Figure 1-1). Continuous traffic count locations need to be grouped based on identified traffic patterns for time of travel consistency such as hour-of-day distributions, day-of-week distributions, and monthly factors. The 2016 and 2022 versions of the *Traffic Monitoring Guide* both present several methods that transportation agencies may follow for grouping count locations, including the 'traditional approach' (which uses general knowledge of the road system and monthly graphs to identify patterns and define groupings), cluster analysis, and volume factor groups. The *Traffic Monitoring Guide* also provides guidance on data collection for vehicle classification (e.g., axles and length), speed, weight, and lane occupancy.

Short-duration data programs furnish most of the geographic coverage for traffic monitoring programs. Recommended durations of counts can vary from 48 hours to a week or even more, with longer-duration counts having been shown to produce more accurate volume estimates (2016 *Traffic Monitoring Guide* Page 3-77, 2022 *Traffic Monitoring Guide* page 3-43). Since short-duration count locations are by definition not continuous, the frequency of counts needs to be set, and the TMG recommends that counts be conducted at least once every six years (2016 *Traffic Monitoring Guide* Page 3-80, 2022 *Traffic Monitoring Guide* page 3-43), and potentially more often for some roads such as those experiencing growth in traffic volumes or that belong to a higher functional class. Some short-duration count locations should also have vehicle classification equipment to allow for estimation of annual average daily truck traffic (AADTT) in addition to other purposes. At least 25-30% of short-duration count locations should include classification counting equipment, or more when agency resources allow (2016 *Traffic Monitoring Guide* Page 3-78, 2022 *Traffic Monitoring Guide* Table 3-18). The *Traffic Monitoring Guide* recommends 48 consecutive hours for vehicle classification counts with variation on a case-by-case basis (2016 *Traffic Monitoring Guide* Page 3-56, 2022 *Traffic Monitoring Guide* Page 3-87).

Calculations for Deriving AADT

Estimating AADT and AADTT from short-duration traffic count locations requires using factors to adjust for traffic patterns by time of day, day of week, and seasons. Figure 2-3 shows an example of the day-of-week traffic distributions that may be observed at count locations, showing both traditional urban traffic patterns with higher weekday than weekend travel and traffic patterns more typical of roads used for recreation, which usually show slightly higher weekend than weekday travel. For extremely short traffic counts (less than 24 hours), there are additional adjustments to be made using data collected from continuous count locations (2016 *Traffic Monitoring Guide* Page 3-95). The 2016 *Traffic Monitoring Guide* provides procedures for computing the following statistics. They are not summarized here since none of these are directly related to calculating AADT on off-system roads.

- Average daily truck traffic (ADTT)
- Annual average daily truck traffic (AADTT)
- Axle correction factors
- Factors for converting daily truck traffic counts into estimates of AADTT (by class)
- Factors that allow conversion of AADTT estimates (by class) into average day of week estimates for use in the draft National Cooperative Highway Research Program (NCHRP) 1-37A Pavement Design Guide
- Sum of FHWA heavy vehicle classes 4-13 for 24 hours (Vehicle classes are shown in Figure 2-4)
- % Single Unit
- % Combination Unit

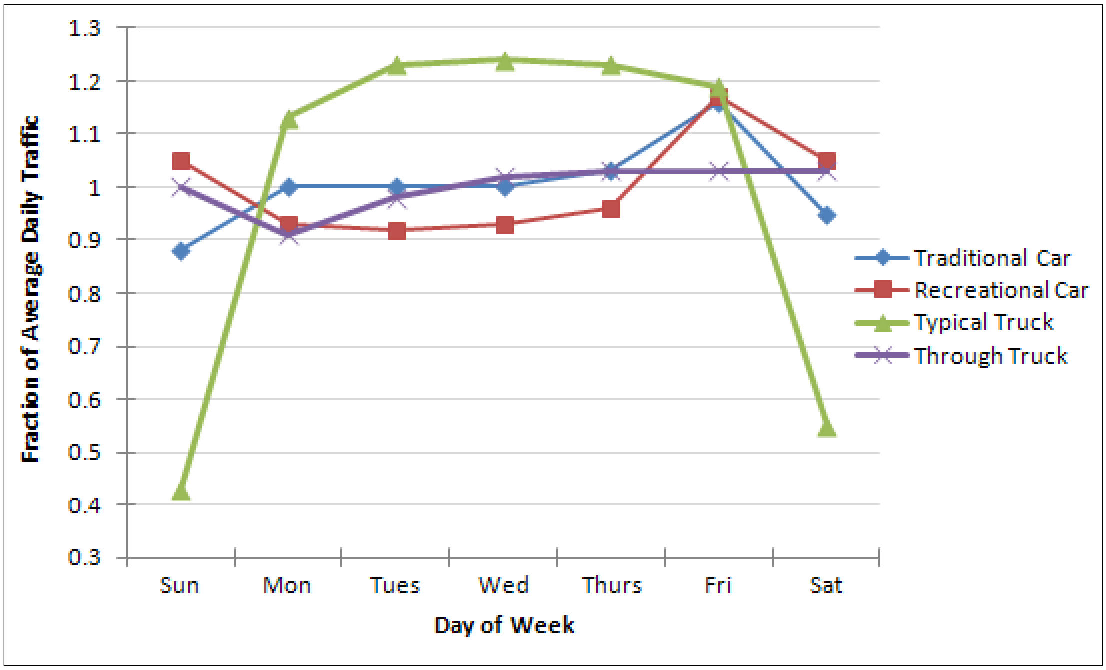


Figure 2-3. Example of Day of Week Travel Patterns

Source: Traffic Monitoring Guide, Figure 1-4, Federal Highway Administration (2016)

Note: This graphic shows example travel patterns observed in many locations in the United States and is not specific to any particular location in Idaho.

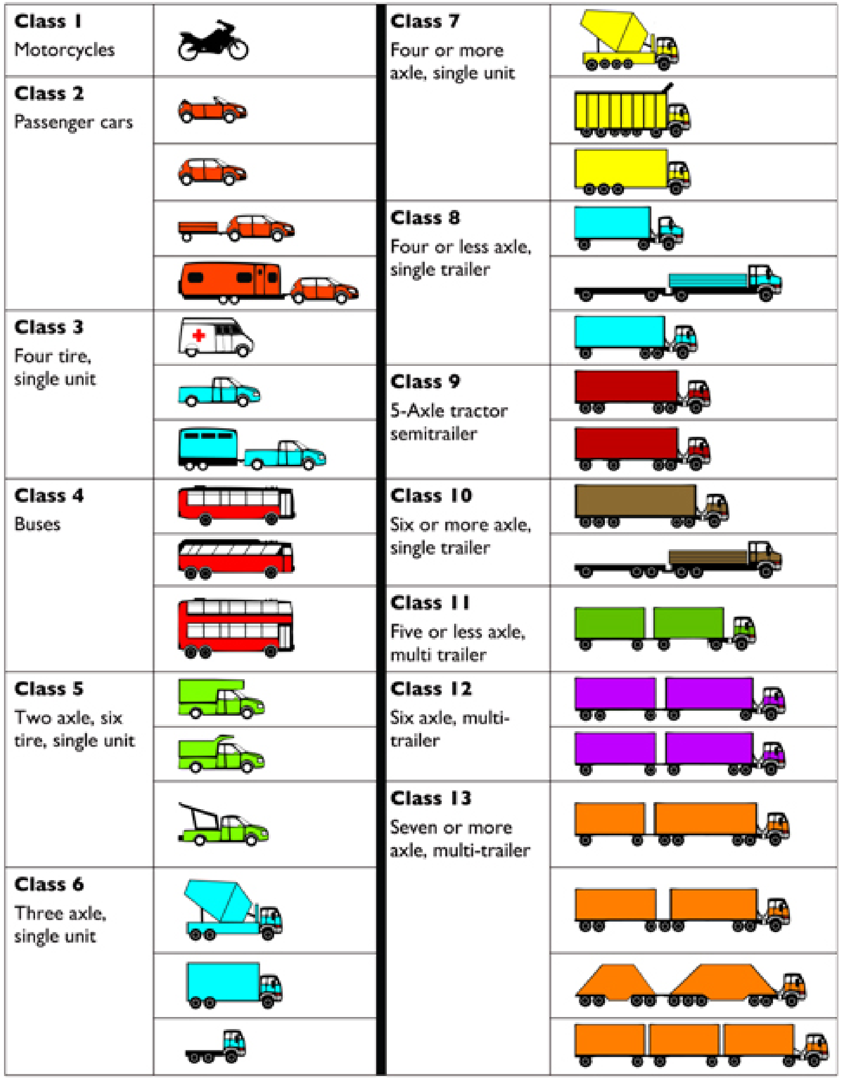


Figure 2-4. FHWA Vehicle Category Classification

Source: *Traffic Monitoring Guide*, Figure C-1, Federal Highway Administration (2016)

Additionally, the 2022 *Traffic Monitoring Guide* provides commonly used equations for estimating AADT based on continuous and short-duration counts, including Equation 1 and Equation 2 for calculating AADT based on continuous counts, and Equation 3 for estimating AADT from short-term counts. While these equations or closely related versions of them are often used to calculate AADT where traffic counts are available, they cannot be used in their existing forms to estimate AADTs in the absence of traffic counts, which are not available for most off-system public roads in Idaho.

Equation 1. Monthly average daily traffic calculation from continuous counts

$$MADT_m = \frac{\sum_{j=1}^{7} w_{jm} \sum_{h=1}^{24} \left[\frac{1}{n_{hjm}} \sum_{i=1}^{n_{hjm}} VOL_{ihjm} \right]}{\sum_{j=1}^{7} w_{jm}}$$

Equation 2. AADT calculation from continuous counts

$$AADT = \frac{\sum_{m=1}^{12} d_m * MADT_{HP_m}}{\sum_{m=1}^{12} d_m}$$

Where,

- $AADT$ is annual average daily traffic.
- $MADT_{HPm}$ is monthly average daily traffic for month m.
- VOL_{ihjm} is total traffic volume for *i*th occurrence of the *h*th hour of day within *j*th day of week during the *m*th month.
- i is the occurrence of a particular hour of day within a particular day of the week in a particular month (i=1,…n_{hjm}) for which traffic volume is available.
- h is the hour of the day (h=1,2,…24) – or other temporal interval.
- j is the day of the week (j=1,2,…7).
- m is the month (m=1,…12).
- n_{hjm} is the number of times the *h*th hour of day within the *j*th day of week during the *m*th month has available traffic volume (n_{hjm} ranges from 1 to 5 depending on hour of day, day of week, month, and data availability).
- w_{jm} is the weighting for the number of times the *j*th day of week occurs during the *m*th month (either 4 of 5); the sum of the weights in the denominator is the number of calendar days in the month (i.e., 28, 29, 30, or 31).
- d_m is the weighting for the number of days (i.e., 28, 29, 30, or 31) for the *m*th month in the particular year.

Equation 3. AADT calculation from short-term counts

$$AADT_{hi} = VOL_{hi} \times M_h \times D_h \times T_h \text{ x } A_i \times G_h$$

Where,

- $AADT_{hi}$ is the annual average daily travel at location *i* of factor group *h.*

- VOL_{hi} is the 48-hour axle volume at location *i* of factor group *h.*
- M_h is the applicable monthly (seasonal) factor for factor group *h.*
- D_h is the applicable DOW factor for factor group *h* (if needed).
- T_h is the applicable TOD factor for factor group *h* (if needed for any partial day counts).
- A_i is the applicable axle-correction factor for location *i* (if not a traffic volume or class count, i.e., for counts collected using a single pneumatic road tube).
- G_h is the applicable yearly change (i.e., growth or decline) rate factor for factor group *h* (if needed).

For traffic counts submitted to HPMS, FHWA recommends longer counts for lower-volume roads. Specifically, "the TMG recommends a minimum of a 24-hour monitoring period for roads with traffic volumes of greater than 5,000 AADT and a 48-hour monitoring period for lower-volume roads" (page 5-10). The longer recommended duration for lower-volume roads is intended to increase the accuracy of AADT estimates (page 3-44). FHWA does not recommend counts below 24 hours, though if they are collected then hour-of-day adjustment factors can be used to estimate AADT (page 3-56) (Federal Highway Administration 2016).

Data Submission

While the *HPMS Field Manual* provides authoritative requirements for data submission, the *Traffic Monitoring Guide* also provides guidance to states in meeting the HPMS program traffic reporting requirements. One broad recommendation is to make the state's traffic monitoring program mirror HPMS data reporting so that information published by FHWA and the state remain as similar as possible.

Third-Party Data Sources

The 2022 *Traffic Monitoring Guide* includes new guidance for obtaining third-party traffic data, which include consideration of agency needs motivating the data acquisition, assessing data equality, assessing contractual limitations from third-party vendors on data ownership and usage, and costs. It recommends the following questions to assess data quality (2022 *Traffic Monitoring Guide* Page 6-1).

- What is the data methodology?
- What is the source of benchmark (the ground truth) data?
- How are the data evaluated against the benchmark data?
- What are the conclusions of the evaluation and how they will be used? Is there persistent over or under estimations comparing to benchmark data? Is there a statistically significant difference between the benchmark data and their corresponding third-party data at no lower than an 85%

level of significance? What are the various percentages of differences between benchmark data and the third-party data?

- What data are to be used to calibrate the result (e.g., permanent, portable)?
- What is the acceptable margin of errors for different factor groups (e.g., FHWA roadway functional classes 1 to 7) and area types (urban and rural)?
- What is the percentage median or mean error as compared with the benchmark data (e.g., AADT from third party vs. AADT from continuous count data)?
- What are the limitations of the method?
- What traffic data quality control procedures does the agency have in place for the third-party data vendor to follow?

Literature Review

This section reviews academic literature and documented practice related to estimating AADT on off-system public roads and roads with similar characteristics to comprehensively identify and summarize the methods, pros and cons, and the associated data requirements. This set of methods serves as a foundation for developing a method for ITD to use to estimate traffic volumes on roads off of the Federal-Aid System in Idaho.

The methods summarized in this section present a comprehensive overview of methods for estimating off-system AADT researched in the United States or implemented by state DOTs. The project team referenced Transportation Research Board research, including NCHRP, state DOT documents, and FHWA guidance to ensure it comprehensively identified efforts over the last 20 years regarding AADT estimation, off of the state-owned highway system or off of the Federal-Aid System. The project team also conducted searches on academic search engines (e.g., Google Scholar) that included journals such as the *Transportation Research Record* and the *Journal of Transport Geography*, and it reviewed the chain of citations for published work to identify relevant peer-reviewed academic research.

As data availability and analysis capacity have increased, several methods have come to the forefront as they are able to provide estimates with higher accuracy at lower agency costs. These methods, which are described later in this section, include improvements to existing sampling methods, regression models, geospatial interpolation models, machine learning supported techniques, travel demand modeling, and network centrality analysis.

State-Level Analyses & Practices

Surveys and Assessments in Literature

Publicly available information on current state DOT practices for estimating AADT is primarily focused on facilities that are counted for short periods (1-3 days) on a rotating basis. However, a 2018 report by Portland State University in collaboration with the Oregon DOT reported on several other states' practices for estimating AADT on roads that lack AADT estimates, including methods for many states with practices that estimate AADT for roads off the of the state-owned system. This information was gathered as part of a survey (Unnikrishnan, et al. 2018) and is summarized below, with the "In Network Only" designation indicating that estimation is only done on state-owned facilities.

Summary of Methods for Estimating Off-System AADT

- **Sample procedures:** This method represents current practice for AADT estimation at ITD and many other agencies and involves conversions of observed counts to AADTs. Enhanced sampling procedures seek to expand the traffic count program strategically to make off-system AADT estimation more accurate.
- **Regression models:** Regression models use count-derived AADTs along with independent variables such as roadway, demographic, and economic characteristics to derive regression models that can be applied to estimate AADT for roadway segments without counts.
- **Geospatial interpolation models:** This method relies primarily on proximity to count locations to estimate AADT at locations without counts. Interpolation of AADT is based on AADT from nearby or neighboring count locations along with other variables in some cases.
- **Machine learning:** This method develops computer models that use data to improve the performance of algorithms and statistical models in estimating AADT. This improvement happens without explicit guidance from analysts.
- **Travel demand models:** This method uses either existing or specially built travel demand models to estimate AADT on facilities of interest, typically calibrated with existing counts.
- **Network centrality:** Network centrality encompasses a set of methods derived from network theory. Network analysis allows for a road's location within a larger network to be quantified through such metrics as centrality and betweenness, which can then be combined with other approaches.

- **Alaska (In Network Only):** Uses annual growth factors for facilities with counts and statewide averages where counts were unavailable.
- **Arizona (In Network Only):** Groups facilities by functional class, calculates growth rates for permanent count stations, and uses proprietary software to calculate missing AADT values.
- **Arkansas:** Applies AADT values to segments from "like" facilities based on county, rural vs. urban classification, functional class, paved vs. unpaved status, number of lanes, and one-way vs. two-way configuration.
- **Florida:** Uses the statewide transportation model to estimate missing values based on housing units, employment sites, observed AADT, AADT values generated by trips on major roadways, and street-level data.
- **Georgia:** Applies growth factors when previous years' data is available, interpolates when other segments along a facility have actual values from the current or previous years, or estimates counts based on functional class, urban code, pavement type, and location.

- **Illinois:** Utilizes nearby AADT counts, aerial imagery for accessing traffic pattern factors, and information from roadways with similar characteristics.
- **Iowa:** Uses proximity to housing, businesses, and other traffic generators to estimate traffic volumes.
- **Kansas:** Uses one of the following five methods: route flows if adjacent segments have counts, route averages, city functional class averages, population group functional class average, or county functional class average.
- **Montana:** Default values are assigned based on functional class, urban vs. rural status, and paved vs. unpaved surface type.
- **Mississippi:** Uses blanket counts for each functional class and county with local routes being sampled.
- **Nevada:** Uses AADT counts from nearby roads of similar functional class.
- **South Carolina:** Groups counts by functional class and urban vs. rural classification, and then applies annual growth factors as needed.
- **Vermont:** Uses a proprietary software package to produce estimates.
- **Washington:** Cities and counties collect AADT and calculate growth factors, and local access is assumed to account for 7% of rural VMT and 11% of urban VMT.
- **Wisconsin:** DOT routes have growth factors applied to previous counts and take counts as needed. Non-DOT routes have estimates provided by the local government.

These methodologies indicate that state DOTs' approaches to estimating local access AADT are typically ad-hoc with reliance on grouping facilities by functional characteristics or land use intensities.

More recently, Huynh et al. (2021) surveyed 17 state DOTs over the summer of 2020 on their methods for estimating AADT on facilities that are not directly counted. Although the names of the participating states are not listed in the publication, the most common processes for locations with no recent counts within the prior 10 years on the facility in question or on nearby facilities were multiple linear regression, visual estimation, and default values (Huynh, et al. 2021).

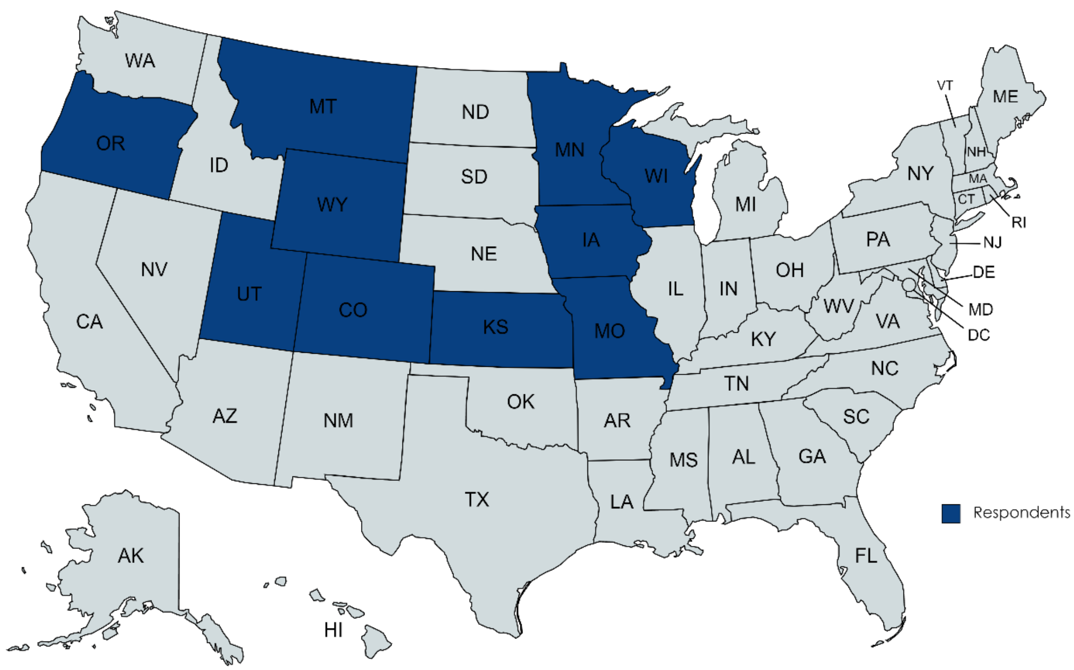
Concurrence on the state of practice between the two reports summarized above indicates that there is little industry consensus in estimating AADT on off-system public roads.

ITD Survey of State DOTs

As part of this project, ITD distributed a survey on AADT estimation to state DOT members of a Jackalope users group. The survey was distributed to the following 14 state DOTs that are part of the Jackalope users group: Colorado, Iowa, Kansas, Kentucky, Minnesota, Missouri, Montana, Nevada, New

Mexico, Oregon, Utah, Washington, Wisconsin, and Wyoming. Responses were received from ten states, which are shown in Figure 2-5.

Figure 2-5. States Responding to Survey



The survey asked states the following six questions. The raw responses are included in Appendix A on page 84.

1. Does your organization estimate AADT for facilities without traffic volume counts? If so, what methodology is used?

2. Does your organization estimate AADT for facilities without traffic volume counts? If not, what are the reasons for this?

3. Does your organization take traffic counts on off-system roads? If so, what methodology is used?

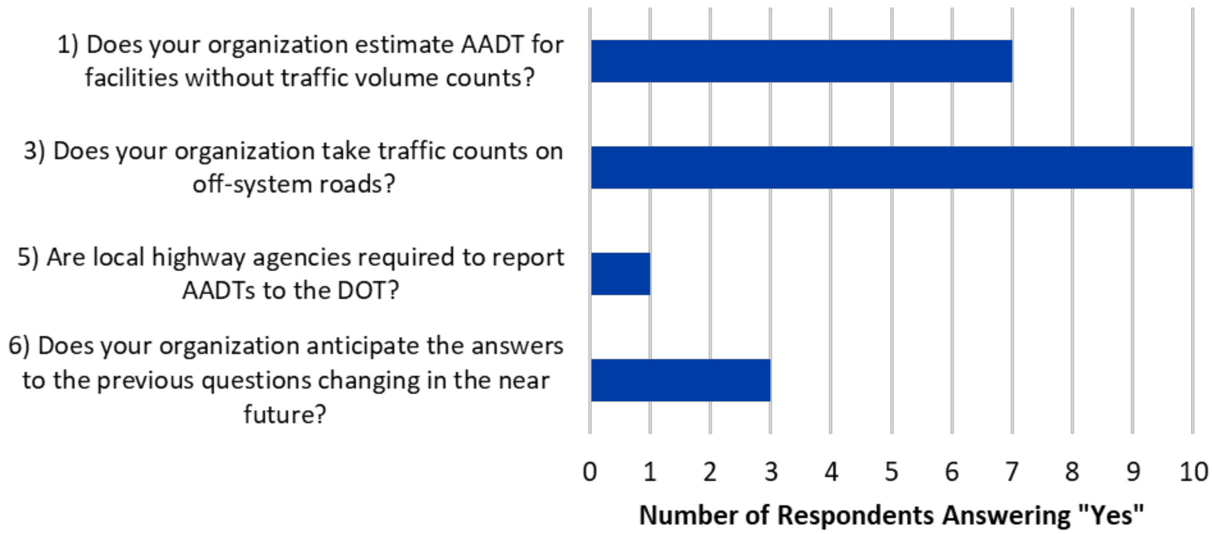
4. Does your organization take traffic counts on off-system roads? If not, what are the reasons for this?

5. Are local highway agencies required to report AADTs to the state DOT?

6. Does your organization anticipate the answers to the previous questions changing in the near future, if so how?

Figure 2-6 summarizes the responses for all the questions except #2 and #4, which are not yes/no questions but instead ask for reasons.

Figure 2-6. Summary of Survey Results



Question #1 - Does your organization estimate AADT for facilities without traffic volume counts? If so, what methodology is used? While most respondents report that their organization estimates AADT for at least some facilities without traffic volume counts, the reasons for which they do this, the methods that they use, and the extent of the system covered vary. For instance, two states report that they estimate AADT for facilities without traffic counts only for the purpose of completing statewide VMT summary tables that are reported annually to the HPMS, whereas other states ostensibly use segment-level AADT estimates for other purposes. While some states (i.e., Utah and Wyoming) estimate AADT for all public roads, many others do not specify the exact network for which AADT is estimated. The three states that do not regularly estimate AADT on roads without traffic counts report using ad hoc methods in the very limited instances where it is necessary.

Question #2 - Does your organization estimate AADT for facilities without traffic volume counts? If not, what are the reasons for this? Two of the three states provided their reasons for not estimating AADT on roads without traffic counts, and those reasons relate to the funding needed for necessary data collection (Minnesota) and the inability to achieve satisfactory accuracy (Oregon).

Question #3 - Does your organization take traffic counts on off-system roads? If so, what methodology is used? All respondents report that their state takes traffic counts on at least some off-system roads, but—as for question #1—the extent of the system that is covered varies widely. States may do so in response to specific requests (Wisconsin), to the extent needed for HPMS reporting (Colorado), where new economic and development activity is occurring (Missouri), for a carefully selected network sample (Montana), for roads that are designated to draw state funding (Minnesota) or for roads whose functional classification has recently increased (Utah). Other respondents do not specify the network that is covered.

Question #4 - Does your organization take traffic counts on off-system roads? If not, what are the reasons for this? Only one state provided a reason for not doing more extensive traffic counts on off-system roads, and the reason relates to the amount of funding and staff that would be required (Minnesota).

Question #5 - Are local highway agencies required to report AADTs to the state DOT? Metropolitan planning organizations (MPOs) are required to report AADT in one state (Montana), while in Minnesota it is only necessary in a select number of cases where metropolitan areas collect traffic counts needed for state-aid funding, in which case, the state DOT then processes and publishes the traffic count data.

Question #6 - Does your organization anticipate the answers to the previous questions changing in the near future, if so how? Three states expect changes to the off-system count or AADT estimation processes, and all three states cite new data as the reason for the expected change. New data may be for lower (6 and 7) functional class roads (Colorado), probe data (Montana), and crowdsourced traffic data (Missouri).

AADT Estimation Research

This section summarizes efforts by various state DOTs and academic institutions to estimate traffic volume on off-system public roads across a wide range of climates, terrains, population densities, and other state characteristics occurring since 2000. Table 2-1 summarizes the locations of prominent research efforts reviewed for this chapter and shows that the literature addresses six main techniques.

Table 2-1. Summary of Estimation Method by Analysis Location

State	Sampling	Regression	Geospatial	Machine Learning	Travel Demand Model (TDM)	Network Analysis
Alabama	--	X	--	--	--	--
Florida	--	X	X	--	X	--
Georgia	X	--	--	--	--	--
Idaho	--	--	--	X	--	X
Indiana	--	X	--	--	--	--
Kentucky	--	X	--	--	--	--
Louisiana	--	--	--	X	--	--
North Carolina	--	--	X	--	--	--
Ohio	X	--	--	--	--	--
South Carolina	--	--	X	--	--	X
Tennessee	--	--	--	X	--	--
Texas	--	--	X	--	--	--
Vermont	--	--	--	X	--	--
Washington	--	--	X	--	--	--
Wyoming	--	X	--	--	--	--

Note: A dashed line indicates that the estimation method has not been tested in the state, and an "X" indicates that the estimation method has been tested in the state.

Review of Methods

This section details the primary categories of methods for estimating off-system AADT described in the literature. DOTs have traditionally focused on the use of sample counts of traffic volumes to estimate AADT. However, researchers and practitioners are also refining methods to increase accuracy without increasing associated expenses. The application of geospatial interpolation and network theory to regression analysis has shown promising results, but there is little consensus on the best path forward. Instead, methods appear to be chosen based on the states' available data sources and technical expertise.

Sampling Procedures

Sampling is a well-established method for estimating AADT on large road networks and is representative of the current state of practice at ITD, as described starting on page 28. "Control counts" are continuously taken by permanently installed equipment at a small set of locations representative of the entire network. In addition to producing accurate counts for these segments, the resulting data distribution is used to identify variations in traffic volume over various time periods, such as season, month, or day of the week.

"Coverage counts" are taken at a wider set of locations, generally 20-50 times the number of control counts for a state-level network. Locations are selected to cover all facility classes, a variety of through-traffic to local-traffic ratios, and different land use characteristics. Counts usually last 1-3 days and are conducted annually or on a rotating basis. Growth and time-period factors are derived from control counts by dividing average values by the subset average. Relevant factors are then applied to these short-term measurements to estimate AADT on the facility.

As this method has been well established and studied, it is often the simplest for state transportation agencies to implement. At its most fundamental level, the process could be achieved without any form of automation. Agencies can quickly obtain reasonable estimations of AADT across their entire network with commonly available equipment, staff expertise, and software. The methodology's simple assumptions and strong basis in concrete measurement facilitate its communication to elected officials, local stakeholders, and the public.

However, this technique does have weaknesses. Local differences between facilities such as number of lanes, frequency of curb cuts, nearby land uses, or pavement condition can impact traffic volumes. This can be offset with more granular grouping of counts, but this results in less statistical accuracy. Short-term counts may also occur during periods of unusually high or low traffic volumes. These and other statistical issues could be resolved with more counts, but increasing the number of counts quickly becomes cost prohibitive for a method that already requires recurrent observations on a significant portion of the network. Table 2-2 summarizes pros and cons.

Table 2-2. AADT Estimation Techniques from Literature – Improved Sampling Procedures

Types	Pros	Cons
Sample locations for traffic counts may be selected randomly or as part of groups to ensure representation of important characteristics ("stratified"),	Simple implementation. Equipment or services often already available to agencies. Focus on real world measurements of traffic patterns.	Ignores local differences between facilities. Vulnerable to anomalies in observations. Requires recurrent observations on a significant portion of the network.

References for Improved Sampling: Seaver, Chatterjee and Seaver (2000), Jiang, McCord and Goel (2006), Yang, Wang and Bao (2014)

Several research efforts have made improvements on the method over the last two decades, including supplementing direct counts with information from aerial imagery and using statistical methods to better stratify samples. It should also be noted that this method is rarely used to estimate AADT on low-volume roads, in part due to the large number of such facilities and high variability in volumes between facilities.

Regression Models

Regression models—whether based on a linear, quadratic, logarithmic, or some other mathematical expression—are a staple of statistical analysis and provide a simple and effective way of examining relationships between variables. This technique has been extensively researched, has a wide variety of existing tools in commonly used software packages, and often forms the basis of other analytical methods. Most data analysis programs such as Excel come with basic functionality for regression analysis, and coding languages like Python and R have packages that automate many of the statistical processes, returning tidy summaries of relevant information. Overall, regression analysis is robust enough to handle a wide variety of data types and scenarios.

As with sampling procedures, facilities can quickly be grouped together for analysis by using categorical variables such as facility classification or ranges of continuous variables such as population density. Regressions are then developed for each subgroup. However, there is an inverse relationship between group size and statistical accuracy because small subgroups provide less information to develop predictive factors. The inverse is also true: as the volume of data provided to the model grows, predictions will better represent reality.

The most basic regression models utilize linear relationships between independent variables and the dependent variable; however, real-world phenomena often do not follow linear trends. Logarithmic and polynomial transformations can account for non-linear relationships, but these quickly increase model complexity. While this is not typically an issue for the computational assessment of regression models, it can make interpretation less intuitive. Similarly, relationships among independent variables may need to be accounted for, and the addition of these components poses the same challenges.

Although a necessary process, balancing the need for clarity against sample size can eventually lead to differences between facilities in the same subgroup not being captured by the model or apparent in the subsequent results. Determining the relative weights of these interests, with this technique and others discussed in this document, is an iterative procedure that should engage a wide variety of experts and team members to produce a useful product.

Finally, data sets used in regression models must be preprocessed to remove observations or correct for trends that may skew the predictions toward values that do not actually reflect real-world conditions. In regression models with multiple independent variables, independent variables should be tested for correlation. If two variables are found to be highly correlated, they are considered “multicollinear.” The result is that the model cannot differentiate between their impacts on the dependent variable. Often, one of the variables must be removed from the model. There are other variable or model characteristics that can violate regression assumptions, including heteroskedasticity, which occurs when the variance in the dependent variable changes along the range of an independent variable. Transformations such as a power or log transformation may make the data homoskedastic, but these are not always successful in distributing prediction residual values. Should this be the case, the variable should not be used in a regression analysis. Outlier observations should be identified during the process and checked for errors (data entry mistakes, incorrect labels, etc.). However, they should not be removed for statistical analyses. Additionally, endogeneity, which occurs when error in the regression model is correlated with explanatory variables, can also undermine the statistical validity of regression models.

Table 2-3 below summarizes some of the pros and cons of using regression models to estimate AADT.

Table 2-3. AADT Estimation Techniques from Literature – Regression Models

Types	Pros	Cons
Several common forms for regression models include linear, logarithmic, and power.	Easy to group facilities for analysis. Can be quickly implemented in a wide range of software packages. Flexible enough to accommodate a wide variety of input categories. Perform more efficiently as number of observations and predictions increase.	Difference in roads among the same group may not be reflected in results. Complex relationships can be difficult to model accurately. Input data needs to be preprocessed before use to identify statistical anomalies.

References for Regression Models: Mohamad et al. (1998), Xia et al. (1999), Zhao and Chung (2001), Zhao and Park (2004), Anderson, Sharfi and Gholston (2006), Pan (2008), Lowry and Dixon (2012)

Geospatial Models

Underpinning the use of geospatial data in statistical analysis is summarized by Waldo Tobler’s first law: “Everything is related to everything else, but near things are more related than distant things” (Tobler 1970). The statistic which typically represents this phenomenon is Moran’s I, calculated as the correlation coefficient between a variable for a given geography and the same variable for neighboring geographies. Neighbors can be classified with any set of arbitrary rules, but some common methods include adjacency, k-nearest neighbors, and distance thresholds. If an independent variable can be confirmed to have statistically significant spatial autocorrelation, it can be used in geospatially weighted regression models and supplemented with various interpolation techniques.

Geospatial models and their interpolative methods fall into two general categories: deterministic models which use arbitrary values, and statistical models which choose parameters based on data metrics. Regardless of the category, spatial analysis uses weighted averages to predict values where measurements are absent.

A simple form of estimating spatial values across a geography from a set of discrete samples is proximity interpolation. The method generates polygons where edges lie along the midpoints between sampled locations. These "Thiessen" polygons then inherit the attributes of the sample point which they enclose, but, without a large sample size, the resulting surface does not generally reflect natural distributions of data.

The next expansion of this process is to include multiple sample points for an unsampled location. For example, the five nearest points can be averaged (with or without weighting) to estimate a value. This is referred to as "k-nearest neighbor" interpolation.

The inverse distance weighted technique calculates values for unsampled locations with weighted averages from nearby locations based on a power coefficient. Larger coefficients create greater drop-offs in influence as the distance to other locations increases. The power coefficient selection is a subjective process unless the analyst optimizes to a selected evaluator measure, e.g. RMSE for existing AADT counts.

Trend surfaces use polynomials in a similar way to linear regression to estimate two-dimensional relationships between sampled points. A "0^{th}" order trend surface is the mean value of all sampled points. A first order trend surface improves this to a trend direction, and a second order trend surface generates a parabolic relationship. The continued addition of higher order terms may increase accuracy, but the success of this method greatly depends on the phenomena under consideration.

Named after Danie Krige, a statistician and mining engineer from South Africa, Kriging is a Gaussian interpolation process governed by prior covariances. In practice, this is implemented by removing spatial trends via the trend surface process described previously, computing a semivariogram to measure spatial autocorrelation, selecting a model to characterize the semivariogram (illustrated in Figure 2-7), and predicting values at unsampled locations. Several varieties exist, but the most common are Ordinary Kriging (constant unknown mean) and Universal Kriging (general polynomial trend model). Although it requires several more steps, this technique addresses weaknesses due to subjectivity in other spatial interpolation methods. Equation 4 shows the equation for a semivariogram (Mathew 2020).

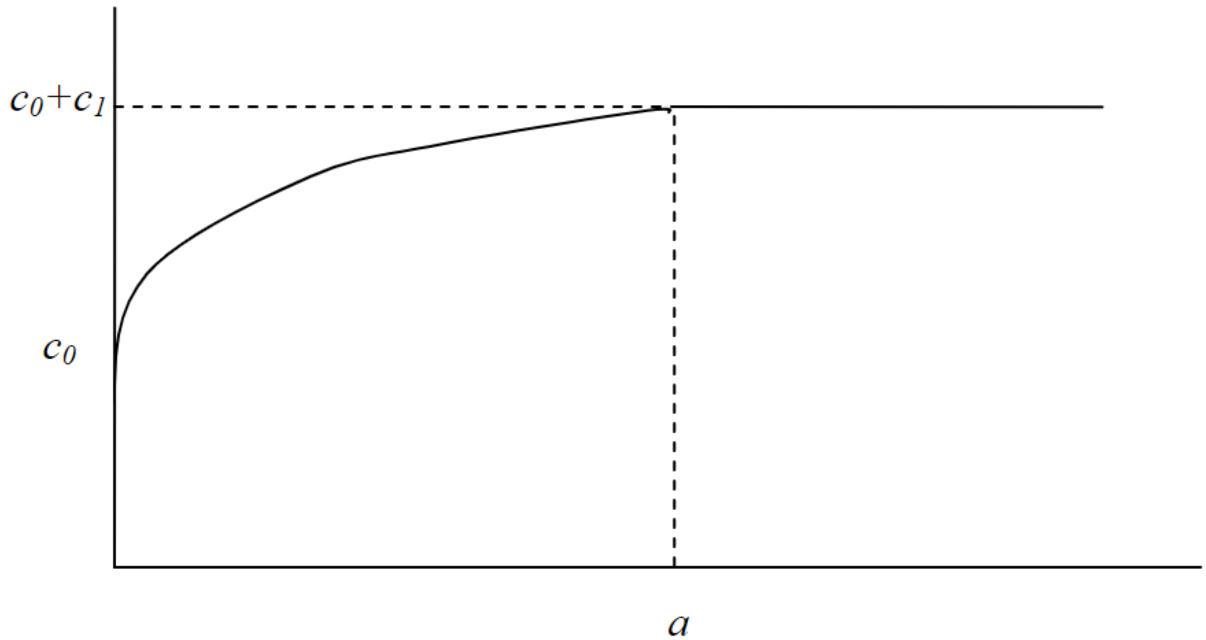


Figure 2-7. Illustration of a Semivariogram

Source: Wang and Kockelman (2009)

Equation 4. Semivariogram formula

$$S(h) = \frac{1}{2 \times N(h)} \times \sum (z_i - z_j)^2$$

Where,

- $S(h)$ is the semivariance for a given distance or lag, h.
- $N(h)$ is the number of pairs of observations separated by the distance h.
- z_i and z_j are the values of the variable being studied at two locations i and j that are separated by the distance h.

Most geospatial analysis software packages have the capacity to perform any of these processes. In general, research has found that the integration of spatial interpolation into regression analysis produced higher accuracies. However, some cases found that larger sample counts were required to realize these benefits. This is due in part to the need to consider facility classes separately during the interpolation process. Additionally, spatial calculations can quickly become computationally intensive over large regions such as statewide road networks. Table 2-4 summarizes these and other approach pros and cons.

Table 2-4. AADT Estimation Techniques from Literature – Geospatial Interpolation and Weighted Regression

Types	Pros	Cons
There are several methods for describing the influence of nearby points, including the use of Thiessen polygons, k-nearest neighbor methods, inverse distance weighting, and Kriging interpolation,	More accurate than non-spatial regression. Most types of geospatial interpolation can be implemented in common geographic information systems (GIS) applications (e.g., Esri ArcGIS). Reflect the spatial nature of transportation systems.	May require higher sample counts per unit area to realize accuracy improvements over non-spatial methods. May ignore differences between facility classifications. Spatial calculations over large areas can be computationally intensive.

References for Geospatial Interpolation and Weighted Regression: (1999), Zhao and Chung (2001), Zhao and Park (2004), Eom et al. (2006), Wang and Kockelman (2009), Pulugurtha and Kusam (2012), Selby and Kockelman (2013), Shamo, Asa and Membah (2015)

Machine Learning

Computer artificial intelligence applications such as machine learning (ML) are a rapidly evolving field that leverage modern computing capacity to identify complex relationships among large sets of data that cannot be evaluated through human expertise on reasonable timescales. This advantage is available whether the data set's size comes from large numbers of observations or from a large set of variables. In practice, this allows analysts to use more granular information that is more closely related to the dependent variable in question. Although their construction typically requires more expertise than other methods considered in this document and is more resource-intensive, their application is highly efficient. As the rapid evolution of ML techniques and algorithms persists, the challenges of implementation will fall drastically while their analytical power continues to increase. Table 2-5 summarizes these and other ML pros and cons.

Table 2-5. AADT Estimation Techniques from Literature – Machine Learning

Types	Pros	Cons
There are several types of machine learning structures, including support vector machines for regression (SVM/SVR), recurrent neural networks, decision trees, clustering, and k-nearest neighbor methods	Can identify important relationships among large groups of variables. Once trained, models can be applied with small amounts of computational resources. Can be applied to improve regression results. Rapid evolution of techniques may offer great potential for improvements in processes and accuracy over the coming years.	Extensive technical expertise required to implement effectively, including fluency in a data science-related programming language. Connections and weights are not readily visible or communicable. Accuracy of estimates may be reduced if trends change over time such that the past is less predictive. Training models requires large amounts of granular data.

References for Machine Learning: Sharma (1999), Sharma et al. (2001), Dixon (2004), Jiang, McCord and Goel (2006), Castro-Neto et al. (2009), Sun and Das (2015), Islam (2016), Fu, Kelly and Clinch (2017), Sfyridis and Agnolucci (2020)

ML algorithms come in several categories. Supervised learning utilizes labeled inputs and outputs to train models. They are simpler to implement but often require tuning and parameter decisions from the analyst implementing them. Examples include Support Vector Machines (SVM) and decision forests. Deep learning models like recurrent neural networks are a subset of supervised learning that mimic the way human brains learn, reinforcing successful neural connections. Unsupervised learning uses unlabeled data to find obscured patterns in large data sets; examples include clustering and dimensionality reduction. Unsupervised models require much more data and more powerful tools to achieve desired outcomes.

Initially developed at AT&T Bell Laboratories in the 1990s, SVMs are non-probabilistic binary linear classifiers that map training examples to an n-dimensional space such that a hyperplane divides categories with the maximum gap between them or that a hyperplane follows the data trend with minimal error. The mathematics used allow SVMs to be effective in data sets with large numbers of variables. This can be achieved even when the number of samples is less than the number of variables, but precautions must be taken to avoid over-fitting. However, the processes used to not directly predict probabilities, instead requiring computationally expensive cross-validation.

Classification and regression trees are sets of conditions and edges that split observations based on the "cost" of splitting a categorical variable by a group or a continuous variable by a given value. Over-fitting is managed by limiting training inputs, setting maximum distance from the "root" to the "leaf," and removing low-importance branches that do not contribute to predictive accuracy. To further correct for overfitting, ensemble methods, such as Random Forest, construct multiple decision trees with random samples of the data set to find and select commonly observed conditions.

Cluster analysis or clustering is an exploratory method that uses scaling and distance functions to find densely populated areas of a dataspace. Depending on the category of algorithm used, practical applications often approximate solutions either to reduce processing time or because complete solutions would take longer than is reasonable given current processing capacity. Some notable clustering algorithm categories include hierarchical clustering, centroid-based clustering (e.g., k-means), distribution-based clustering, and density-based clustering. These analytical processes are often run using purpose-built packages or libraries in programming languages like R and Python.

K-nearest neighbors is similarly concerned with meta-distances between observations but operates on a much simpler premise of limiting comparison to a given number of closest neighbors (k) provided by the analyst. Dimensional and data reductions are typically used to further reduce computation time and to reduce the influence of outliers.

Recurrent neural networks are representative of deep learning principles. A set of inputs are fed through a series of interrelated weight multipliers and bias additions to develop complex relationships between inputs and desired outputs. Compared to the supervised models discussed previously, the resulting models are almost impossible for analysts to communicate as they quickly obscure relationships between inputs as the number of "neurons" increase. To train these networks, large, labeled data sets are used to reinforce existing knowledge, often over tens of thousands of iterations on high-powered graphical processing units. However, after training has been completed, the resulting model is small enough to run on a computer with less processing capacity than most smartphones.

Travel Demand Models

Although travel demand models (TDMs) are more broadly used for system-wide forecasts, they can be used to estimate current AADT on existing facilities. The most traditional model structure is a four-step model with trip generation, trip distribution, mode choice, and route assignment. Trip generation utilizes information about work, shopping, and other origins and destinations to estimate how many trips will have a given location as their origin or destination. This is often based on an estimated factor for trips per square foot. Distribution allocates trips between analysis zones based on origins and destinations available in each, typically using gravity models. Mode choice splits trips by purpose between single occupancy vehicles, transit, and other modes through functions such as nested logit models. Finally, trips are iteratively assigned to a route until the congestion on that route makes an alternative more attractive.

The principal alternative to the four-step model is an activity-based model; however, these are highly resource-intensive from the perspectives of technical knowledge and computation time. Extensive surveying of local trip patterns is matched with demographic data to simulate individuals and households in a region and how they choose to travel. Common characteristics include household size, number of vehicles, number of workers, gender, age, and student status. While these models can produce greater accuracy than their four-step counterparts, they require far more investment in technical expertise and computational time to do so.

Overall, TDMs have better accuracy than regression models as they integrate road network relationships that can be updated in response to capital improvements and socio-demographic shifts. The information they output can also be used in other contexts such as economic development and housing. However, they often require technical expertise alongside specialized software to develop, run, and calibrate. Table 2-6 summarizes pros and cons of using travel demand models for off-system AADT estimation.

Table 2-6. AADT Estimation Techniques from Literature – Travel Demand Models

Types	Pros	Cons
The two most common types of travel demand models are four-step and activity-based models.	More robust than regression models in its ability to simulate traffic volume given future conditions and to estimate project impacts. Can be updated annual in response to road capacity and / or travel behavior changes. Information can be utilized in a wide variety of other transportation contexts.	Require specialized software and knowledge to implement. Need preprocessed data at small scales. Extracting relevant data can be time consuming.

References for Travel Demand Models: Zhong and Hanson (2009), Wang, Gan and Alluri (2013)

Network Centrality

Network theory is a subset of the mathematical study of graphs: graphs here are defined as sets of pairwise relationships between objects represented by vertices and edges. The principles of network theory have been implemented in disciplines as disparate as biology, linguistics, social structures, computer science, and transportation. In analyzing traffic volumes, a road network can easily be converted to a graph where intersections are represented as nodes and other road segments are represented as edges. Then, measurements of nodes' and edges' relationships to the rest of the network, or "centrality", can be used in the regression analyses techniques discussed previously. These measures include how many connections a node has to other nodes (degree centrality), how often a node is used in a shortest path (betweenness centrality), how much influence a node has on the network (eigenvector centrality) and how far a node is from other nodes in the network (closeness centrality).

Application of network centrality measures to AADT estimation in regression analysis has been shown to improve accuracy substantially, and there is a large body of work outside the transportation field that can be leveraged to discover novel techniques. Graph theory also adds network relationships at a far lower technical and computational cost when compared to TDMs. However, calculating centrality measures can still be computationally expensive on large networks, especially when origin-destination matrices are required. Therefore, networks may need to be subset for better efficiency. Easy-to-use tools for calculating centrality are available for popular geographic information systems (GIS) software packages like Esri and QGIS as well as in coding languages such as R and Python. Table 2-7 summarizes these and other pros and cons of use of network centrality for estimating off-system AADT.

Table 2-7. AADT Estimation Techniques from Literature – Network Centrality

Types	Pros	Cons
Degree Closeness Betweenness	Utilize networks like TDM methods but can be analyzed more quickly. Contribute major accuracy improvements to regression models. Large body of work from outside the transportation field that can be applied in novel ways.	Determining shortest paths between all network nodes is resource intensive. Incomplete networks can cause major inaccuracies in results. Requires methodology to account for differences between central portions of the network and its extremities.

References for Network Centrality: Lowry (2014), Kehan (2017) Jayasinghe et al. (2019)

Explanatory Factors Considered in Literature

The real-world information provided to the models reviewed varied widely in the literature. However, the information referenced by these methods can be divided into four broad groups: information about the physical condition and design of the facility, the relationship between the facility and the wider network, socio-economic data about the surrounding area, and the demographics of nearby populations. A summary of methodologies' minimum data requirements is shown in Table 2-8, and recurrent data points from the literature reviewed is displayed in Table 2-9.

Table 2-8. Minimum Data Requirements by Methodology

Estimation Method	Observed AADT	Count Site Location	Road Network	Functional Classes	Population	**Workers**
Sampling	X	--	--	X	--	--
Regression	X	--	--	X	--	--
Geospatial	X	X	--	--	--	--
Machine Learning	X	--	--	X	--	--
TDM	X	X	X	X	X	X
Network Analysis	X	X	X	X	--	--

Note: A dashed line indicates that the data type is not strictly required to minimally use the method, and an "X" indicates that the data type is strictly required to minimally use the method.

Table 2-9. Common Explanatory Factors Used in Literature

Road Characteristics	Network Characteristics	Socio-Economic Data	Demographic Factors
Functional Classification Number of Lanes Speed Limits Surface Material	Distance to intersection Accessibility (to primary or secondary roads) Centrality Measures Road Mileage Density	Urban / Rural Designation Workers Employment by Industry Median Income Poverty Rates	Population Dwelling Units Vehicle Registration

Common Challenges

For models to perform well in real-world scenarios, their underlying processes should minimize statistical errors and be developed through an iterative process. Some examples of prominent challenges facing most transportation models are listed below

- Data availability & granularity
- Institutional technical knowledge
- Obtaining feedback from ground-truth-ed assessments
- Lag in data sets being updated (for example, when an employment center has been closed but data reflects the closure only after a substantial delay)

Validation of AADT Estimates for Local Roads

Most practitioners and researchers validate their results by reserving a portion of their locations from the model calibration process and using it to compare to the estimated AADT produced by their process. From the literature reviewed that estimates AADT for off-system public roads, reserving 25% of off-system road locations is most common (Mathew 2020, Staats 2016, Tsapakis, Holik, et al., Informational Guide on Data Collection and Annual Average Daily Traffic (AADT) Estimation for Non-Federal Aid-System (NFAS) Roads 2020, Pulugurtha and Mathew 2020), with one study reserving 30% for validation (Raja, Doustmohammadi and Anderson 2018).

There are several plots and statistics that are commonly used for validation, including visually examined scatter plots (Raja, Doustmohammadi and Anderson 2018), mean absolute error (Staats 2016), MAPE (Tsapakis, Holik, et al., Informational Guide on Data Collection and Annual Average Daily Traffic (AADT) Estimation for Non-Federal Aid-System (NFAS) Roads 2020, Mathew 2020, Staats 2016, Pulugurtha and Mathew 2020), mean percent error (Mathew 2020, Pulugurtha and Mathew 2020), root mean square error (Mathew 2020, Pulugurtha and Kusam 2012), R-square (Raja, Doustmohammadi and Anderson 2018), and the Nash-Sutcliffe coefficient (Raja, Doustmohammadi and Anderson 2018). While less used in the literature, an R-squared value or information criterion (such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC)) may also be used.

Current ITD AADT Estimation Process

This section summarizes ITD's existing process for estimating AADT and aligning segment-level AADT estimates with statewide VMT estimates. The information is based on interviews with ITD staff that occurred in June 2022 (Hanson and Calderon 2022, Coladner, Laib and Calderon, Modernization Processing Tools 2022, Pridmore and Calderon, Historical Method/Modernization 2022), ITD's *Traffic Smoothing Documentation Guide* (Idaho Transportation Department 2020), emails (Pridmore 2022), and review of Python script in the Traffic Smoothing Toolbox. ITD's current processes have grown in response to challenges that it has faced and resolved in the traffic volume estimation processes prior to the current tenure of personnel. For instance, in 2018, ITD's principal research analyst for traffic counts retired, and the resulting challenges due to institutional knowledge loss resulted in nearly a decade of counts being taken with little direction or rigor regarding location or timing (Pridmore and Calderon, Historical Method/Modernization 2022). With roadway data management in 2018, issues such as count inconsistencies and high process labor-hours prompted an effort to revise and automate the count and AADT estimation process. Work to automate and streamline information processes revealed substantial gaps in data and some facilities with no data at all. After examining resources, ITD realized that it had the tools to automate significant portions internally of the process of recording information and the estimation smoothing processes (Pridmore and Calderon, Historical Method/Modernization 2022).

ITD's current AADT estimation process occurs primarily on the state highway system. It supports reporting of HPMS samples, which do not cover the entire Federal-Aid System but are generated by the FHWA provided software. ITD has primarily conducted counts on Federal-Aid System along with some on the state highway system and off-highway system.

Traffic Volume Count Program

Historically, AADT estimation methodologies have used a combination of traffic counts from permanent stations and short-term counts (1-3 days) to predict values by functional classification or geographic location. Traffic volume counts are done at set station locations identified by ITD. Traffic volume counts at active stations can either be from permanent count equipment (also known as 'control' and includes equipment such as magnetic loops embedded in pavement) or from short-term counts done by ITD personnel or by temporary count installations (also known as 'coverage' and includes equipment such as pneumatic tube laid across a road). Field staff use pneumatic road tubes (produced by International Road Dynamics Inc.) and video cameras (produced by Miovision) to conduct coverage counts. Control count equipment types are either Automatic Traffic Recorders (ATRs) or Weigh-in-Motion (WIM) devices that are embedded in or below the roadway surface. These counts are validated annually with hand counts (Genlogs) conducted by ITD field staff. Additional traffic volumes may be provided by MPOs and external agencies (Idaho Transportation Department 2020), while some localities do not want ITD to collect data within their boundaries, requiring ITD to estimate AADT on relevant roads using informed qualitative assessment (Pridmore and Calderon, Historical Method/Modernization 2022). Table 2-10 summarizes the number of units and types of data collected for permanent and short-term data collection devices managed by ITD.

Table 2-10. ITD-Management Traffic Data Collection Devices

Equipment Type	Number of Units of Equipment	Data Collected	Additional information
Weigh-in-motion (WIM)	20+	Volume, vehicle type, speed, and weight	Permanent count equipment
Automatic traffic recorders (ART)	200+	Volume for sites with one loop. Volume, vehicle type, and speed for sites with two loops.	Permanent count equipment
Two pneumatic tubes	300-450	Traffic volume and vehicle type	Short-term count equipment
One pneumatic tube	2,000-3,000	Traffic volume	Short-term count equipment

Source: Derived from *Traffic Smoothing Document Guide*

Before the current methodology was implemented in 2020, there were 62,000 designated station locations. However, the number of field and office staff—four total—was insufficient for collecting and maintaining data resources on such a large system. To begin reducing the number of locations to a more manageable size, ITD retired between twenty and thirty thousand locations where counts had not occurred in the past 10-20 years. Of the remaining ~40,000 station locations that can still be used for counts if needed, ITD has designated 8,179 sections as active count locations (Hanson and Calderon 2022).

A traffic count can be influenced by its point location relative to the rest of the segment. Previously, count locations were inconsistent from year-to-year, which made establishing trends difficult since the inconsistent location along a segment added noise. This inconsistency was addressed with the creation of locations of record for each segment with a count location (Pridmore and Calderon, Historical Method/Modernization 2022).

The frequency of coverage counts depends on the most recent AADT estimate and whether it is an HPMS sample location (Pridmore and Calderon, Historical Method/Modernization 2022) as summarized below.

- HPMS sample locations or locations with AADT above 3,000 are counted once during a three-year period.
- Locations with AADT of 1,000-3,000 are counted once during a four-year period.
- Locations with AADT less than 1,000 are counted once during a five-year period.
- Exceptions occur most often at intersections, where count frequency is based on the highest volume leg.

Unfortunately, there are several issues with the count program currently. Some communities are averse to ITD conducting business on their roads, whether they are state-owned or not; no unpaved facilities have continuous count locations; and there are currently only three ITD field staff conducting short-term counts and related activities. The funding needed to expand the number of count locations or the frequency of counts is not expected within the foreseeable future, so ITD will likely need to estimate volumes on facilities with no direct observations to meet regulatory requirements (Pridmore and Calderon, Historical Method/Modernization 2022).

Count data and additional information that is relevant to AADT estimation (e.g., roadway section descriptions, historical counts, historical AADT and commercial AADT (CAADT)) is accessed via ITD's relevant data storage and management systems. Jackalope is the ITD Roadway Data Unit's traffic data management system, and TRADAS serves as Jackalope's back-end database, meaning that it contains count data and the additional relevant information. ITD uses Oracle SQL Developer to extract data from TRADAS (Idaho Transportation Department 2020).

Traffic Smoothing Process for Count-based AADT Estimation

While traffic volumes on primary arterials and major collectors are often recorded directly with permanently installed equipment allowing for straightforward calculation of the road's AADT, volumes on off-system public roads are typically unknown and must be estimated when funds are not apportioned to conduct direct observation. Moreover, even when counting equipment is present, count data must be checked for reasonableness, and in some instances data from multiple count locations on the same segment must be consolidated (Idaho Transportation Department 2020). Once count data has been obtained for a year, the ITD data analytics team processes the information to adjust for factors that impact the estimation of AADT across the entire state road network. The traffic smoothing process converts raw traffic counts and in some cases vehicle type classifications to AADT and CAADT. This subsection summarizes the traffic smoothing data inputs and process.

The Traffic Smoothing Toolbox is a custom-made toolbox scripted in Python and run in Esri ArcMap Desktop (Coladner, Laib and Calderon, Modernization Processing Tools 2022). ITD developed the scripts for the tool over the last three to five years, and as of summer 2022 ITD is in the process of updating them from Python 2 to Python 3 in anticipation of an eventual change to Esri's ArcGIS Pro software package (Hanson and Calderon 2022).

The Traffic Smoothing Toolbox contains steps that users follow to generate AADT estimates. Raw data is cleaned and processed. While some steps require the user to specify input and to review outputs such as flagged roadway sections, the toolbox is designed to automate data inputs and processing as much as possible. Each step is supported by one or more Python scripts that are called when the user runs the toolbox step. These primary steps are summarized in Table 2-11 below in order of their operation, and the corresponding Python scripts are listed. The steps progress from preparing a data set containing all the information that the tool requires to checking that data set for errors, using it to estimate AADT and CAADT, flagging sections for manual review, and preparing a final data set incorporating the automated and manually reviewed AADT estimates (Idaho Transportation Department 2020).

Table 2-11. Steps in Traffic Smoothing Toolbox

Traffic Smoothing Toolbox Steps	Description	Python Scripts
Data Setup Location Tool	Creates a JSON artifact of 'master' location data.	DataSetupGeospatial_1.py DataSetupLocation_2.py
Data Setup Station	Creates a JSON artifact file containing time-slicing station location. Stations are where portable short-term counters and continuous ATR and WIM sites are located.	DataSetupStation_3.py
Data Setup Count	Creates a JSON artifact file with current time-slicing count data. Requires connection to TRADAS database.	DataSetupCount_4.py
Data Setup AADT	Creates a JSON artifact file containing current time-slicing AADT data.	DataSetupAADT_5.py
Data Setup Create Master	Performs work to combine location, station, count, and AADT data into a single JSON artifact file	DataSetupCreateMaster_6.py
Data Checks	Performs a series of data checks on the JSON artifact files to ensure that sections and routes are not being inadvertently added or dropped.	TrafficSmoothingDataChecks.py
AADT Data Deconstructor	Performs the AADT forecasting and writes the forecasted data to a CSV artifact file containing location, forecasted AADT, and other section data.	AADTDataDeconstructor.py
CSV to Feature Table	Transforms the AADTOutput.csv file into a feature table with matching LRSE.AADT schema.	AADTCsvToFeatureTable.py
Final Feature Class	Creates a feature class using a subset of valid LRSN_RoadNetwork's route feature and the feature table created in step 5.CsvToFeatureTable.	CreateFinalFeature.py
Flagged Feature Class	Creates a feature class containing only the flagged records, along with a feature class for historical data required for the revising procedure.	InitializingFlaggedFeatureTool.py FlaggedFeature.py
Find StationID	Establishes a connection to TRADAS Historical AADT database and retrieves the valid StationIDs and the corresponding information.	FindStationID.py
Distribute Tiers of Data	Divides the initial estimates feature class into several tiers of AADT records.	DistributeTiers.py
Join All Records	It creates a joined data set after final review of final smoothing process.	JoinAllRecords.py
Final_AADT_csv	Creates a .csv table as the final output of this toolset.	FinalAADTcsv.py

Source: Derived from *Traffic Smoothing Documentation Guide*

The key technical steps for AADT estimation occur in the AADT Data Deconstructor step. This is where counts are converted to AADTs for road sections with current-year counts, and current-year AADT estimates are produced based on historical data for road sections without current-year counts (Idaho Transportation Department 2020). The way in which AADT is estimated depends on the availability of current-year and historical count data.

- **Sections for which current-year count data is available:** If current-year count data is available at a count station that corresponds with the roadway section, the tool converts that data to

AADT/CAADT.[1] When count data is available from multiple sources, the process uses in order of decreasing priority WIM data, ATR data, and short-term count data (e.g., pneumatic tubes). When short-term counts are used, the process uses factors to account for seasonal, day-of-week, and hour-of-day bias (Idaho Transportation Department 2020). These seasonal, day-of-week, and hour-of-day factors are derived for individual clusters of short-term count locations that ITD creates based on similarity among the count locations (Pridmore and Calderon, Historical Method/Modernization 2022). In their second year of use as of summer 2022, growth factors allow counts that occurred in a prior year to be updated to the current year. These growth factors are also utilized for the statewide VMT estimation process (Pridmore and Calderon, Historical Method/Modernization 2022).

- **Sections for which current-year count data is not available, but historical data is available:** The process uses linear regression to extrapolate from historical data to the current year using the last five years of data. When there are multiple traffic counts in a given historical year, the tool generates a single value by taking the average of those counts. The tool also uses historical data to fill in data that is missing from a year of historical data. The tool removes outliers in the historical data before running the regression (Idaho Transportation Department 2020).
- **Sections with neither current-year count data nor historical data:** These roadway sections are flagged for manual review. The process also flags AADT and CAADT estimates that deviate substantially from the prior-year estimates, as described in the Table 2-12 and Table 2-13.

Table 2-12. Conditions for Flagging AADT Estimates

Prior Year AADT	New Estimated AADT Within
100,000 or more	+5%
50,000 – 99,999	+10%
10,000 – 49,999	+20%
5,000 – 9,999	+30%
1,000 – 4,999	+40%
Less than 1,000	+50%

Source: *Traffic Smoothing Documentation Guide*

[1] The process for assigning count stations to sections is not detailed here because it is unlikely to change for off-system AADT estimation. ITD resolves differences on the same section by balancing counts according to internal methods based on count location (as described in interview with ITD staff).

Table 2-13. Conditions for Flagging CAADT Estimates

Prior Year CAADT	New Estimated CAADT Within
10,000 or more	+/-05%
5,000 – 9,999	+/-10%
1,000 – 4,999	+/-20%
500 – 999	+/-100 (CAADT estimate)
100 – 499	+/-60 (CAADT estimate)
Less than 100	+/-20 (CAADT estimate)

Source: *Traffic Smoothing Documentation Guide*

Rounding of estimated AADT and CAADT occurs as summarized Table 2-14. Values under 1,000 are rounded to the nearest ten, values between 1,000 and 10,000 are rounded to the nearest hundred, and values over 10,000 are rounded to the nearest five hundred. The *American Association of State Highway and Transportation Officials (AASHTO) Guidelines for Traffic Data Programs* (2009) suggests rounding values over 10,000 to the nearest thousand, but ITD has found it easier to balance high and low volume roads with the higher degree of precision.

Table 2-14. Rounding Rules for AADT and CAADT

AADT Range	Rounding Precision
Less than 10	Convert to 10
Less than 1,000	Round to nearest ten
Equal to or greater than 1,000 and less than or equal to 10,000	Round to nearest hundred
More than 10,000	Round to nearest five hundred

Source: Script in Smoothing Toolbox called AADTDataDeconstructor.py

After the tool has been run for a year's counts, a geospatial file is generated and sent to the linear referencing system (LRS) administrator to be added to the Esri Roads and Highways system that houses ITD's roadway geospatial data. The LRS does not cover 100% of Idaho's roads; however, new facilities are being added throughout the year as new roads are built, existing ones are changed, or previously unrecorded facilities are included. Within the Local Road Inventory (LRI) program, local highway districts inform ITD of new roads that have been added or extended (Hanson and Calderon 2022). New facilities may also be added when the Crash Analyst finds that an accident has occurred on a road that is not part of the LRS.

Statewide VMT Estimation

Because no data source directly measures VMT, statewide VMT is estimated from contextual data, including facility AADT where available. In the past, ITD has considered data sets such as population growth, gas tax revenues, and vehicle registrations in calculating statewide VMT (Pridmore and Calderon 2022), and would combine these with its observations on AADT change that it had measured for the Federal-Aid System to establish generalized AADT for functionally classified local and rural minor collector roads, which allowed it to estimate VMT for the non-Federal-Aid System. These generalized AADTs mean that every low-functionally classified road in a given county has the same VMT (Pridmore 2022).

ITD decided to change this process for several reasons. Since the COVID-19 pandemic the indicators of population growth, gas tax revenues, and vehicle registrations had become less reliable and less representative of vehicle-miles traveled on Idaho roads (Pridmore and Calderon 2022, Pridmore 2022). For example, ITD saw increased use of recreational vehicles during 2020 and 2021, which increased fuel consumption by more than VMT, and observed data sources like gas tax were too disconnected from VMT to be relied on for estimation purposes (Pridmore and Calderon, Historical Method/Modernization 2022). Additionally, ITD observed that these indicators were not adequately attuned to VMT changes even independently of factors related to the COVID-19 pandemic (Pridmore 2022).

ITD developed a new method for estimating statewide VMT. The statewide VMT has two components. It calculates the first component, VMT on the Federal-Aid System, from the AADT estimates for roadway sections derives from the Traffic Smoothing Toolbox as previously described. The second component is VMT for the non-Federal-Aid System, which includes 80% of the road mileage but only about 20% of the VMT on average. ITD derives county-level VMT and creates growth factors for groupings of geographically close counties in which VMT on the Federal-Aid System has changed in similar ways, and it creates a growth factor for each urban area and small urban areas. The growth factors are derived from changes in VMT for Federal-Aid System roads. The way in which these growth factors are applied to prior-year AADT estimates differs based on roads' location (Pridmore 2022).

- **For roads in urban areas**, ITD applies the appropriate growth factor to prior-year VMT for non-Federal-Aid System roads.
- **For roads in small urban areas**, the process is the same as for urban areas except that ITD accounts for college towns—which saw larger-than-usual COVID-related traffic changes—separately from other small urban areas.
- **For roads in rural areas**, ITD applies the appropriate growth factors after weighting prior-year VMT by mileage per county.

ITD checks the reasonableness of these estimates by comparing against FHWA Traffic Volume Trend reports (Federal Highway Administration 2022), growth rates for the Federal-Aid System, and growth observed by ITD's continuous count sites (Pridmore 2022).

Challenges for Estimating AADT on Off-System Public Roads

The requirement for state DOTs to estimate AADT on all local public roads comes in part from HSIP, which is requiring every public paved road to have AADTs made available by 2026. There are several challenges facing ITD in preparing estimates for this requirement, many of which align with the challenges noted in the literature such as institutional knowledge and data availability.

At 22.3 people per square mile in 2020, Idaho is the 44th least population dense state in the U.S., and a few rural roads on state highway system are unpaved (< 0.1%). However, unpaved local roads are often used for going to a site (like a campground) at the end of the road and for passing through areas where a paved road cannot be maintained. Therefore, use patterns on unpaved state system roads are not analogous to paved local roads.

Additionally, methodologies for estimating AADT primarily focus on National Highway System (NHS) or state highway systems. Local, off-system, and low-volume roads have only seen attention in academic literature in the last five to ten years. ITD will not only need to evaluate several potential methodologies for calculating estimates, but it will also need to find ways to evaluate a methodology and have systems in place to adjust methodologies as federal requirements and socio-demographic trends shift over the coming years.

From a practical standpoint, ITD expects to switch its GIS processing to Esri's ArcGIS environment in the next 2-4 years (Hanson and Calderon 2022). Any chosen methodology will need to be implemented in that software, and this does impose limitations on the techniques available for a given level of staff expertise, experience, and time investment.

Ultimately, the chosen methodology will supplement the Roadway Data and Data Analytics team's current practices in estimating AADT as well as the efforts to estimate statewide VMT. As such, it will need to take those projects into consideration with the goals of improving overall accuracy and better serving Idaho residents.

3. Data Inventory

This chapter inventories all readily available data sources suitable for estimating AADT on public roads in Idaho. It describes each data source, in terms of its geospatial scale and network coverage, and compares them with the requirements of the AADT estimation methods documented in this project's task 2 report *Literature Review, FHWA Requirements, and Documentation of Existing AADT Estimate Methodologies*.

The chapter has four sections, including this Introduction. The "Data Source Gathering Methodology" section describes how data sources were identified and how information about each data source was obtained. The "Data Sources" section inventories data sources related to each data category introduced in the literature review including:

- Count data
- Roadway data
- Network data
- Economic data
- Demographic data

Data Source Gathering Methodology

The project team gathered Idaho-specific data sources corresponding with the variables and data sources described in the literature review from several sources.

- **National Data Sets –** The project team identified relevant national economic and demographic data sets, which include the Decennial Census, the American Community Survey (ACS), the Longitudinal Employer Household Dynamics (LEHD) Origin Destination Employment Statistics (LODES) by the U.S. Census Bureau, and data sets from the U.S. Geological Survey.

- **State Data Sets –** ITD shared relevant data sets along with information about the project team such as field names, data types, and feature types. ITD's data sets are listed in Table 8-1 of Appendix B on page 90. The project team interviewed ITD subject matter experts when additional information was needed to fully understand data extent, geospatial scale, network coverage, and other characteristics. In addition, the team added data from the ITD Department of Motor Vehicles (DMV) and examined ITD's Open Data portal (Idaho Transportation Department 2022) to ensure no major data sources had been omitted.

- **MPO Traffic Counts –** The project team examined traffic count data produced by each Idaho Metropolitan Planning Organization (MPO). While MPOs' websites summarize their count data collection processes, none provide a complete picture of such practices; to fill in gaps, the project team contacted MPO staff identified by ITD and held virtual meetings to clarify details about the purpose, extent, recency, and other characteristics of MPO-collected traffic counts.

- **Local Government Traffic Counts –** The project team sampled information about all highway districts and several randomly selected county governments to determine the extent to which highway districts and local governments conduct traffic counts and what kinds of data collection, processing, and formatting standards are used. Highway districts build, maintain, repair, and acquire highways and rights of way for their highway systems, whose boundaries often coincide with county boundaries (Idaho Statues 40-1302 and 40-1310). The project team focused its review of highway districts on members of the Idaho Association of Highway Districts (IAHD) with websites listed on the IAHD's membership page (Idaho Association of Highway Districts 2022). The project team found that the local agencies sampled do not use common data standards for traffic counts, so using locally collected traffic count data across regions is not practical.

Data Sources

This section inventories the data sources that most closely correspond with the types of data used in the AADT estimation methodologies described in the literature review. The more data sources are available for the larger share of Idaho's network of public roads, the more AADT estimation methodologies have the potential to estimate AADT on Idaho's off-system public roads. The ways in which the data would be used to estimate AADT vary across the methods. There are six subsections, each describing the availability of a different type of data.

Traffic Count Data Sets

Traffic count data sets record the number and sometimes types of vehicles using roads or intersections. Normally, counts distinguish among direction of travel, and they may distinguish times of day. This subsection inventories traffic counts collected by ITD, states, and localities.

Idaho Transportation Department (ITD) Traffic Count Data

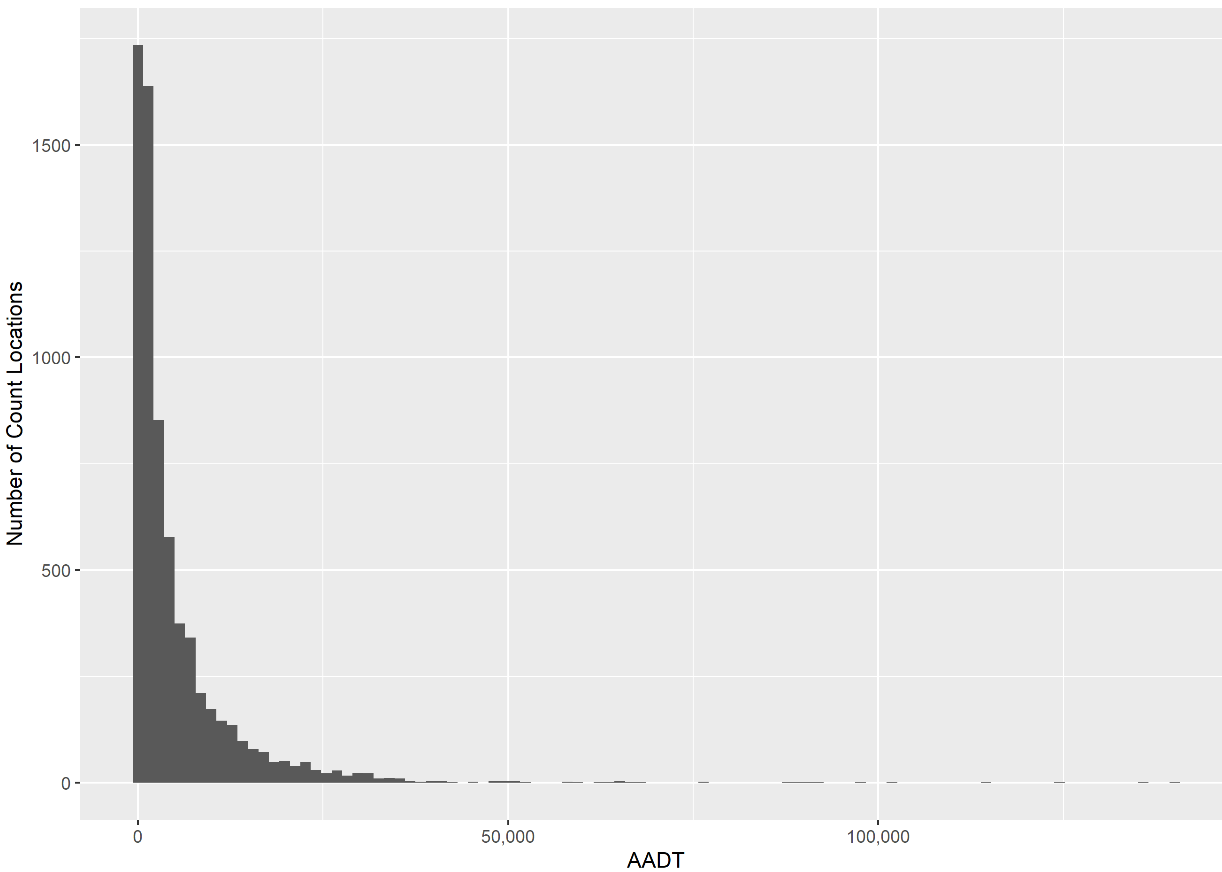
As discussed in the literature review, ITD collects traffic volume data using a combination of continuous and short-duration counts, which it combines with other data to estimate AADT on all state-owned and Federal-Aid roads. ITD keeps internal records of count data, and its website provides some continuous count data collected by ATRs and WIM devices (Idaho Transportation Department 2022).

There are 7,768 count locations of which 3% are continuous count sites (ATR or WIM sites), 73% are on count cycle (which are on the master list to be scheduled), and 24% are assigned stations (which are identified as an assigned station for one of ITD's AADT sections without being on the master list to be scheduled). When examined at the county level, the median county has 105 count stations of any type, while the maximum county has 1,064 and the minimum has 24.

There are many more low- and medium-volume roads than high-volume roads, and the amount of traffic measured by count locations reflects a similar distribution of data with more count locations measuring low- and medium-volume roads than high-volume locations. Traffic volume was associated with each count location by matching the most recent AADT estimate with each traffic count location

(Idaho Transportation Department 2022). Figure 3-1 below is a histogram showing the number of count locations with different levels of AADT, which shows skewed traffic volumes with many more count locations on roads with lower traffic volumes than higher traffic volumes. This skew is important for estimating off-system AADT since many off-system segments have lower traffic volumes than the state highway system or Federal-Aid roads.

Figure 3-1. Histogram of AADT and Count Locations



Metropolitan Planning Organization Traffic Count Data

All five MPOs in Idaho collect some traffic count data and provide information about it on their websites, but none make the information publicly available for download in a geospatial format. Additionally, most of the data provided is more than two or three years old, and filtering out these older counts would reduce network coverage. The count data that is available requires additional processing to account for hour of day, day of week, and seasonal trends before it can be used to derive AADT estimates. Table 3-1 summarizes the MPOs' count data characteristics, and Table 3-2 provides the corresponding data sources.

Table 3-1. MPO Traffic Count Data

MPO Name	Years	Notes
Community Planning Association of Southwest Idaho (COMPASS)	Not specified	Counts are collected from member agencies, particularly the Ada County Highway District, which collects counts on a three-year cycle and by several localities in Canyon County. Data is stored using segmentation that has a pavement management ID (PMID), which could be cross referenced with the LRS. All counts are non-continuous.
Kootenai MPO	2010-2022 (depending on count source)	Counts are collected from localities and highway districts. Frequency of counts depend on the organization providing them. The highway districts generally provide the most frequent counts and the most counts on local roads, with localities generally providing counts for higher functional class roads.
Bannock Transportation Planning Organization	2017-2022	The MPO collects counts for Interstates, arterials, collectors, and some local side streets approximately every three years. The data showcased on website is for prior 4-5 years, though the MPO has more historical records available. Counts are based on short-duration observations (1-3 days) and are collected using road tubes, radar, and video.
Bonneville MPO	1993-2019	Counts are taken primarily on collectors and higher functional classifications. Local roads may be included if they are near a collector or are the site of a local traffic study. Counts are used for travel demand modeling. Counts are obtained primarily with tube counters.
Lewis-Clark Valley Metropolitan Planning Organization (LCVMPO)	Not specified	Counts cover peak periods only, though 24-hour counts will be added in the near future. The MPO is currently collecting counts for most roads in the MPO area starting with functionally classified roads for its long-range transportation plan, and counts are expected to be complete in 2023.

Table 3-2: MPO Data Sources

MPO Name	Website Link	Interviewee Name and Interview Date
Community Planning Association of Southwest Idaho (COMPASS)	https://www.compassidaho.org/prodserv/traffic_counts.htm	MaryAnn Waldinger and Mitch Skiles, September 26, 2022
Kootenai MPO	https://www.kmpo.net/traffic-counts/	Ali Marienau, October 12, 2022
Bannock Transportation Planning Organization	https://www.bannockplanning.org/traffic-counts/	Mori Byington, September 23, 2022
Bonneville MPO	https://www.bmpo.org/traffic-counts	DaNiel Jose, September 26, 2022
Lewis-Clark Valley Metropolitan Planning Organization (LCVMPO)	https://lewisclarkmpo.org/2209/Intersection-Counts	Shannon Grow, September 29, 2022

Localities

Almost none of the local or highway districts the team reviewed have traffic count data collection programs with the exception of the Ada County Highway District which provides 24-hour and, in some cases, directional AM and PM peak traffic for some locations (Ada County Highway District 2022). The lack of local traffic counts supports the assumption that local traffic counts are too sparse to provide adequate geospatial coverage for a statewide AADT estimation process and that collection techniques and data formats are likely to be too disparate to use without excessive cleaning and processing.

Roadway Data Sets

ITD collects many types of data describing roadways, such as functional classification, whether or not shoulders are present, the terrain that the roadway traverses, and speed limits. The subset of roadway data types that are needed to support the AADT estimation approaches detailed in the literature review are shown in Table 3-3 and discussed in more detail in the subsections following the table. The HPMS coding for the HPMS data sets, which describes the data's range of possible values, is listed in Appendix B.

Table 3-3. ITD Roadway Data Sets*

Data	Data Set Name	Manager / Source
Functional Classification	rhgdb.ITDRH.LRSE_FunctionalClass	Headquarters (HQ) Planning
Number of Lanes	rhgdb.ITDRH.LRSE_HPMS_ThroughLanes	HPMS Coordinator
Lane Width	rhgdb.ITDRH.LRSE_HPMS_LaneWidth	HPMS Coordinator
Surface Material	rhgdb.ITDRH.LRSE_HPMS_SurfaceType	HPMS Coordinator
Speed Limits	rhgdb.ITDRH.LRSE_HPMS_OHSpeedLimits	HPMS Coordinator
Parking	rhgdb.ITDRH.LRSE_HPMS_Parking	HPMS Coordinator
Shoulders	rhgdb.ITDRH.LRSE_HPMS_Shoulders	HPMS Coordinator
Terrain Type	rhgdb.ITDRH.LRSE_HPMS_TerrainType	HPMS Coordinator
Interchanges	rhgdb.ITDRH.LRSE_Interchanges	HQ GIS
Local Road Inventory (LRI)	rhgdb.ITDRH.LRSE_LocalRoadInventory	HQ GIS

** All data is collected at the road segment level.*

These data sets are primarily concerned with state highway system facilities. Their coverage of the statewide highway system (which is described below in the Network Data section) is summarized in Table 3-4, which reports the percentage of route identifiers and the percentage of total miles with non-null data. There are 48,746 unique route identifiers and 72,593 miles of roadway in the ITD road network. Off-system facilities account for 42,484 route identifiers and 44,358 miles of roadway. Existing coverage of ITD roadway data across the network shows that only the number of lanes has widespread coverage, indicating that AADT estimation will likely rely largely on external data sources. There are two surface type data sets. The LRI data set for surface type covers local roads where the vast majority of unpaved roads are expected to be located, and it provides nearly complete data on surface type for that subset of the road network. It should be possible to identify essentially all unpaved roads between the two data sets for surface type.

Table 3-4. ITD Roadway Data Set Network Coverage

Data	Percent of All Route Identifiers	Percent of Total Network Miles	Percent of Off-System Route Identifiers	Percent of Off-System Network Miles
Functional Classification	7.52%	32.72%	1.30%	4.52%
Number of Lanes	97.24%	90.15%	97.17%	71.90%
Lane Width	1.51%	15.56%	0.002%	0.017%
Surface Type – LRI	28.12%	44.37%	27.58%	35.19%
Surface Type – HPMS	1.58%	15.05%	0.002%	0.017%
Speed Limits	1.13%	0.98%	0.002%	0.017%
Parking	1.50%	14.98%	0.002%	0.017%
Shoulders	1.48%	13.09%	0.002%	0.017%
Terrain Type	1.49%	13.43%	0.002%	0.017%

Note: Where functional classification is not specified, it should be assumed to be class 7 (local).

Functional Classification

Functional classification groups facilities based on the type of service they provide within the road network. According to FHWA's HPMS, group 1 includes Interstates, group 2 includes other freeways and expressways, group 3 includes principal arterials, group 4 includes minor arterials, group 5 includes major collectors, and group 6 includes minor collectors. Local roads either receive a functional classification of 7 or no functional classification. States are also required to designate facilities as either urban or rural. Common practice is to either create urban area limits for roadways or to expand the functional classification to have separate sets of numbers for urban and rural facilities. Updates to functional classifications normally occur after the release of decennial census data.

Lane Characteristics

The ITD roadway data set describes the number of through lanes in both directions carrying through traffic in the off-peak period and the through-lane width to the nearest whole foot.

Surface Material

The ITD roadway data set describes roadway construction materials, which dictate ease of travel and, by extension, typical driver speeds and volumes. The three most common surface materials are asphalt, concrete, and gravel (or unpaved). A full list of surface materials is included in Appendix B. The local roadway inventory also provides information about roadway improvements for off-system facilities.

Speed Limits

Where available, speed limits are recorded by ITD in miles per hour.

Shoulders

Facility shoulders are recorded by ITD in the following categories:

1. None
2. Surfaced shoulder exists – bituminous concrete
3. Surfaced shoulder exists – Portland Cement Concrete surface (PCC)
4. Stabilized shoulder exists (stabilized gravel or other granular material with or without admixture)
5. Combination shoulder exists (shoulder width has two or more surface types; e.g., part of the shoulder width is surfaced and a part of the width is earth)
6. Earth shoulder exists
7. Barrier curb exists; no shoulder in front of curb

Parking

ITD records the availability of parking along a facility's edge. The presence of vehicle parking can help communicate characteristics about the facility's use and how drivers perceive available space along the roadway. Parking status is allowed on one side, allowed on both sides, or no parking allowed/none available.

Terrain Type

ITD categorizes the surrounding terrain for a facility as either level, rolling, or mountainous. Rapid changes in elevation and tighter turns necessitated by roadway design in rough topography require lower speeds to navigate safely, which may deter drivers and result in lower traffic volumes.

Network Data Sets

Network data can be derived from three primary ITD data sets: rhgdb.ITDRH.LRSE_RoadNetwork, rhgdb.ITDRH.Intersections, and rhgdb.ITDRH.LRSE_FunctionalClass. Most network characteristics used in estimation methodologies the team studied in the literature review require additional calculation based on these data sets. These characteristics and the associated data sets are listed in Table 3-5 and the subsections after the table briefly describe how each data set could be converted to a network characteristic.

Table 3-5. Network Data

Data	Data Set Name	Manager / Source
Road Network	rhgdb.ITDRH.LRSE_RoadNetwork	HQ GIS
Distance to Intersection	rhgdb.ITDRH.Intersections	HQ GIS
Accessibility (to Primary or Secondary Roads)	rhgdb.ITDRH.LRSE_FunctionalClass	HQ Planning
Total County Arterial Mileage	rhgdb.ITDRH.LRSE_FunctionalClass	HQ Planning
Centrality Measures	rhgdb.ITDRH.LRSE_RoadNetwork	HQ GIS
Road Mileage Density	rhgdb.ITDRH.LRSE_RoadNetwork	HQ GIS

Road Network Data

The road network is represented as lines and points which correspond to road segments and intersections respectively. In addition to data about the physical characteristics of the roadway already discussed, analysis of network connections in a system can quantify differences in facility traffic volumes that are attributable to destination accessibility.

Distance to Intersection Data

The distance from any segment to an intersection can be calculated through a GIS application such as Esri ArcMap. This value can be calculated as either the spatial distance to the nearest intersection point or as the distance along the network. Alternatively, the value can be a statistic derived from distances to multiple intersections within a subset of the network.

Accessibility to Primary or Secondary Roads Data

Accessibility can also be gauged by measuring the distance from a given road segment to the nearest facility of a given class, such as an Interstate or arterial road. Proximity to these high traffic volume facilities may suggest higher volumes on smaller facilities than on those isolated from major elements of the network. This distance to the nearest facility of a given functional class can be measured as a straight line to the nearest feature or by using the network to find the shortest driving distance to a qualifying facility.

Proximate Arterial Mileage Data

Arterial mileage or mileage for any other functional class can be calculated at the county scale or any other scale using GIS and a measure of functional class. If the functional class data set contains a centerline mileage field, then it can be used as the source of segment lengths. Otherwise, segment length can be derived from the segment geometry. If the functional class data set also contains a count of through lanes or can be joined with a count of through lanes, it can also be multiplied by centerline mileage to estimate lane mileage.

Road Mileage Density Data

Road mileage can be calculated from a road network as the ratio of the length of roads in a given area to the land area. While road length can come from a road network data set, the land area must come from another data set, either derived from a polygon of the area in question (e.g., census block group, census tract) using GIS tools to calculate area based on the geometry or from a polygon attribute containing land area, such as is the case in many of the U.S. Census Bureau's cartographic boundary files (U.S. Census Bureau 2022), whose "ALAND" attribute describes the land area in square meters (U.S. Census Bureau 2022).

As scale increases, metric values will become more consistent, especially in rural areas with lower population and road facility densities. However, this smoothing may make the metric less effective after exceeding a certain scale. Some evaluation will be needed to identify potential issues and optimal scale.

Centrality Measures Data

Several possible centrality measures can be calculated from road network information. Centrality comes from the field of network theory or graph theory. Applied to transportation networks, segments are termed "links" or "edges" and intersections are termed "nodes" or "vertices". Common GIS applications such as QGIS or Esri products include tools to calculate centrality measures directly from road network geospatial data. Centrality measures rank or assign numeric values to nodes or edges based on their relationship to the larger network. Potential measures of centrality that are calculable from the road network include the following:

- **Degree Centrality:** Number of links connected to a given node or adjacent nodes (Golbeck, Analyzing Networks 2015).
- **Closeness Centrality:** "Indicates how close a node is to all other nodes in the network" (Golbeck 2013).
- **Betweenness Centrality:** Reflects how often a node or edge is used as part of the shortest route between any two other nodes or edges (Marsden 2005).
- **Eigenvector centrality:** "Measures a node's importance while giving consideration to the importance of its neighbors" (Hansen, et al. 2020).

Economic Data Sets

Economic data can be acquired from public federal sources, such as the U.S. Census Bureau and the U.S. Geological Survey. Nearly all the most common data sources cited in the literature for the relevant AADT estimation approaches are available using data from one of these two agencies and these data sets are typically reported at small spatial scales, such as census block, block group, or tract. The exception is land use, which does not exist in a national database but is generally tracked by localities. Here, National Land Cover is the best national alternative. Table 3-6 lists key economic data types.

Table 3-6. Economic Data

Data Type	Name	Source	Year	Smallest Unit
Surrounding Land Uses	National Land Cover Database (NLCD)	U.S. Geological Survey	2019	Raster
Urban / Rural Designation	Urban Areas	U.S. Census Bureau, Tiger/LINE	2021	Polygon
Urban / Rural Designation	Core Based Statistics Areas (CBSAs)	U.S. Census Bureau, Cartographic Boundaries	2018	Polygon
Urban / Rural Designation	Metropolitan Planning Organization (MPO) boundaries	U.S. Department of Transportation, National Transportation Atlas Database	2019	Polygon
Workers	LODES 7 Workplace Area Characteristics (WAC)	U.S. Census Bureau	2017	Block (2010 boundaries)
Employment by Industry	LODES 7 Workplace Area Characteristics (WAC)	U.S. Census Bureau	2017	Block (2010 boundaries)
Median Income	B19001 | Household income in the past 12 months (in 2020 inflation-adjusted dollars)	U.S. Census Bureau, American Community Survey	2020 5-Year Estimate	Block Group
Median Income	B19013 | Median household income in the past 12 months (in 2020 inflation-adjusted dollars)	U.S. Census Bureau, American Community Survey	2020 5-Year Estimate	Block Group
Poverty Rates	B17010 | Poverty status in the past 12 months of families by family type by presence of related children under 18 years by age of related children	U.S. Census Bureau, American Community Survey	2020 5-Year Estimate	Block Group

Surrounding Land Uses

No statewide or regional data set was identified that distinguishes among common land use types such as residential, commercial, and industrial land. Idaho does have a land use technical working group that is seeking to standardize land use classifications and to facilitate data integration among localities (Idaho Geospatial Office 2022). However, it does not provide or reference any data on its website. Therefore, the closest data set is the National Land Cover Database (NLCD) maintained by the U.S. Geological Survey. It is a nationwide raster data set that distinguishes categories such as high-, medium-, and low-intensity developed; open space developed; cultivated crops; grassland; evergreen forest; shrub/scrub; and woody wetland (U.S. Geological Survey 2018).

Urban / Rural Designation Data

Urban and rural areas can be designated in several ways; for instance, areas inside MPO boundaries can be designated as urban while areas outside MPO boundaries are designated as rural, or they can be designated based on U.S. Census Bureau urban area boundaries or core-based statistical areas (CBSAs), which include metropolitan and micropolitan statistical areas. There is adequate data to support all of these designations. ITD designates urban and rural locations based on functional class and MPO boundaries (Calderon and Laib 2022).

Number of Workers Data

The number of workers is available at the Census block level as recently as 2019 using workplace area characteristics (WACs) in the LODES data set. Upward from the block level, the number of workers can be aggregated and associated with road segments using GIS functions such as buffers and intersections. The LODES data set also includes origin-destination pairs for workers, which could be used to estimate trip distribution.

Employment by Industry Data

Employment at the two-digit North American Industry Classification System (NAICS) code level is available at the Census block level as recently as 2019 through the LODES data set. It can be summarized and associated with segments similarly to number of workers (U.S. Census Bureau n.d.).

Median Income Data

Several U.S. Census Bureau data sets provide estimates associated with median income at the block group level, including “household income in the past 12 months” (table number B19001) and “median household income in the past 12 months” (table number B19013). This block group-level data can be associated with road segments using GIS functions.

Poverty Rate Data

The U.S. Census Bureau data set “poverty status in the past 12 months of families by family type by presence of related children under 18 years by age of related children” (table number B17010) provides poverty rates at the block group level, which can be associated with road segments using GIS functions.

Demographic Data Sets

Table 3-7 lists the primary demographic data categories cited in the literature for AADT estimation referenced in the literature review. All the demographic data cited in the literature is available for recent years at the block or block group scale except for school enrollment (only available at or above the census tract level) and for data associated with vehicle registration and driver licenses (only available at the county level). Higher-scale data may be less impactful as it does not distinguish local nuances that can differentiate nearby road segments.

Table 3-7. Demographic Data

Data Type	Name	Source	Year	Smallest Unit
Nearby Population	P1 | Race	U.S. Census Bureau, 2022 Decennial Census	2020	Block
Population Density	Census Block	U.S. Census Bureau, Cartographic Boundary	2020	Block
Dwelling Units	H1 | Occupancy status	U.S. Census Bureau, DEC Redistricting Data	2020	Block
School Enrollment	S1401 | School enrollment	U.S. Census Bureau, American Community Survey	2020 5-year estimate	Census Tract
Vehicle Usage	Total Vehicle Registrations	ITD, DMV Data, "Vehicle Registration" tab	2010-2020	County
Vehicle Usage	Driver Licenses in Force by County	ITD, DMV Data, "Driver Licenses" tab	2017-2021	County

Population Data

Data on local population can be obtained from the U.S. Census Bureau, either in the Decennial Census count data or from estimates in their American Community Survey product. Densities can then be calculated using geographic areas.

Dwelling Units Data

The U.S. Census Bureau defines “housing units” or “dwelling units” as “A housing unit is a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied (or if vacant, is intended for occupancy) as separate living quarters. Separate living quarters are those in which the occupants live and eat separately from any other persons in the building and which have direct access from the outside of the building or through a common hall” (U.S. Census Bureau 2022).

When used in conjunction with population estimates, the number and density of units can serve to better estimate vehicle usage in the local area.

School Enrollment Data

The number and density of children enrolled in K-12 education can dictate major travel trends in local areas as parents/guardians and students drive to and from schools. Although enrollment is not available at a fine spatial resolution compared with some other data sets discussed here, it is still valuable in identifying patterns and estimating volumes in residential areas.

Vehicle Usage Data

County-level vehicle registration is available from ITD’s DMV data. Additionally, ITD provides the number of driver licenses in each county, which could serve as a proxy variable for vehicle registration.

Non-Government Data Sources

Several companies estimate traffic volumes using non-traditional means, such as cellular network pings, anonymized credit card transactions, Bluetooth signals, or smartphone location records tracked by mobile applications. These third-party sources for traffic-volume estimates are included here because they could either provide AADT estimates on some portions or all of the system for which ITD does not currently estimate traffic volume for a fee. This data could be used either as AADT estimates or as reference points for the AADT estimation validation process. Four companies were identified based on prior research, and these companies and sources are summarized below. The project team has not yet requested cost quotes.

INRIX / National Performance Management Research Data Set (NPMRDS)

INRIX provides directional traffic volume data optionally segmented by time of day in 15-minute intervals (INRIX 2022) which comes from global positioning system (GPS) probes in connected vehicles, commercial fleets, and mobile GPS applications (Datarade 2022). INRIX data is available for two types of segments. The first is traffic message channel (TMC) segments, which are typically divided at significant decision points for drivers along a road, such as interchanges and intersections. The second is XD segments, which provided greater granularity and more coverage onto lower functional classification roads than TMC segments (INRIX 2014). INRIX data is also the basis for the FHWA's National Performance Management Research Data Set (NPMRDS) data set. While INRIX data is available for purchase, state DOTs have access to the NPMRDS data set for free for a select number of TMCs, generally on higher functional classifications (INRIX 2017). ITD has purchased INRIX data at the XD scale for all available roads since 2017. However, traffic volume is not one of the purchased fields, limiting applications for this study (Coladner 2022).

AirSage

AirSage collects a sample of anonymized cell phone and smartphone application ("app") location records to understand travel behavior and to extrapolate travel volume. It inflates the sample that is collected to approximate the universe of trips on each road segment by aligning with traditionally collected traffic counts and/or automated counters within traffic signal systems. Traffic volume can be estimated for roads outside of cell phone coverage zones through devices' internal records of location while outside coverage areas. These records are then reported to mobile applications when network connection is re-established. AirSage processes its data by anchoring it to select ITD traffic counts to estimate AADT on the entire network. The accuracy of estimates depends on the number of records of people traveling along a given road in AirSage data (Silverberg 2022).

StreetLight Data

StreetLight Data comes from similar sources to NPMRDS and AirSage, meaning that it is largely derived from phone location data. StreetLight provides AADT estimates out of the box, meaning that it has already expanded its sample of locations to approximate a universe without anchoring to ITD's traffic counts in the way that AirSage does. This allows it to quickly provide AADT coverage for "nearly every road in the U.S. using the OpenStreetMap network" (StreetLight Data 2022). Instead of anchoring to ITD's traffic counts, it trains and validates its estimates using 3,000 permanent counter locations in 22 states. Although a cost quote has not been obtained, StreetLight claims its count data costs 56% less than tube counters (StreetLight Data 2022).

Replica

Replica data uses a combination of point of sales data, phone location data, and probe data to create a synthetic population for a region for use in an activity-based travel demand model. The model then simulates where people are traveling and how their travel would change under user-input conditions. This travel demand data can be used to estimate AADT (Replica 2022).

Data Inventory Conclusions

Comparison with Needs of Methods

The data inventory demonstrates that the essential data needs for estimating AADT on Idaho roads and most additional useful data needs can be met using data already possessed by ITD or readily available from other public sources. These minimum necessary data requirements are summarized in Table 3-8. Moreover, many of the data sets that can be helpful in improving the methods' accuracy are available for at least part of the road network, as summarized in Table 3-9. The only data needs that cannot be met are land use data (for which land coverage data may be substituted) and vehicle registrations (which are only available at the county level). Further exploration of these data sets will show the extent of the road network that they cover to determine not only if the segment-level data is available but also if the data is present on enough of the network to provide a meaningful contribution to the method.

Table 3-8. Data Needs and Sources by AADT Estimation Method

Data Need	Need Category	Data Source(s)	Sampling	Regression	Geospatial	Machine Learning	TDM	Network Analysis
Observed AADT		ITD, Third Party	Yes	Yes	Yes	Yes	Yes	Yes
Count Site Location		ITD	--	--	Yes	--	Yes	Yes
Road Network	Network	ITD	--	--	--	--	--	Yes
Functional Classes	Road	ITD	Yes	Yes	--	Yes	Yes	Yes
Population	Demographic	Public	--	--	--	--	Yes	--
Workers	Economic	Public	--	--	--	--	Yes	--

'Yes' signifies that the method requires that data type. Two dashes signify that the method does not require that data type. The "Source" column shows sources identified in this report for that data need.
TDM refers to 'travel demand model.'

Table 3-9. Additional Useful Data Needs and Sources by AADT Estimation Method

Data Need	Need Category	Data Source(s)	Sampling	Regression	Geospatial	Machine Learning	TDM	Network Analysis
Number of Lanes	Road	ITD	--	Yes	--	Yes	Yes	--
Speed Limits	Road	ITD	--	Yes	--	Yes	Yes	--
Surface Material	Road	ITD	--	Yes	--	Yes	--	--
Distance to Intersection	Network	ITD	--	--	Yes	Yes	--	Yes
Accessibility (to Primary or Secondary Roads)	Network	ITD	--	--	Yes	Yes	--	Yes
Total County Arterial Mileage	Network	ITD	--	--	Yes	Yes	--	Yes
Centrality Measures	Network	ITD	--	--	Yes	Yes	--	Yes
Road Mileage Density	Network	ITD	--	--	Yes	Yes	--	Yes
Surrounding Land Uses	Economic	Not available	--	Yes	--	Yes	Yes	--
Urban / Rural Designation	Economic	ITD	--	Yes	--	Yes	Yes	--
Workers	Economic	Public	--	Yes	--	Yes	Yes	--
Employment by Industry	Economic	Public	--	Yes	--	Yes	Yes	--
Median Income	Economic	Public	--	Yes	--	Yes	--	--
Poverty Rates	Economic	Public	--	Yes	--	Yes	--	--
Nearby Population	Demographic	Public	--	Yes	--	Yes	Yes	--
Population Density	Demographic	Public	--	Yes	--	Yes	--	--
Dwelling Units	Demographic	Public	--	Yes	--	Yes	--	--
School Enrollment	Demographic	Public	--	Yes	--	Yes	--	--
Vehicle Registration	Demographic	Public (county level)	--	Yes	--	Yes	--	--

Notes: 'Yes' signifies that the method may benefit from that data type. The "Source" column shows sources identified in this report for that data need. TDM refers to 'travel demand model.'

4. Methodology

This chapter presents the method for estimating off-system AADT that was selected for Idaho. Originally, three methods were derived from the assessment of the literature, practices in other states, and FHWA requirements documented in chapter 2 (“Literature Review, FHWA Requirements, and Documentation of Existing AADT Estimate Methodologies”), as well as the availability and coverage of data documented in chapter 3 (“Data Inventory”). These three methods were based on regression, Idaho’s statewide TDM, and geospatial interpolation. The methods were implementable using the tools and data sets that ITD already possesses or that are publicly available. Additionally, each method met or exceeded ITD’s minimum requirements for an AADT estimation model, namely that it:

- Incorporate Idaho’s urban vs rural designations.
- Incorporate Idaho’s paved vs unpaved roadway characteristics.
- Balance with the statewide VMT.
- Take into consideration any other inputs that the survey and the review of literature may suggest.
- Be feasible to implement using geospatial tools that are accessible to ITD.
- Span the entire state network.
- Produce estimates in conformance with the FHWA *Traffic Monitoring Guide*.
- Be possible to validate.

The geospatial interpolation method was selected based on the TAC’s feedback (detailed in Appendix D on page 104). Specifically, the geospatial interpolation method provides lower data collection requirements, a high degree of flexibility and adaptability, high ease of explanation about the process, and relatively high accuracy.

This chapter describes the geospatial interpolation method in five sections including this Introduction. The “Geospatial Interpolation” section provides step-by-step guidance for executing the methods. In many cases the guidance from the literature documented in chapter 2 is specific enough to make firm methodological decisions. In other cases, the literature provides guidance for making those decisions based on analysis of the data. The “Validation Methodologies” section describes how results can be validated using the same processes for all three methods. Additional details and implementation steps of this method are described in chapter 5 (“Implementation and Validation Plans”).

Geospatial techniques use proximity to count-derived AADTs to interpolate AADTs at locations for which AADT is otherwise unknown. They are appropriate for spatially correlated data when the value of one location can provide useful information about the likely value of nearby locations (Staats 2016). Geospatial techniques encompass several approaches, including Kriging interpolation, k-nearest neighbors, and inverse distance weighted interpolation. While each approach has been tested in the literature, this section focuses on two methods. The first is Kriging interpolation, which is one of the most widely used (Mathew 2020, DeVine 2020, Selby and Kockelman, Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression 2013) and has been found to be the most precise in at least some contexts in estimating AADT (Mathew 2020). The second is inverse distance weighting, which is a methodologically simpler deterministic method that has also achieved high accuracy in the literature (Ben-Gurion University of the Negev 2022).

Some studies show geospatial techniques to be more accurate than geographically weighted regression (GWR) (Selby and Kockelman 2013), while others show them to be less accurate (Pulugurtha and Mathew 2020). While this is a robust method that requires little additional data, there are several obstacles to implementing it for off-system public roads statewide in Idaho. Notably, some studies have found high error when distances from count sites exceed one or two miles (Gadda, Magoon and Kockelman 2007), which will be the case for many off-system public roads in Idaho. Additionally, Kriging interpolation often produces larger error at low-volume roads than high-volume roads (Staats 2016) or overestimates their AADT (Wang and Kockelman 2009), and close attention will need to be paid to ensure that its estimates for low-volume roads are reasonable. The method described in this chapter seeks to address these risks by incorporating variables such as population and employment density that correlate with AADT and can therefore provide useful data for the geospatial interpolation models to use in estimating AADT.

As shown in Table 4-1, geospatial interpolation can meet the need of an approach for estimating AADT, including differentiating among urban and rural roads, and among paved and unpaved roads provided that there are adequate count locations on each. The parsimonious nature of the method allows for a range of data inputs from only observed AADT values to a full gamut of demographics, economics, land use, roadway, and network characteristics.

Table 4-1. Summary of Kriging-Interpolation Characteristics

Characteristic	Details
Incorporate Idaho's urban vs rural designations	The interpolation should be run separately for urban and rural roads or include a variable distinguishing urban and rural roads.
Incorporate Idaho's paved vs unpaved roadway characteristics	The interpolation includes a variable distinguishing paved and unpaved roads.
Balance with the statewide VMT	The raw AADT estimates will be scaled to balance with statewide VMT.
Take into consideration any other inputs that the survey and the review of literature may suggest	An advantage of geospatial interpolation is its ability to estimate AADT with few data inputs. However, this limits the ability to account for additional factors.
Be feasible to implement using geospatial tools that are accessible to ITD	The method primarily requires GIS tools such as Esri's ArcGIS or Python.
Span the entire state network	The approach can be used on the entire network.
Produce estimates in conformance with the FHWA *Traffic Monitoring Guide*	No inconsistencies
Be possible to validate	The validation methods discussed in this report, notably related to comparison with select count locations, can work for this method.

The following subsections detail the steps for using geospatial interpolation to estimate off-system AADT.

Pre-Processing

Step 1: Import and Clean Data

Geospatial interpolation estimates AADT at locations where unknown based on AADT at nearby locations where it is derived from counts. Therefore, the first step is to collect count-derived AADT. Counts from prior years can be used if they are updated with a growth factor to estimate current year AADT. For instance, DeVine (2020) used ten years of AADT data. Data from prior years should be updated to the current year based on growth rates and seasonality factors if applicable. Only the most recent AADT should be used at a given location. If duplicate geometries exist within the observed data set, errors will likely arise from the underlying statistical functions. It may also be necessary to remove extreme AADT outliers, which can be done using the interquartile method described by DeVine (2020), who removed outliers from short-term counts that skewed toward high AADTs.

Off-system public roads without count-derived AADTs are imported into an estimation data set. Any additional data included in these files that will be utilized in the analysis process should be cleaned at this point, substituting "NA" or "NULL" values with appropriate assumptions (e.g., unpaved for the surface type of remote facilities), removing or fixing invalid geometry, and checking for data entry errors.

Additional data sets are also cleaned and prepared for conflation with the roadway data set at this stage. These additional data sets may include roadway characteristics (e.g., number of lanes, functional class, surface type), economic and demographic characteristics (e.g., employment density, population density), and network characteristics (e.g., network centrality).

Step 2: Prepare Data Set for Estimation

Off-system public roads without count-derived AADTs are imported into an estimation data set. Some studies have conducted cleaning of this data set, such as removing dead-end roads and driveways, and combining multiple sequential segments into a single segment (DeVine 2020). AADT will be estimated for the segment mid-point.

Additional characteristics are also associated with roadway segments in this step. Each data set that may be used as a variable in the regression should be associated with the entire network. The data sets should be converted to variables and only the variable values associated with the network rather than the raw values. For instance, raw roadway surface types are condensed to just two values: paved and unpaved. Similarly, raw population values within Census block groups may need to be converted to population density by dividing by land area. Appendix C on page 100 describes variables that may be calculated, identifying priority variables that are most likely to be useful to produce an accurate model. If these priority variables prove to be inadequate in producing an accurate model, then return to this step and calculate remaining variables.

The entire network includes both on-system and off-system public roads since there may not be enough off-system traffic counters to derive the relationship with AADT purely from off-system public roads. This means that segment-based data whose extent or segmentation differs from the final network should be assigned to the network for which AADT will be estimated. Location-based variables and variables derived from network analysis should also be calculated for segments in that network. Table 4-2 below summarizes how to compile data into a single data set aligned with roadway segments. All variables that will be considered for the analysis should be included in this final data set, including AADTs that are directly derived from traffic counters. AADTs produced through the ITD's existing traffic smoothing process should be omitted so that the relationship between AADT and other variables is derived from known AADTs only. The result of this step is a geospatial file (e.g., geodatabase, shapefile) with the dependent variable (i.e., AADT) where derived from count locations and with all independent variables that may be used in the model.

Table 4-2. Steps for Compiling Data by Data Type

Data Type	Applicable Data Types	Approach
Segment-based data with a common unique identifier field and the same segmentation as the final network	Roadway Characteristics	Tabular join using shared unique identifiers as a join field
Segment-based data without a common unique identifier field but with route IDs and measures	Roadway Characteristics	Use the Overlay Route Events tool in Esri ArcMap to join based on route IDs and measures.
Area-based data (such as data for Census tracts or block groups)	Demographic and economic characteristics	Geospatial assignment (e.g., intersection, buffers)
Variables derived from network analysis (e.g., betweenness, centrality)	Network characteristics	Calculate directly from the network

In addition to associating variables with segments, variables for economic and demographic characteristics should also be rasterized, since the raster data may be needed for the spatial predictions based on the model.

Part of calculating variables will depend on treating missing data correctly when data is missing for some road segments. Data can be filled in if it is possible to deduce what the data would likely be. For instance, missing numbers of through lanes may be assumed to be 2 (one in each direction), and roads without surface type data may be assumed to be unpaved if in a rural area and paved if in an urban area. Table 4-3 below provides guidance for filling in missing data for key variables. The entries should be left null for variables where there is no default value or where the value of missing values cannot be reasonably deduced by examining the pattern of missing data.

Table 4-3. Guidance for Filling in Missing Values

Variable	How to Fill in Missing Data
Number of through lanes	Assume two through lanes (one in each direction) if data is missing.
Paved vs. unpaved	If data on surface material is missing, assume unpaved if the segment is in a rural area and paved if it is in an urban area.
Functional class	Assume “local” designation for all segments for which functional class is unknown or that have not been functionally classified.
Urban vs. rural	If missing, use MPO boundaries, assuming urban inside of MPO boundaries and rural outside.

Although not strictly required, network characteristics have shown the capacity in the literature to greatly increase AADT estimation accuracy (Kehan 2017) . Simple measures such as “miles of functional class 6 within five miles” can be calculated directly from a network shapefile without further data cleaning. However, more complex network measures such as “betweenness centrality” or “travel distance to the nearest primary arterial facility” require more extensive data preparation. A network graph cannot be created by assessing facility overlap since some facilities that overlap are not connected (e.g., highway overpasses). Therefore, a careful network assessment process for the entire state may be needed to calculate these values.

Once network characteristics have been prepared, routes should be segmented such that each portion is associated with a unique value from each characterizing data set. For example, routes that have paved and unpaved portions should be split by surface type, and routes that span multiple counties should be split to have unique numbers of registered vehicles at the county level. Once all routes have been split according to facility characteristics and external geographies, all external data sets can be joined to the facilities of interest. If splitting routes is too onerous to attempt, then characteristics may be assigned based on the extent of overlap.

Finally, the midpoint for each segment can be calculated with the resulting midpoints retaining the information associated with their respective segments. This allows for much faster extraction of predicted AADT values later in the process.

Inverse Distance Weighting

Inverse distance weighting is based on the sum of a weighted average where weights are calculated based on the distance from the sampled point to the point of interest and a “power parameter,” according to Equation 5 and Equation 6 below (Esri 2021).

Equation 5. Inverse distance weighting equation for a value at an unknown point

$$u(x) = \begin{cases} \dfrac{\dfrac{\sum_{i=1}^{N} w_i(x) u_i}{\sum_{i=1}^{N} w_i(x)}}{u_i}, \dfrac{if\ d(x, x_i)\ \neq 0\ for\ all\ i}{if\ d(x, x_i) = 0\ for\ some\ i} \end{cases}$$

Where,

- $u(x)$ is the estimate value of the variable at an unknown point x.
- $w_i(x)$ is the weight assigned to the ith known point, based on the distance between the unknown point x and the ith known point.
- $d(x, x_i)$ is the distance function between the unknown point x and the ith known point.
- u_i is the value of the variable at the ith known point.
- N is the number of known points used in the interpolation.

Equation 6. Weighting equation for inverse distance weighting

$$w_i(x) = \frac{1}{d(x, x_i)^p}$$

Where,

- $w_i(x)$ is the weight assigned to observed value u_i.
- x denotes a point of interest.
- x_i is a known point.
- d is a given distance from the known point to the point of interest.
- N is the total number of known points.
- p is a positive real number referred to as the "power parameter." The larger the value of p, the more emphasis there is on observations closer to the known location.

The power parameter p changes the impact of distance to a known point on its contribution to the weighted value. Larger values of p prioritize known points further away from the point of interest. This relationship allows for optimization of a model to best fit a testing subset of the known points.

Although inverse distance weighting only describes the spatial relationship between points' AADT, stratification of the known values and facilities of interest can be used to better model these relationships. Characteristics of known locations can then be compared to the overall population to create associations between groups of count locations and off-system facilities.

Bin sizes for various factors could be determined based on a range of divisors such as quantiles, regulatory designations, analyst or decision maker interest, or maintenance requirements. An example using orders of magnitude for AADT and population density is shown in Table 4-4, and a similar estimate for AADT and roadway surface type is shown in Table 4-5. Ideally, these resulting divisions should contain enough samples for statistical analysis, at least 30 and preferably more. However, some locations may not be of interest as they are already directly analyzed or are already being directly observed, thus they can collect in small groups along the periphery of the sampling space.

Table 4-4. Count Station Distribution for AADT and Population Density Ranges

Residents per Square Mile	0-100	100-1,000	1,000-10,000	10,000+
0-10 AADT	9	-	-	-
10-100 AADT	285	24	5	-
100-1,000 AADT	1275	329	227	1
1,000-10,000 AADT	1213	897	1417	9
10,000+ AADT	153	238	587	-

Table 4-5. Count Station Distribution for AADT and Roadway Surface

Roadway Surface Type	Paved	Unpaved
0-100 AADT	117	206
100-1,000 AADT	985	847
1,000-5,000 AADT	993	1,572
5,000-10,000 AADT	276	695
10,000+ AADT	267	711

Step 3: Conduct Spatial Interpolation via Inverse Distance Weighting

Creating multiple models based on characteristics shared between count sites and facilities of interests can greatly enhance inverse distance weighting's ability to accurately estimate AADT. As discussed previously, these groupings can be based on any available factors, but it is recommended that the analyst begin with factors directly related to facilities (e.g., surface type, functional class, through lanes, network centrality) rather than location-based factors (e.g., population density, registered vehicles, median income). Once a set of factors is established, groups can be selected via random sampling or through more involved methods such as stratified sampling, Latin Hypercube sampling, or cluster sampling. There should be at least 40 observations for statistical significance after accounting for the fact that some observations will be used for validation.

Once groups have been established, they should be split into "training" and "testing" sets. The literature has often used approximately 80% of count locations for training the model and retained the remaining 20% used for testing or 'validating' estimates, although training sets have been observed to vary between approximately 75% to 90% of the data (Staats 2016, DeVine 2020, T. Wang, Improved Annual Average Daily Traffic (AADT) Estimation for Local Roads using Parcel-Level Travel Demand Modeling 2012). Subsequent analysis is performed on the training set, and the testing set is to be reserved for evaluation of accuracy in the Validation Plan described on page 65. Locations included in the training and testing data sets should be randomly selected to avoid the possibility of bias.

Once a training data set with 80% of count locations has been split from a testing data set with the remaining 20% of count locations, inverse distance weighting can now be performed on the training set of points. Although details vary among GIS software packages, inverse distance weighting tools typically require two inputs: a point layer with a field representing the variable of interest, and an extent (either a raster or polygon). GIS software packages that conform with ITD's needs are discussed on page 64. For this process, inverse distance weighting is performed on each group in the training set sequentially. Values are extracted for facilities of interest from the inverse distance weighting results for their respective groups. Any initial transformation of the actual AADT values is reversed at this point.

A transformation of AADT values (e.g., log, power) may improve accuracy, but these transformations should be approached on a group-by-group basis.

Inverse distance weighting is optimized by adjusting the power parameter. This requires an iterative process comparing the actual AADT values to the predicted values at those same locations. A measure of statistical accuracy, such as RMSE, is used to evaluate each input value for the parameter, and the value that minimizes error is selected. This is then repeated for each group's model. Optimization may need to be performed using scripts unless sufficient tools are already present in the analyst team's ArcGIS toolboxes.

Kriging Interpolation

Much like regression analysis, Kriging interpolation utilizes 'best linear unbiased predictions' to compute weighted averages in spatial contexts. Although the underlying mathematics are more complex than the inverse distance weighting technique, the practical implementation has been made available in a wide variety of accessible tools.

The analysis process also follows a similar format to the inverse distance weighting analysis. First, a formula of dependent and independent variables is used to calculate an empirical semivariogram. The resulting values are then associated spatial distance and semivariance (describing how random the relationship appears to be). A model can be fitted to this relationship and then used to make predictions given a broader set of points of interest.

The creation of such a model without independent variables is referred to as "Ordinary Kriging." This method only differs from inverse distance weighting during preparation in the development of the model. Equation 7 summarizes Ordinary Kriging. Explanatory variables such as population density, network connectivity, and roadway surface type can be used more directly in "Universal Kriging" also known as "regression Kriging." Here, the independent variables are incorporated into the model development, and then predictions are made based on a raster of values for those variables, rather than a raster which only represents a spatial extent as in Ordinary Kriging. Equation 8 summarizes Universal Kriging. The first summation of Equation 8 models the deterministic component and the second models the stochastic component.

Equation 7. Generalized method for Ordinary Kriging

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i \, Z(s_i)$$

Where,

- $Z(s_i)$ is the measured value at a known location.
- $\hat{Z}(s_0)$ is the estimated value at the prediction location s_0.
- λ_i is an unknown weight for the measured value at a known location.

- N is the number of measured values.

Equation 8. Generalized method for Universal Kriging

$$\hat{Z}(s_0) = \sum_{k=0}^{P} \widehat{\beta_k}\, q_k(s_0) + \sum_{i=1}^{N} \lambda_i\, e(s_i)$$

Where,

- $\widehat{\beta_k}$ are the estimated deterministic model coefficients.
- q_k is the matrix of predictors at the known location.
- $\lambda_{i'}$ are the Kriging weights determined by the spatial dependence structure of the residual.
- $e(s_i)$ is the residual at a known location.

Step 3a: Prepare Data for Spatial Interpolation via Kriging

The stratification or clustering of AADT count locations and facilities of interests as discussed in the inverse distance weighting section may be beneficial, especially if Ordinary Kriging interpolation is used. Stratification or clustering may also be used for Universal Kriging, but the use of independent variables may allow there to be fewer groups because the independent variables allow for accounting for these factors outside of stratification or clustering. Rather, grouping can be focused on data sets that are not well suited to rasterization. These can include facility specific information or linear systems where proximity or density have low spatial autocorrelation.

Once groups have been established, they should be split into “training” and “testing” sets. The literature has often used approximately 80% of count locations for training the model and retained the remaining 20% used for testing or ‘validating’ estimates, although training sets have been observed to vary between approximately 75% to 90% of the data (Staats 2016, DeVine 2020, T. Wang, Improved Annual Average Daily Traffic (AADT) Estimation for Local Roads using Parcel-Level Travel Demand Modeling 2012). Subsequent analysis is performed on the training set, and the testing set is to be reserved for evaluation of accuracy in the Validation Plan described on page 65. Locations included in the training and testing data sets should be randomly selected to avoid the possibility of bias.

A transformation of AADT values (e.g., log, power) may also improve accuracy, but these transformations should be approached on a group-by-group basis.

Step 3b Option 1: Conduct Spatial Interpolation via Ordinary Kriging

Although the Ordinary Kriging process is similar to the process for inverse distance weighting, the complexity of the implementation can vary widely depending on the software that is used. A software package with a graphical user interface may simply require a point layer of observations and a raster or polygon layer for the extent, as with the inverse distance weighting implementation. Some scripting language packages may require further inputs such as selecting a semivariogram model or the output resolution for the raster.

Once the data is prepared, grouped, and split, a semivariogram is calculated for the training set based solely on actual AADT values. A semivariogram is a graphical tool for examining spatial dependence between observations. It depicts variance as distance between pairs of observations changes, where distance between pairs of observations is shown on the x-axis and variance is shown on the y-axis. A semivariogram can be used to account for autocorrelation, which is the tendency for nearby observations to be more similar than more distant observations (Mathew 2020). A transformation may be applied based on an evaluation of each group. A model is then fit to the semivariogram. Fitting processes vary across packages, with some having more explicit functions (PyGIS) while others streamline it within the model selection process (PyKrige). Some other packages, such as an implementation of Gaussian processes through SciKit-Learn may require manual selection and fitting.

The final model and a raster or polygon extent layer are then fed into the predictive functionality of the software to produce a raster of AADT estimates. Any initial transformation of the actual AADT values is reversed at this point.

Step 3b Option 2: Conduct Spatial Interpolation via Universal Kriging

Universal Kriging interpolation has the capacity to integrate independent variables directly into the semivariogram and modeling processes. This means that ordinal factors that would be used for grouping with other techniques, such as number of through lanes, can be incorporated directly. Categorical variables, such as functional class, should be integrated with caution to ensure that the model is treating them as categorical variables rather than interval or continuous variables. Conversely, having a single group with functional class as an independent, it is also possible to split the data into multiple groups by functional classification and omit it as an independent variable.

The project team recommends beginning the analysis with the priority variables designed in Appendix C on page 100. As additional variables are added, it is necessary to conduct statistical analysis to ensure that regression assumptions are met. For regression analysis has three general assumptions: correlation between the variable and AADT, collinearity between independent variables, and heteroskedasticity (uneven or heterogeneous variance). Forward and backward model selection can also be used on a large set of factors to identify which factors typically affect the overall landscape most.

If there is no relationship between a variable and AADT counts, it should not be included unless the analysts or other team members identify a reasonable relationship between that factor and underlying trends that influence AADT. If multiple variables are found to be colinear, the variable with the greatest correlation to AADT can be retained and the others discarded. The presence of heteroskedasticity in any factor indicates that a transformation (e.g., log, power) may be necessary or that the variable cannot be used.

Spatial autocorrelation is a statistical measurement of how the physical distances between observations impact trends. Typically measured with Moran's I or Geary's C, high spatial autocorrelation does not necessarily mean that the variable will be predictive of AADT, but a low value does indicate that the variable's impact on the interpolation will be minimal (Fortin, Drapeau and Legendre 2012, Bivand 2009). The strength of the measure may also help prioritize the selection of a variable for use in the final model.

Once factors have been selected, a semivariogram is calculated for the training set based solely on actual AADT values. A transformation may be applied based on an evaluation of each group. A model is then fit to the semivariogram. Some packages have auto-fitting functionality that can select the optimal model and fit it accordingly. Some packages may require manual selection and fitting.

The final model and a raster or polygon extent layer are then fed into the predictive functionality of the software to produce a raster of AADT estimates. Any initial transformation of the actual AADT values is reversed at this point.

This process may be iterated several times to test different combinations of independent variables and their impact on AADT estimates. Other decision points, such as which semivariogram model to use, can also be tested at this point.

Post Processing

Step 4: Scale Resulting AADT Forecasts to Sum to Statewide VMT

The resulting AADTs should be adjusted so that they sum with on-system AADT estimates to equal the statewide VMT estimate. A multiplier is applied to all off-system AADT estimates to adjust them enough so that they sum with on-system AADT estimates to equal the statewide VMT estimate. Equation 9 and Equation 10 describe how to calculate that multiplier.

Equation 9. Off-system VMT

$$off\ system\ VMT = \sum_{i \in I} AADT_i \times length_i$$

Where,

- i is an off-system road segment.
- $AADT_i$ is the unadjusted AADT estimate for segment i.
- $length_i$ is the length in miles of segment i.

Equation 10. Scaling multiplier

$$multiplier = \frac{statewide\ VMT}{on\ system\ VMT\ + off\ system\ VMT}$$

Where,

- $on\ system\ VMT$ is the estimated VMT for road segments currently estimated through the count program and smoothing process.
- $statewide\ VMT$ is the estimate of statewide VMT through existing processes.

Step 5: Round Off-System AADT Estimates

Off-system AADT estimates should be rounded based on existing rounding rules described in Table 2-14 on page 33 with the exception that values below 10 are permitted. Off-system roads may in some cases have very low average traffic volumes, and rounding them up to ten may significantly overstate the true average traffic volume. Table 4-6 shows the rules for rounding off-system AADT estimates.

Table 4-6. Rounding Rules for AADT

AADT Range	Rounding Precision
Less than 10	Round to the nearest whole number
Less than 1,000	Round to nearest ten
Equal to or greater than 1,000 and less than or equal to 10,000	Round to nearest hundred
More than 10,000	Round to nearest five hundred

Source: Based on script in Smoothing Toolbox called AADTDataDeconstructor.py, with modifications for rounding of estimates below 10.

5. Implementation and Validation Plans

This chapter expands on the description of geospatial interpolation discussed in chapter 4 (“Methodology”) to develop an implementation and validation plan. The implementation and validation plan provides technical details for developing and selecting a geospatial interpolation model to estimate AADT on off-system public roads, assumptions underlying the analysis, tools that can be used to develop estimating scripts to be run in Esri ArcGIS Pro, guidance on validating the results, roles within ITD for estimating off-system AADT, and a draft schedule.

Assumptions

The analysis rests on several assumptions, including those related to missing data which were documented in Table 4-3 on page 55. At the beginning of the project implementation, these assumptions should be checked and refined to the extent possible. The assumptions underlying the analysis are listed below.

- ITD currently estimates AADT for all Federal-Aid roads through its traffic smoothing process.
- Functional classification is assumed to be class 7 (local) where it is unknown.
- The number of through lanes is assumed to be two (one in each direction) where it is unknown.
- Surface type is assumed to be unpaved in rural areas and paved in urban areas when it is unknown.
- If roads are not designated as either urban or rural, then they are assumed to be urban if inside of MPO boundaries and rural otherwise.

Implementation Tools

Geospatial interpolation is a common function of many GIS software packages. However, some more advanced methods may require analysis through external scripts in Python. As ITD transitions to ArcGIS Pro, new tools and functionality will need to be developed in that context (e.g., Python 3.7+). Additional library options that can be utilized in the ArcGIS Pro Python environment are presented here for both inverse distance weighting and Kriging.

Inverse Distance Weighting Tools

Built-in tools for inverse distance weighting are available in both Esri and QGIS products. If the stratification of sample sites is automated via a Python script or other automated tool, it may be desirable to also perform inverse distance weighting within the same tool or context.

Some examples of Python libraries for spatial analysis include ArcPy (directly compatible with Esri products), geopandas, GDAL/OGR (broadly used by Esri, QGIS, GRASS, and other GIS software), SciKit-Learn, Rasterio, PyProj, Shapely, Fiano, and RSGISLib. The implementation team may evaluate available packages at the time of analysis to determine which will provide the best fit for the available data and desired outputs.

Kriging Tools

Ordinary Kriging is typically available in popular GIS software suites, but their capacity for developing, tuning, and evaluating complex models for Universal Kriging may be limited. Due to these limitations and inconsistencies, the project team recommends developing a Python script to handle Kriging procedures as greater control can be exerted and intermediary results can be evaluated.

Three well-documented Python libraries with the capacity for geospatial interpolation through Kriging include: GSTools, PyKrige, and SciKit-Learn (using the available Gaussian estimation processes). The first two libraries both offer Ordinary and Universal Kriging. The implementation team may evaluate available packages at the time of analysis to determine which will provide the best fit for the available data and desired outputs.

Validation Plan

Validation is part of the data process for estimating AADT on off-system public roads. Validation occurs both to help select an optimal model for the geospatial interpolation approach to estimating AADT and then to confirm that the selected model produces AADT estimates that meet acceptable levels of accuracy. Since validation involves comparing estimated AADTs with AADTs calculated from observed counts, this plan also addresses the possibility of ITD eventually expanding its count program to collect counts in locations that are currently underrepresented among off-system counts. This expansion is unlikely to happen in the near future because there is a lack of additional resources with which to expand the count program.

Statistics

This subsection provides guidance for validating AADT estimates with measures assessing accuracy and bias. Fundamentally, to assess accuracy is to compare each AADT estimate that is part of the validation group with AADT calculated from observed counts. While this can be done individually, there are statistics that exist to summarize accuracy for a set of data points. While accuracy refers to how close the estimated AADT is to the AADT calculated from observed counts, bias refers to whether AADT is generally overestimated or underestimated. Bias exists when errors are not comprised of overestimation and underestimation fairly equally. Validation is expected to occur using only traffic counts that ITD already regularly collects. Table 5-1 below presents these statistics.

Table 5-1. Key Statistics for Validation

Name	Purpose	Interpretation	Sources
Mean error (ME)	To quantify bias in estimates	Zero signifies no bias in the estimates regardless of accuracy. A negative value indicates that observed AADT generally exceeds estimates, while a positive value indicates the reverse.	Morley (2016)
Mean absolute percentage error (MAPE)	To quantify accuracy of estimates	Zero signifies that estimates perfectly match observed AADT. A higher number means that there is more difference between estimated and observed AADT for an average location regardless of direction. Negative numbers are not possible.	Staats (2016), Pan (2008)
Root-mean-square error (RMSE)	To quantify accuracy of estimates	Values of zero or negative values are not possible. Values closer to zero signify more accurate estimates than higher values. Compared with MAPE, RMSE adds an extra penalty for estimates being farther away from observed AADT.	Moody (2019)

While both mean absolute percentage error (MAPE) and root-mean-square error (RMSE) can be used to examine error, RMSE is more sensitive to the size of the difference between estimates and observed AADT. Thus, if there are two models whose estimates receive the same MAPE value, the one that has more uniform deviations between estimates and observed AADT will likely have a lower RMSE value while the one with a mix of very low and very high deviations will receive a greater penalty for those very high values and have a higher RMSE value (Morley 2016).

As shown in Equation 11, ME can account for bias because it treats instances of overestimation differently from instances of underestimation. This contrasts with MAPE and RMSE, which use absolute values and square the difference between estimates and observed AADT respectively to ensure that instances of overestimation and underestimation are both positive. Equation 12 and Equation 13 describe MAPE (Pan 2008) and RMSE respectively (Moody 2019), and Python code facilitating the calculation of each statistic is included directly under each equation.

Mean Error (ME)

Equation 11. ME

$$ME = \frac{1}{n}\sum_{i=1}^{n}(AADT_{Mi} - AADT_{Fi})$$

Where,

- $AADT_{Fi}$ is the ith count-derived AADT value.
- $AADT_{Mi}$ is the ith AADT value estimated by the estimation method.
- n is the number of count locations retained for validation.

```
Code to calculate ME in Python
# Source: Statology (2020)
# Creating a Function for ME
import numpy as np

def mape(y_test, pred):
   y_test, pred = np.array(y_test), np.array(pred)
   me = np.mean(y_test - pred)
   return me
```

Mean Absolute Percentage Error (MAPE)

Equation 12. MAPE

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{AADT_{Mi} - AADT_{Fi}}{AADT_{Fi}}\right|$$

```
Code to calculate MAPE in Python
# Source: datagy (2022)
# There is no built-in Python package to calculate MAPE in Python, so this function
# can be defined in the numpy package to calculate MAPE, using the following code.
# Creating a Function for MAPE
import numpy as np


def mape(y_test, pred):
   y_test, pred = np.array(y_test), np.array(pred)
   mape = np.mean(np.abs((y_test - pred) / y_test))
   return mape
```

Root-Mean-Square Deviation (RMSE)

Equation 13. RMSE

$$RMSE = \sqrt{\sum_{i=1}^{n}\left(\frac{(AADT_{Mi} - AADT_{Fi})^2}{n}\right)}$$

Code to calculate RMSE in Python

```
# Source: Statology (2020)
#import necessary libraries
from sklearn.metrics import mean_squared_error
from math import sqrt

#calculate RMSE
sqrt(mean_squared_error(actual, pred))
```

Relative Accuracy

The purpose of assessments of relative accuracy is to determine which model or which version of a model produces the most accurate results and merits being retained for further development and ultimately for calculating the final estimates of off-system AADT. When, in the course of estimating off-system AADT, different versions of the geospatial interpolation models are tested containing different variables or with different model types (e.g., Universal Kriging vs. inverse distance weighting), the versions' accuracy can be assessed by calculating these statistics for all versions and selecting the version that is most accurate according to the statistics described in Table 5-1 for further development. This may be an iterative process that occurs in several rounds. If there are certain types of count locations that are particular concerns for accuracy because of relatively the small number of count locations on these roads (e.g., low-volume roads, unpaved roads), then they can be split off from the rest of the data set after AADT estimates are produced and accuracy statistics can be calculated for them separately from other roads.

Absolute Accuracy

Absolute accuracy demonstrates if there is bias in the final AADT estimates (i.e., the model is consistently underestimating or overestimating AADT) and how close estimates are to observed AADTs. The same statistics used for relative accuracy (shown in Table 5-1) are also used to assess absolute accuracy. There is no universal threshold for what constitutes a 'good' or acceptable score for each statistic for all use cases. What is acceptable depends on the use case. Table 5-2 below summarizes the values that other similar studies using geospatial or regression-based approaches have found as a basis for deciding what is acceptable. Regression-based models are included since they have performed with similar levels of accuracy to geospatial techniques in some research (Selby and Kockelman, Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression 2013, Pulugurtha and Mathew 2020). A larger list of errors associated with different studies by other methods is included in research conducted by DeVine (2020).

Table 5-2. Error Ranges from Prior Similar Analysis

Statistic	Range
ME	ME is not assessed in any of the reviewed studies. However, Wang and Kockelman (2009) obtained a median percentage error value of 33% using Kriging approaches, indicating slight overestimation of AADTs.
MAPE	Ranging from 32% to 160% depending on road type, with higher values for lower-volume roads (applied to all roads statewide from regression-based model) (Pan 2008).
	Ranging from 85% to 100% depending on the region (applied to local roads from regression-based model) (Staats 2016).
RMSE	73% for low-volume roads (ADT < 400) (based on regression) (Apronti, et al. 2016).
	56% to 95% depending on the year (based on Kriging techniques) (Shamo, Asa and Membah 2015).

In addition to reviewing accuracy for all roads, it may also be useful to assess accuracy for road types with relatively few count locations, such as low-volume roads and unpaved roads. To do this, count locations in the validation set can be observed separately for unpaved roads and for roads by AADT ranges, such as 0-99, 100-999, 1,000-9,999, and 10,000 or more. If bias or unacceptable inaccuracies are observed, then model adjustments may be required such as adding a variable to a Universal Kriging model that is closely associated with AADT, such as roadway intersection density, roadway mileage density, or population and employment density if not already included.

While the right model can produce reasonable and useful results with a high degree of accuracy, there will inevitably remain differences between estimates and observed AADT for reasons relating to statistical limitations, data availability, inaccuracies in input data, and the number of variables that it is possible to include in the model. These differences do not undermine the model results or their utility but rather point to the limits of estimates compared with counts. In percentage terms, these differences are likely to be greatest for low-volume roads where smaller deviations in absolute terms produce larger percentage variations. Additionally, the state of the art has not yet achieved perfect accuracy. Even the most recent research retains moderate deviations between predictions and observations, especially for lower-volume roads (Mathew 2020).

Even when results are highly accurate, it is important to properly explain them so that there are not apparent inaccuracies that simply reflect misunderstanding of what the results are intended to show. AADT represents average traffic over an entire year, while many counters and observers will be basing their observations on a snapshot in time (e.g., a given day, time of day, or week). Non-continuous counts and observations necessarily capture a snapshot of traffic that may not account for annual trends until seasonality, day-of-week, time-of-day, and other factors are accounted for. Therefore, if a person observes a higher or a lower volume of a roadway during the summer on a weekday, their observation will not reflect how traffic volumes might change on other days of the week or in other seasons. Single observations could also be distorted by special events or occurrences that change traffic volumes away from what they would typically be at that time, on that day, or in that season. It is important to explain the difference between AADT estimates' purpose and more casual observations that may not account for factors that could cause variation (e.g., season, day, time of day, etc.) to avoid the appearance in inaccuracies in estimates where they are not truly present.

Count Locations

This subsection identifies the types of locations that could be added to ITD's count program to improve validation. If in the future it is possible to count traffic volumes in new locations, these locations could be selected to improve the accuracy of AADT estimates for roads with the least accuracy. In general, it is expected that the roads with the least accurate estimates will be the ones that are least represented among off-system count locations.

The project team analyzed the characteristics of off-system count locations and found the least representation in low-density rural areas below 100 people per square mile and for unpaved roads. If count locations can be added on more roads with these characteristics in the future, it will increase the amount of data available for geospatial interpolation and for validation, improving estimates' accuracy where it is likely to be needed most.

Next Steps

This report has provided guidance for implementing a geospatial interpolation approach to estimate off-system AADT. This process and its next steps can be understood as belonging to several broad categories, which are summarized below, and for which roles for execution with ITD and a draft schedule are presented in the following subsections.

- Checking assumptions
  - Check the reasonableness of assumptions that underlie this analysis and refine them to the extent possible. Assumptions are listed on page 64.
- Data collection
  - Collect the off-system road network, historical AADTs, count locations, and data for calculation of other variables using process described in "Step 1: Import and Clean Data" on page 53. Data sets maintained by ITD can be collected from the teams within ITD that are listed in Appendix C on page 100. Links to collect economic and demographic data sets are provided in Appendix B on pages 98 and 99 respectively.
  - Convert raw data to the variables listed in Appendix C. Start with the variables marked as priority variables and then proceed to next step. Priority variables are the ones that are most likely to be used in the final model. If these priority variables are inadequate to estimate AADT with sufficient accuracy, then return to this step and calculate the remaining variables. Variables for roadway and network characteristic will be already assigned to roadway segments while economic and demographic variables will be assigned to geospatial units such as counties and Census tracts. Whether aligned with

segments or geospatial areas, all variables will be joined with the off-system road network in the “data conflation” step.

- Data conflation
    - Assign roadway data to the off-system roadway network using the process described in “Step 2: Prepare Data Set for Estimation” on page 54.
    - Rasterize economic and demographic variables as described in “Step 2: Prepare Data Set for Estimation” on page 54.
- Pre-model data processing
    - Establish training and testing groups as described in “Step 3: Conduct Spatial Interpolation via Inverse Distance Weighting” and “Step 3a: Prepare Data for Spatial Interpolation via Kriging” on pages 58 and 60 respectively.
- Modeling
    - Produce several models and compare the accuracy of their estimates using the validation process described on page 65 to identify the best variables. Vary these models based on the model aspects shown in Table 5-3.

Table 5-3. Model Aspects

Model Aspects	Description
Technique	Inverse distance weighting (page 58) vs. ordinary Kriging (page 60) vs. Universal Kriging (page 61)
Variables included in model	See Appendix C on page 100 for a list of priority and additional variables.
Transformations of variables	If one or more of the variables does not produce a good fit in the model, they can be transformed using a logarithmic or power transformation to attempt to obtain a better fit.

- Validation and iteration
    - For each set of model results, produce the statistics for the testing (validation) group described in the validation plan (page 65), namely ME, MAPE, and RMSE. Select models with the best performance on these statistics and continue refining until not further improvements are achieve.
    - When a final model is obtained, calculate the ME, MAPE, and RMSE for the testing (validation) group to ensure that they are within expected ranges (described in Table 5-2 on page 69)
- Scaling

    - Scale results for alignment with the statewide VMT estimate as described in “Step 4: Scale Resulting AADT Forecasts to Sum to Statewide VMT” on page 62.

- Rounding
    - Round AADT as described in “Step 5: Round Off-System AADT Estimates” on page 63.

- Data set finalization
    - Finalize data set by merging AADT estimates for off-system public roads with ITD’s data set and system for storing AADT estimates.

Roles and Responsibilities

The following roles and responsibilities are intended to mimic the roles and responsibilities used during ITD’s existing traffic smoothing process. The first time the model is run while scripts are being developed, the data analytics team have a larger role to develop the scripts for the process, as shown in Table 5-4. Once scripts are developed, then the roadway data team will have primary responsibility, as shown in Table 5-5.

Table 5-4. Roles and Responsibilities within ITD During Scripting Process (First Run)

Step	Lead Responsibility	Assistance
Checking assumptions	Roadway data	GIS
Data collection	Roadway data	GIS
Data conflation	GIS	Roadway data
Pre-model data processing	Data analytics	Roadway data
Modeling	Data analytics	Roadway data
Validation and iteration	Data analytics	Roadway data
Scaling	Data analytics	Roadway data
Rounding	Data analytics	Roadway data
Data set finalization	GIS	Roadway data

Table 5-5. Roles and Responsibilities within ITD After Scripting Complete (Subsequent Runs)

Step	Lead Responsibility	Assistance
Checking assumptions	Roadway data	GIS
Data collection	Roadway data	GIS
Data conflation	GIS	Roadway data
Pre-model data processing	Roadway data	Data analytics
Modeling	Roadway data	Data analytics
Validation and iteration	Roadway data	Data analytics
Scaling	Roadway data	Data analytics
Rounding	Roadway data	Data analytics
Data set finalization	GIS	Roadway data

Schedule

The schedule for developing processing the final data, developing scripts, and estimating AADT is shown in Table 5-6. The proposed schedule is intended to show a reasonable path based on expected staff availability within ITD, and implementation could move more quickly or more slowly than shown in the schedule depending on the ITD staff's final availability. The schedule does not specify a start time and can begin as staff availability allows. Specifically, the schedule makes the following assumptions.

- ITD staff will be available for each of the steps, including the Python scripting and creation of an ArcGIS Pro toolbox to automate many of the methodological steps.
- ITD intends to switch to ArcGIS Pro in the second half of 2023 or the first half of 2024. Therefore, Python scripts should be compatible with ArcGIS Pro. Because of this, all scripts should be developed in Python version 3. ArcGIS Pro is not compatible with Python version 2, which contrasts with ArcGIS Desktop, which is compatible with version 2 (Esri n.d.).

- Python scripts for the traffic smoothing process took several years to develop. Part of this development time is associated with staff availability and the fact that processes were being refined while scripting was occurring. It is assumed that scripting will occur concurrently with each step such that by the end of the process both a complete set of scripts and a complete data set will be present.

Table 5-6. Proposed Schedule by Month Since Start of Implementation

Step	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7
Checking assumptions	X	-	-	-	-	-	-
Data collection	X	-	-	-	-	-	-
Data conflation	-	X	-	-	-	-	-
Pre-model data processing	-	-	X	X	-	-	-
Modeling	-	-	-	X	X	-	-
Validation and iteration	-	-	-	-	X	-	-
Scaling	-	-	-	-	-	X	-
Rounding	-	-	-	-	-	X	-
Data set finalization	-	-	-	-	-	-	X

Note: 'X' indicates that the given step is expected to occur during a given month. '-' indicates that the given step is not expected to occur during a given month.

6. Works Cited

AASHTO. 2009. "AASHTO Guidelines for Traffic Data Programs."

Ada County Highway District. 2022. *Traffic Counts.* Accessed September 28, 2022. http://www.achdidaho.org/Departments/Engineering/Traffic/trafficCounts.aspx.

Anderson, Michael D, Khalid Sharfi, and Sampson E Gholston. 2006. "Direct demand forecasting model for small urban communities using multiple linear regression." *Transportation Research Record* 1981 (1): 114-117. Accessed June 8, 2022. http://dx.doi.org/10.3141/1981-18.

Apronti, Dick, Khaled Ksaibati, Kenneth Gerow, and Jamie Jo Hepner. 2016. "Estimating traffic volume on Wyoming low volume roads using linear and logistic regression methods." *Journal of Traffic and Transportation Engineering* 493-506. Accessed June 2, 2022. doi:http://dx.doi.org/10.1016/j.jtte.2016.02.004.

Ben-Gurion University of the Negev. 2022. *Chapter 12: Spatial Interpolation of Point Data.* February 21. Accessed November 4, 2022. http://132.72.155.230:3838/r/spatial-interpolation-of-point-data.html.

Bivand, Roger. 2009. "Applying Measures of Spatial Autocorrelation: Computation and Simulation." *Geographical Analysis* 41 (4): 375-384. doi:10.1111/j.1538-4632.2009.00764.x.

Calderon, Vicky, interview by Peter Hylton and Bryce Miller. 2022. *Email* (Deember 20).

Calderon, Vicky, and Amanda Laib, interview by Peter Hylton and Bryce Miller. 2022. *ITD Off-System Roads AADT Estimation Study - Progress Meeting #9* (September 27).

Castro-Neto, Manoel, Youngseon Jeong, Myong K. Jeon, and Lee D. Han. 2009. "AADT prediction using support vector regression with data-dependent aprameters." *Expert Systems with Applications* 2979-2986. Accessed May 27, 2022. doi:10.1016/j.eswa.2008.01.073.

Coladner, David, interview by Peter Hylton. 2022. (October 20).

Coladner, David, Amanda Laib, and Vicky Calderon, interview by Peter Hylton and Bryce Miller. 2022. *Modernization Processing Tools* (June 23).

Das, Subasish, and Ioannis Tsapakis. 2020. "Interpretable machine learning approach in estimating traffic volume on low-volume roadways." *International Journal of Transportation Science and Technology* 76-88. Accessed June 2, 2022. doi:https://rosap.ntl.bts.gov/view/dot/29681.

datagy. 2022. *How to Calculate MAPE in Python.* Feburary 11. Accessed January 13, 2023. https://datagy.io/mape-python/.

Datarade. 2022. *INRIX Real-time Traffic Flow from GPS data (for Europe and North America).* Accessed October 7, 2022. https://datarade.ai/data-products/inrix-real-time-traffic-flow.

DeVine, Ryan Andrew. 2020. "Use of Geospatial Technique to Estimate Traffic Volume on South Carolina Roads (Doctoral Dissertation)." Masters Thesis. Accessed June 2, 2022. https://scholarcommons.sc.edu/etd/6125 .

Dixon, Michael. 2004. "The effects of errors in annual average daily traffic forecasting: study of highways in Rural Idaho." Accessed June 8, 2022. https://apps.itd.idaho.gov/apps/research/Completed/RP167.pdf.

Eom, Jin Ki, Man Sik Park, Tae-Young Heo, and Leta F Huntsinger. 2006. "Improving the prediction of annual average daily traffic for nonfreeway facilities by applying a spatial statistical method." *Transportation Research Record* 1968 (1): 20-29. Accessed June 8, 2022. https://doi.org/10.1177%2F0361198106196800103.

Esri. n.d. *FAQ: What version of Python is used in ArcGIS?* Accessed January 13, 2023. https://support.esri.com/en/technical-article/000013224.

—. 2021. *How IDW works.* Accessed January 13, 2023. https://desktop.arcgis.com/en/arcmap/latest/tools/3d-analyst-toolbox/how-idw-works.htm#:~:text=IDW%20relies%20mainly%20on%20the%20inverse%20of%20the,real%20number%2C%20and%20its%20default%20value%20is%202.

Federal Highway Administration. 2016. "Appendix C. Vehicle Types." In *Traffic Monitoring Guide*. Accessed January 17, 2023. https://www.fhwa.dot.gov/policyinformation/tmguide/tmg_2013/vehicle-types.cfm.

Federal Highway Administration. 2018. "Highway Performance Monitoring System Field Manual." Accessed September 23, 2022. https://www.fhwa.dot.gov/policyinformation/hpms/fieldmanual/page00.cfm.

Federal Highway Administration. 2016. "Traffic Monitoring Guide." https://www.fhwa.dot.gov/policyinformation/tmguide/.

Federal Highway Administration. 2022. "Traffic Monitoring Guide." Accessed January 23, 2023. https://www.fhwa.dot.gov/policyinformation/tmguide/.

—. 2022. *Traffic Volume Trends.* Accessed July 18, 2022. https://www.fhwa.dot.gov/policyinformation/travel_monitoring/tvt.cfm.

Fortin, Marie-Josée, Pierre Drapeau, and Pierre Legendre. 2012. "Spatial autocorrelation and sampling design in plant ecology." In *Progress in theoretical vegetation science*, by G. Grabherr, L. Mucina, M. B. Dale and C. J. F. Braak, 209–222. doi:10.1007/978-94-009-1934-1_18.

Fu, Miao, J. Andrew Kelly, and J. Peter Clinch. 2017. "Estimating annual average daily traffic and transport emissions for a national road network: A bottom-up methodology for both nationally-aggregated and spatially-disaggregated results." *Journal of Transport Geography* 186-195. Accessed June 2, 2022. http://dx.doi.org/10.1016/j.jtrangeo.2016.12.002.

Gadda, Shashank, Atul Magoon, and Kara M. Kockelman. 2007. "Estimates of AADT: Quantifying the Uncertainty." *86th Annual Meeting of the Transportatin Research Board.* Washington, DC.

Golbeck, Jennifer. 2015. "Analyzing Networks." In *Introduction to Social Media Investigation*, by Jennifer Golbeck. Accessed September 23, 2022. https://www.sciencedirect.com/topics/computer-science/degree-centrality.

Golbeck, Jennifer. 2013. "Network Structure and Measures." In *Analyzing the Social Web*, by Jennifer Golbeck. Accessed September 23, 2022. doi:https://doi.org/10.1016/C2012-0-00171-8.

Hansen, Derek L., Ben Shneiderman, Smith Marc A., and Itai Himelboim. 2020. "Social network analysis: Measuring, mapping, and modeling collections of connections." In *Analyzing Social Media Networks with NodeXL*, by Derek L. Hansen, Ben Shneiderman, Smith Marc A. and Itai Himelboim, edited by Second. Accessed September 23, 2022. doi:https://doi.org/10.1016/C2018-0-01348-1.

Hanson, Nicole, and Vicky Calderon, interview by Peter Hylton and Bryce Miller. 2022. *GIS and AADT Temporality* (June 23).

Holik, William, Ioannis Tsapakis, Anita Vandervalk, Shawn Turner, and John Habermann. 2017. *Innovative Traffic Data QAQC Procedures and Automating AADT Estimation.* Federal Highway Adminstration Office of Safety. Accessed June 16, 2022. https://safety.fhwa.dot.gov/rsdp/downloads/fhwasa17035.pdf.

Huynh, Nathan, Jing Wang, Ryan DeVine, Mashurur "Ronnie" Chowdhury, Weimin Jin, Pronab Biswas, and Gurcan Comert. 2021. *Estimating AADT on Non-coverage Roads.* South Carolina Department of Transportation. Accessed June 16, 2022. https://scdot.scltap.org/wp-content/uploads/2022/01/SPR-749-Final-Report-2.pdf.

Idaho Association of Highway Districts. 2022. *Highway District Members.* Accessed September 28, 2022. https://iahd.com/highway-district-members/ .

Idaho Geospatial Office. 2022. *Land Use & Land Cover TWG.* Accessed September 23, 2022. https://gis.idaho.gov/land-use-land-cover-twg.

Idaho Transportation Department . 2022. *Road Data.* Accessed September 28, 2022. https://itd.idaho.gov/road-data/.

Idaho Transportation Department. 2022. "Historical Count History: 1960-2021." Accessed October 25, 2022.

Idaho Transportation Department. 2016. "Idaho Transportation Department Systems Procedures." 22. https://apps.itd.idaho.gov/apps/plan/ITDSystemsProcedures.pdf.

—. 2022. *ITD Open Data.* Accessed August 25, 2022. https://data-iplan.opendata.arcgis.com/.

Idaho Transportation Department. 2020. "Traffic Smoothing Documentation Guide."

INRIX. 2014. "I-95 Vehicle Probe Project II." https://tetcoalition.org/wp-content/uploads/2015/02/I-95-VPP-II-INRIX-Inteface-Guide-v1-Dec-2014.pdf.

—. 2017. *INRIX Selected by The U.S. Federal Highway Administration for National Traffic Data Set.* June 5. Accessed October 7, 2022. https://inrix.com/press-releases/npmrds/.

—. 2022. *INRIX Volume Profile.* Accessed October 7, 2022. https://inrix.com/products/volume/.

INSIDE Idaho. 2020. "Attributes for Idaho Transportation." https://insideidaho.org/data/ago/uofi/library/roads/roadTransportationAttributesDescription.pdf.

Islam, Sababa. 2016. *Estimation of Annual Average Daily Traffic (AADT) and Missing Hourly Volume Using Artificial Intelligence.* Masters Thesis, Clemson University. Accessed June 2, 2022. https://tigerprints.clemson.edu/all_theses/2562.

Jayasinghe, Amila, Kazushi Sano, C. Chethika Abenayake, and P.K.S. Mahanama. 2019. "A novel approach to model traffic on road segments of large-scale urban road networks." *MethodsX* 1147-1163. Accessed June 2, 2022. https://doi.org/10.1016/j.mex.2019.04.024.

Jiang, Zhuojun, Mark R McCord, and Prem K Goel. 2006. "Improved AADT estimation by combining information in image-and ground-based traffic data." *Journal of Transportation Engineering* (American Society of Civil Engineers) 132 (7): 523-530. Accessed June 8, 2022. https://doi.org/10.1061/(ASCE)0733-947X(2006)132:7(523).

Keehan, McKenzie. 2017. "Annual Average Daily Traffic (AADT) Estimation with Regression Using Centrality and Roadway Characteristic Variables." Thesis, Clemson University. Accessed September 26, 2022. https://tigerprints.clemson.edu/all_theses/2644.

Kehan, McKenzie. 2017. "Annual Average Daily Traffic (AADT) Estimation with Regression Using Centrality and Roadway Characteristic Variables." *All Theses.* Clemson University. Accessed June 2, 2022. https://tigerprints.clemson.edu/all_theses/2644.

Lowry, Michael. 2014. "Spatial interpolation of traffic counts based on origin--destination centrality." *Journal of Transport Geography* 36: 98-105. Accessed June 8, 2022. https://doi.org/10.1016/j.jtrangeo.2014.03.007.

Lowry, Michael, and Michael Dixon. 2012. *GIS Tools to Estimate Average Annual Daily Traffic.* Final Report, US Department of Transportation Research and Special Programs Administration. Accessed May 27, 2022. https://rosap.ntl.bts.gov/view/dot/24942.

Marsden, Peter V. 2005. "Network Analysis." In *Encyclopedia of Social Measurement*, by Kimberly Kempf-Leonard. Elsevier. Accessed September 23, 2022. https://www.sciencedirect.com/topics/computer-science/centrality-measure.

Mathew, Sonu. 2020. *Modeling Annual Average Daily Traffic for Local Functionally Classified Roads.* PhD Dissertation, University of North Carolina at Charlotte. Accessed June 2, 2022. https://repository.charlotte.edu/islandora/object/etd%3A2066.

McCord, Mark R., and Prem Goel. 2009. *Estimating AADT from combined air photos and ground based data: System design, prototyping, and testing.* NEXTRANS Project, USDOT Region V Regional Univeristy Transportation Center. Accessed June 2, 2022. https://rosap.ntl.bts.gov/view/dot/5707.

Mohamad, Dadang, Kumares C Sinha, Thomas Kuczek, and Charles F Scholer. 1998. "Annual average daily traffic prediction model for county roads." *Transportation Research Record* 1617 (1): 69-77. Accessed June 8, 2022. https://doi.org/10.3141%2F1617-10.

Moody, James. 2019. *What does RMSE really mean?* September 5. Accessed November 1, 2022. https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e.

Morley, Steve Karl. 2016. *Alternatives to accuracy and bias metrics based on percentage errors for radiation belt modeling applications.* Los Alamos National Laboratory. https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-16-24592.

Pan, Tao. 2008. "Assignment of Estimated Average Annual Daily." Thesis, University of South Florida. https://digitalcommons.usf.edu/cgi/viewcontent.cgi?httpsredir=1&article=1441&context=etd#:~:text=Assignment%20of%20Estimated%20Average%20Annual%20Daily%20Traffic%20Volumes,fulfillment%20of%20the%20requirements%20for%20the%20degree%20of.

Pan, Tao. 2008. *Assignment of estimated average annual daily traffic volumes on all roads in Florida.* Master's Thesis, University of South Florida. Accessed June 8, 2022. https://digitalcommons.usf.edu/etd/442/.

Park, Nokil. 2004. *Estimation of Average Annual Daily Traffic (AADT) Using Geographically Weighted Regression (GWR) Method and Geographic Information System (GIS).* PhD Dissertation, Florida Internaional University. Accessed June 2, 2022. https://digitalcommons.fiu.edu/dissertations/AAI3128612/.

Pridmore, Margaret. 2022. *Email.* August 30.

Pridmore, Margaret, and Vicky Calderon, interview by Peter Hylton and Bryce Miller. 2022. *Historical Method/Modernization* (June 21).

Pulugurtha, Srinivas S, and Prasanna R Kusam. 2012. "Modeling annual average daily traffic with integrated spatial data from multiple network buffer bandwidths." *Transportation Research Record* 2291 (1): 53-60. Accessed June 8, 2022. https://doi.org/10.3141%2F2291-07.

Pulugurtha, Srinivas S., and Sonu Mathew. 2020. *Develop Local Functional Classification VMT and AADT Estimation Method.* Final Report, North Carolina Department of Transportation Research and Analysis Group. Accessed June 2, 2022. https://connect.ncdot.gov/projects/research/RNAProjDocs/Pulugurtha_NCDOT-RP2019-12_FinalReport.pdf.

Raja, Prithiviraj, Mehrnaz Doustmohammadi, and Michael D. Anderson. 2018. "Estimation of Average Daily Traffic on Low Volume Roads in Alabama." *International Journal of Traffic and Transportation Engineering* 1-6. Accessed June 2, 2022. doi:10.5923/j.ijtte.20180701.01.

Replica. 2022. *Solutions.* Accessed October 7, 2022. https://replicahq.com/solutions/.

Robinson, T. P., and Graciela Metternicht. 2006. "Testing the performance of spatial interpolation techniques for mapping soil properties." *Computers and electronics in agriculture* 50 (2): 97-108. 10.1016/j.compag.2005.07.003.

Seaver, William L, Arun Chatterjee, and Mark L Seaver. 2000. "Estimation of traffic volume on rural local roads." *Transportation Research Record* 1719 (1): 121-128. Accessed June 8, 2022. https://doi.org/10.3141%2F1719-15.

Selby, Brent, and Kara M Kockelman. 2013. "Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression." *Journal of Transport Geography* 29: 24-32. Accessed June 8, 2022. https://doi.org/10.1016/j.jtrangeo.2012.12.009.

Selby, Brent, and Kara M. Kockelman. 2013. "Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression." *Journal of Transport Geography* 29: 24-32.

Sfyridis, Alexandros, and Paolo Agnolucci. 2020. "Annual average daily traffic estimation in England and Wales: An application of clustering and regression modelling." *Journal of Transport Geography* 83: 102658. Accessed June 8, 2022. https://doi.org/10.1016/j.jtrangeo.2020.102658.

Shamo, Benedict, Eric Asa, and Joseph Membah. 2015. "Linear spatial interpolation and analysis of annual average daily traffic data." *Journal of Computing in Civil Engineering* 29 (1). Accessed June 8, 2022. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000281.

Sharma, Satish C, Pawan Lingras, Fei Xu, and Guo X Liu. 1999. "Neural networks as alternative to traditional factor approach of annual average daily traffic estimation from traffic counts."

Transportation Research Record 1660 (1): 24-31. Accessed June 8, 2022. https://doi.org/10.3141%2F1660-04.

Sharma, Satish, Pawan Lingras, Fei Xu, and Peter Kilburn. 2001. "Application of neural networks to estimate AADT on low-volume roads." *Journal of Transportation Engineering* 127 (5): 426-432. Accessed June 8, 2022. https://doi.org/10.1061/(ASCE)0733-947X(2001)127:5(426).

Silverberg, Jonathan and Wichman, Chris, interview by Peter and Miller, Bryce Hylton. 2022. (September 15).

Staats, William Nicholas. 2016. *Estimation of Annual Average Daily Traffic on Local Roads in Kentucky.* Master's Thesis, University of Kentucky. Accessed May 27, 2022. https://uknowledge.uky.edu/ce_etds/36.

Statology. 2020. *How to Calculate MAPE in Python.* July 7. https://www.statology.org/mape-python/.

—. 2020. *How to Calculate RMSE in Python.* September 3. Accessed January 13, 2023. https://www.statology.org/rmse-python/.

StreetLight Data. 2022. *Get more data, at finer resolution, for 1/3 the cost.* Accessed September 25, 2022. https://www.streetlightdata.com/streetlight-cost-analysis/.

—. 2022. *StreetLight AADT.* Accessed September 26, 2022. https://www.streetlightdata.com/aadt-average-annual-daily-traffic-count/.

Sun, Xiaoduan, and Subasish Das. 2015. "Developing a Method for Estimating AADT on All Louisiana Roads." Final Report, Louisiana Department of Transportation and Devleopment. Accessed May 27, 2022. https://rosap.ntl.bts.gov/view/dot/29681.

Tobler, Waldo R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 234-240.

Tobler, Waldo R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* (Taylor & Francis, Ltd.) 46: 234-240. doi:https://doi.org/10.2307/143141.

Tsapakis, Ioannis, William Holik, Subasish Das, Edgar Kraus, and Paul Anderson. n.d. *Data Collection and Annual Average Daily Traffic (AADT) Estimation for Non-Federal Aid System (NFAS) Roads.* FHWA-SA-20-064, Washington, DC: Federal Highway Administration. https://safety.fhwa.dot.gov/rsdp/downloads/fhwasa20064.pdf.

Tsapakis, Ioannis, William Holik, Subasish Das, Edgar Kraus, and Paul Anderson. 2020. "Informational Guide on Data Collection and Annual Average Daily Traffic (AADT) Estimation for Non-Federal Aid-System (NFAS) Roads." Accessed June 2, 2022. https://safety.fhwa.dot.gov/rsdp/downloads/fhwasa20064.pdf.

U.S. Census Bureau. 2022. *Cartographic Boundary Files.* Accessed September 23, 2022. https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.html#:~:text=Cartographic%20Boundary%20Files%20The%20cartographic%20boundary%20files%20are,are%20specifically%20designed%20for%20small%20scale%20thematic%20mapping.

—. 2022. *Land Area and Persons Per Square Mile.* Accessed September 23, 2022. https://www.census.gov/quickfacts/fact/note/US/LND110210.

U.S. Census Bureau. n.d. "LEHD Origin-Destination Employment Statistics (LODES) Dataset Structure, Fromat Version 7.5." Accessed September 23, 2022. https://lehd.ces.census.gov/data/lodes/LODES7/LODESTechDoc7.5.pdf.

U.S. Geological Survey. 2018. *National Land Cover Data Base.* September 11. Accessed September 23, 2022. https://www.usgs.gov/centers/eros/science/national-land-cover-database#overview.

Unnikrishnan, A., M. Figliozzi, M.K. Moughari, and S. Urbina. 2018. *A Method to Estimate Annual Average Daily Traffic for Minor Facilities for MAP-21 Reporting and Statewide Safety Analysis.* Final Report, Oregon Department of Transportation. Accessed June 16, 2022. https://www.oregon.gov/ODOT/Programs/ResearchDocuments/SPR_804_Final_Report.pdf.

Wang, Tao. 2012. *Improve Annual Average Daily Traffic (AADT) Estimation for Local Roads Using Parcel-Level Travel Demand Modeling.* PhD Dissertation, Florida International University. Accessed May 27, 2022. https://digitalcommons.fiu.edu/dissertations/AAI3517050/.

Wang, Tao. 2012. *Improved Annual Average Daily Traffic (AADT) Estimation for Local Roads using Parcel-Level Travel Demand Modeling.* Dissertation, Florida International University. Accessed 25 October, 2022. https://digitalcommons.fiu.edu/dissertations/AAI3517050/.

Wang, Tao, Albert Gan, and Priyanka Alluri. 2013. "Estimating annual average daily traffic for local roads for highway safety analysis." *Transportation Research Record* 2398 (1): 60-66. Accessed June 8, 2022. https://doi.org/10.3141%2F2398-07.

Wang, Xiaokun, and Kara M. Kockelman. 2009. "Forecasting network data: Spatial interpolation of traffic counts from texas data." *Transportation Research Record* 2105 (1): 100-108.

Wang, Xiaokun, and Kara M. Kockelman. 2009. "Forecasting Network Data: Spatial Interpolation of Traffic Counts Using Texas Data." *Transportation Research Rcord* 100-108. Accessed May 27, 2022. https://www.caee.utexas.edu/prof/kockelman/public_html/trb09aadtkriging.pdf.

Xia, Qing, Fang Zhao, Zhenmin Chen, L David Shen, and Diana Ospina. 1999. "Estimation of annual average daily traffic for nonstate roads in a Florida county." *Transportation Research Record* 1660 (1): 32-40. Accessed June 8, 2022. https://doi.org/10.3141%2F1660-05.

Yang, Bingduo, Sheng-Guo Wang, and Yuanlu Bao. 2014. "New Efficient Regression Method for Local AADT Estimation via SCAD Variable Selection." *IEEE Transactions on Intelligent Transportation Systems* (IEEE) 15 (6): 2726-2731. Accessed June 8, 2022. http://dx.doi.org/10.1109/TITS.2014.2318039.

Zhao, Fang, and Nokil Park. 2004. "Using geographically weighted regression models to estimate annual average daily traffic." *Transportation Research Record* 1879 (1): 99-107. Accessed June 8, 2022. https://doi.org/10.3141%2F1879-12.

Zhao, Fang, and Soon Chung. 2001. "Contributing factors of annual average daily traffic in a Florida county: exploration with geographic information system and regression models." *Transportation Research Record* 1769 (1): 113-122. Accessed June 8, 2022. https://doi.org/10.3141%2F1769-14.

Zhong, Ming, and Brody L Hanson. 2009. "GIS-based travel demand modeling for estimating traffic on low-class roads." *Transportation Planning and Technology* 32 (5): 423-439. Accessed June 8, 2022. https://doi.org/10.1080/03081060903257053.

7. Appendix A. Survey Responses

This appendix provides the raw responses to the survey of state DOTs about traffic count and off-system AADT estimation practices.

Question 1: Does your organization estimate AADT for facilities without traffic volume counts? If so, what methodology is used?

- **Colorado:** For Statewide-HPMS purposes. Functional classification 7's and rural 6's; we estimate that 10% of the VMT is traveling on functional class (FC) 7 and 5% on rural 6's. For specific traffic stations that need an AADT; we estimate the AADT based off of the land use, adjacent counts and the functional classification of the roadway.
- **Iowa:** Rarely. Roads and facilities over or adjacent to DOT jurisdictional roads sometimes require estimating of traffic. And in those cases, there might not be a count available. One method, would be to use Origin-Destination data and available volume estimation data. Another would be to look at similar roads in the area and estimate upon the similar conditions/circumstances. These occurrences are rare, with one to two estimations of this nature a year.
- **Kansas:** We count a sample of off-system roads, and apply averages to the un-counted aggregate road network. Rural Major Collectors are sampled, at least 1 site on each route, every 3 years. Rural Minor Collectors are sampled every 6 years. Rural Locals are sampled every 9 years. The local sample includes a non-corporate branch which is used to develop county-wide traffic estimates, and a corporate branch which is used to develop regional traffic estimates stratified by city population. Urban classified streets are sampled every 3 years, Urban Locals every 9 years, to develop traffic estimates for each urban area.
- **Minnesota:** No, for system level VMT defaults from the 1990s are used, but point/project specific values are not available.
- **Missouri:** Segments without sites are given an estimated AADT based on surrounding segments and aerial imagery. These segments are then grouped with other segments that do have sites so that when those sites are collected, the grouped sites grow/shrink by an equivalent percentage.
- **Montana:** We have defaults that vary by our factor groups. These defaults are modified with our growth (or lack of) rates each year. We use fuel receipts and the FHWA VM1 (I think, it is the report that estimates mileage by vehicle type) to make sure our total VMT (including defaults) is within reason.
- **Oregon:** Rarely. Most roads that we need, we either collect ourselves or get from the local jurisdictions. We know that some of their estimations as part of the functional class process are a bit sketchy. For the small number that we estimate, we look at counts farther along that road; turning movement counts on the road; or similar roads.

- **Utah:** The Utah Department of Transportation collects traffic counts for all functionally classified roadways except FC 7 and Rural FC 6. We have a system in place for updating the estimated AADT values for the lower functional classification roads that we do not collect traffic volume counts for, as well for federally owned roads such as Forest Service, Fish & Wildlife, etc. We use two growth factors from our CCS's lowest functional class groups for Urban and Rural to update these estimated AADT's annually. The local roadways are lumped summed into categories by County, City, and Rural/Urban. These AADT values are only used to help calculate the VMT summary tables for HPMS annually, and not reported as individual AADT's with our annual submission.

  The issue In this data is, we have no record of where the original AADT originated, nor a process in place to update the original AADT with some sort of validation that can statistically be applied to all roadways in each summed group.

- **Wisconsin:** When a traffic count is not collected. We will use the next closest count taking and use the AADT calculated for a non-collected site. We will take the hourly data from the surrounding collected site and apply the monthly generated factors and growth.

- **Wyoming:** Yes we do. We growth factor most of it the roadways. All of our roadways, State-owned and non-state owned are counted about every 10 years minimum (higher classified roads are counted within *Traffic Monitoring Guide* /HPMS guidelines) we then use growth factors based on functional classification to estimate those roadways that aren't counted in a given year.

Question 2: Does your organization estimate AADT for facilities without traffic volume counts? If not, what are the reasons for this?

- **Minnesota:** Funding to collect the sample data which would be needed each year. A pilot project done with TTI (Texas A&M Transportation Institute) indicated around 1,200 counts would be needed each year. Also the complexity of applying samples to the non-Federal-Aid System. The pilot project with TTI also recommended the use of employment density census information, which is not an attribute already tracked in our LRS.

- **Oregon:** The variability and inaccuracy would be very great.

Question 3: Does your organization take traffic counts on off-system roads? If so, what methodology is used?

- **Colorado:** Traffic counts on off-system roads are based off the needs of HPMS. Counts are collected either on a 3-year of 6-year basis depending on if the count is collected on the NHS and/or the functional classification of the road is 3.

- **Iowa:** Yes, for the short term traffic collection program, where tube counters are deployed.
- **Kansas:** 24-hour cumulative counts with roadhose counters
- **Minnesota:** For roads in MN that are designated to draw state aid funding but are not non-Federal-Aid System roads roughly 12,000 locations are collected on a 4-12 year cycle. However, these are generally higher in volume and not the random sampling needed for overall non-Federal-Aid System AADT estimation.
- **Missouri:** We do collect traffic on off system roads where where RPCs and MPOs have interest in traffic data due to a new business, neighborhood, etc. We add these sites to our regular count program so we have historical AADTs when we are trying to upgrade the functional classification of a roadway.
- **Montana:** We have samples on all road types. I'm not sure that counts as a methodology but we try to have a similar amount in each county/urban area.
- **Oregon:** Yes. Virtually the same as on system, except for timing. We collect our state road volumes during the Spring and Autumn, when the seasonal factors are closer to average. We usually manage to fit some other roads in then also. But most of our non-state roads get counted in the Summer, and consequently have a much more variable seasonal factor.
- **Utah:** We do have a small sample selection of counts that remain on our schedule for local roads that have recently functionally classified higher that met the criteria. This typically goes through a house cleaning and removal process during each census update that adds/removes roadways.
- **Wisconsin:** Traffic counts taken on the off-system roads are collected by a 24-48 Short term special counts as a request. The factors used for the short-term count are based on the continuous count taken in the same area.
- **Wyoming:** Yes, 48-hour hose counts.

Question 4: Does your organization take traffic counts on off-system roads? If not, what are the reasons for this?

- **Minnesota:** Lack of funding and staff.
- **Oregon:** We only don't, when we can get some local jurisdiction counts on the necessary roads.
- **Utah:** With the exception of the locations in questions 3, Utah Department of Transportation doesn't budget or plan to collect data for roadways that aren't specifically required for HPMS, i.e., AADT, Single Unit, Combination Unit, etc.

Question 5: Are local highway agencies required to report AADTs to the DOT?

- **Colorado:** No
- **Iowa:** No
- **Kansas:** No. There is some urban counting done by cities in the large urbanized areas of Kansas City and Wichita – they publish their counts and we may use some of that information, but there is no requirement to report.
- **Minnesota:** No, except for in the situation where the MN local government screening boards have granted certain larger metropolitan areas permission to collect the traffic counts needed for state aid funding, which MnDOT processes and publishes. These total around 8,500 on the Federal-Aid System and 1,300 on the non-Federal-Aid system.
- **Missouri:** No, but some do provide counts and turning movements ahead of designed projects
- **Montana:** We have MPOs and some municipalities that share traffic count data. MPOs are required to report AADTs
- **Oregon:** No
- **Utah:** Not for AADT. The planning group does however work with local agencies to generate Future AADT based on their available model forecasting.
- **Wisconsin:** They are not required to report their collected data to DOT. The City of Madison also uses Jackalope for storage and processing of their collected data. They are not required to share the data collected.
- **Wyoming:** No

Question 6: Does your organization anticipate the answers to the previous questions changing in the near future, if so how?

- **Colorado:** The only thing I could see changing would be the methodology in determining the Statewide VMT estimates for FC's 6's and 7's. With the ever changing landscape of third-party data there could be opportunities to use this data to provide AADT/VMT estimates.
- **Iowa:** No
- **Kansas:** No. We have statutory requirements (State Statute) to develop annual travel estimates by county and certain cities for the purpose of fuel tax disbursement. This is independent of, and pre-dates, federal MIRE requirements. We anticipate that the MIRE requirements can be met under the existing program with minor adjustments.

- **Minnesota:** No, the cost to buy equipment and hire staff is prohibitive for most smaller cities and counties. MnDOT routinely reaches out to local agency's to ask for any count data they have, but in most cases they do not collect any because they relay on MnDOT. If they do have data it is often time consuming to translate into a format DOT staff can load to our traffic monitoring system. We have found more might be done to partner with other MnDOT offices that are in need of AADT on the non-Federal-Aid System, such as the Rail Safety office, who may be willing/have funding to hire a contractor to collect count data.

- **Missouri:** Yes. We are in the process of acquiring crowdsourced traffic data. This data is very different then the factored short counts we traditionally use and will change the way we analyze and process our traffic data.

- **Montana:** We are looking into using probe data to generate off-system AADT values.

- **Oregon:** Not much. We are looking at the possibility of using non-traditional methods. But so far have not seen good enough results to certify them.

- **Utah:** We've had internal discussions on reviewing the issue of not having knowledge of where the AADT value for the local roadways comes from, and determining how to correct this. We're unsure on how to proceed to fix this, as it's a complex system to replace, and requires research as well as determining whether it should be UDOT or Utah's local agencies to be the data owners and who should manage that process.

- **Wisconsin:** Not at this time, we do not anticipate any changes.

- **Wyoming:** No

Respondents

- **Colorado:** Aaron Moss, Traffic Analyst, aaron.moss@state.co.us

- **Iowa:** Jeff von Brown, Transportation Planner IV, jeff.vonbrown@iowadot.us

- **Kansas:** Bill Hughes, P.E., bill.hughes@ks.gov

- **Minnesota:** Christy Prentice, Supervisor Traffic Monitoring Total Volume Program, Christy.Prentice@state.mn.us

- **Missouri:** Spencer Robinson, TMS Administrator, spencer.robinson@modot.mo.gov

- **Montana:** Peder Jerstad, Transportation Planner, Traffic Data Collection & Analysis Section, pjerstad@mt.gov

- **Oregon:** Don R. Crownover, P.E., State Traffic Count Engineer, don.r.crownover@odot.oregon.gov

- **Utah:** Nicolas Black, Traffic Data Analytics Manager, nfblack@utah.gov
- **Wisconsin:** Rhonda McDonald, Statewide Data Analyst, rhonda.mac41@gmail.com
- **Wyoming:** Chad Mathews, Transportation Surveys Supervisor, chad.mathews@wyo.gov

8. Appendix B: Data Inventory

Table 8-1. List of ITD Data Sets

GIS Feature Class	Feature Type
aadts2021_assigned_stations_final	Point
rhgdb.ITDRH.CalibrationPoint	Point
rhgdb.ITDRH.Centerline	Line
rhgdb.ITDRH.CenterlineSequence	Table
rhgdb.ITDRH.Intersections	Point
rhgdb.ITDRH.Lrs_Locks	Table
rhgdb.ITDRH.Lrs_Metadata	Table
rhgdb.ITDRH.LRSE_AADT	Line
rhgdb.ITDRH.LRSE_BLM_Roads	Line
rhgdb.ITDRH.LRSE_Bridges	Point
rhgdb.ITDRH.LRSE_CityLimits	Line
rhgdb.ITDRH.LRSE_CountHistory	Point
rhgdb.ITDRH.LRSE_CountyLimits	Line
rhgdb.ITDRH.LRSE_DesignatedRoutes129k	Line
rhgdb.ITDRH.LRSE_DistrictLimits	Line
rhgdb.ITDRH.LRSE_Equations	Point
rhgdb.ITDRH.LRSE_ExtraLengthComboRoutes	Line
rhgdb.ITDRH.LRSE_FederalAid	Line
rhgdb.ITDRH.LRSE_FlowMapNode	Point
rhgdb.ITDRH.LRSE_Fsroads	Line
rhgdb.ITDRH.LRSE_FunctionalClass	Line
rhgdb.ITDRH.LRSE_FutureFacility	Line
rhgdb.ITDRH.LRSE_HighwayInvestmentCorridors	Line
rhgdb.ITDRH.LRSE_HighwaySafetyCorridors	Line
*rhgdb.ITDRH.LRSE_HPMS_AADTCombination	Line
*rhgdb.ITDRH.LRSE_HPMS_AADSingleUnit	Line
*rhgdb.ITDRH.LRSE_HPMS_AccessControl	Line
*rhgdb.ITDRH.LRSE_HPMS_AlternativeRouteName	Line
*rhgdb.ITDRH.LRSE_HPMS_BaseThickness	Line
*rhgdb.ITDRH.LRSE_HPMS_BaseType	Line
*rhgdb.ITDRH.LRSE_HPMS_Capacity	Line
*rhgdb.ITDRH.LRSE_HPMS_Dfactor	Line
*rhgdb.ITDRH.LRSE_HPMS_FacilityType	Line
*rhgdb.ITDRH.LRSE_HPMS_Faulting	Line
*rhgdb.ITDRH.LRSE_HPMS_FutureAADT	Line

GIS Feature Class	Feature Type
*rhgdb.ITDRH.LRSE_HPMS_Kfactor	Line
*rhgdb.ITDRH.LRSE_HPMS_LaneWidth	Line
*rhgdb.ITDRH.LRSE_HPMS_LastOverlay	Line
*rhgdb.ITDRH.LRSE_HPMS_Median	Line
*rhgdb.ITDRH.LRSE_HPMS_OHSpeedLimits	Line
*rhgdb.ITDRH.LRSE_HPMS_Ownership	Line
*rhgdb.ITDRH.LRSE_HPMS_Parking	Line
*rhgdb.ITDRH.LRSE_HPMS_PassingSight	Line
*rhgdb.ITDRH.LRSE_HPMS_Pavement	Line
*rhgdb.ITDRH.LRSE_HPMS_PctPeakCombination	Line
*rhgdb.ITDRH.LRSE_HPMS_PctPeakSingle	Line
*rhgdb.ITDRH.LRSE_HPMS_PeakLanes	Line
*rhgdb.ITDRH.LRSE_HPMS_PSR	Line
*rhgdb.ITDRH.LRSE_HPMS_SectionMovement	Line
*rhgdb.ITDRH.LRSE_HPMS_Shoulders	Line
*rhgdb.ITDRH.LRSE_HPMS_StructureType	Line
*rhgdb.ITDRH.LRSE_HPMS_SurfaceType	Line
*rhgdb.ITDRH.LRSE_HPMS_TerrainType	Line
*rhgdb.ITDRH.LRSE_HPMS_ThicknessFlexible	Line
*rhgdb.ITDRH.LRSE_HPMS_ThicknessRigid	Line
*rhgdb.ITDRH.LRSE_HPMS_ThroughLanes	Line
*rhgdb.ITDRH.LRSE_HPMS_TOPS	Line
*rhgdb.ITDRH.LRSE_HPMS_TurnLanes	Line
*rhgdb.ITDRH.LRSE_HPMS_Widening	Line
*rhgdb.ITDRH.LRSE_HPMS_YearLastConstruction	Line
*rhgdb.ITDRH.LRSE_HPMS_YearLastImprovement	Line
rhgdb.ITDRH.LRSE_IndianReservations	Line
rhgdb.ITDRH.LRSE_Interchanges	Line
rhgdb.ITDRH.LRSE_Jurisdiction	Line
rhgdb.ITDRH.LRSE_LocalRoadInventory	Line
rhgdb.ITDRH.LRSE_LRI_SummerFW	Line
rhgdb.ITDRH.LRSE_MilepointDescription	Point
rhgdb.ITDRH.LRSE_MilepostSigns	Point
rhgdb.ITDRH.LRSE_NationalTruckNetwork	Line
rhgdb.ITDRH.LRSE_NHS	Line
rhgdb.ITDRH.LRSE_OTIS_Linear	Line
rhgdb.ITDRH.LRSE_OTIS_Point	Line
rhgdb.ITDRH.LRSE_Oversized_Routes	Line
rhgdb.ITDRH.LRSE_RailRoadCrossings	Point

GIS Feature Class	Feature Type
rhgdb.ITDRH.LRSE_RoadMode	Line
rhgdb.ITDRH.LRSE_RoadNames	Line
rhgdb.ITDRH.LRSE_RoadNetworkPrimary	Line
rhgdb.ITDRH.LRSE_RoadType	Line
rhgdb.ITDRH.LRSE_RouteDominance	Line
rhgdb.ITDRH.LRSE_RouteQualifier	Line
rhgdb.ITDRH.LRSE_SegmentCode	Line
rhgdb.ITDRH.LRSE_SHSPrimary	Line
rhgdb.ITDRH.LRSE_SpeedZones	Line
rhgdb.ITDRH.LRSE_StateHighwaySystem	Line
rhgdb.ITDRH.LRSE_StationBook	Point
rhgdb.ITDRH.LRSE_STRAHNET	Line
rhgdb.ITDRH.LRSE_TrendFileData	Point
rhgdb.ITDRH.LRSE_UrbanLimits	Line
rhgdb.ITDRH.LRSE_VerticalClearance	Line
rhgdb.ITDRH.LRSE_WARSNetwork	Line
rhgdb.ITDRH.LRSE_WeightCapacity	Line
rhgdb.ITDRH.LRSE_WinterRoadReportSeg	Line
rhgdb.ITDRH.LRSE_RoadNetwork	Line
rhgdb.ITDRH.LRSE_Redline	Line
rhgdb.ITDRH.tblConfig_HPMS_backbone	Table

Count Stations by County

Table 8-2. Number of Count Stations in Each County

County Name	Continuous Count Sites	On Count Cycle	Assigned Station	Total
Ada	41	784	239	1,064
Adams	3	31	20	54
Bannock	23	284	88	395
Bear Lake	4	41	15	60
Benewah	1	39	11	51
Bingham	2	213	77	292
Blaine	2	76	25	103
Boise	6	34	10	50
Bonner	8	162	23	193
Bonneville	7	270	171	448
Boundary	3	48	11	62
Butte	3	23	14	40
Camas	-	17	7	24
Canyon	27	554	134	715
Caribou	1	41	5	47
Cassia	5	143	53	201
Clark	4	27	17	48
Clearwater	-	37	3	40
Custer	2	32	8	42
Elmore	3	144	52	199
Franklin	2	39	10	51
Fremont	4	74	16	94
Gem	3	74	17	94
Gooding	3	72	57	132
Idaho	5	118	20	143
Jefferson	3	80	43	126
Jerome	6	83	74	163
Kootenai	20	492	183	695
Latah	7	177	47	231
Lemhi	3	26	4	33
Lewis	3	31	11	45
Lincoln	1	31	9	41
Madison	1	118	51	170
Minidoka	1	104	28	133
Nez Perce	3	243	50	296

County Name	Continuous Count Sites	On Count Cycle	Assigned Station	Total
Oneida	2	66	22	90
Owyhee	4	68	4	76
Payette	7	106	35	148
Power	5	83	42	130
Shoshone	1	107	86	194
Teton	3	37	5	45
Twin Falls	5	251	67	323
Valley	2	74	4	80
Washington	4	78	25	107
Total	**243**	**5,632**	**1,893**	**7,768**

Roadway Characteristics Data Sets

This section describes the data in key ITD data sets provided to the FHWA within HPMS submissions and describes their coding based on the *HPMS Field Manual* (Federal Highway Administration 2018). Additionally, for the rhgdb.ITDRH.LRSE_LocalRoadInventory data set it is supplemented surface type codes reported by ITD (Calderon 2022).

Rhgdb.ITDRH.LRSE_HPMS_ThroughLanes

- The number of through lanes in both directions carrying through traffic in the off-peak period.

Rhgdb.ITDRH.LRSE_HPMS_LaneWidth

- The predominant through-lane width to the nearest whole foot.

Rhgdb.ITDRH.LRSE_HPMS_SurfaceType

- 1: Unpaved
- 2: Bituminous (Asphalt Pavement)
- 3: JPCP – Jointed Plain Concrete Pavement (includes whitetopping) (Jointed Concrete Pavement)
- 4: JRCP – Jointed Reinforced Concrete Pavement (includes whitetopping) (Jointed Concrete Pavement)
- 5: CRCP – Continuously Reinforced Concrete Pavement (CRCP)
- 6: Asphalt-Concrete (AC) Overlay over Existing AC Pavement (Asphalt Pavement)
- 7: AC Overlay over Existing Jointed Concrete Pavement (Asphalt Pavement)
- 8: AC (Bituminous Overlay over Existing CRCP) (Asphalt Pavement)
- 9: Unbonded Jointed Concrete Overlay on PCC Pavement (Jointed Concrete Pavement)
- 10: Bonded PCC Overlay on PCC Pavement (Jointed Concrete Pavement)
- 11: Other (e.g., plank, brick, cobblestone, etc.)

rhgdb.ITDRH.LRSE_HPMS_Parking

- 1: Parking allowed on one side.
- 2: Parking allowed on both sides.
- 3: No parking allowed or none available.

Rhgdb.ITDRH.LRSE_HPMS_Shoulders

- 1: None
- 2: Surfaced shoulder exists – bituminous concrete
- 3: Surfaced shoulder exists – Portland Cement Concrete surface (PCC)
- 4: Stabilized shoulder exists (stabilized gravel or other granular material with or without admixture)
- 5: Combination shoulder exists (shoulder width has two or more surface types; e.g., part of the shoulder width is surfaced and a part of the width is earth)
- 6: Earth shoulder exists
- 7: Barrier curb exists; no shoulder in front of curb

rhgdb.ITDRH.LRSE_HPMS_TerrainType

- 1: Level
- 2: Rolling
- 3: Mountainous

rhgdb.ITDRH.LRSE_LocalRoadInventory

- Earth (C)
- F (Dust Suppressant Treated Gravel)
- G-1 (Paved)
- G-2 (Paved)
- Gravel (E)
- J (Paved)

MPO Counts

Table 8-3. MPO Count Data Sources

Organization	Link
Community Planning Association of Southwest Idaho (COMPASS)	https://www.compassidaho.org/prodserv/traffic_counts.htm
Kootenai MPO	https://www.kmpo.net/traffic-counts/
Bannock Transportation Planning Organization	https://www.bannockplanning.org/traffic-counts/
Bonneville MPO	https://www.bmpo.org/traffic-counts
Lewis-Clark Valley Metropolitan Planning Organization (LCVMPO)	https://lewisclarkmpo.org/2209/Intersection-Counts

Economic Data

Table 8-4. Economic Data Sources

Name	Source	Link
National Land Cover Database (NLCD)	U.S. Geological Survey	https://www.usgs.gov/centers/eros/science/national-land-cover-database#overview
National Land Cover Database (NLCD)	Land Use and Land Cover Technical Working Group	https://gis.idaho.gov/land-use-land-cover-twg
Urban Areas	U.S. Census Bureau / Tiger/LINE	https://www.census.gov/cgi-bin/geo/shapefiles/index.php
Core Based Statistics Areas (CBSAs)	U.S. Census Bureau, Cartographic Boundaries	https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html
Metropolitan Planning Organization (MPO) boundaries	U.S. Department of Transportation, National Transportation Atlas Database	https://data-usdot.opendata.arcgis.com/datasets/metropolitan-planning-organizations/explore?location=44.550377%2C-113.322503%2C5.64
LODES 7 Workplace Area Characteristics (WAC)	U.S. Census Bureau	https://lehd.ces.census.gov/data/#lodes
LODES 7 Workplace Area Characteristics (WAC)	U.S. Census Bureau	https://lehd.ces.census.gov/data/#lodes
B19001 | Household income in the past 12 months (in 2020 inflation-adjusted dollars)	U.S. Census Bureau, American Community Survey	https://data.census.gov/cedsci/table?t=Income%20%28Households,%20Families,%20Individuals%29&g=0400000US16%241500000&tid=ACSDT5Y2020.B19001
B19013 | Median household income in the past 12 months (in 2020 inflation-adjusted dollars)	U.S. Census Bureau, American Community Survey	https://data.census.gov/cedsci/table?t=Income%20%28Households,%20Families,%20Individuals%29&g=0400000US16%241500000&tid=ACSDT5Y2020.B19013
B17010 | Poverty status in the past 12 months of families by family type by presence of related children under 18 years by age of related children	U.S. Census Bureau, American Community Survey	https://data.census.gov/cedsci/table?t=Poverty&g=0400000US16%241500000&tid=ACSDT5Y2020.B17010

Demographic Data

Table 8-5. Demographic Data Sources

Name	Source	Link
P1 | Race	U.S. Census Bureau, 2022 Decennial Census	https://data.census.gov/cedsci/table?q=population&g=0400000US16%241000000&d=DEC%20Redistricting%20Data%20%28PL%2094-171%29&tid=DECENNIALPL2020.P1
Census Block	U.S. Census Bureau, Cartographic Boundary	https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.html
H1 | Occupancy status	U.S. Census Bureau, DEC Redistricting Data	https://data.census.gov/cedsci/table?q=dwelling%20units&t=Housing%20Units&g=0400000US16%241000000&y=2020&tid=DECENNIALPL2020.H1
S1401 | School enrollment	U.S. Census Bureau, American Community Survey	https://data.census.gov/cedsci/table?q=dwelling%20units&t=School%20Enrollment&g=0400000US16%241400000&tid=ACSST5Y2020.S1401
Total Vehicle Registrations	ITD, DMV Data, "Vehicle Registration" tab	https://itd.idaho.gov/dmvdata/
Driver Licenses in Force by County	ITD, DMV Data, "Driver Licenses" tab	https://itd.idaho.gov/dmvdata/

9. Appendix C: Variables for Potential Use in Geospatial Interpolation Model

Table 9-1 summarizes the variables that are likely to be used in a geospatial interpolation model. Not all variables will be used in the final model. The priority variables that are most likely to be useful are designated with an asterisk (*).

Table 9-1. Variables and Data Sets

Variable Category	Variable	Comments	Data Name	Data Source	Page for More Information about Data Source
Traffic Volume*	Average Annual Daily Traffic (AADT)	AADT values derived from as far back as 10 years ago can be used as traffic patterns in rural areas are less likely to change quickly.	historical_count_history_1960-2021	ITD	53
Roadway*	Number of through lanes	Data not available for all facilities. Assume two lanes when unknown.	rhgdb.ITDRH.LRSE_HPMS_ThroughLanes	ITD, HMPS Coordinator	41
Roadway*	Surface type	Only distinguishes paved and unpaved roads on Federal Aid System roads. Assume unpaved in rural areas and paved in urban areas when information not available from Local Road Inventory.	rhgdb.ITDRH.LRSE_HPMS_SurfaceType • Use where data is available. rhgdb.ITDRH.LRSE_LocalRoadInventory • Use to supplement HPMS surface type data.	ITD, HMPS Coordinator	41
Roadway*	Functional class	Classes must be treated in geospatial interpolation models as categorical data rather than ordinal or interval.	rhgdb.ITDRH.LRSE_FunctionalClass	ITD, HQ Planning	41
Network	Degree centrality	Indicates the number of connected facilities. Calculate using centrality tools in GIS applications.	rhgdb.ITDRH.LRSE_RoadNetwork	ITD, HQ GIS	43
Network	Closeness centrality	Indicates proximity to the rest of the network. Calculate using centrality tools in GIS applications.	rhgdb.ITDRH.LRSE_RoadNetwork	ITD, HQ GIS	43

Variable Category	Variable	Comments	Data Name	Data Source	Page for More Information about Data Source
Network	Betweenness centrality	Indicates how often a segment or intersection is part of a shortest route. Calculate using centrality tools in GIS applications.	rhgdb.ITDRH.LRSE_RoadNetwork	ITD, HQ GIS	43
Network	Distance to intersection	Distance from each road segment to the nearest intersection.	rhgdb.ITDRH.LRSE_RoadNetwork	ITD, HQ GIS	43
Network	Intersection density	Intersections per square mile. Calculated for a given radius or as a moving average because administrative boundaries such as counties or Census tracts vary in size.	rhgdb.ITDRH.LRSE_RoadNetwork	ITD, HQ GIS	43
Network	Accessibility (to Primary or Secondary Roads)	Network distance to the nearest facility of functional class 1-2 (Primary) or 3-5 (Secondary).	rhgdb.ITDRH.LRSE_RoadNetwork	ITD, HQ GIS	43
Network*	Road mileage density	Miles of road per square mile. Calculated for a given radius or as a moving average because administrative boundaries such as counties or Census tracts vary in size.	rhgdb.ITDRH.LRSE_RoadNetwork	ITD, HQ GIS	43
Network*	Urban / rural designation	Determined by ITD.	rhgdb.ITDRH.LRSE_UrbanLimits	ITD	45
Economic*	Employment density	Use residence area characteristics. Aggregate at the Census tract level. Use land area (ALAND) from Tiger/LINE shapefiles to calculate density.	LEHD LODES • Used to calculate total employment. Tiger/LINE • Use for land area (ALAND is the attribute for land area in square meters)	U.S. Census Bureau	46 (description)
Economic	Employment by Industry	Use two-digit NAICS codes to define industries. Use workplace area	LEHD LODES	U.S. Census Bureau	46 (description),

Variable Category	Variable	Comments	Data Name	Data Source	Page for More Information about Data Source
		characteristics. Aggregate at the Census tract level.			98 (link to data set)
Economic*	Median income	Aggregate at the Census tract level.	American Community Survey	U.S. Census Bureau	46 (description), 98 (link to data set)
Economic	Poverty rates	Can be calculated as a rate (x per 1000 residents) or as a percentage of residents. Aggregate at the Census tract or county levels.	American Community Survey	U.S. Census Bureau	46 (description), 98 (link to data set)
Economic	Nearby population	Although necessary for other calculations, this factor is likely to have high correlation to population density and to be less informative. If used, add an approximately ¼-mile buffer around each road segment and calculate the population density within that buffer.	Decennial Census, P1 | Race • Use for population.	U.S. Census Bureau	47 (description), 98 (link to data set)
Demographic*	Population density	Calculate at the Census tract level using population data from the Decennial Census and land area (ALAND) from Tiger/LINE shapefiles.	Decennial Census, P1 | Race • Use for population. Tiger/LINE • Use for land area (ALAND is the attribute for land area in square meters)	U.S. Census Bureau	47 (description), 99 (link to data set)
Demographic	Dwelling units	Dwelling units are likely correlated with population variables but may be informative relative to the ratio of vehicles to population.	DEC Redistricting Data, H1 | Occupancy status	U.S. Census Bureau	47 (description), 99 (link to data set)
Demographic	School enrollment	Can be calculated as a rate (x per 1000 residents) or as a percentage of residents. This should be calculated at	ACS, S1401 | School enrollment	U.S. Census Bureau	47 (description),

Variable Category	Variable	Comments	Data Name	Data Source	Page for More Information about Data Source
		the Census tract level and assigned to roads within the Census tract.			99 (link to data set)
Demographic	Vehicle registration	Vehicle registrations are only available down to the county level and may therefore be less useful than more granular data. If used, this variable should capture the number of vehicles registered in the county.	DMV Data, "Driver Licenses" tab	ITD	47 (description), 99 (link to data set)

10. Appendix D: TAC Committee Feedback on Methodological Approaches for Estimating Off-System AADT

This appendix provides the raw answers that were received from the TAC to a survey gathering their feedback on the three primary methods that were considered for estimating AADT on off-system roads: regression analysis, geospatial interpolation, and travel demand modeling. These three approaches are briefly summarized below.

- **Regression analysis:** Evaluates statistical relationships between AADT and explanatory variables.
    - **Types:** Linear, geographically weighted, Poisson
- **Geospatial interpolation:** Estimates traffic flows based on spatial proximity to other observations.
    - **Types:** Inverse distance weighted, k-nearest neighbors, ordinary Kriging, universal Kriging
- **Travel demand modeling:** Distributes expected trips along local roadways based on origin-destination pairs.
    - **Types:** ITD TDM output analysis (which disaggregates traffic from centroid connectors onto relevant off-network roads), simplified local models (which replicate some TDM functions for off-system roads in simplified form)

The following sections provide the TAC committee member's responses to each of the survey questions.

1. Please enter your name
    a. Vicky Calderon
    b. David Coladner
    c. Nicole Hanson
    d. Kevin Kuther
    e. Margaret Pridmore
    f. Matthew Syphus
2. Rank the three methods discussed during the TAC meeting on 9 November 2022 (shown in Figure 10-1).

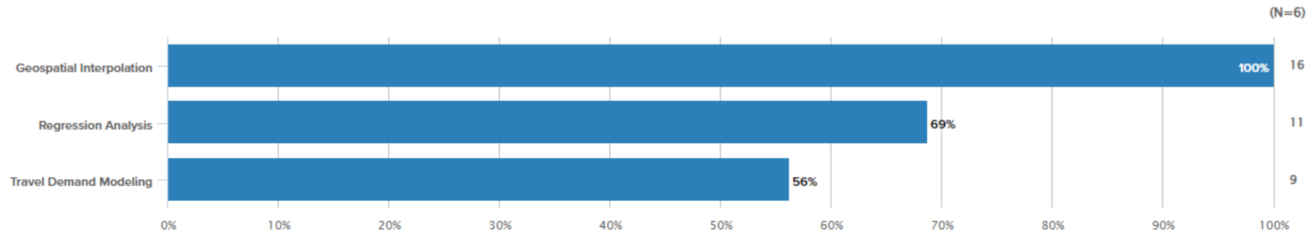


Figure 10-1. TAC Member Ranking

3. Please briefly explain your ranking.
   a. See below for complete responses.
   b. Themes include—
      i. Geospatial interpolation (five 1st place choices)
         1. Pros: related to AADT's local character; it can work based on conversations with other DOTs; very comprehensive.
         2. Cons: People didn't like the risk of overestimating AADT.
      ii. Regressions (one 1st place choice)
         1. Pros: Familiar and well documented; simple and flexible
      iii. TDM (no 1st place choices)
         1. Pros: Familiar

4. Within the methods that you ranked most highly, which variations of those methods do you find most compelling (for instance, disaggregating output from the statewide TDM versus developing simplified local travel demand models)?

Table 10-1. Responses to Questions 3 and 4

Name	Ranks	3) Please briefly explain your ranking.	4) Which variations within highest ranked method
Vicky Calderon	1. Geospatial Interpolation 2. Regression Analysis 3. Travel Demand Modeling	I like the idea of being able to produce estimates geospatially but do have concerns with accuracy.	I like the simplicity/flexibility of the regression models. Model type #2 catches my attention. I'm torn between geospatial interpolation and regression analysis. Specifically, concerned with the geospatial interpolation of estimates on low volume off-system roads (which is our main star for this research).
David Coladner	1. Geospatial Interpolation 2. Travel Demand Modeling 3. Regression Analysis	I think spatial aspect is most directly relating to local character of average daily traffic. TDM can be good though.	Your example is a great one! Also for geospatial, targeted counts on a sample basis (according to example in NY state) where they essentially sampled 10% of all local roads county by county to establish a flavor for each locality.
Nicole Hanson	1. Geospatial Interpolation 2. Regression Analysis 3. Travel Demand Modeling	From talking with other DOTs, sounds like the geospatial will work,	would just want to make sure that the data isn't EXTREMELY over or under guessed
Kevin Kuther	1. Geospatial Interpolation 2. Regression Analysis 3. Travel Demand Modeling	No answer	No answer
Margaret Pridmore	1. Geospatial Interpolation 2. Travel Demand Modeling 3. Regression Analysis	Geospatial Interpolation will provide the most comprehensive look, even if it's more complicated.	Using two variants would not be inappropriate - one for the higher volume large urban areas, and another for the low volume rural roads.
Matthew Syphus	1. Regression Analysis 2. Travel Demand Modeling 3. Geospatial Interpolation	• Regression analysis is flexible and well documented. • TDM is familiar. • Geospatial interpolation is prone to wrong numbers on low volume roads.	No response