Final Report

TRC1801

Evaluation of WIM Auto-Calibration Practices and Parameters

By
Sarah Hernandez
Principal Investigator
College of Engineering, Department of Civil Engineering

Josh Baker
Fu Zhang Durandal
Graduate Research Assistants
College of Engineering, Department of Civil Engineering

September 2019

TECHNICAL REPORT DOCUMENTATION

1. Report No.	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Evaluation of WIM Auto-Calibration Practices and Parameters		**5. Report Date** September 2019 **6. Performing Organization Code**
7. Author(s) Sarah Hernandez		**8. Performing Organization Report No.**
9. Performing Organization Name and Address Department of Civil Engineering University of Arkansas 4183 Bell Engineering Center Fayetteville, Arkansas 72701		**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** TRC1801
12. Sponsoring Agency Name and Address Arkansas Department of Transportation PO Box 2261 Little Rock, Arkansas 72203-2261		**13. Type of Report and Period Covered** FINAL **14. Sponsoring Agency Code**
15. Supplementary Notes Project performed in cooperation with the Federal Highway Administration		

16. Abstract

Weigh-in-Motion (WIM) systems capture weight and axle configurations of vehicles using the state highway network and serve as valuable and essential input for evaluating the performance of our transportation infrastructure. To produce accurate weights, sensors must be calibrated at regular intervals. Since on-site calibration can be costly, several states have adopted auto-calibration procedures. Auto-calibration is an algorithmic procedure by which weights measured by the WIM sensor are adjusted using reference weights, e.g., assumed front axle weights (FAW) of five-axle tractor-trailers. This project developed a new form of auto-calibration based on Automatic Vehicle Identification (AVI) data, specifically truck Global Positioning System (GPS) data. Two data collection efforts were undertaken to gather WIM, static scale, and video data for algorithm validation resulting in a sample of approximately 500 trucks to use for algorithm performance evaluation. Without any form of auto-calibration, we observed FAW errors ranging from 24% to 85% with Gross Vehicle Weight (GVW) error in the same range. Site-specific tuning of the user-specified values required in the Arkansas Department of Transportation (ARDOT) and Minnesota Department of Transportation (MnDOT) auto-calibration algorithms resulted in errors for FAW between 9% and 11% and for GVW between 14% and 18%. A limitation of developing site-specific, user-specific values like FAW references is that it would require very detailed and time-consuming data collection efforts. The AVI approach, on the other hand, alleviates the need to perform manual field data collection by leveraging AVI truck tracking technologies such as GPS. The AVI-based method reduces errors to between 10% and 35% for FAW and 16% and 35% for GVW. If auto-calibration using AVI were to replace on-site calibration efforts, ARDOT could realize up to 63%. Future work should examine performance increases resulting from a larger AVI sample, evaluation license plate matching as a basis for AVI data, and the use of AVI data to prioritize WIM site sensor upgrades.

17. Key Words Weigh in Motion (WIM), Calibration, Quality Control, Weight	**18. Distribution Statement** This document is available to the US public through the National Technical Information Service, Springfield, Virginia 22161		
19. Security Classification (of this report) Unclassified	**20. Security Classification (of this page)** Unclassified	**21. No of Pages**	**22. Price**

METRIC CONVERSIONS

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
^oC	Celsius	1.8C+32	Fahrenheit	^oF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS, ACRONYMS, AND SYMBOLS

AASHTO – American Association of State and Highway Transportation Officials
APE – Absolute Percent Error
ARDOT – Arkansas Department of Transportation
ASTM – ASTM International, formerly known as American Society for Testing and Materials
AVI – Automatic Vehicle Identification
CF – Calibration Factor
CMR – Correct Match Rate
DOT – Department of Transportation
ELD – Electronic Logging Device
ER – Error Rate
ESAL – Equivalent Single Axle Load
FAW – Front Axle Weight
FHWA – Federal Highway Administration
GPS – Global Positioning System
GVW- Gross Vehicle Weight
HPMS – Highway Performance Monitoring System
ILD – Inductance Loop Detector
LTPP – Long Term Pavement Performance Program
MAPE – Mean Absolute Percent Error
MdAPE – Median Absolute Percent Error
MEPDG – Mechanistic Empirical Pavement Design Guide
MnDOT – Minnesota Department of Transportation
PVR – Per Vehicle Record
QC – Quality Control
ROI – Return on Investment
TMG – Traffic Monitoring Guide
TMR – True Match Rate
TRB – Transportation Research Board
VC – Vehicle Class
WIM – Weigh-in-Motion

EXECUTIVE SUMMARY

Weigh-in-Motion (WIM) is the process of measuring dynamic tire forces of a moving vehicle and estimating the corresponding tire loads (ASTM, 2009). Unlike static scales that require select trucks to exit the highway mainlines to be weighed, WIM systems capture vehicle characteristics while the vehicle is moving at full highway speeds in the mainline highway lanes. Since the WIM sensor operates continuously throughout the year and measures all passing trucks, WIM provide a means to estimate current and historical trends in truck volumes and weights. Through proper calibration procedures, such as those outlined in American Society of Testing and Materials (ASTM) Standard E1318-09 (*ASTM, 2009*), measurement errors commonly observed within WIM measured weights can be addressed. However, calibration protocols outlined in ASTM E1318-09 require on-site testing which can be cost prohibitive.

As a result, many states have adopted lower cost protocols that do not require on-site testing. One such practice, called auto-calibration, uses software to periodically calculate calibration factors based on presumed traffic characteristics. Weight data collected by the WIM sensor are then adjusted based on these calibration parameters. Although auto-calibration has significant value for budget-constrained state agencies, it has several drawbacks that limit its effectiveness. The proposed work sought to evaluate the auto-calibration practices used in the WIM systems in Arkansas and to propose an alternative method based on Automatic Vehicle Identification (AVI).

The proposed AVI-based auto-calibration method consisted of first, matching AVI-tracked trucks to WIM Per Vehicle Records (PVR), and second, applying a calibration procedure. This procedure measures the weights of the same truck, tracked by AVI across multiple WIM sites, to generate a reference weight and calibration factor. The approach currently used by ARDOT generates calibration factors based on the measured Front Axle Weight (FAW) averaged for a sample of 50 five-axle tractor trailers and compares it to a pre-defined reference weight. A more robust method, developed by the MnDOT, expands on that approach by defining three FAW references based on Gross Vehicle Weight (GVW) bins and applying correction factors when sample sizes are small.

The proposed AVI-based approach was compared to the ARDOT and MnDOT approaches for a set of six Arkansas WIM sites at Lamar, Lonoke, Bald Knob, Glen Rose, Arkadelphia, and Texarkana. During two data collections in March of 2018 and 2019, we collected WIM PVR at each WIM site, AVI data from a national truck GPS data provider, and static weight recordings at Arkansas Highway Police weight enforcement sites at Alma and Hope. An extensive data preprocessing methodology was developed and applied to provide data necessary for auto-calibration algorithm performance evaluation. In total, we identified approximately 500 truck matches from WIM to static scale locations which were used to evaluate performance of the three auto-calibration approaches.

Without any form of auto-calibration, we observed FAW errors ranging from 24% to 85% with GVW error in the same range. Site-specific tuning of the user-specified values required in the ARDOT and MnDOT auto-calibration algorithms resulted in errors for FAW between 9% and 11% and for GVW between 14% and 18%. Overall, through site-specific tuning of parameters like FAW reference values and GVW bin thresholds used in the ARDOT and MnDOT algorithms, we can potentially reduce measurement errors by 1% to 23%. Due to the fourth power relationship between measured weight and Equivalent Single Axle Load (ESAL) used in pavement design, an improvement in only 1% can result in 4% increase in ESAL. A limitation of developing site-specific, user-specific values like FAW references is that it would require detailed, time-consuming, and potentially expensive data collection efforts to gather necessary data. The AVI approach, on the other hand, alleviates some of the need to perform manual field data collection by leveraging AVI truck-tracking technologies such as GPS.

Comparison of FAW and GVW estimation accuracy across all three auto-calibration methods and study sites are summarized as follows:

- The ARDOT method reduced errors to between 12% and 16% for FAW and 14% to 29% for GVW.
- The MnDOT method reduced errors to between 11% and 26% for FAW and 11% to 41% for GVW.
- The AVI-based method reduced errors to between 10% and 35% for FAW and 16% and 35% for GVW.

In general, the AVI-based method worked well for sites with higher measurement error as seen at Lonoke and Glen Rose but maintained similar performance as the ARDOT and MnDOT algorithms in most other cases. Performance of the AVI-based algorithm was also found to correlate with the volume of trucks tracked by the AVI system, in this case a GPS tracking system. When more trucks were present, lower FAW and GVW estimation errors were observed. Using the ARDOT method, there was a correlation between higher GVW ranges and increased error. The AVI-based approach did not exhibit this same trend.

Future improvements to the proposed AVI-based approach include the following: (1) Although we saw performance in line with existing auto-calibration algorithms, we anticipate performance improvements as we increase the size of the AVI sample. We propose gathering AVI data, specifically GPS tracking data, from multiple providers to ensure a large and robust sample. (2) The manual process to generate data for model development and evaluation was time-consuming which limited our ability to increase our test sample size beyond 500 truck samples. We propose allowing license plate matching technology in place of side-fire video as was used here, to increase sample size and speed up data pre-processing. To be clear, video and license plate matching are only needed for model evaluation, not for real-time deployment. AVI-based auto-calibration using license plate readers instead of GPS tracking has the potential to greatly increase the sample size of trucks matched across sites as well as to alleviate possible biases in the sample of GPS trucks. (3) AVI can be used as a guide to prioritize WIM sensor improvements so that system-wide weight measurement accuracy can be improved. The AVI-based auto-calibration algorithm which tracks and compares truck weights across multiple sensors can be used to prioritize WIM site sensor upgrades. For example, we can consider a set of “anchor” sites which are commonly crossed by all trucks. Then, as we track trucks from these sites to “satellite” WIM sites (e.g., those with lower volume or lower quality sensors), we can use the measured weight at the “anchor” site as a reference by which to calculate a calibration factor for the “satellite” location.

The estimated Return on Investment (ROI) for this project was estimated at a 63% cost savings over anticipated expenditures related to on-site calibration requirements that involve test trucks of known weight or use of static scales for weight comparisons. The tradeoff between on-site calibration costs and measurement accuracy can be balanced by replacing on-site calibration methods with AVI-based solutions like that proposed in this work.

CHAPTER 1: PROJECT OVERVIEW

BACKGROUND

Weigh-in-Motion (WIM) is the process of measuring dynamic tire forces of a moving vehicle and estimating the corresponding tire loads (ASTM, 2009). A WIM system consists of embedded roadway sensors that collect vehicle arrival time and date, axle weights and Gross Vehicle Weight (GVW), axle spacing, and speed. WIM system software takes axle count, axle spacing, and weight measurements and predicts the vehicle class. Typically, vehicles are classified using the classification sieve for FHWA Scheme F that includes 13 axle-based classes (Figure 1). Unlike static scales that require select trucks to exit the highway mainlines to be weighed, WIM systems capture vehicle characteristics while the vehicle is moving at full highway speeds in the mainline highway. Since the WIM sensor operates continuously throughout the year and measures all passing trucks, WIM provides a means to estimate current and historical trends in truck volumes and weights. There are 42 WIM stations in Arkansas (Figure 2).

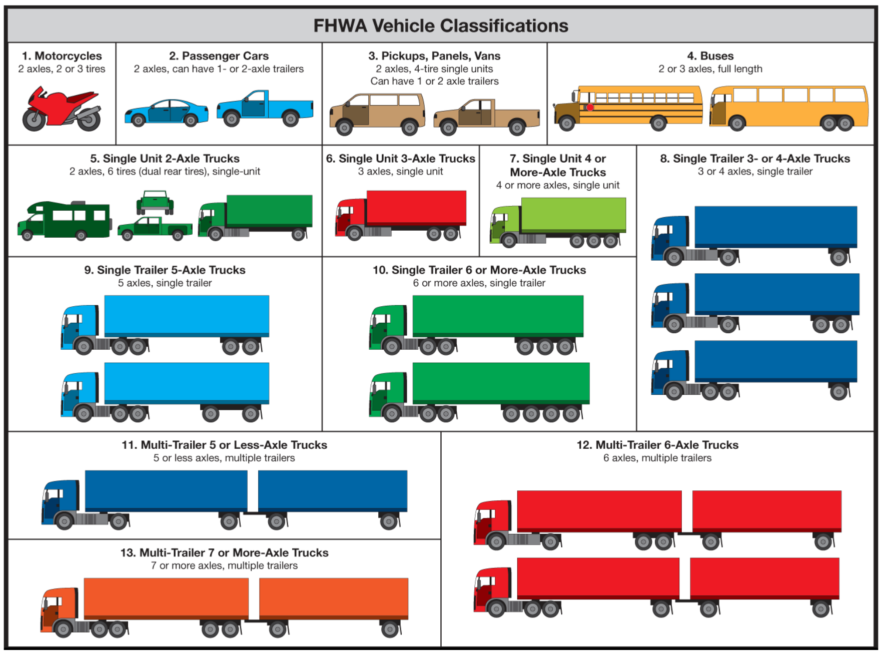


Figure 1. FHWA Classification Scheme (REF: TxDOT, 2017)

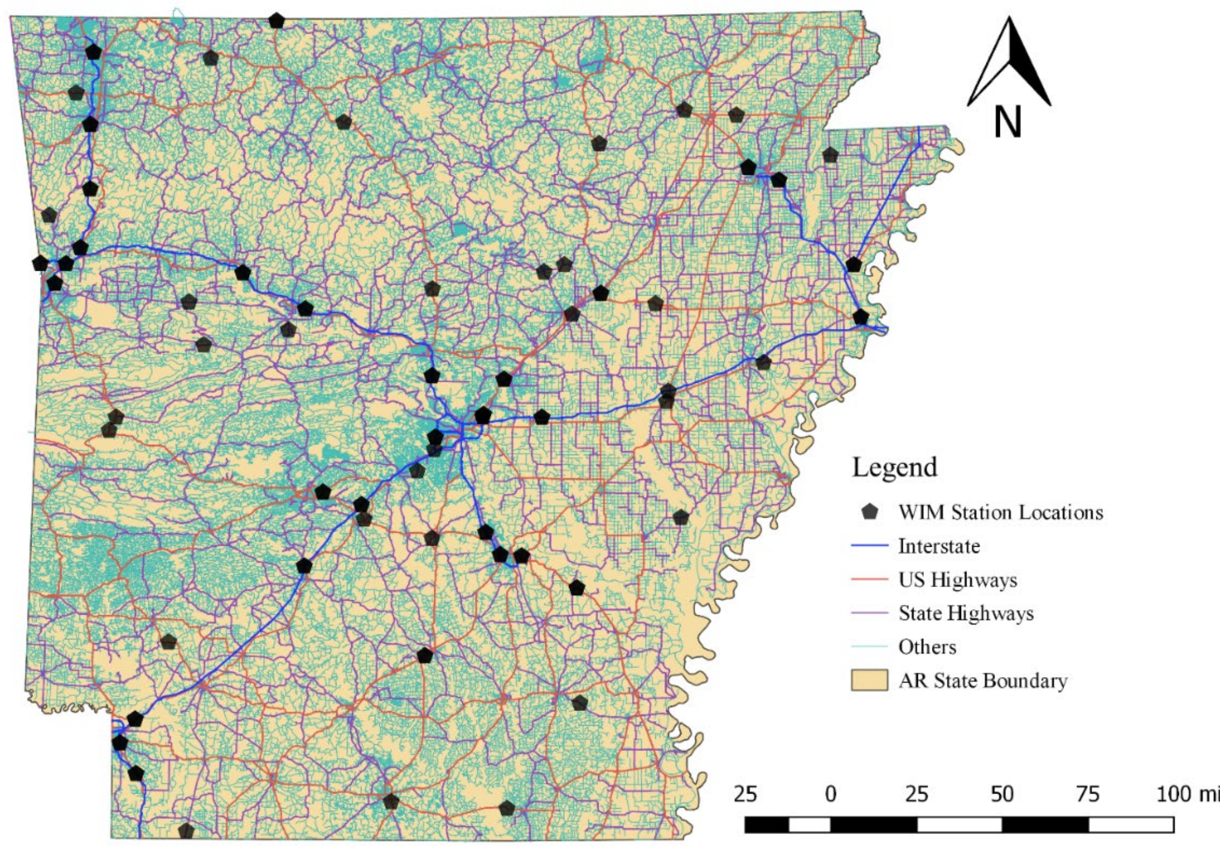


Figure 2. WIM Sites in Arkansas

WIM systems have been used since the 1950s to collect data for highway traffic monitoring, truck size and weight enforcement, infrastructure preservation and design, traffic safety, and transportation policy (*Nichols and Bullock, 2004; FHWA, 2016*). In most states, the main use of WIM data is for federal report requirements such as the Highway Performance Monitoring System (HPMS) database (NCHRP Synthesis 386). In pavement design, under the Mechanical Empirical Pavement Design Guide (MEPDG), truck weight data is needed at a high level of spatial resolution. Several reports detail the use of WIM data for pavement and bridge design:

- Long Term Pavement Performance Program (LTPP)
- Commercial Vehicle Information System and Network (CVISN) Program
- Mechanistic Empirical Pavement Design Guide (MEPDG) (NCHRP Project 1-37A)
- Protocols for Collecting and Using Traffic Data in Bridge Design (NCHRP Project 12-76)

Freight forecasting models such as the Arkansas Statewide Travel Demand Model (AR STDM) use truck weight and volume data from WIM for model calibration and validation (*Battelle, 2011; Alliance, 2015*). Emissions estimation tools use vehicle classification schemes based on truck weights to apply emissions rates (*CARB, 2016*). Truck weight data from WIM has also been used to assess the impact of truck loads on the performance and durability of bridges (*FHWA, 2016; Caltrans, 2016*). Beyond public agency uses, WIM sensors are an essential component of truck pre-clearance and bypass programs like PrePass™ and Drivewyze (*PrePass, 2017; Drivewyze, 2017*). These pre-clearance systems allow trucks that pass specific safety checks to bypass static enforcement scales. Every truck is weighed as it passes over the mainline WIM sensor located several hundred feet upstream of the static scale and an in-cab bypass signal is given if the truck measures within a specified tolerance of state weight limits.

Agencies using WIM data are aware that WIM data is prone to accuracy errors in speed, spacing, and weight measurements (*FHWA, 2001*). WIM measurement inaccuracies are the result of several possible factors: (1) vehicle dynamics such as speed, acceleration, tire condition, load, and body type; (2) site conditions such as pavement smoothness; (3) environmental factors such as temperature and precipitation (*Lee, 1998; NCHRP, 2008*). On the other hand, systematic errors, more commonly referred to as calibration errors, are persistent inaccuracies in which the true weight is either consistently over- or under-estimated. Figure 3 shows the effects of WIM sensor calibration error on calculated Equivalent Single Axle Loads (ESAL) values commonly used for pavement design (*FHWA, 1998*). In this example, calibration error is the difference between the weight measured by the WIM sensor and the true weight of the truck. In general, for every 1% of error that a scale is under-calibrated, ESALs are approximately 3% underestimated; for every 1% of error that a scale is over-calibrated, ESALs are approximately 4.5% overestimated *(FHWA, 1998)*. If a WIM station produced 10% over-estimated truck weights, this would result in 45% error in estimating damage to the pavement. Since ESALs and truck loads are critical inputs to pavement and bridge designs, a 45% over-estimate of ESAL would result in unnecessarily more costly pavements and bridges.

Through proper calibration procedures such as those outlined in American Society of Testing and Materials (ASTM) Standard E1318-09 (*ASTM, 2009*), systematic error can be addressed. However, calibration protocols outlined in ASTM E1318-09 require on-site testing which can be cost prohibitive. Many states have adopted lower cost protocols that do not require on-site testing. One such practice, called auto-calibration, uses software to calculate adjustment parameters based on presumed traffic characteristics. Weight data collected by the WIM sensor are then adjusted based on these calibration parameters. Although auto-calibration has significant value for budget-constrained state agencies, it has several drawbacks that limit its effectiveness.

The proposed work seeks to evaluate the auto-calibration practices used in the WIM systems in Arkansas. The following sections describe WIM system hardware and provide an overview of WIM calibration and auto-calibration practices.

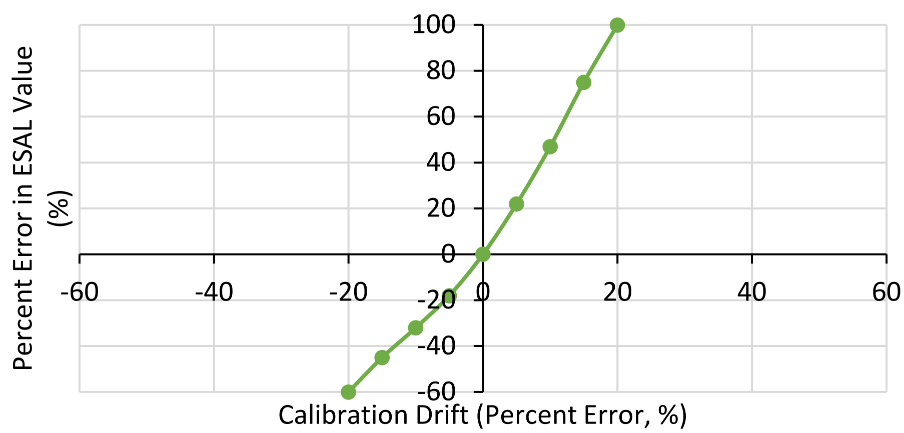


Figure 3. Effects of Sensor Calibration on ESAL Value (REF: FHWA, 1998)

PROJECT OBJECTIVES

The goal of this study was to develop a method to improve auto-calibration of WIM sites in Arkansas. The approach detailed in this report makes use of AVI systems, specifically truck GPS data, to estimate site and lane specific calibration factors. This research goal was carried out under three objectives which are summarized below.

Objective 1: Coordination with WIM Sensor Vendors

The research team worked with Peek Traffic, the WIM system provider used by ARDOT, to detail the current auto-calibration procedures and system characteristics. Peek was able to provide information on the auto-calibration algorithm and reference values used within the algorithm.

Objective 2: Field Test Calibration

Two field data collection efforts were carried out in March 2018 and 2019 to provide data for model development and validation. Video recordings taken at WIM sites were used to track trucks as they passed over the static enforcement scales. Each data collection consisted of two to three WIM sites and one static scale at an enforcement weigh station. The objective was to determine a WIM weight and a static weight for each truck in the sample for algorithm development and validation.

Objective 3: Automatic Vehicle Identification (AVI)-Based Calibration

The research team worked with Drivewyze, a truck pre-clearance program, to collect AVI data for the Arkansas WIM sites. Truck GPS data was provided to the research team monthly and used to develop and validate the proposed AVI-based auto-calibration algorithm. The algorithm has two main parts: (1) a matching algorithm to assign AVI trucks to the respective WIM record, and (2) an algorithm to calculate hourly, site, and lane specific calibration factors based on the GPS tracked trucks. Results of the proposed AVI-based auto-calibration algorithm, the current algorithm used by ARDOT, and an enhanced algorithm developed by MnDOT were compared using data from the two field experiments.

STRUCTURE OF THE REPORT

Following the Project Overview in Chapter 1, this report is organized as follows:

- Chapter 2 briefly describes WIM calibration challenges and calibration procedures,
- Chapter 3 presents the proposed AVI-based auto-calibration methodology,
- Chapter 4 summarizes the data collection efforts,
- Chapter 5 describes modifications to the existing ARDOT methodology and summarizes performance metrics resulting from applying the ARDOT and MnDOT methods to the collected data,
- Chapter 6 presents the performance metrics for the AVI-based auto-calibration method and compares performance to existing auto-calibration algorithms, and
- Chapter 7 summaries key findings, addresses limitations, and suggests avenues for future work.

CHAPTER 2: WEIGH-IN-MOTION (WIM) SYSTEMS

This chapter describes WIM system components and calibration procedures. Three general calibration procedures are discussed: on-site calibration using test trucks or traffic stream parameters, off-site calibration using data quality control procedures, and auto-calibration algorithms.

WEIGH-IN-MOTION (WIM) SYSTEM COMPONENTS

WIM system components include a controller, communications equipment, and roadway weight and presence sensors for all monitored lanes (*NCHRP Synthesis 386*). The ASTM E1318-09 classifies WIM systems according to their applications and functional performance:

- Type I systems are equipped with bending plates, load cell plates, or quartz piezoelectric sensors. Type I systems weigh the right- and left- hand side axles individually. Type I sensors can measure load data at speeds of 10 to 80 mph.
- Type II systems are equipped with ceramic or polymer piezoelectric sensors in various configurations. Piezoelectric sensors can be rated as Class 1 (for weighing) or Class 2 (for axle detection only). Type II systems weigh entire axles and can measure load data at speeds of 15 to 80 mph.
- Type III systems are used for load enforcement screening or sorting. They are installed on the approaches to truck inspection stations. Type III sensors can measure load data at speeds up to 10 mph.

A typical WIM station includes a weight sensor, e.g., piezoelectric or bending plate, straddled by Inductive Loop Detectors (ILD) (Figure 4). ILDs act as presence detectors while the piezoelectric or bending plate sensor measures axle or tire weight. The weight sensor measures the dynamic forces applied by the vehicle as it passes over the sensor. The type of weight sensor varies in application and desired accuracy.

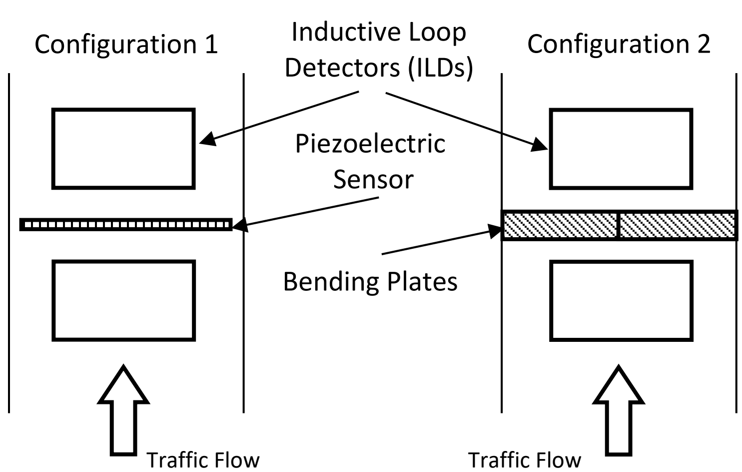


Figure 4. Typical WIM System Configuration

Bending plate sensors (Figure 5a) use strain gauges mounted to the underside of high-strength steel plates called weight pads. For traffic data collection, two weight pads are installed in a staggered configuration to allow speed estimation. This configuration also allows the left and right wheels to be measured individually. Load-cell WIM weight sensors use transducers to convert dynamic tire forces to proportional electrical signals. Load-cell installations are costly due to the requirements for a reinforced concrete vault to support the scales in the pavement, however they are the most accurate. Piezoelectric sensors translate the voltage generated by a tire force acting on a piezoelectric material to a dynamic load (Figure 5b).

In Arkansas, all of the WIM sites use piezoelectric sensors. Three types of piezoelectric sensors are used for WIM systems: piezoceramic, pizeopolymer, and piezoquartz. The main distinction lies in the data quality and cost of the sensors. Piezoquartz are the most expensive at $20,000 per installation compared to $9,000 for piezoceramic sensors, but provide the highest quality data (*FHWA, 2016*). In Arkansas, only one of the WIM sites is equipped with a piezoquartz sensor. The remainder of the sites contain piezoceramic sensors.

(a) Bending Plate Weight Sensors in a Staggered Configuration (Type I)

(b) Piezoelectric Weight Sensors (Type II)

Figure 5. Examples of Weight Sensor Types (REF: FHWA, 2016)

Accuracy standards are set by ASTM E1318-09 and are dependent of the sensor type, e.g. Type I, II, or III, as shown in Table 1. Accuracy is measured in terms of the probability that individual axle load measurement errors are within prescribed limits and is calculated according to Equation 1. Error can be calculated using individual axle loads, axle-group load (e.g. tandem or single axles), or GVW. ASTM standards are more stringent for Type I than Type II sensors, typically used for traffic monitoring. Type III, which are used for enforcement, are subject to the highest accuracy standards.

$$error = \frac{WIM - static}{static} x100\% \qquad \text{Equation 1}$$

Where
'error' is the percent error between the static and WIM measured loads
'static' is the load measurement of a vehicle as measured by a static scale
'WIM' is the load measurement of a vehicle as measured by a WIM sensor

Table 1. WIM System Accuracy Tolerances (REF: ASTM E1318-09)

Function	Tolerance for 95% Compliance		
	Type I	Type II	Type III
Wheel Load	±25%	NA	±20%
Axle Load	±20%	±30%	±15%
Axle-Group Load	±15%	±20%	±10%
Gross Vehicle Weight	±10%	±15%	±6%
Speed	±1 mph		
Axle-Spacing	±0.5 ft		

CALIBRATION AND AUTO-CALIBRATION PROCEDURES

WIM sensors can contain both random and systematic measurement errors due to vehicle dynamics over the sensor, site conditions, and environmental factors (*Nichols and Bullock, 2004; Papagiannakis et al., 2008; Prozzi and Hong, 2007*). Random errors are difficult to detect, and no procedures have been introduced to correct for random error. Systematic errors, sometimes called calibration errors, are consistent inaccuracies, which exhibit as over- or under- estimations of the true weights.

Once a site has been identified to have calibration issues, the system should be recalibrated. Calibration involves adjusting the system's calibration factors by setting the mean error measurements to zero. As mentioned in the previous section, error can be measured against the individual axle weights, the axle-group loads, or the GVW. The factors that minimize errors between known weights and WIM-measured weights are subsequently applied within WIM system calculations to correlate measurements taken by the WIM system to those taken at static scales. Although uncommon, it is ideal to compute multiple calibration factors for different speed bins and temperature ranges. The data for determining error are obtained from one of two on-site sources: (A) test trucks or (B) traffic stream trucks of known static weights. Alternatively, instead of on-site calibration, protocols have been developed for off-site data quality control procedures and auto-calibration methods. This project, e.g. TRC 1801, focused on these latter two approaches as they provide more cost-effective means of calibrating the WIM systems.

Calibration Using Test Trucks

The ASTM E1318-09 standard describes on-site calibration methods using test trucks of known static weight and dimensions. Under the ASTM standard, two test trucks are required to make multiple runs over the WIM system sensors at prescribed speeds in each lane. Most states use a FHWA Class 5 (two-axle, single-unit truck) and FHWA Class 9 (five-axle tractor-trailer) as test trucks. These trucks should have a suspension type that is typical of the trucks operating at that site. Weight and axle spacing of the trucks should be taken a minimum of three times at static scales prior to site testing. The ASTM standard calls for routine calibration at least annually for sensors over one year in operation (*NCHRP Synthesis 386*). According to the survey findings of *NCHRP Synthesis 386*, 22 of the 34 DOTs that responded to the survey reported use of test trucks as part of their calibration protocol when the system is used for data collection.

The *Traffic Monitoring Guide (TMG)* enhances procedures outlined in ASTM E1318-09. Enhancements include (1) testing more than two test vehicles, (2) testing at various speeds, (3) testing under varying environmental conditions (e.g. temperatures), and (4) performing data quality control (QC) procedures. Basic QC procedures involve monitoring traffic stream data over time. Traffic stream parameters of interest are: (1) front axle weights of five-axle tractor-trailers, (2) GVW distribution of five-axle tractor-trailers, (3) spacing of tandem drive axles, and (4) vehicle classification counts. The *Long Term Pavement Performance Program (LTPP)* protocol requires the following in addition to the ASTM requirements: (1) one of the test trucks must be a five-axle tractor trailer with air ride suspension and weigh between 72 and 80 kips, (2) the second test truck must be of a different configuration or different suspension type and cannot be a three or four axle truck, (3) both test trucks must have conventional tread patterns, and (4) at least 20 passes of each truck must be made at highway speeds. Like the TMG, the LTPP protocol also recommends ongoing monitoring using data QC procedures such as calibrating using the GVW distribution pattern (*LTPP, 1998*).

The *Long Term Pavement Performance Specific Pavement Study Traffic Data Collection Pooled Fund Study* refined the LTPP protocol by adding the following: (1) a minimum of two test trucks, (2) one test truck must be a five-axle tractor-trailer with GVW between 76 and 80 kips and the second truck must be another five-axle tractor-trailer with GVW between 60 and 64 kips and the second, (3) the heavier test truck must have air ride suspension for both tractor and trailer tandem axles, and (4) ASTM tolerance limits must be met by the entire data set and also for subsets of the dataset based on temperature and speeds.

Calibration Using Traffic Stream Vehicles

On-site calibration using traffic stream vehicles involves comparing measurements of vehicles passing the WIM system to the measurements taken at a static scale site for that same vehicle. This requires tracking vehicles as they pass both the WIM and static scale sites. Tracking can be performed manually using video recording or automatically using license plate readers or other Automated Vehicle Identification (AVI) technologies. According to the survey findings of *NCHRP Synthesis 386*, seven of the 34 DOTs that responded to the survey reported use of traffic stream data as part of their calibration protocol when the system is used for data collection.

For on-site testing using traffic stream vehicles, either a fixed number of trucks must be sampled, or a fixed time interval is used. The results of *NCHRP Synthesis 386* show that when a fixed number of trucks is used, the average sample size was 40 vehicles. The *Long Term Pavement Performance Program (LTPP)* protocol for calibrating WIM located near enforcement scales is based on direct comparisons of WIM and static scale weights for a test set of at least 150 vehicles. Due to the sample size requirements, this method is usually only applicable to WIM sites located near static scale sites. When using a fixed time interval, the majority of state agencies used a time interval between one and four hours (*NCHRP Synthesis 386*).

NCHRP Project 3-39(02) On-Site Evaluation and Calibration of Weigh-in-Motion Systems provides guidelines for on-site calibration using traffic stream vehicles equipped with AVI systems, specifically license place readers. This method is based on comparing static weight data from enforcement stations for a subset of vehicles which have been identified though AVI to have also driven over mainline WIM systems. Similarly, the Montana DOT developed a traffic stream calibration procedure using AVI data from the truck bypass program, PrePass^{TM} (*NCHRP Synthesis 386*). In this procedure, PrePass^{TM} static axle load data are obtained for 25 five-axle tractor-trailers for each of the lanes being monitored. WIM measurements are collected for the same 25 trucks by manually (visually) matching the truck from the WIM sensor and static scale.

Calibration Using Data Quality Control Procedures

WIM data QC procedures involve analysis of historical WIM data to determine calibration issues. Several methods have been developed to assess whether a WIM sensor may contain errors due to calibration (*Dhalin, 1992; Hand et al, 1995*). A common method of data QC to compare the measurements of traffic stream vehicles to some reference value. A typical reference values include the front axle weight (FAW) of five-axle semi-tractor trailers. This value is restricted to a very narrow range around 10 kips. Therefore, if a WIM system is reporting FAWs of around 11 kips, it is likely that the scale has calibration problems. Similarly, the GVW distribution can be used as a reference measure. The GVW distribution of five-axle tractor-trailers typically follow a distinct bimodal distribution due to the presence of loaded and unloaded trucks (*NCHRP Synthesis 386*). If the peaks of the GVW distribution are shifted, it is likely that the scale has calibration problems. Specific examples of data QC checks are provided in Table 2.

Several state agencies have adopted data QC procedures to support or replace on-site calibration protocols (*NCHRP Synthesis 386*). In many cases, data QC is used to monitor when a site needs to be calibrated using on-site testing procedures. According to the survey findings of *NCHRP Synthesis 386*, 20 of the 34 DOTs that responded to the survey reported use of data quality control (QC) as part of their calibration protocol when the system is used for data collection.

Table 2. Examples of Data QC Elements

Document	Data QC Elements
Traffic Monitoring Guide (TMG)	• FAW of five-axle tractor-trailers • GVW distribution of five-axle tractor-trailers • Spacing between tandem drive axles • Vehicle classification counts
NCHRP Report 509: Equipment for Collecting Traffic Load Data	• Distribution of loaded and unloaded GVW peaks for five-axle tractor-trailers • Consistency of mean FAW for loaded five-axle tractor-trailers • Consistency of percentage of weekday five-axle tractor-trailers • Changes in percentage of unclassified vehicles • Increases in equipment counting errors • Consistency in load-relative magnitudes between right and left wheel path weighing sensors • Consistency of spacing between tandem drive axles • Total number of vehicles within expected load ranges • Changes to time-of-day traffic patterns • Changes in hourly data volumes

Auto-Calibration Algorithms

Auto-calibration is a mechanism built into WIM software effecting automatic calibration adjustments when certain measurements fall outside prescribed limits (*NCHRP Synthesis 386*). The ASTM E1318 standard describes auto-calibration as the ability for WIM software to automatically adjust the calibration factor (ASTM, 2017). Auto-calibration borrows the same general principles from off-site data QC protocols in that traffic stream measurements are compared to reference values to determine calibration status. The main difference is that auto-calibration routinely updates calibration factors and adjusts output measurements whereas data QC is used only to monitor calibration status. The survey conducted in *NCHRP Synthesis 386* revealed that auto-calibration is used primarily for Type II systems but did not detail how many states rely on auto-calibration methods or how those methods are implemented.

Generally, state-of-the-practice auto-calibration algorithms vary in their calibration factor adjustment procedure. For instance, the FAW can be used as a target measurement such that the calibration factor is adjusted to ensure FAWs are within range of the target value. Other methods may use GVW as the target measure. The Minnesota DOT, for instance, developed a procedure which compares the peaks of the GVW distribution of five-axle tractor-trailers. The loaded peak should be between 74 and 78 kips and the unloaded peak should be between 28 and 30 kips. If these data are within range, then an auto-calibration factor is estimated based on the FAW of five-axle tractor-trailers for each GVW group. The auto-calibration algorithm determines differences between the average FAW for each GVW class and calculates weighted adjustment factors. The adjustment factors are weighted by the number of sampled vehicles used to compute their value. Florida DOT, Texas DOT, and Caltrans monitor the average GVW, while Indiana DOT and MnDOT monitor both the GVW and the FAW (*Papagiannakis and Quinley 2008*).

In addition to variability in the reference metric used, there is also variability in the target value of the metric, target vehicle type, the sample size used to update the calibration parameter, and the time period over which to routinely perform auto-calibration. The auto-calibration algorithm implemented by Peek Traffic (the vendor of WIM system equipment in Arkansas), for example, allows the following settings to be defined at each site and for each lane: (1) type of vehicle to use for reference, (2) which axle to use for reference, (3) target value for reference axle, and (4) number of vehicles to include in the sample used to calculate parameter updates. For a four-lane site (two lanes in each direction), the auto-calibration may be set as shown in Table 3.

Auto-calibration can be specified for different volumes of the same vehicle type over all the lanes or for higher volumes of different types of vehicles. An issue facing several of the sites in Arkansas is that of low volumes of trucks to use when sampling to adjust the calibration factor. For instance, to update the calibration parameter each hour (recommended to compensate for temperature fluctuations) a large enough sample of five-axle tractor-trailers is needed to ensure consistency. While the number of sampled trucks can range from 1 to 100, the majority of state DOTs use 50 vehicles (*NCHRP Synthesis 386*). In some locations in Arkansas, within a given hour, truck volumes can be too low to provide consistent data to reliably update the calibration factors. In these cases, a more common vehicle, such as a passenger car, can be used as the reference vehicle to achieve a larger sample size.

Another form of auto-calibration attempts to compensate for the effects of temperature changes on WIM sensor measurements (*Burnos, 2008*). As pavements heat and cool due to ambient weather conditions, relative variation in the weight data can be observed (Figure 6). To account for this variation, a temperature compensation curve is defined and is referenced to update the calibration parameters as temperatures shift throughout the day. A temperature probe at the site is required if auto-calibration is supported or replaced by a temperature compensation curve. In communications with Peek Traffic, the use of temperature probes and compensation curves have been shown to have poor performance due to the difficulty in maintaining temperature probes. Also, when only one temperature probe is installed at a site, it may not accurately reflect the temperature conditions across all travel lanes. As a result, the WIM systems in Arkansas do not use temperature probes and instead rely on auto-calibration based on comparing measured to reference FAWs .

NCHRP Report 509: Equipment for Collecting Traffic Load Data states that auto-calibration should not be used to replace data QC procedures. However, it seems to be common practice to rely on the auto-calibration procedures and parameters set by the WIM system vendors. Often, instead of determining site-specific auto-calibration settings like reference vehicle type or volume, state agencies use default values based on state or regional averages. This can be problematic because there is variability in the target values not only by site, but also across lanes at the same site. If auto-calibration values are not correctly defined, auto-calibration can actually force scales to become uncalibrated (*LTPP, 1998*). Thus, for auto-calibration to be effective, a state must determine what procedure should be used, how to determine site-specific parameters, and how many test trucks cross the sensor during the calibration process to allow for proper calibration (*LTPP, 1998*).

Table 3. Auto-Calibration Settings for a Four Lane WIM Site
(REF: Personal Correspondence with Peek Traffic)

Lane	(1) Type of Vehicle	(2) Reference Axle	(3) Target Value	(4) No. of Vehicles
1 (outer lane)	Five-axle tractor-trailers	Front axle	10.2 kips	50
2 (inner lane)	Passenger cars	Front axle	2.2 kips	50
3 (inner lane)	Passenger cars	Front axle	2.3 kips	75
4 (outer lane)	Five-axle Tractor-trailers	Front axle	10.8 kips	25

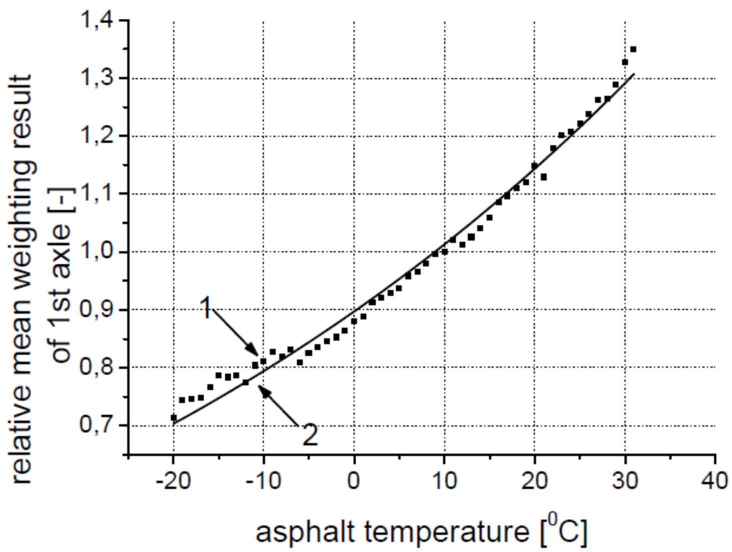


Figure 6. Relative Variations in Weight at Varying Pavement Temperatures (REF: Burnos, 2008)

CHAPTER 3: AVI-BASED AUTO-CALIBRATION

This chapter describes the AVI-based auto-calibration method developed for this project. Following a brief overview of the AVI-based approach, a description of the AVI data obtained from Drivewyze and WIM Per Vehicle Record (PVR) data are provided. The two main components of the AVI-based auto-calibration are discussed: (1) a matching algorithm to assign AVI truck records to WIM PVR records, and (2) an algorithm to estimate hourly, lane, and site specific calibration factors that uses AVI data to track trucks across multiple WIM sensors.

OVERVIEW OF AVI-BASED APPROACH

Instead of using reference values to calculate calibration factors, it is possible to track trucks across WIM sites and compare their weights to generate calibration factors. Automatic Vehicle Identification (AVI) is a method of using computers to identify and track vehicles. Common examples of AVI technologies include vision-based systems like cameras, license plate readers, and Radio Frequency Identification (RFID). AVI is used for traffic enforcement in border and customs checkpoints, electronic toll collection, intersection violations, and for transportation analysis (*Ozbay et al., 2007*). AVI has the capability of recording path flow information and tracking a vehicle's trip from origin to destination. In addition, AVI methods may have the capability to capture a larger sample than traditional surveys and traffic counts if desired.

The two most common methods using visual recordings in AVI are tag-based, and license plate-based recognition. Other forms of AVI are cellular phone based (*Dixon and Rilett, 2002*), GPS-based (*Hyun K., Tok A., Ritchie S. G., 2017*), transponder number (*Nichols & Cetin, 2015*), and inductive signature (*Jeng & Chu, 2015 & Hyun K., Tok A., Ritchie S. G. 2017*). Inductive loops and transponders create unique records using inductive signatures and transponder numbers, uniquely identifying trucks with their respective time stamps when they travel over these sensors. These trucks may then be identified in other sensor sites of the network with their unique IDs or signatures. Hyun K., Tok A. and Ritchie S. G. (*2017*) and Jeng and Chu (*2015*) used ILDs along with WIM sensors to collect vehicle attributes of shared trucks for their studies.

The type of AVI technology applied depends on traffic sensor, infrastructure and equipment available, and budget. For example, California uses advanced ILDs to track trucks across multiple scales while Oregon uses transponder identification numbers (*Cetin and Nichols, 2015*). The state of Arkansas does not have ILD available and automatic license plate recognition is not permitted. Thus, a suitable form of AVI for Arkansas is to use passive GPS tracking. There are many private data providers that collect and share GPS tracking data with public agencies for various applications including INRIX, HERE, and Drivewyze. Each of these companies collect data from only a small percentage of the total traffic. However, this type of probe vehicle data is often enough to estimate performance measures like travel time and speed.

For this study, the GPS AVI data was provided by a Drivewyze, a private company that operates as an app providing pre-clearance for trucks through weigh and enforcement stations. This AVI data was used to derive calibration factors (CFs) by comparing the weight of the same truck as measured by different WIM sites (Figure 7).

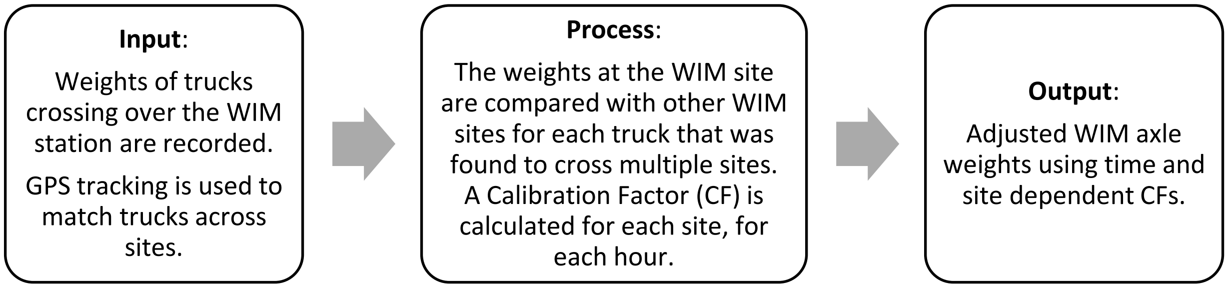


Figure 7. Overview of AVI-Based Auto-Calibration Methodology

DESCRIPTION OF ALGORITHM INPUT DATA

The main data used by the AVI-based auto-calibration algorithm include (1) truck AVI data from Drivewyze and (2) WIM PVR data. Each source is briefly described in this section.

Truck AVI Data

Each AVI truck record contained a unique ID that remains constant across traveled sites, time of day, and day of week. AVI records can be represented as an array for each unique truck d_k , with the origin WIM site and time stamp at which the truck crossed the site, S_w, and $t_{k,s}$, and the next term $S_{w'}, t_{k,s'}$ corresponding to site and timestamp of the next WIM site crossed:

d_k = [$(S_w , t_{k,s})$, $(S_{w'} , t_{k,s'})$…]

Where:

d_k is the AVI record, where *k* denotes the unique id.

S_w is the ID of the WIM site that the truck traveled over.

$t_{k,s}$ is the timestamp of truck *k* at site *s*.

WIM PVR Data

WIM Per Vehicle Record (PVR) files include data of each vehicle detected by the WIM sensors. PVR data contains a record number, traveled lane, direction of travel, speed, vehicle class (according to the FHWA 13 class scheme), axle and gross vehicle weights, and axle spacing. Note that the WIM sensor may not detect all vehicles and may also produce duplicate records (e.g., when vehicles change lanes over the sensors). WIM records for each vehicle may be represented in an array for vehicle parameters of interest with WIM site D_s with vehicle parameter, W_i:

$$D_s = [W_{s,1}, W_{s,2}, W_{s,3} \dots W_{s,n}]$$

Where:

D_s is WIM site s.

$W_{s,i}$ is a vehicle parameter data record at s for vehicle parameter I, e.g. axle weight, axle spacing, vehicle length, etc.

AVI-BASED AUTO-CALIBRATION METHODOLOGY

The WIM auto-calibration model consisted of two parts: (1) a truck-matching algorithm and (2) an auto-calibration algorithm. Each is described in the following sections.

Truck-Matching Algorithm

As part of data pre-processing, we manually determined the time offsets between the AVI and WIM records and matched AVI truck records to WIM records using video data. In this section, we describe the algorithm, called "Truck-Matching", used to perform the same task automatically. The Truck-Matching algorithm followed three steps (Figure 8): (1) Time Offset Calculation, (2) Match Filtering, and (3) Data Pairing. Each is described below.

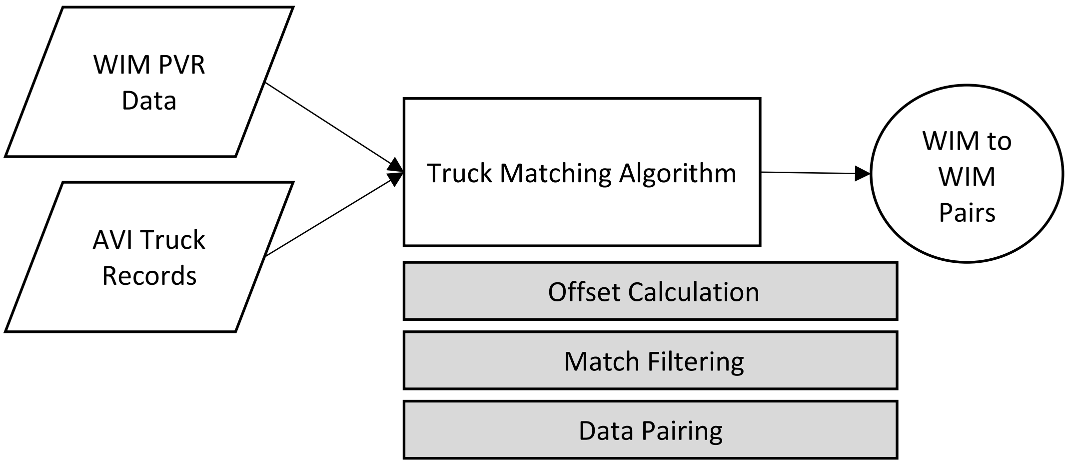


Figure 8. Truck-Matching Inputs, Processes, and Outputs

Offset Calculation

Since the GPS data provides unique identifiers (IDs) for each truck record, trucks can be tracked across WIM sites solely based on their ID (e.g., advanced truck re-identification was not necessary). However, it was necessary to find the time offset between the AVI records and the WIM PVRs via an automated process. The algorithm steps were as follows:

1. Query the AVI truck records for a given day and year to produce a list of stations and trucks crossing those stations.
2. For each AVI truck, query WIM PVRs within a specified time window around the AVI truck timestamp (the optimal time window was found to be three minutes, or 180 seconds).
3. For each WIM PVR returned by the query, calculate the time offsets between the PVRs and the AVI truck timestamp.
4. Find the 'mode' (e.g., the most frequently occurring value) among all AVI trucks and PVR offsets.
5. If no mode exists or there are multiple modes within the initial time window, then widen the window and repeat Steps 3-6. The time window was widened by 10 seconds each iteration and allowed to increase to five minutes. Note that no mode exists if all offset values occur only once.
6. When a mode is found, assign it as the offset for the station.

Match Filtering

AVI records of trucks tracked across multiple WIM sites were subjected to a match filter**.** A travel time filter was applied first to reduce the possibility of truck weight differences due to pick-up and deliveries between sites. A maximum travel time threshold ensured the same truck with the same cargo and trailer was found. The maximum threshold was based on observed travel time distributions among WIM sites. A minimum travel time threshold controlled for recording errors inherent in the GPS data. A temporal window of one day was applied to the AVI-matched trucks (d_k) to filter out potential variation in weights due to drop-off, pick-up, cargo, and trailer changes.

Data Pairing

Data pairing between the AVI and WIM records was necessary because there are typically several candidates within the time window of each AVI record when matching AVI to WIM records, even after filtering out records outside the match filter. There are many more WIM records than AVI records, e.g. the AVI data represents less than 10% of the total truck population. Therefore, the set of candidate matches was reduced by examining the timestamps and axle spacing recorded by the WIM. The objective of data pairing was to assign each AVI record uniquely to a WIM record. The algorithm was carried out in two steps: (1) identify candidate WIM records for each AVI record, and (2) assign a unique WIM record to each AVI record.

Step 1. Identify Candidate WIM Records. A time buffer, Δ, around the AVI timestamp ($t_{k,s} + \Delta$) for each site of interest, for each truck d_k was established based on time offsets and was used to obtain candidate WIM records. The set of candidate matches for truck d_k was:

$C(t_{k,s}) = [\ W((t+x)-\Delta)_{s,i}, \ldots, W(t+x)_{s,i,+n}, \ldots W((t+x)+\Delta)_{s,m}]$

Where,

$C(t_{k,s})$ is the set of WIM records corresponding to the time stamp of an AVI truck at site *s* at time $t_{k,s}$.

$W(t)_{s,i}$ is the WIM record of the vehicle at site *s*, timestamp *t* such that the set of candidates is within a buffer, Δ, around $t_{k,s}$ (t-Δ, t, t+Δ), i = 1… m.

Step 2. Assign WIM Record to AVI Record. Finally, for an AVI truck d_k crossing stations s and s', the set of candidate WIM records were filtered to find a unique match such that the timestamp and vehicle parameters from each corresponding WIM record were minimized. A matrix $D_{s,s'}$ representing all pairwise combinations of WIM to WIM pairs ($W_{s,i}$ to $W_{s',j}$) contained the candidate sets $C(t_{k,s})$ and $C(t_{k,s'})$ for sites s and s'. The sum of absolute differences of vehicle parameters was used as a metric to find the unique match. The WIM records that produce the minimum difference, e.g., $argmin(D_{s,s'}, \{W_{s,i}, W_{s',j}\})$, and were within all temporal constrains are selected as a unique match. The vehicle parameters compared in the study were inter-axle spacing values.

$$D_{s,s'} = \begin{bmatrix} |W_{s,1} - W_{s',1}| & \ldots & |W_{s,1} - W_{s',m}| \\ \ldots & \ldots & \ldots \\ |W_{s,n} - W_{s',1}| & \ldots & |W_{s,n} - W_{s',m}| \end{bmatrix}$$

Where,

$D_{s,s'}$ is the matrix of differences between all WIM records at sites s and s' pertaining to $t_{k,s}$ and $t_{k,s'}$

$W_{s,i}$ is WIM record *i* at site *s* contained in $C(t_{k,s}) = [\ W(t-\Delta)_{s,i}, \ldots, W(t)_{s,i,+n}, \ldots W(t+\Delta)_{s,n}]$, i = 1…n

$W_{s',j}$ is WIM record *j* at site s' contained in $C(t_{k,s'}) = [\ W(t-\Delta)_{s',j}, \ldots, W(t)_{s',j,+n}, \ldots W(t+\Delta)_{s',m}]$, j = 1… m

$\left|W_{s,n} - W_{s',m}\right| = \sum_{p=1}^{P}\left|y_{p,s} - y_{p,s'}\right|$, the sum of the absolute differences between vehicle parameters, y, for sites s and s'.

Auto-Calibration Algorithm

The output of the Truck-Matching Algorithm was WIM records paired to each AVI truck record. This data was then used to compare weights of the same vehicle at different WIM sites as a form of auto-calibration. This auto-calibration process requires calibration factors be determined based on the weight of the same truck measured at different WIM sites (Figure 10).

The proposed auto-calibration method computed hourly, site-specific calibration factors. Calibration factors were calculated as the ratio of the measured FAW of FHWA Class 9 tractor-trailers to a likely FAW based on the set of FAW measurements for the same truck across multiple sites. The algorithm first calculated the deviation among *front axle weights* (also referred to as FAW) for the same AVI truck. Pairwise differences between *front axle weights* were calculated and the differences were measured against a predefined threshold, δ_S. If percent of sites with *front axle weights* above δ_S was greater than a predefined threshold on the number of sites in agreement, P_S, the calibration factor for each site for the specified hour was set to 1.0, e.g., all sites were in agreement on the *front axle weight* and thus are in calibration. Otherwise, for sites with *front axle weights* in "disagreement", a calibration factor was calculated as follows.

First, high-volume sites were used to determine the *likely weight* (ω_L) of the truck's front axle. The *likely weight* was found through a cluster analysis in which *front axle weights* corresponding to the AVI truck across multiple WIM sites were compared to find a common measurement (Figure 9). The inputs to clustering were the *front axle weight* and the GVW.

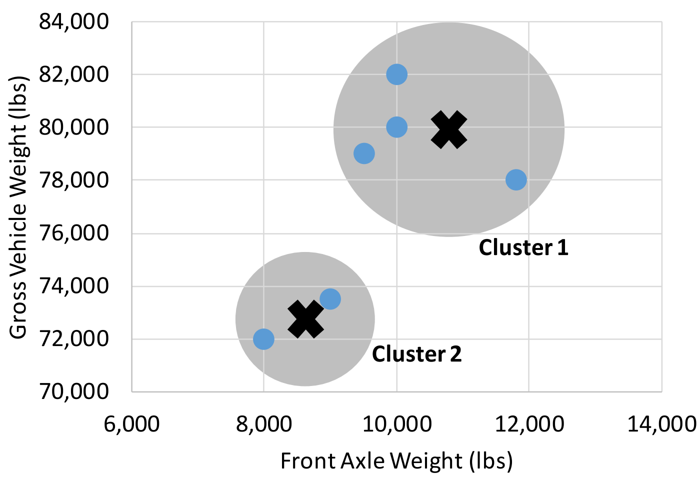


Figure 9. Example of Cluster Analysis for AVI Auto-Calibration

Then, the *likely weight* was compared to a reference weight (W_R), e.g., the same used in the traditional ARDOT auto-calibration method of 10 kips, to assess its reasonableness. The *likely weight* was used to compute the calibration factor if it fell within a certain deviation, δ_W, of the reference weight; otherwise the reference weight was used. The *likely weight* found through the cluster analysis of the high-volume sites were used to compute the calibration factors for the high- and low-volume sites. Thus, each truck, *k*, produced a calibration factor corresponding to each site, *l*, during the corresponding hour, *h*, it was detected at the site:

$$CF_{k,h,i} = \frac{W_{k,h,i}}{\omega_L} \qquad \textbf{Equation 2}$$

Where,

$W_{k,h,i}$ is the WIM recorded weight of the truck *k* at site *i* in hour h

Next, calibration factors for each truck ($CF_{k,i,h}$) were averaged for each site to determine the average calibration factor, $\widehat{CF}_{i,h}$, for each site for each hour.

$$CF_{h,i} = \sum_{k}^{K} W_{k,h,i} / \omega_L \qquad \textbf{Equation 3}$$

Finally, the adjusted weights for every truck at each WIM site were computed as:

$$\widehat{W}_{k,i} = CF_{h,i} \times W_{k,h,i} \qquad \textbf{Equation 4}$$

The key distinction between traditional and the proposed auto-calibration was the method to incorporate reference axle weights to compute calibration factors. In traditional auto-calibration algorithms, WIM-measured weights are compared to a predetermined, non-changing reference weight, such as a reference FAW, to compute a calibration factor. In the proposed auto-calibration method the use of a reference weight was replaced by a *likely weight*, ω_L, defined from the AVI-WIM pairs.

Rather than averaging, choosing the mode, or using a median *front axle weight* from the set of WIM *front axle weights* for a truck, the clustering approach was adopted to ensure that the *likely weight* reflected the majority among all measurements. The clustering approach also allowed the ability to detect weight discrepancies caused by trucks making pick-ups or deliveries between WIM sites resulting in different FAW and/or GVWs. Moreover, the proposed auto-calibration algorithm distinguished between high- and low-volume sites when estimating the *likely weight.* This was an important distinction because FAW measurements taken at high-volume WIM sites, e.g., sites with more than 50 FHWA Class 9 trucks per hour, tended to be more accurate than those collected at low-volume sites as more values could be used for clustering. Differences in accuracy can be attributed to the increased auto-calibration frequency at high-volume sites that, in turn, tracks with temperature changes. Ambient and pavement temperatures significantly affect WIM piezo-sensor accuracy.

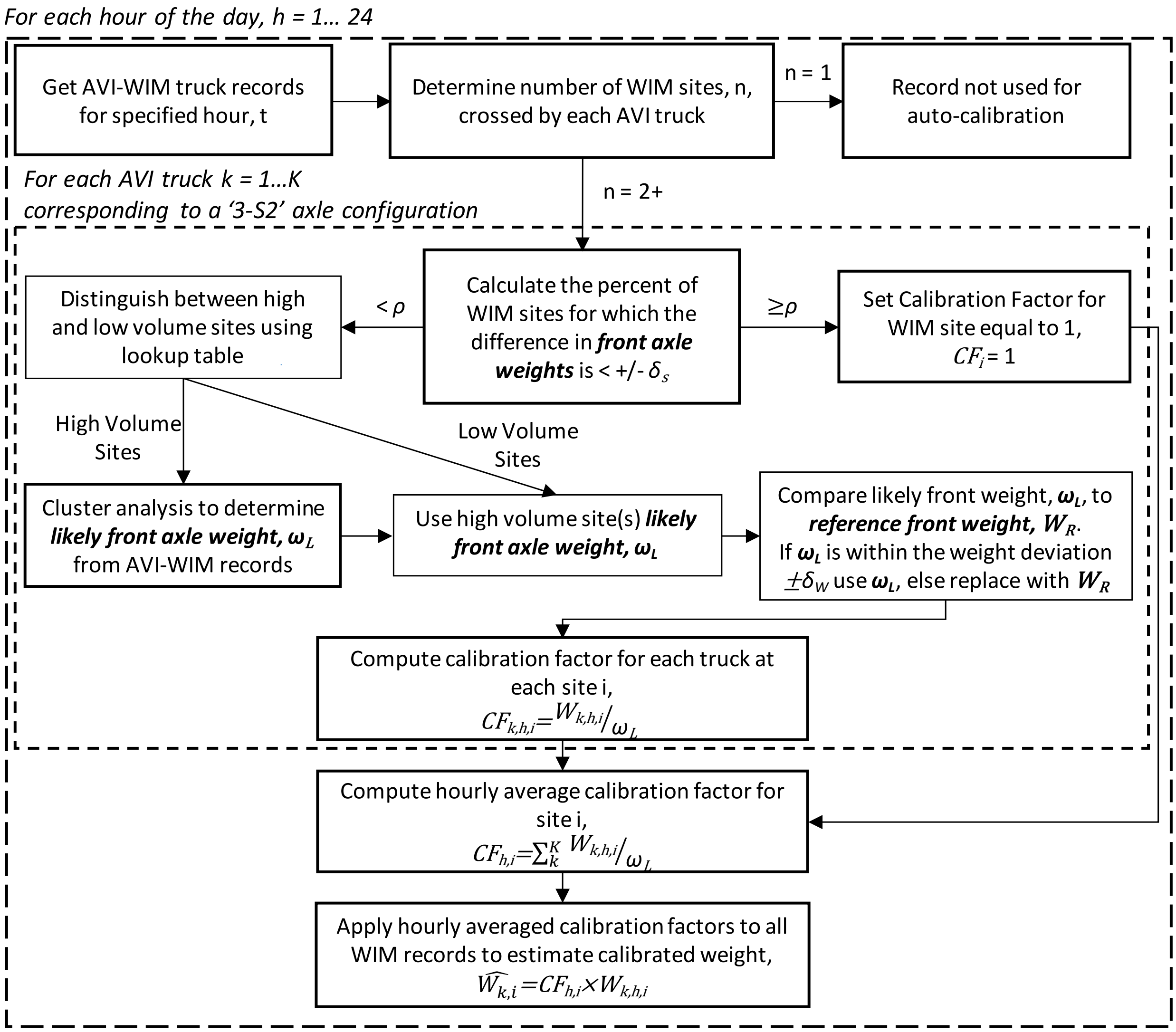


Figure 10. AVI Auto-Calibration Procedure

CHAPTER 4: DATA COLLECTION

This chapter reviews the data collection efforts carried out to provide data for model development and validation. A description of the site selection process, field data collection procedures, and summary of collected samples is presented in this chapter. Equations for the performance metrics to compare auto-calibration methods are also provided.

Data include WIM PVR, AVI truck-tracking data, and video recordings at select sites. Still images taken at selected static enforcement sites and video footage of trucks crossing selected WIM sites were recorded and used for model development and validation.

SITE SELECTION

Static scale and WIM sites used for data collection were selected based on an analysis of the common truck paths, e.g., 'shared traffic', observed in Arkansas using historical AVI data from Drivewyze. Traffic flows for the month of March 2018 among WIM sites show a large portion of truck volumes from Texarkana to Malvern along I-30 and from Little Rock to West Memphis along I-40 (Figure 11).

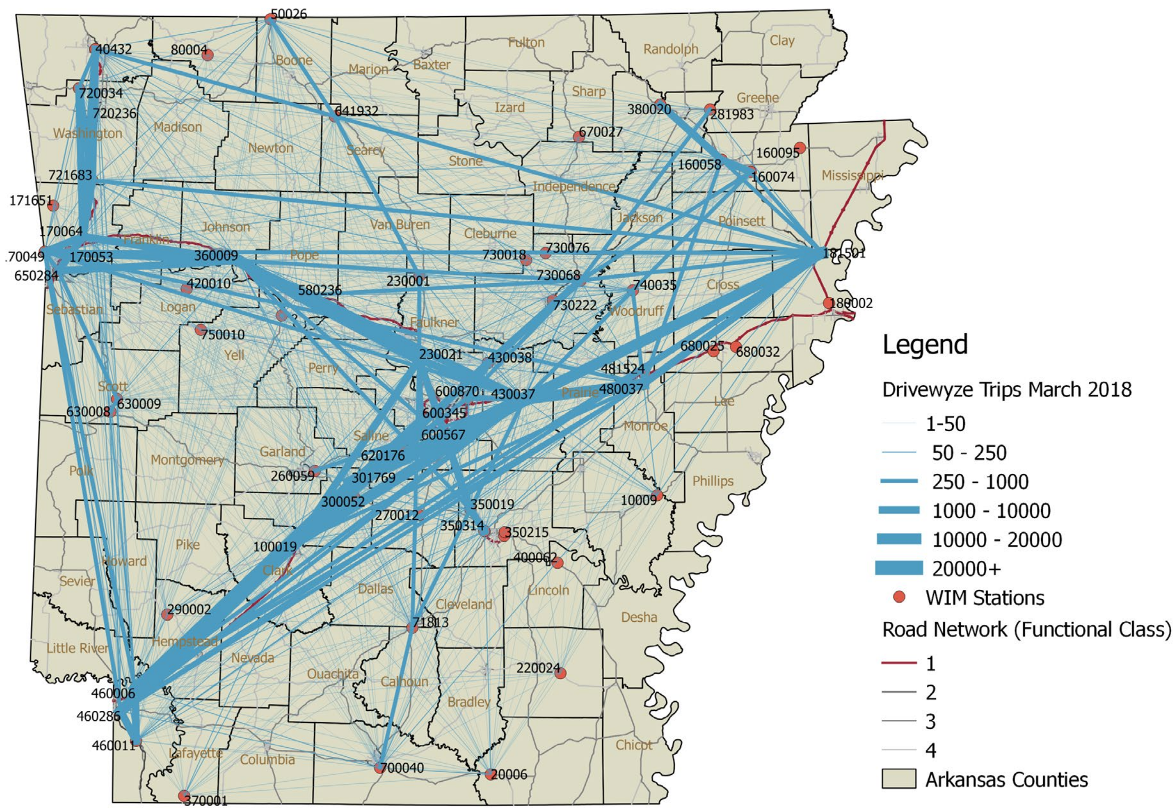


Figure 11. Drivewyze Truck Traffic Patterns for March 2018

For the first round of data collection on March 2018, the selected static scale was the Alma Eastbound weigh station along I-40 and the selected WIM sites were Lamar and Lonoke along I-40 and Bald Knob along Highway 67. In this case, Lamar (WIM 360009) and Lonoke (WIM 430037) were higher volume sites with a significant proportion of shared truck volume, and Bald Knob (WIM 730068) was selected as a low-volume site. For the second round of data collection on March 2019 the WIM sites selected where Glen Rose (WIM 301769) and Arkadelphia (WIM 100019) on I-30 and Texarkana (WIM 460286) on I-49.

The selected static scale for this instance was Westbound Hope on I-30 located between the Glen Rose and Arkadelphia WIM sites. Data collection sites are highlighted in Figure 12.

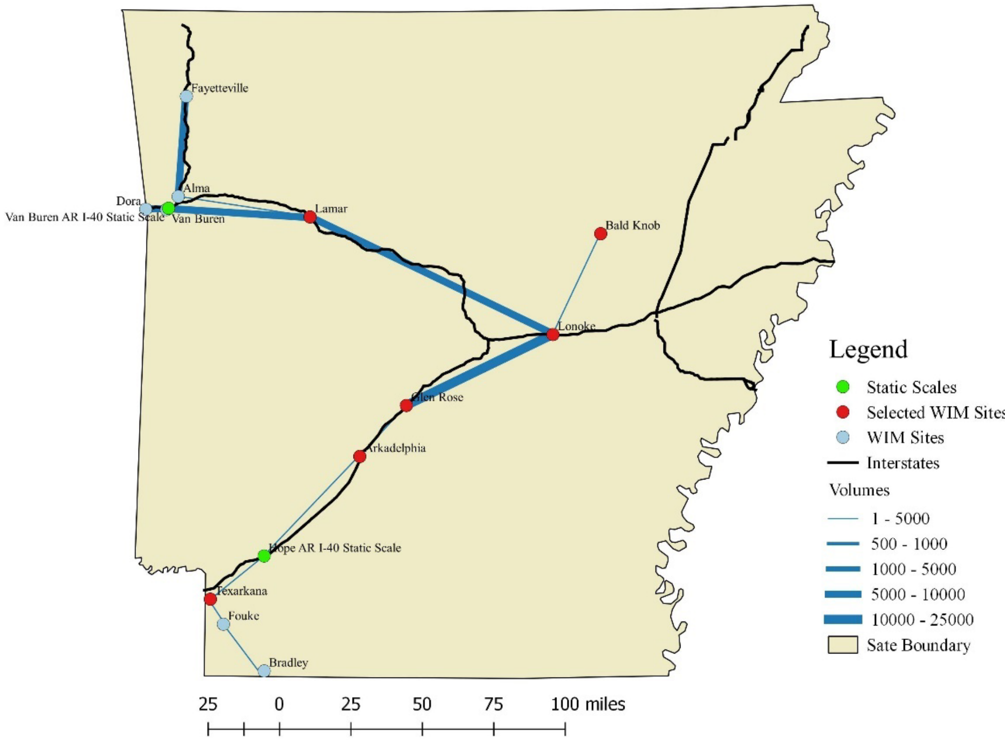
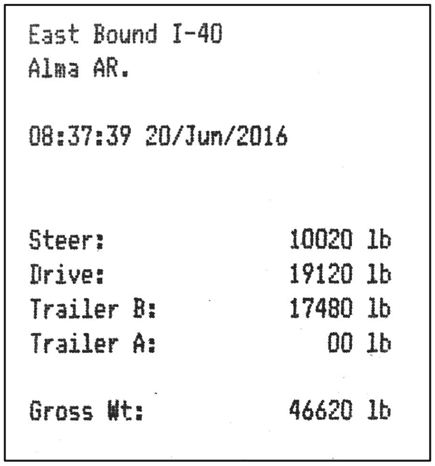


Figure 12. March 2018 and 2019 Data Collection Sites

FIELD DATA COLLECTION PROCEDURES

Data collection was performed on March 15th, 2018 and March 19th, 2019. In the 2018 data collection, raw or "un-calibrated" weights at WIM sites were recorded because auto-calibration was turned off at the selected sites. In the 2019 data collection, auto-calibration was running for all the selected sites, thus the PVR records for the 2019 sites have calibrated axle and gross vehicle weights per the ARDOT calibration method. The static weight data collection and visual data recordings procedure in 2018 was repeated in 2019 where cameras were set up at each static and WIM site to record passing trucks. Static weights were collected from trucks that stopped at the static enforcement scales using weight receipts that recorded axle and Gross Vehicle Weights (Figure 13). The FAW was categorized as 'Steer', the second and third axles where weighed as 'Drive', and third and fourth axle were weighed as 'Trailer', further distinguished as 'TrailerA' and 'TrailerB' if multiple trailers were present. Still images and video recordings were collected at the static sites for traffic that exited the interstate into the weigh station. Video was recorded for all the selected WIM sites for the study. GPS records were gathered for all WIM sites in the Arkansas network. Figure 14 and Figure 15 show examples of still images recorded at the weight enforcement station.

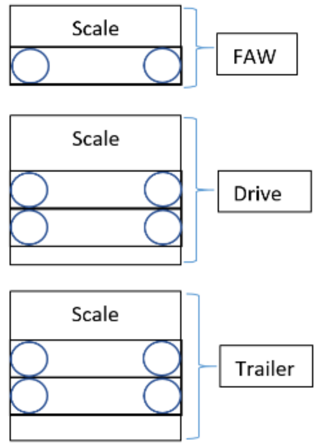
It should be noted that during the March 2018 data collection, the WIM sites at Lonoke and Bald Knob were not located at the latitude/longitude positions indicated in the WIM site specifications. Instead, they had been moved about 1 mile upstream of their current locations. Since the incorrect positions were shared with the GPS data provider, the AVI screen line point and the WIM station did not correspond to the same location. The camera was set up at the correct WIM station. Further complicating the data collection, a traffic incident occurred upstream of the WIM site. This reduced our ability to match WIM and AVI data at the Lonoke site.


East Bound I-40
Alma AR.

08:37:39 20/Jun/2016

Steer:	10020 lb
Drive:	19120 lb
Trailer B:	17480 lb
Trailer A:	00 lb
Gross Wt:	46620 lb

(A)



(B)

Figure 13. Example of a Weight Receipts (A) and Weight Recording Configuration (B)

Figure 14. Examples of Still Camera Images at Weight Enforcement Station

Figure 15. Examples of Video Frames from Camcorders at WIM Stations

DATA PRE-PROCESSING

Pre-processing of the field data required two main tasks: (1) Manual matching of trucks from WIM to static scale locations, and (2) Manual matching of WIM PVR records to AVI truck records. Prior to these tasks, the time offset between the cameras and the WIM sensors was determined. This was done by looking at truck sequencing patterns from the WIM vehicle records and then finding that same truck sequence in the traffic videos using timestamps as reference, e.g., one-minute buffer of video watching around the WIM record timestamps. First, we compared vehicle headways by vehicle class (Table 4). This was repeated for the morning, noon, and afternoon at each study site to find an average time offset between the WIM and video. The example below pertains to the time offset for the Texarkana video and WIM for the morning, which was a difference of 17 seconds.

The video processing for identifying shared trucks between the static scales and WIM was performed by manually examining pictures of trucks weighed at the static scale and re-identifying them in the video recordings from the WIM sites.

Table 4. Video to WIM Time Offset Calculation Example

WIM Records				Video			Estimated Offset
Record	**Class**	**Timestamp**	**Headway**	**Class**	**Timestamp**	**Headway**	
604	5	10:02:40	0:00:13	5	10:02:22	0:00:12	-0:00:18
605	7	10:02:52	0:00:12	7	10:02:35	0:00:13	-0:00:17
606	9	10:03:05	0:00:13	9	10:02:48	0:00:13	-0:00:17
607	9	10:03:07	0:00:02	9	10:02:50	0:00:02	-0:00:17
608	4	10:03:12	0:00:05	4	10:02:55	0:00:05	-0:00:17
Average Time Offset							***-0:00:17***

To match the AVI and WIM PVR truck records, first 20 minutes of video for each AVI truck, e.g., a 10-minute buffer around the AVI timestamp, was examined. Images and descriptions of each truck were recorded. Then the same process was followed when observing the video at the second WIM site looking for trucks that had previously crossed first WIM site within the 10-minute buffer of the AVI time stamp. After trucks started being successfully matched or re-identified crossing both WIM sites, the video to AVI time offset was determined. At Lamar the offset was 12 to 17 seconds and at Lonoke it was 1 min 45 seconds to 3 minutes. The greater time variability at Lonoke was attributed to traffic congestion caused by a traffic accident during the data collection.

PERFORMANCE METRICS

The Truck-Matching Algorithm was evaluated using three performance indexes: True Match Rates (TMR), Correct Match Rates (CMRs), and Error Rates (ER). The TMR reflects our ability to match all vehicles seen in the video. There were three reasons for being unable to manually match all AVI trucks to their corresponding WIM record to achieve 100% TMR. First, trucks had to be visually confirmed to have passed both stations but with the camera recording from a side-fire position it was not possible to view trucks in the inner lane due to occlusion. Second, trucks were not able to be visually confirmed if a site had a high variability in the time offset between the records and the video. This was found to occur at Lonoke due to traffic congestion upstream of the data collection site. Third, some trucks seen in the video and recorded in the AVI data were not recorded by the WIM sensor. This is likely due to sensor error or the truck traveling off-center to the sensors causing measurement error.

$$\textbf{True Match Rate}, \qquad TMR = \frac{M_{true}}{T_{veh}} \qquad \textit{Equation } 5$$

$$\textbf{Correct Match Rate}, \qquad CMR = \frac{M_{TM}}{M_{true}} \qquad \textit{Equation } 6$$

$$\textbf{Error Rate}, \qquad ER = \frac{M_{miss}}{M_{true}} \qquad \textit{Equation } 7$$

Where,

T_{veh} = total number of vehicles observed at the site
M_{true} = total number of actual true matches from groundtruth
M_{tm} = number of successful matches obtained using algorithm
M_{miss} = number of missmatched trucks selected by algorithm

Absolute Percent Error (APE), Mean Absolute Percent Error (MAPE), and Median Absolute Percent Error (MdAPE) were used to measure the discrepancy between the auto-calibration algorithm outputs and static (or true) weights. Unlike MAPE which averages APE for each truck, the MdAPE, which computes the median APE, is robust to outliers.

$$APE = \frac{|Calibrated\ WIM\ Weight - Static\ Weight|}{Static\ Weight} \times 100 \qquad \textit{Equation } 8$$

$$MAPE = \frac{\sum APE}{n} \qquad \textit{Equation } 9$$

$$MdAPE = Median(APE) \qquad \textit{Equation } 10$$

Where,
$Calibrated\ WIM\ Weight$ is the truck axle or GVW weight adjusted using calibration factors produced by the AVI auto-calibration method
$Static\ Weight$ is the truck axle or GVW weight measured by static scales

Three methods which seek to reduce the discrepancy between weights recorded at WIM sites will be evaluated, however it is important to know that the adjusted weights by a calibration method always tend to have some discrepancy between true static weights due to dynamic weights and scale errors. The performance of three auto-calibration methods were evaluated using the above performance metrics to test how close they got to static weights. The three methods included the current ARDOT method, an enhanced version of the ARDOT method used by MnDOT, and the proposed AVI auto-calibration algorithm. The calibration factors for the proposed study where applied to FAWs and GVWs.

DATA COLLECTION SUMMARY

Sample Size

Data was collected on March 15, 2018 and March 20, 2019 at different sites (Table 5). The WIM sites selected for the 2018 data collection were eastbound along I-40 at Lamar and Lonoke and east/northbound along Highway 167 at Bald Knob. The eastbound Alma weigh station along I-40 was used as the static enforcement site. The WIM PVR records for 2018 show that at around 6 a.m. the auto-calibration algorithm was turned off at Lamar, Lonoke, and Bald Knob and from that time unadjusted weight records were reported. At the Alma static scale, 263 trucks were recorded, from these 106 were re-identified at Lamar, 69 at Lonoke and 2 at Bald Knob. A total of 121 AVI trucks crossed Lamar and Lonoke during video recording hours (8 a.m. – 6 p.m.), 44 of these trucks also crossed the Alma weigh station, and 33 of these were without error at either WIM site.

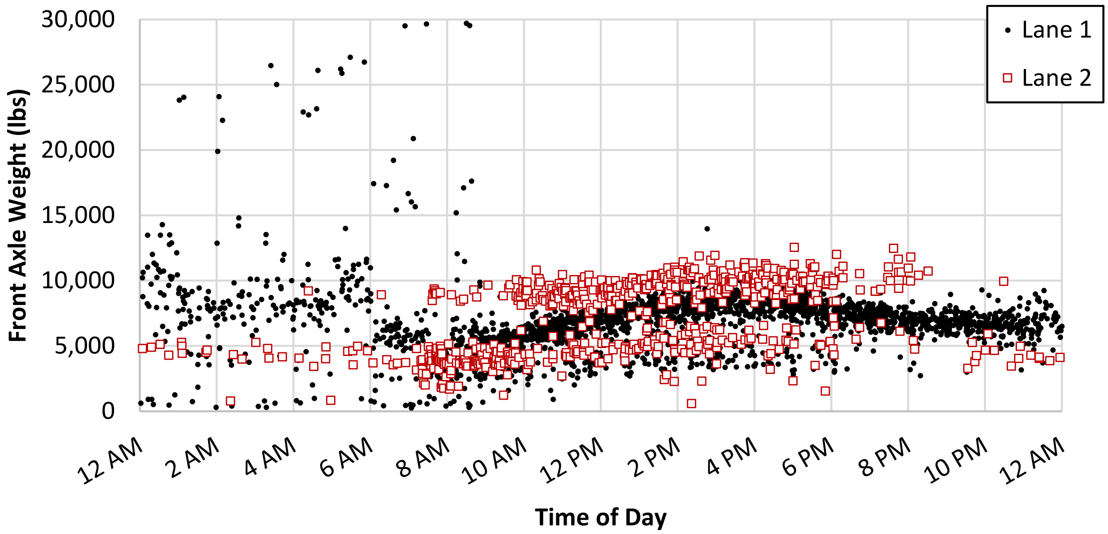
The WIM sites for the data collection in 2019 were southbound on I-30 at Glen Rose and Arkadelphia, and southbound on I-49 at Texarkana. The southbound weigh station on I-30 at Hope was used as the static enforcement site. The WIM PVR for this instance were recorded with the ARDOT calibration method on, therefore the PVR are all adjusted. The number of trucks weighed at Hope weigh station was of 261. From these 88 were re-identified at Glen Rose, 157 at Arkadelphia, and 17 at Texarkana. A total of 157 AVI truck crossed Arkadelphia of which 88 and 17 also crossed Glen Rose and Texarkana. Unfortunately, a data logging error occurred during the data collection and the AVI data for the southbound WIM sites at Glen Rose and Arkadelphia were not available. Therefore, we were not able to evaluate the WIM to AVI truck-matching algorithm. But we were able to replicate the AVI data with video records to evaluate the auto-calibration method at these sites.

Table 5. Data Collection Summary

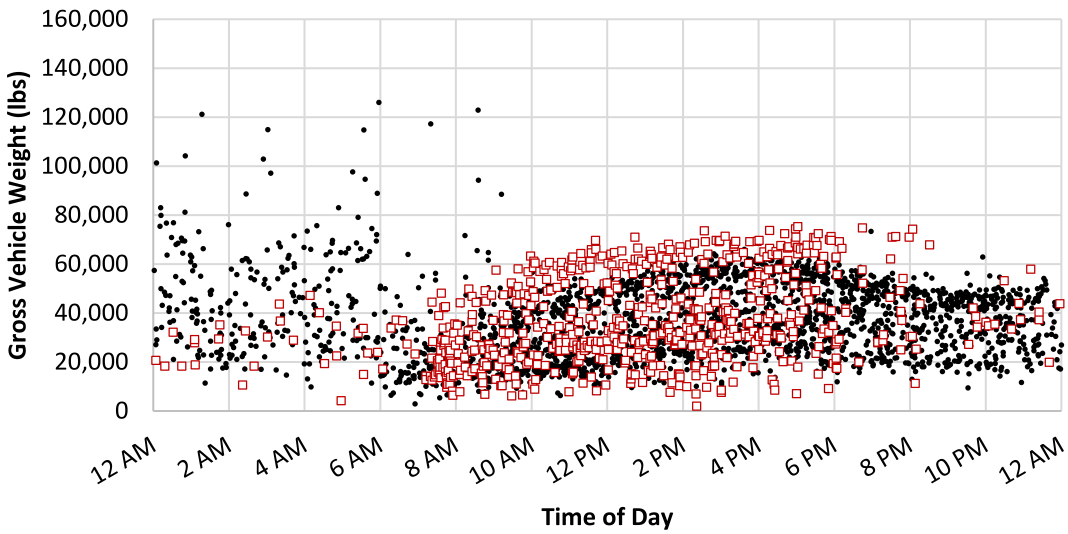
Data Collection	Site	WIM PVRs	AVI Trucks Matched to Static Scale
March 15, 2018 I-40 EB/Hwy 167 NB	Lamar	5,346	106
	Lonoke	10,801	69
	Bald Knob	1,963	2
	Alma (Static Scale)	263	
March 20, 2019 I-30 SB/I-49 SB	Glen Rose	9,905	88
	Arkadelphia	11,007	157
	Texarkana	2,159	19
	Hope (Static Scale)	255	

Data Characteristics

WIM-recorded Front Axle Weight and Gross Vehicle Weights of FHWA Class 9 five-axle tractor-trailers for the March 2018 sites show a distinct drop in weight around 6 a.m. when the auto-calibration program was disabled (Figure 16 and Figure 17). Clear distinctions between lanes are observed for all sites. Disabling the WIM auto-calibration function during the March 2018 data collection also disabled the calibration factors set when the site was initially installed. As a result, we see a distinct shift in recorded weights around 6 a.m. when the auto-calibration function was disabled. Throughout the day, we observe that the recorded weights track with daily temperature changes. While the auto-calibration algorithms proposed in this work can correct calibration as a result of temperature or other minor fluctuations, it may not be possible to fully correct for an initial site calibration factor. Similar trends in shifts throughout the day and differences by lane were also observed for the March 2019 sites (Figure 18, Figure 19, and Figure 20). Recall that for the March 2019 data, the auto-calibration function remained enabled during data collection but that it appears to produce poor results for at least one of the lanes at each site.

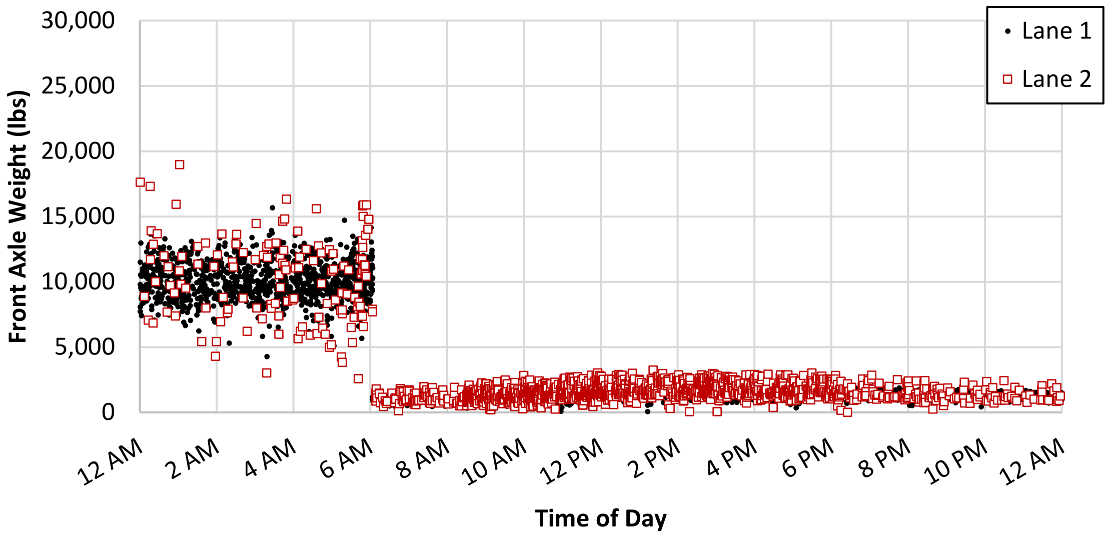


(A) Lamar FAW

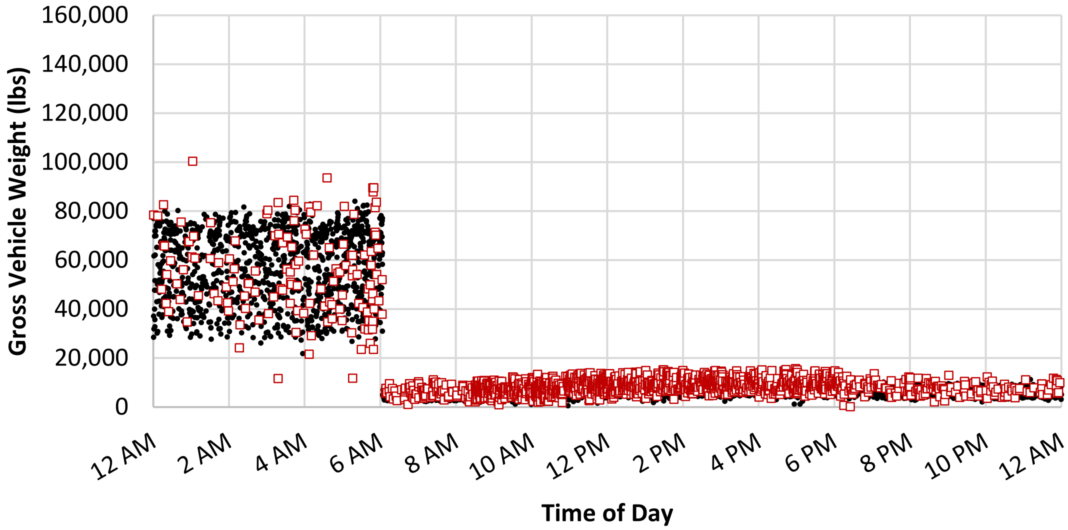


(B) Lamar GVW

Figure 16. Lamar Front Axle and Gross Vehicle Weight by Time of Day

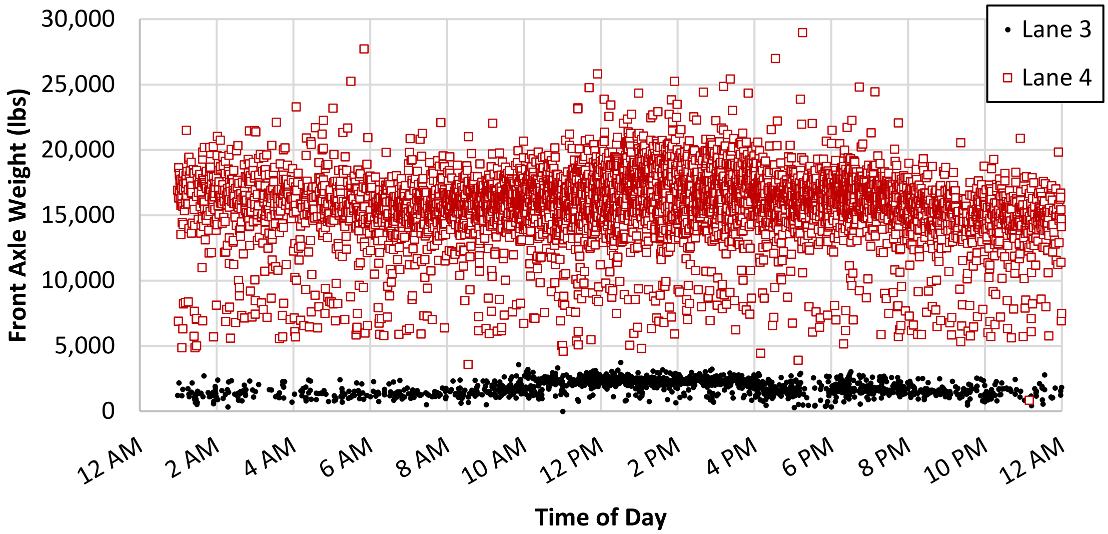


(A) Lonoke FAW

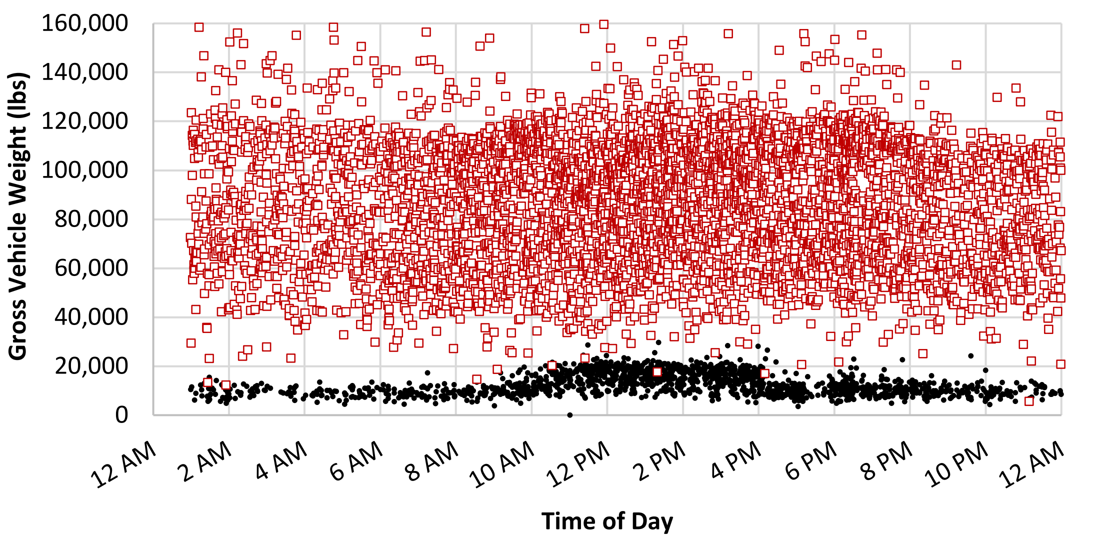


(B) Lonoke GVW

Figure 17. Lonoke Front Axle and Gross Vehicle Weight by Time of Day

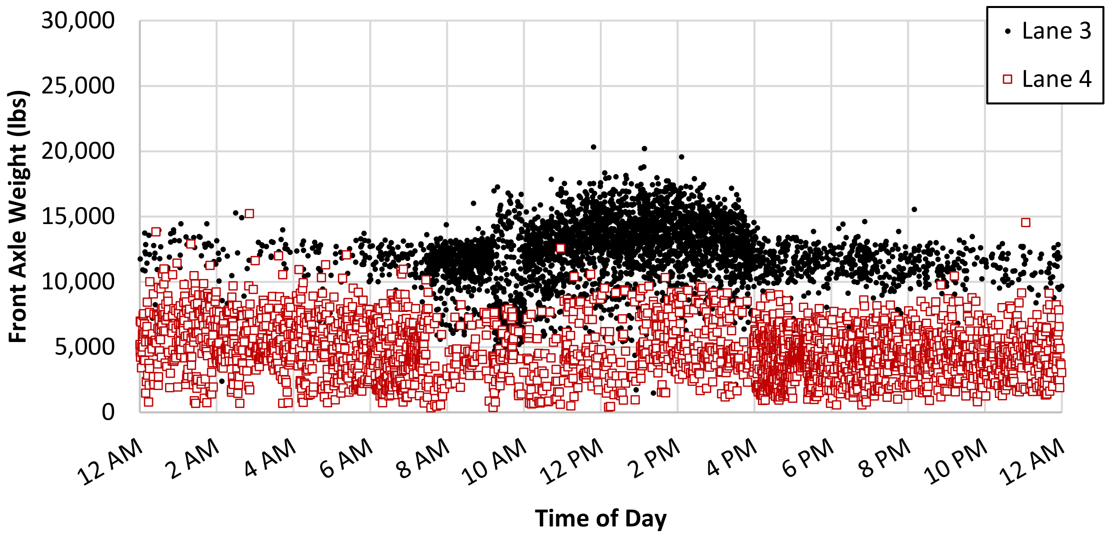


(A) Glen Rose FAW

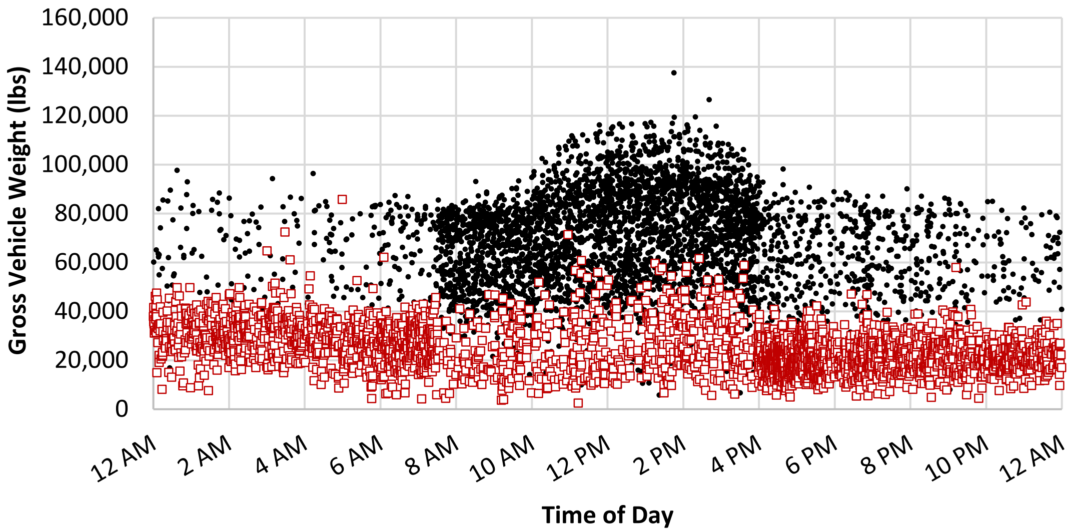


(B) Glen Rose GVW

Figure 18. Glen Rose Front Axle and Gross Vehicle Weight by Time of Day

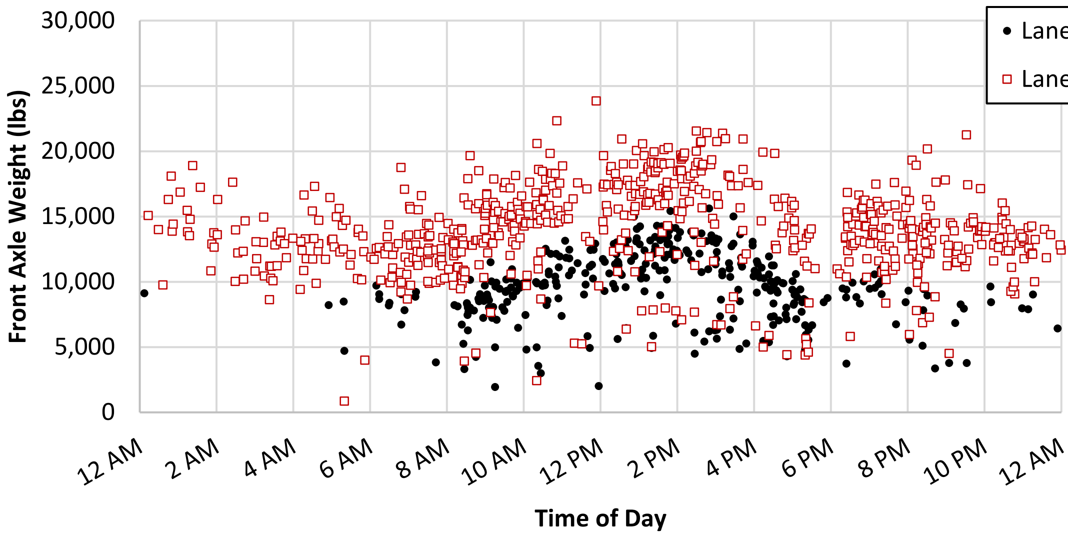


(A) Arkadelphia FAW

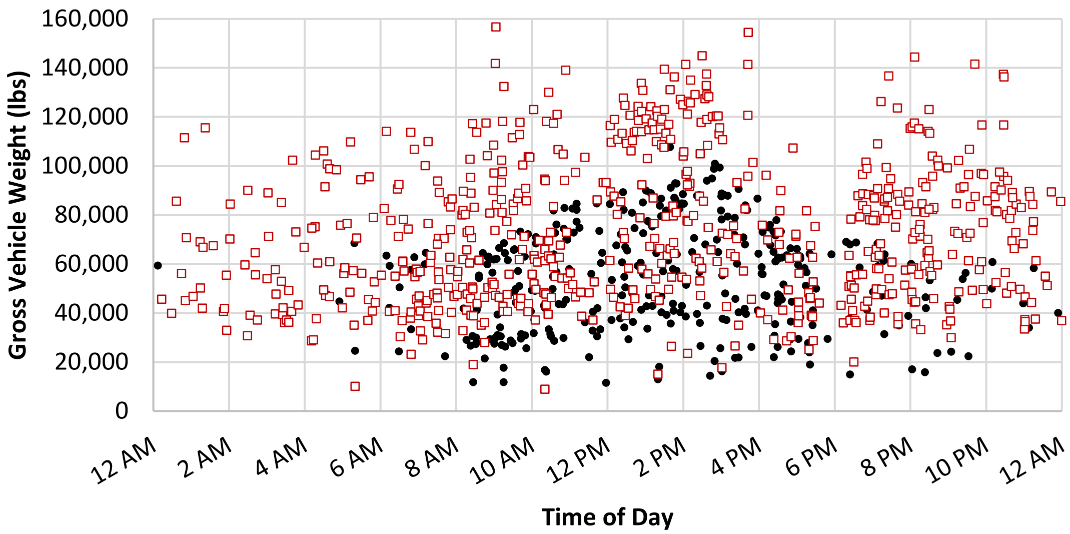


(B) Arkadelphia GVW

Figure 19. Arkadelphia Front Axle and Gross Vehicle Weight by Time of Day



(A) Texarkana FAW



(B) Texarkana GVW

Figure 20. Texarkana Front Axle and Gross Vehicle Weight by Time of Day

FAW and GVW recorded by the WIM and the static scales were used to calculate MAPE and MdAPE for each data collection site. Only samples with the same vehicle class designation at both sites were considered in the comparison. MAPE ranged from 24% to 85% for FAW and 26% to 83% for GVW (Table 6). Lonoke exhibited the highest MAPE and Texarkana exhibited the lowest MAPE.

For static scale measured weights, the Front Axle Weights (FAWs) vary by Gross Vehicle Weight (GVW) for the March 2018 and 2019 data collections (Figure 21 A and B). In comparison to the reference weights used for auto-calibration by ARDOT (10.2 kips) and MnDOT (8.5, 9.3, and 10.4 kips), the March 2018 and 2019 samples exhibit higher FAW.

Table 6. Front Axle Weight and Gross Vehicle Weight Error Summary

Static Scale and Date	WIM Site	Samples	Measure	MdAPE (%)	MAPE (%)
Alma *March 2018*	Lamar	*94*	FAW	30	32
			GVW	26	31
	Lonoke	*56*	FAW	85	85
			GVW	86	83
Hope *March 2019*	Glen Rose	*88*	FAW	53	56
			GVW	54	54
	Arkadelphia	*157*	FAW	23	25
			GVW	26	29
	Texarkana	*19*	FAW	21	24
			GVW	19	26

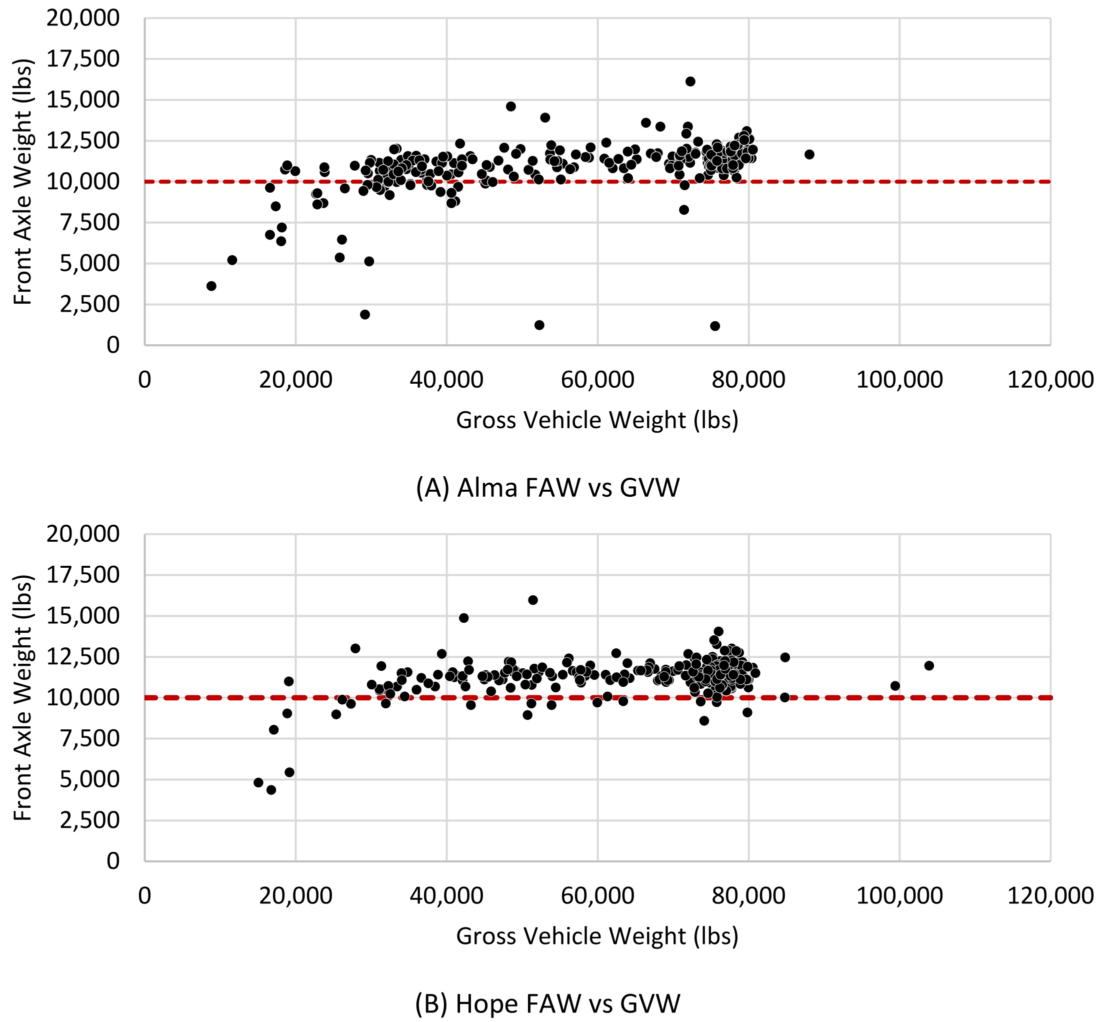
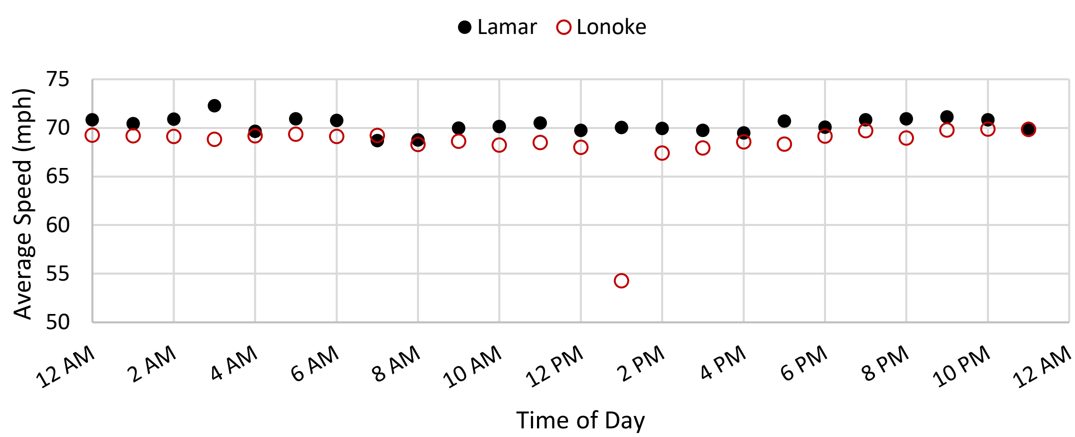


Figure 21. Static Scale Front Axle and Gross Vehicle Weights by Time of Day

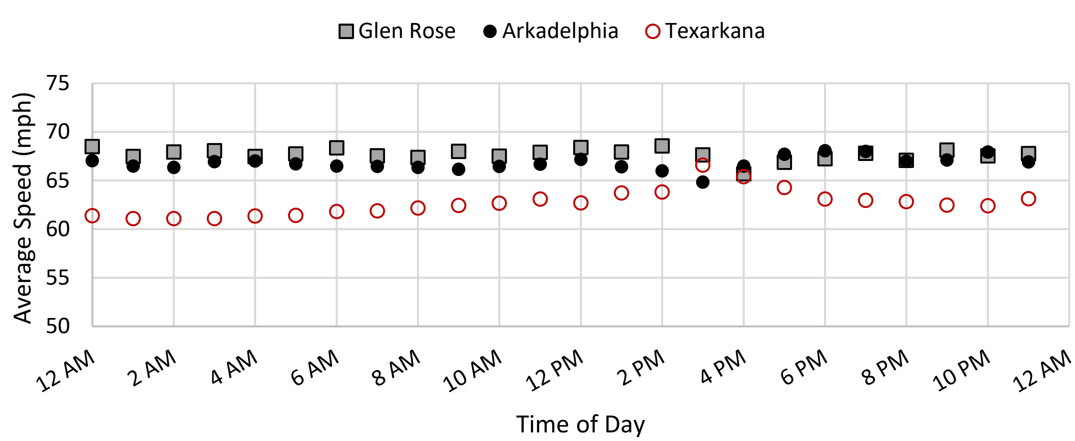
Site Conditions

The average speed of the March 2018 WIM sites ranged from 54 to 72 miles per hour (mph) with a median speed of 70 mph (Figure 22A). The decrease in speed at Lonoke around 1 p.m. was a result of the traffic incident that occurred between the camera location and the WIM site. The average speed of the March 2019 WIM sites ranged from 61 to 69 mph with a median speed of 67 mph (Figure 22b). At the Glen Rose and Arkadelphia sites, the speed dropped to around 65 mph around 4 p.m. but recovered to 68 mph by 5 p.m. At the Texarkana site along I-49, the speed was relatively slower than the I-30 locations, approximately 62 mph, but increased to 65 mph by 3 p.m. before returning to around 63 mph.

Pavement temperature was recorded hourly during the data collection activities using a laser temperature gun. The temperature gun had a sensitivity of +/- 3.6 degrees Fahrenheit and was aimed at the center of the outside lane closest to the observer at a range of approximately 6 feet from the observer. The pavement temperature of the March 2018 WIM sites ranged from 40 to 92 degrees Fahrenheit with an average temperate of 73 degrees Fahrenheit (Figure 23A). The pavement temperature of the March 2019 WIM sites ranged from 46 to 73 degrees Fahrenheit with an average temperate of 62 degrees Fahrenheit (Figure 23B).

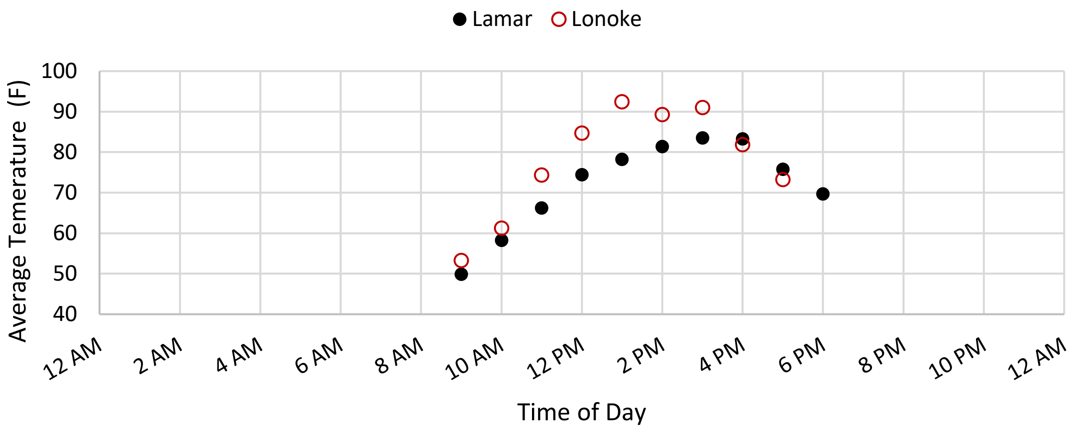


(A) March 2018 WIM Site Speeds

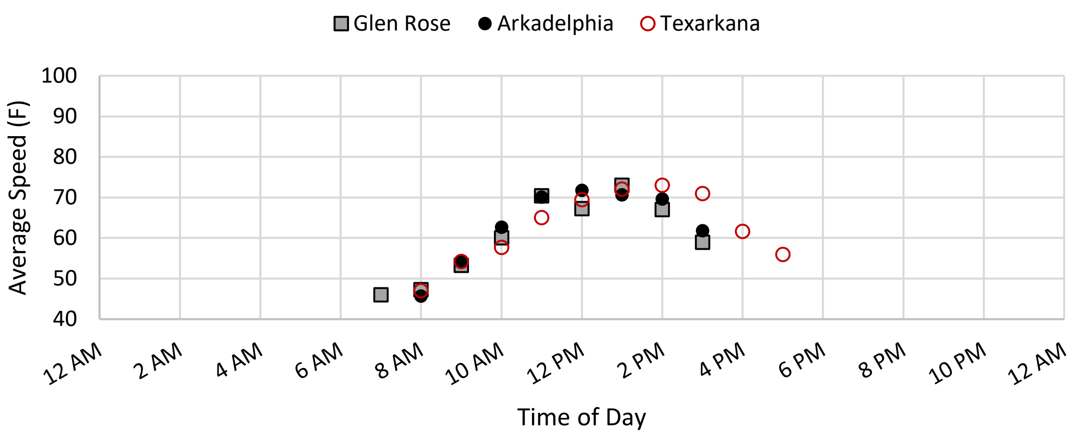


(B) March 2019 WIM Site Speeds

Figure 22. Speeds by Time of Day and Data Collection Period



(A) March 2018 WIM Sites



(B) March 2019 WIM Sites

Figure 23. Pavement Temperature by Time of Day and Data Collection Period

CHAPTER 5: EVALUATION OF CURRENT ARDOT AUTO-CALIBRATION METHODOLOGY

This chapter presents an evaluation of the current ARDOT auto-calibration methodology that uses front axle weight reference values to derive calibration factors. Following a brief description of the current ARDOT methodology and a similar, but more robust methodology developed by MnDOT which use reference weight comparisons to derive calibration factors, an analysis of the reference parameters used by both methods is presented. Finally, performance metrics for each methodology are summarized and limitations of the method are discussed.

AUTO-CALIBRATION USING REFERENCE PARAMETERS

ARDOT's Auto-Calibration Method

The current auto-calibration method employed in Arkansas compares the average FAW of 50 FHWA Class 9 vehicles against the expected average FAW, e.g., 10,200 lbs., and computes a calibration factor so that the average measured FAW is equal to the reference value (Figure 24).

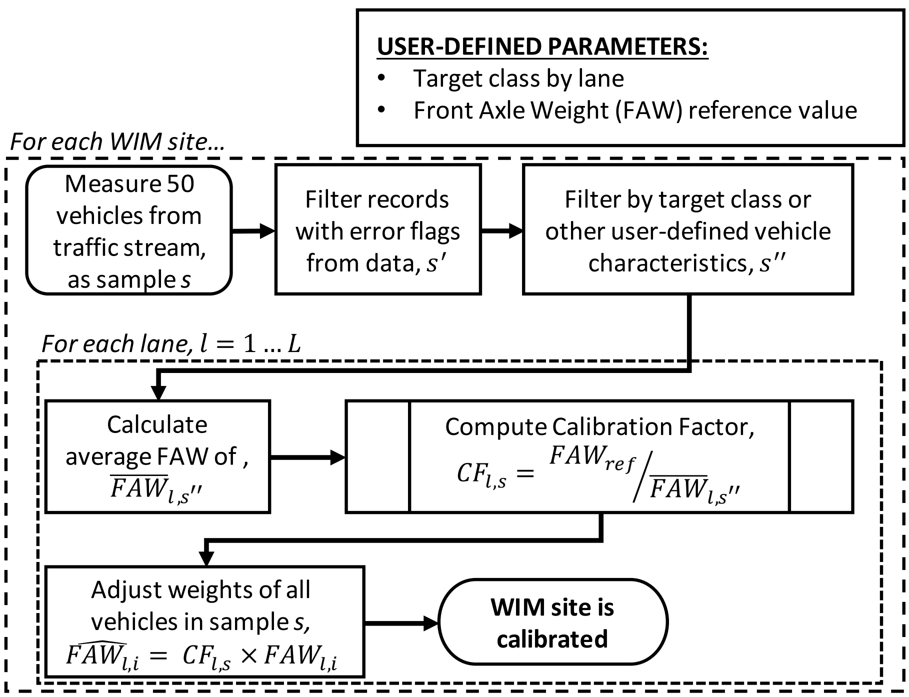


Figure 24. ARDOT Auto-Calibration Procedure

This method has several limitations. The first is that this algorithm relies on a frequent sample of FHWA Class 9 vehicles. To compute a calibration factor, 50 vehicles of the specified class must pass over the sensor. This does not present a problem for areas with high traffic volume, however there are locations that do not receive fifty reference-class vehicles passing for many hours at a time. During the time that it takes to accumulate 50 vehicles, the calibration of the WIM sensor will have drifted considerably due to temperature changes. While the WIM software can utilize a temperature correction curve when computing a calibration factor, this feature is not used in Arkansas.

Another limitation of this method is the variability of the reference value. Often the FAW reference value is a regional average; however, the average FAW can vary between sites depending on local industry, season, the specific lane the sensor is located in, weight laws, and vehicle drivers (Hallenbeck, 1998; Vaziri et al., 2013). Although the FAW distributions may vary between WIM sites, each site's own distribution should remain consistent (Dahlin 1992). Lastly, the accuracy of piezoelectric sensors is heteroscedastic as a function of vehicle weight (Hashemi et al., 2013) and thus reference values should likely vary by weight. The ARDOT method uses only one weight bin when calculating the calibration factor and as a result, heavy vehicles are adjusted by the same factor as lighter vehicles.

MnDOT's Auto-Calibration Method

The MnDOT auto-calibration algorithm assigns unique reference values to each GVW bin (Figure 25) (McCall and Vodrazka Jr 1997). Since the GVW weight bins represent unloaded trucks, partially-loaded trucks, and fully-loaded trucks, a unique reference FAW for each bin more accurately reflects loading characteristics and their effect on FAW.

The MnDOT algorithm preforms calibration every 250 vehicles, specifically FHWA Class 9 trucks ('3-S2' configurations), or every 48 hours. First, a percent deviation is computed via Equation 11 from the average FAW of the vehicles in each GVW weight bin to the reference FAW for each bin (Table 7). If two of the three weight bins are outside acceptable deviation limits, then calibration factors are calculated. Calibration factors are modified based on the number of samples in each GVW bin (Table 8).

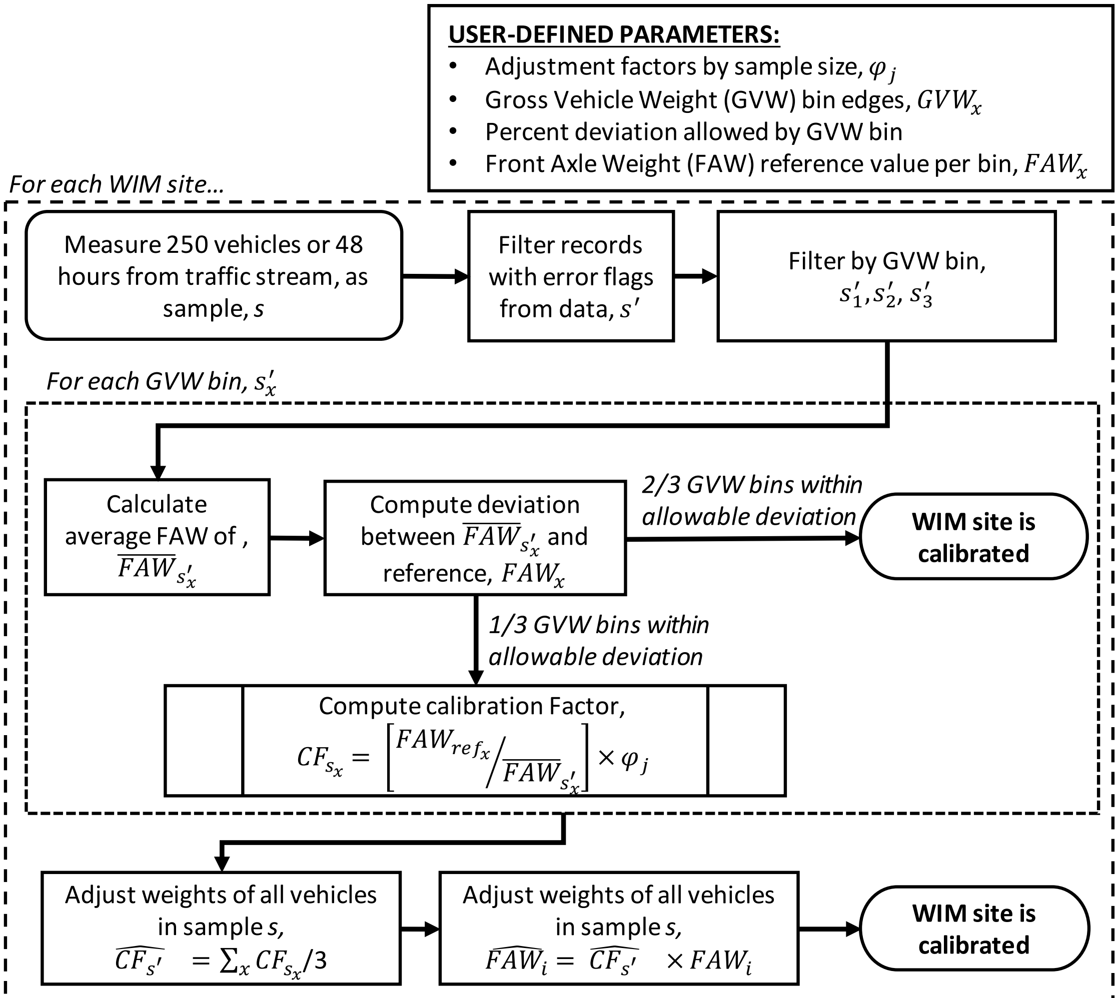


Figure 25. MnDOT Auto-Calibration Procedure

Table 7. MnDOT Reference Parameters (McCall and Vodrazka Jr 1997)

Gross Vehicle Weight Range *(x)*	Reference Front Axle Weight, $FAW_{ref,x}$
< 32,000 lbs.	8.5 kips
32,000 – 70,000 lbs.	9.3 kips
> 70,000 lbs.	10.4 kips

$$\Delta_x = \frac{\overline{FAW}_{s'_x} - FAW_{ref,x}}{FAW_{ref,x}} * 100\% \qquad \textit{Equation 11}$$

Where,
Δ_x = percent deviation for GVW bin x
$\overline{FAW}_{s'_x}$ is the average front axle weight of the filtered sample, s'_x
$FAW_{ref,x}$ is the reference front axle weight of GVW bin x

$$CF_{s_x} = \left[{FAW_{ref,x}} \Big/ {\overline{FAW}_{s'_x}} \right] \times \varphi_j \qquad \textbf{Equation 12}$$

Where,
CF_{s_x} is the calibration factor for GVW bin x that applies to the sample s
φ_i is the sample size adjustment factor for sample s'_x

Table 8. Minnesota DOT Sample Size Adjustment Factors (McCall and Vodrazka Jr 1997)

Number of 5-Axle Semis Weighed	Adjustment Factor Percentage (φ)	Number of 5-Axle Semis Weighed	Adjustment Factor Percentage (φ)
0	0.0	45 - 49	80.0
1	20.0	50 - 54	80.0
2	20.0	55 - 59	90.0
3	20.0	60 - 64	90.0
4	20.0	65 - 69	90.0
5 - 9	30.0	70 - 74	90.0
10 - 14	50.0	75 - 79	90.0
15 - 19	50.0	80 - 84	90.0
20 - 24	60.0	85 - 89	90.0
25 - 29	70.0	90 - 94	90.0
30 - 34	70.0	95 - 99	90.0
35 - 39	70.0	100	95.0
40 - 44	80.0	> 100	95.0

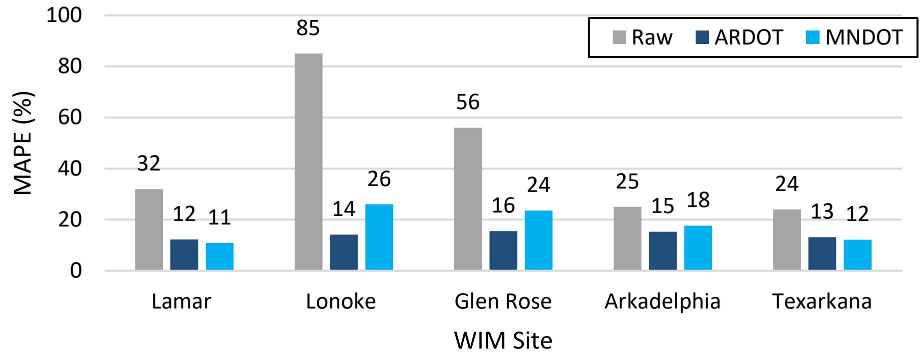
PERFORMANCE EVALUATION

We applied the ARDOT and MnDOT auto-calibration algorithms with original parameters to the 2018 and 2019 datasets to estimate MAPE and MdAPE for FAW and GVW. For the 2018 data, the ARDOT and MnDOT algorithms were applied to raw measurement data since auto-calibration at these sites was disabled during data collection. Although auto-calibration was enabled during the 2019 data collection, the measured weights differed considerably for each lane leading us to suspect that auto-calibration may have been disabled or malfunctioning. Thus, we carried out the ARDOT and MnDOT auto-calibration procedures on the 2019 to provide a best-case performance evaluation.

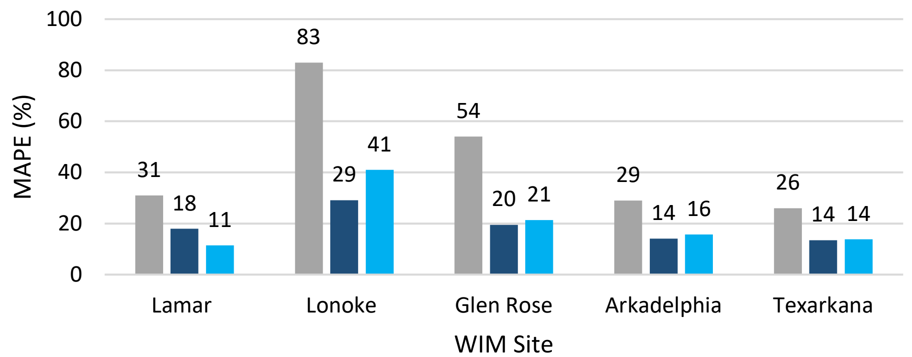
Both the ARDOT and MnDOT methods reduce FAW and GVW MAPE and MdAPE compared to the raw, un-calibrated measurements (Figure 26). MAPE for FAW are generally lower than for GVW for both methods at all WIM sites (Table 9). This is because the algorithms both aim to calibrate FAW. The reduced MAPE and MdAPE for the March 2019 sites confirms the hypothesis that one of the lanes at each site was not calibrated properly even with auto-calibration enabled during data collection.

Table 9. Performance Summary for Baseline Auto-Calibration Algorithms on March 2018 Sites

Static Scale and Date	Method		ARDOT		MnDOT	
	Reference Parameters	Frequency	50		250	
		FAW	10,200 lbs.		8,500, 9,300, and 10,400 lbs.	
		Thresholds	None		<32,000; 32,000-70,000; >70,000 lbs.	
	Error	(%)	**MdAPE**	**MAPE**	**MdAPE**	**MAPE**
Alma *2018*	**Lamar**	FAW	6.6	12.3	2.2	10.9
		GVW	7.6	17.9	2.5	11.4
	Lonoke	FAW	9.9	14.1	21.0	26.0
		GVW	12.9	29.1	28.9	41.0
Hope *2019*	**Glen Rose**	FAW	12.7	15.5	18.1	23.5
		GVW	17.0	19.5	17.9	21.4
	Arkadelphia	FAW	12.9	15.3	12.9	17.6
		GVW	12.2	14.1	12.4	15.7
	Texarkana	FAW	13.3	13.1	34.1	12.2
		GVW	22.3	13.5	26.5	13.8



(A) FAW MAPE



(B) GVW MAPE

Figure 26. Comparison MAPE for Baseline Auto-Calibration Algorithms for FAW and GVW

EVALUATION OF AUTO-CALIBRATION PARAMETERS

The ARDOT and MnDOT auto-calibration contain several tunable parameters including (1) auto-calibration frequency, (2) weight thresholds, and (3) reference weights. The weight threshold refers to the GVW bounds used in the MnDOT method to define loaded and unloaded vehicles. Each parameter individually effects the accuracy of the auto-calibrated weights and combinations of parameter values have compounding effects on accuracy. Using the March 2018 data, we investigated how the selection of site-specific parameter values affects measurement accuracy. First, ranges of feasible parameter values were assessed by comparing resulting MAPE. Second, once individual parameter values were optimized, e.g., the values that produced the lowest MAPE, combinations of parameter values were assessed using an iterative search approach (Figure 27). The analyses are described in this section.

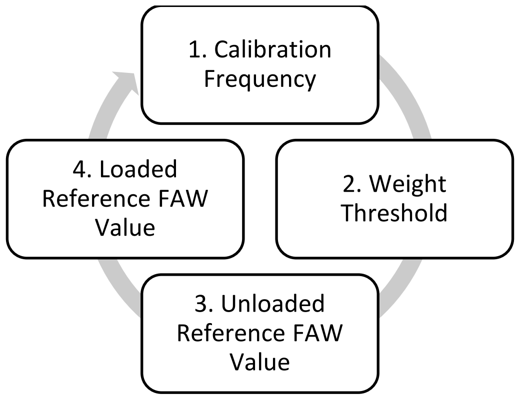
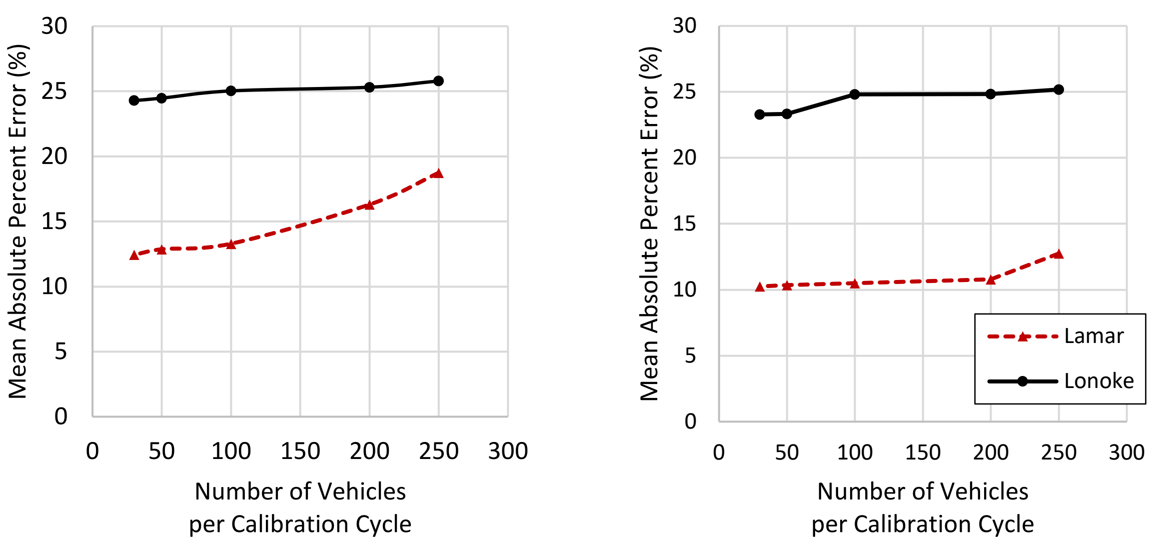


Figure 27. Steps of the Parameter-Tuning Process

Auto-Calibration Frequency

Calibration frequency should ideally keep pace with temperature changes but should also maintain a reasonable number of sample vehicles in each calibration update. When no temperature sensors are installed, more frequent calibrations are needed to keep pace with temperature changes, and therefore smaller sample sizes are required, e.g., 50 vehicles. The ARDOT algorithm calibrates every 50^{th} vehicle while the MnDOT procedure calibrates each 250 vehicles or 48 hours. An error analysis was conducted for calibration frequency at values of 250, 200, 100, 50, and 30 vehicles for the ARDOT method and 250, 200, 100, and 50 vehicles for the MnDOT method.

In general, more frequent calibrations yielded the lowest MAPE for the ARDOT and MnDOT algorithms (Figure 28). For both algorithms, minimal improvement in MAPE was achieved by reducing the frequency below 30 vehicles and increases in MAPE were observed when more than 100 vehicles were needed to trigger an auto-calibration update. We concluded that for both sites and both algorithms, calibration at 50 vehicle intervals was optimal.



(A) ARDOT Auto-Calibration Algorithm (B) MnDOT Auto-Calibration Algorithm

Figure 28. Auto-Calibration Frequency Parameter Tuning

Weight Thresholds

Using the optimal calibration frequency parameters described above, e.g., 50 vehicles for the ARDOT and MnDOT algorithms, we assessed changed in MAPE resulting from changes to GVW thresholds. The MnDOT auto-calibration algorithm uses GVW bins to define FAW reference values. The bin thresholds specified in the MnDOT algorithm defining unloaded, partially-loaded, and fully-loaded trucks were less than 32,000 lbs, 32,000 to 70,000 lbs, and greater than 70,000 lbs, respectively. The two thresholds used in the MnDOT algorithm to separate unloaded, partially-loaded, and fully-loaded trucks should be used when the GVW distribution is tri-modal. These two thresholds can be reduced to a single threshold separating loaded and unloaded/partially loaded trucks if the GVW distribution is observed to be bimodal. The Lonoke and Lamar sites exhibited bi-modal GVW distributions. Therefore, adjustments to the bin threshold value separating loaded and unloaded/partially-loaded trucks was carried out using 5,000 lb increments ranging from 40,000 to 80,000 lbs. Additional iterations were performed at 32,000 lbs for Lamar and 85,000, 90,000, and 95,000 lbs for Lonoke to assess performance improvements for extreme observations seen in the data.

Overall, at both the Lonoke and Lamar sites, the lowest MAPE was achieved using the originally defined weight threshold separating loaded and unloaded/partially-loaded trucks, e.g., 72,000 lbs (Figure 29). For both sites, the GVW bin threshold had only minor effects on the MAPE as observed by the flatness of the lines in Figure 29. The maximum allowable, unpermitted GVW in Arkansas is 80,000 lbs and previous studies cited 72 kips as the threshold that likely separates fully-loaded vehicles from the rest of the traffic stream (Dahlin 1992; McCall and Vodrazka Jr 1997).

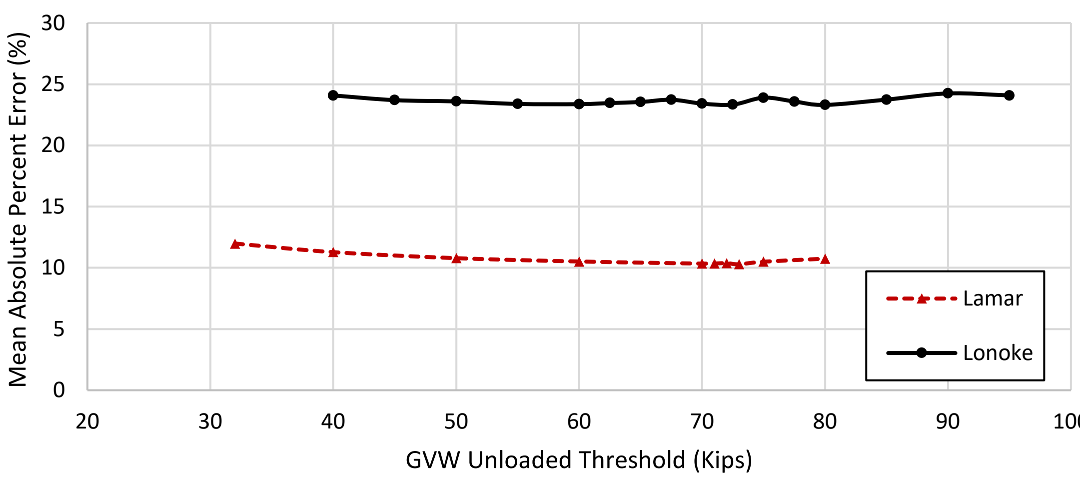


Figure 29. GVW Threshold Parameter-Tuning for MnDOT Auto-Calibration Algorithm

Reference Weights

Holding the calibration frequency at 50 vehicles and the GVW weight bin threshold at 72,000 lbs for both sites and both algorithms, we assessed changes in MAPE resulting from changes to the FAW reference values. The ARDOT autocalibration algorithm uses only one FAW reference value, e.g., 10,200 lbs. We assessed the ARDOT auto-calibration algorithm performance as a result of changing the FAW reference parameter at 200 lb (0.2 kip) increments. For Lamar, a 9,600 lb FAW reference value produced the lowest MAPE while at Lonoke a FAW reference value of 10,600 lbs was optimal (Figure

30). This is indicative of the higher volumes of partially- and fully-loaded vehicles observed at Lonoke compared to Lamar.

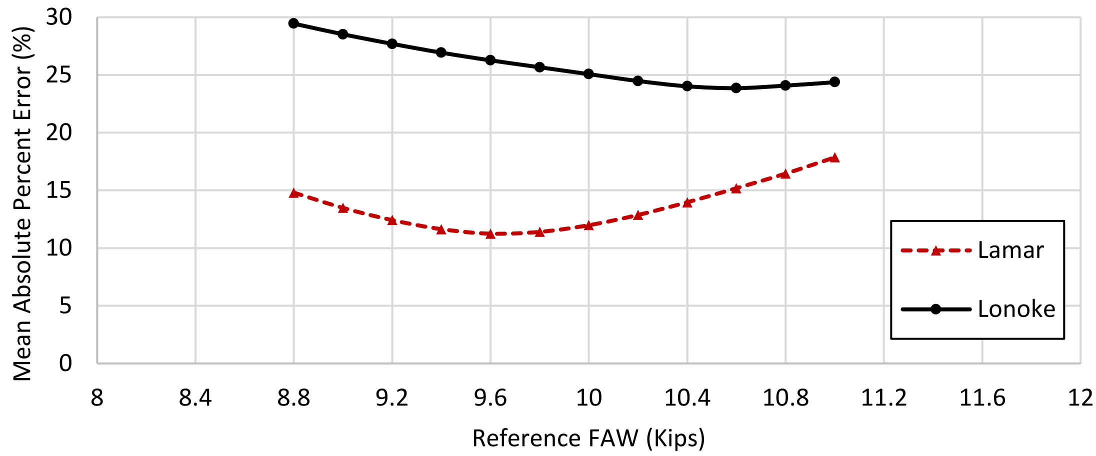
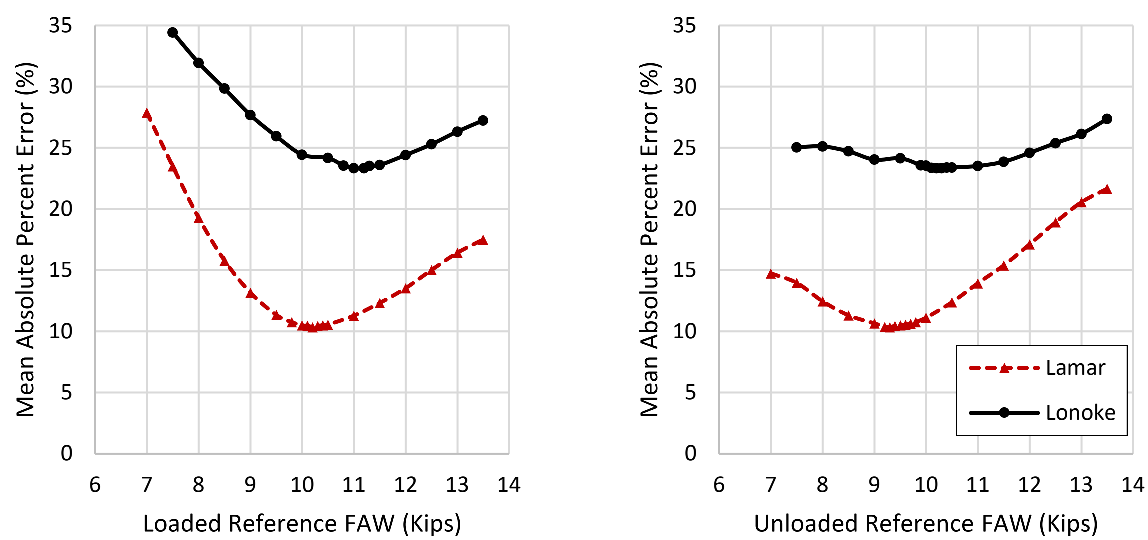


Figure 30. Front Axle Weight Parameter-Tuning for the ARDOT Auto-Calibration Algorithm

The MnDOT auto-calibration algorithm, uses three FAW reference values, one for each GVW bin, e.g., 8,500 lbs, 9,300 lbs, and 10,400 lbs. We assessed the MnDOT auto-calibration algorithm performance as a result of changing the FAW reference parameters at 500 lb increments and then narrowing the increments to 100 lbs around the optimal FAW value. For Lamar, the loaded FAW reference weight that produced lowest MAPE was 10,200 lbs and for Lonoke the FAW reference for loaded trucks was 11,000 lbs (Figure 31A). The unloaded FAW reference weight for Lamar was found to be 9,300 lbs and for Lonoke it was 10,300 lbs (Figure 31B).



(A) Loaded FAW Reference Value (B) Unloaded FAW Reference Value

Figure 31. Front Axle Weight Parameter-Tuning for the MnDOT Auto-Calibration Algorithm

SUMMARY OF FINDINGS

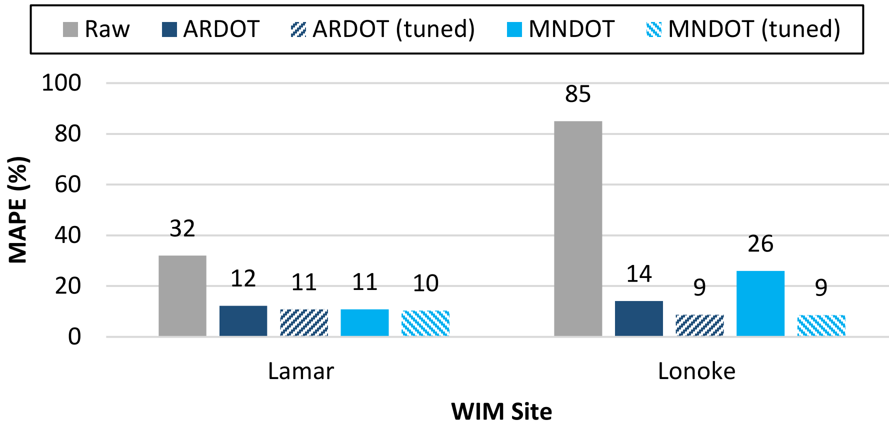
We examined the sensitivity of auto-calibration parameters on algorithm accuracy as measured by the MAPE. We analyzed three parameters, e.g., frequency, weight thresholds, and reference weights, using a tiered, iterative approach applied to the March 2018 data collection sites at Lamar and Lonoke. First, we determined an optimal auto-calibration frequency. Then at that set frequency, we found the optimal GVW bin threshold values. Lastly, for the set frequency and GVW bin value, we found the optimal FAW reference values for each bin. Resulting MAPE for tuned, auto-calibration models show overall improvement for FAW estimation at both sites (Table 10). However, estimation accuracy as measured by MAPE and MdAPE for GVW declined for the MnDOT tuned algorithm at Lamar.

Comparing to the baseline parameters, the optimal (tuned) values of the ARDOT algorithm reduced MAPE for FAW by 3% on average, and for GVW by 6.5% on average (Figure 32). Likewise, for the MnDOT algorithm, the tuned parameters reduced FAW MAPE by 9% on average. However, we see an increase in MAPE for GVW at Lamar using the tuned MnDOT parameters. This may be due to the distinct bimodal FAW measurement distribution seen for Lane 2 at Lamar and the inability to capture this distinction within the same lane using the MnDOT or ARDOT auto-calibration algorithms (see Figure 16A). Although calibration factors are calculated for each lane, we cannot account for differences within the same lane, which is the case for the Lamar WIM site.

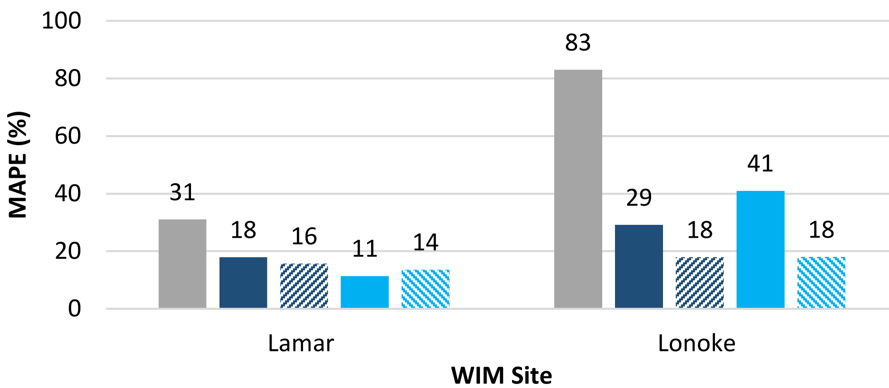
Using data collected from two WIM sites and a static scale we were able to show that the baseline ARDOT auto-calibration method can be improved by accounting for parameters which are unique to each data collection site. It was determined that the loaded threshold value was spatially transferrable between sites with similar traffic stream characteristics; however, this parameter needs to be tested for transferability across sites with differing traffic stream characteristics, e.g., high- and low-volume sites. Moreover, the FAW reference values were shown to be specific to each site indicating that site-specific FAW reference values are needed. Lastly, the use of a temperature correction curve should be investigated, which would allow for larger sample sizes and less frequent calibration. In all, site-specific data needs to include static scale and WIM measured weight comparisons for at least a 12-hour period across all four seasons to understand the temperature effects on the sensor. This is cost prohibitive and time consuming. AVI-based auto-calibration algorithm should considerably reduce these data collection burdens by automating the process of matching trucks across sites.

Table 10. Auto-Calibration Parameter-Tuning Results Summary

Method		ARDOT		MnDOT	
Reference Parameters	Frequency	50		50	
	FAW	9,600 (Lamar) 10,200 lbs. (Lonoke)		9,300 and 10,200 lbs. (Lamar) 10,300 and 11,000 lbs. (Lonoke)	
	Thresholds	None		>70,000 lbs.	
Error	(%)	**MdAPE**	**MAPE**	**MdAPE**	**MAPE**
Lamar	FAW	5.4	10.8	4.7	10.4
	GVW	7.4	15.7	4.9	13.6
Lonoke	FAW	2.5	8.7	3.8	8.6
	GVW	4.1	17.9	2.4	18.0



(A) FAW MAPE



(B) GVW MAPE

Figure 32. Comparison of Baseline and Tuned Auto-Calibration Algorithms for March 2018 Data Collection

CHAPTER 6: EVALUATION OF AVI-BASED AUTO-CALIBRATION METHODOLOGY

This Chapter summarizes the performance of the AVI-based auto-calibration method when applied to the data collecting during the 2018 and 2019 field collection efforts. Performance is evaluated for the Matching Algorithm and the Auto-Calibration Algorithm.

WIM TO AVI MATCHING PERFORMANCE

During the March 2018 data collection, a total of 121 AVI trucks traveled from Lamar to Lonoke. Out of these 121 trucks, 93 (e.g., TMR of 77%) were successfully matched with their respective WIM PVR at Lamar. At Lonoke, matches between PVR and AVI were only sought for the trucks also found at Lamar. Thus, all 93 AVI trucks were successfully matched to their WIM PVR record, e.g., 100% TMR.

The CMR and ER reflect the ability of the matching algorithm to correctly match WIM PVR and AVI truck records. CMR assesses the truck-matching algorithm success rate such that a value closer to 100% is better. We used the 93 successfully matched WIM records that were manually matched to AVI records. The CMR at Lamar was 75% (e.g., 70 of 93 records correctly matched) and at Lonoke 52% (e.g., 48 of 93 records correctly matched). ER captures the same concept as CMR but is represented as error, e.g., the goal is to achieve a low ER. Thus, the ER for Lamar was 25% and Lonoke was 48%.

An initial time window of 180 seconds (3 minutes) was found to produce the highest CMR across all sites. The selected time window was based on trial and error, running the algorithm under different time window settings which yielded the best CMR.

Lower CMR can be attributed to WIM sensor errors like missed detections, ghost detections (detections of vehicles that were not actually there), counting vehicles with two trailers as two separate vehicles, and counting vehicles straddling two lanes as two separate vehicles. Traffic flow was also a contributing factor, as demonstrated in Lonoke. A lower CMR (higher ER) at Lonoke was mostly attributed to an upstream accident that occurred around noon during data collection which caused larger variability in the time offset between the WIM and AVI records and data recording errors. Data recording errors, indicated by a "flagged" record, peaked during the traffic incident (Figure 33). Recall the location of the WIM site and the AVI screen line were about 1 mile apart. Most of the shared AVI trucks between Lamar and Lonoke crossed Lonoke around noon also contributing to the lower CMR (Figure 34). The temporal inputs and sequencing of the records were two central inputs components of the truck-matching algorithm therefore having uninterrupted traffic flow is critical if the WIM and AVI locations differ.

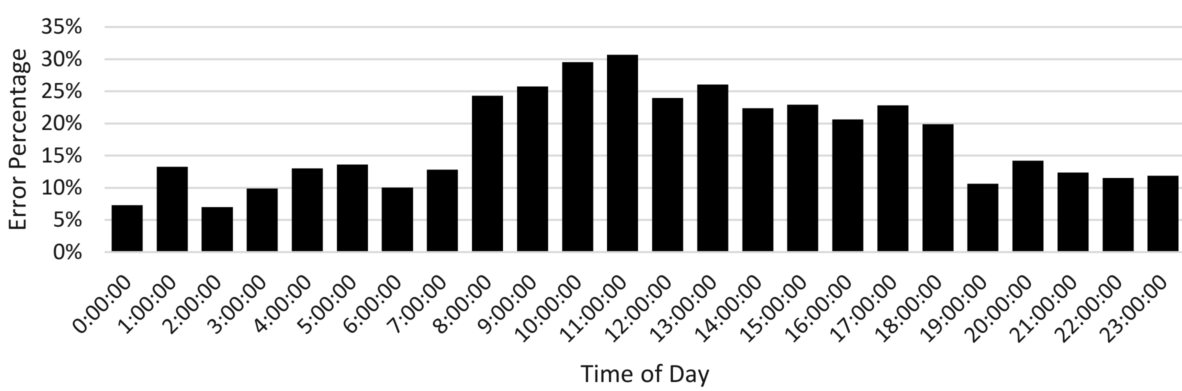


Figure 33. WIM Sensor Error Rate by Time of Day at the Lonoke WIM Site During the March 2018 Data Collection

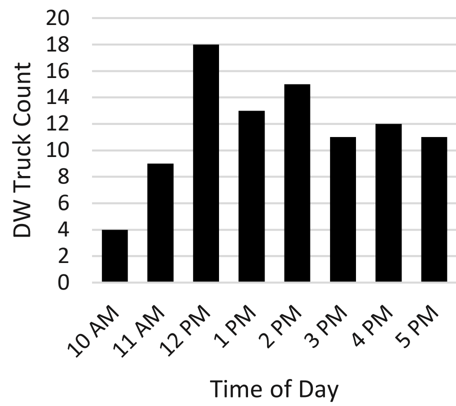


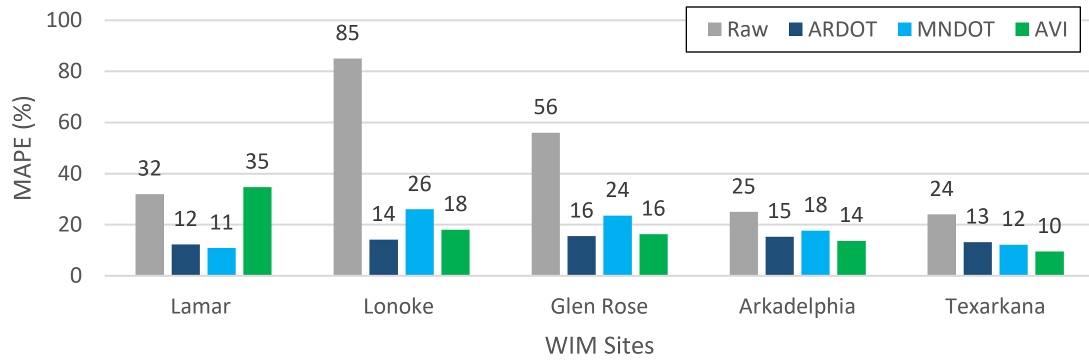
Figure 34. Shared AVI Trucks between Lamar to Lonoke During Field Data Collection Hours

AUTO-CALIBRATION PERFORMANCE

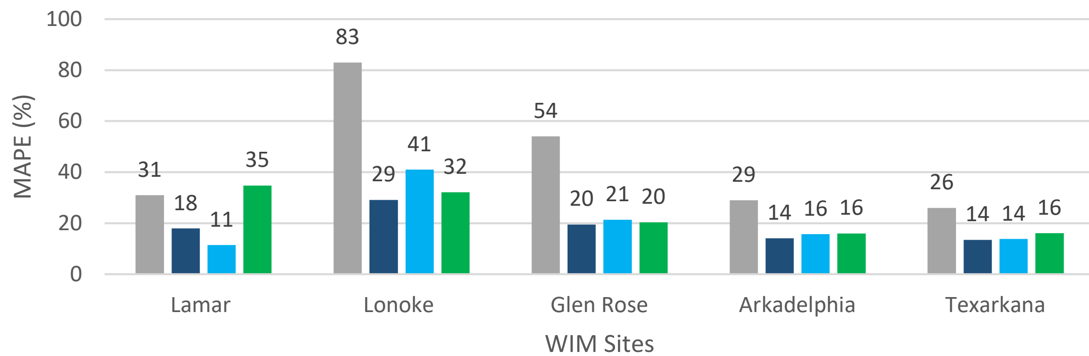
Auto-calibration performance was assessed via the MAPE and MdAPE for FAW and GVW adjustments for the March 2018 and 2019 data collection sites. The AVI-based auto-calibration method applied to the March 2018 and 2019 data resulted in FAW MAPE between 10% and 35% and GVW MAPE between 16% and 35% (Table 11). The results of the proposed AVI algorithm were compared to the ARDOT and MnDOT methods using their baseline parameters (Figure 35). The AVI-based auto-calibration approach presented only slight accuracy gains over the ARDOT and MnDOT methods at most sites. Lamar was an exception where the AVI-method resulted in MAPE higher than that experienced for the unprocessed data.

Table 11. MAPE and MdAPE for FAW and GVW by Data Collection Site

Static Site and Data	Method		AVI	
	Error	(%)	MdAPE	MAPE
Alma *March 2018*	Lamar	FAW	32.6	34.7
		GVW	28.4	34.8
	Lonoke	FAW	14.5	18.0
		GVW	14.0	32.1
Hope *March 2019*	Glen Rose	FAW	13.3	16.3
		GVW	17.4	20.3
	Arkadelphia	FAW	11.6	13.7
		GVW	10.9	16.0
	Texarkana	FAW	7.9	9.5
		GVW	7.3	16.1



(A) FAW



(B) GVW

Figure 35. Comparison of FAW and GVW MAPE by Auto-Calibration Method and WIM Site

PERFORMANCE EVALUATION

The higher MAPE for FAW and GVW produced by the AVI-based auto-calibration algorithm at Lamar provided several insights into the benefits and limitations of the approach. First, the auto-calibration algorithm produced lower MAPE when the original "raw" data exhibited higher error such as at the Lonoke and Glen Rose locations. When applied to locations with more accurate "raw" data, such as at Lamar, Arkadelphia, and Texarkana, the AVI-based auto-calibration algorithm performed just as well as the ARDOT and MnDOT algorithms, reducing FAW MAPE from around 25% to as low as 10%. The different in the degree of calibration produced by the AVI-based algorithm is evident when observing the raw and adjusted FAW and MAPE and comparing Lamar and Lonoke (Figure 36).

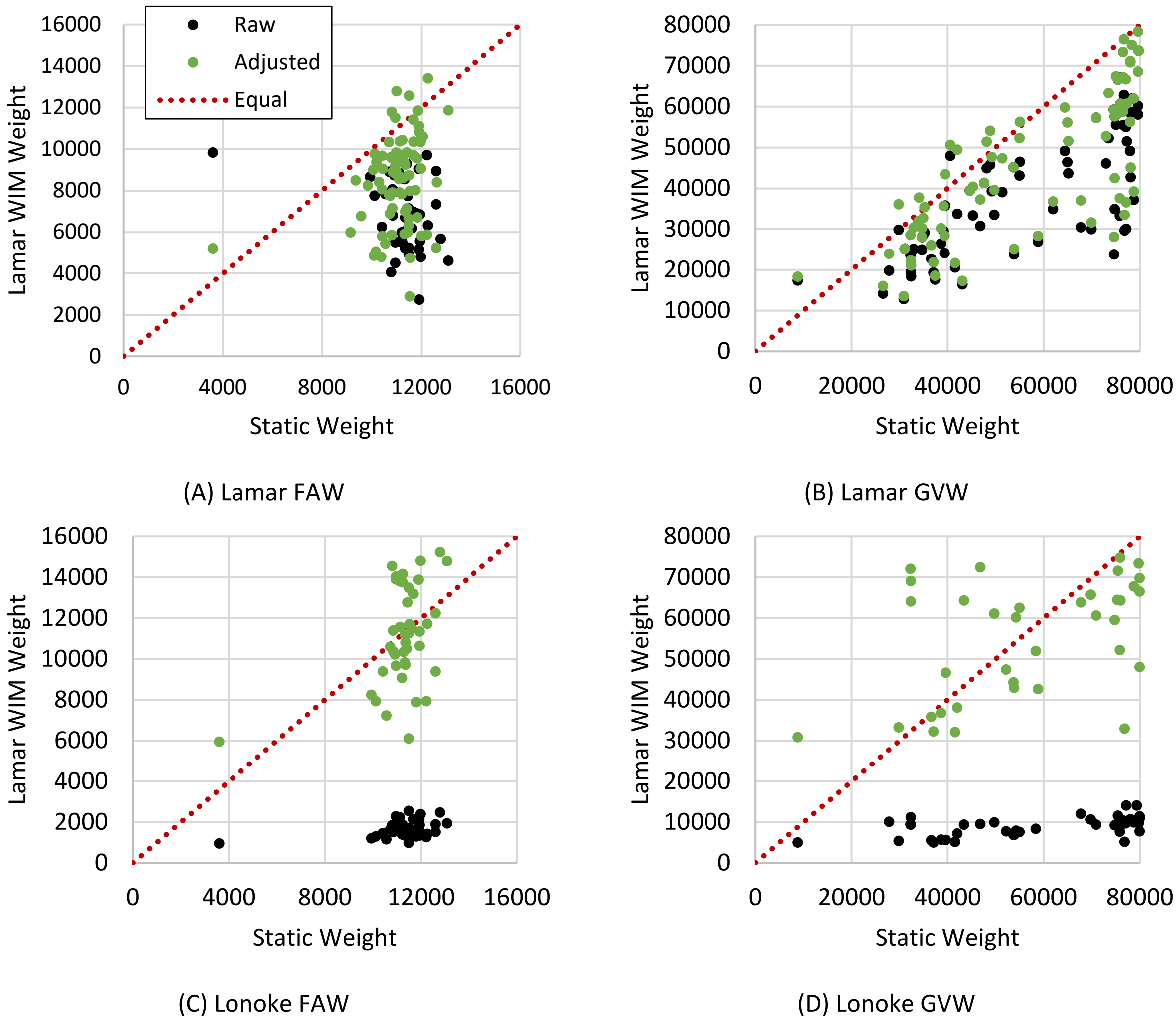
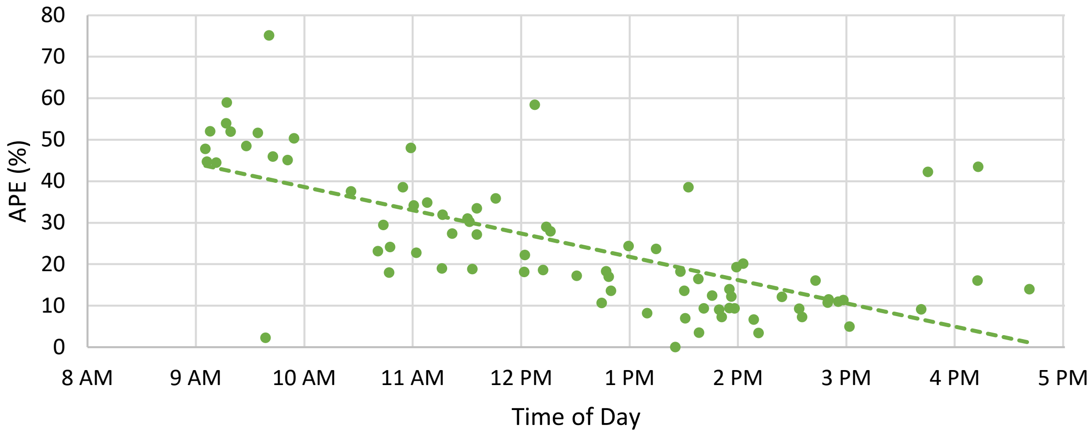
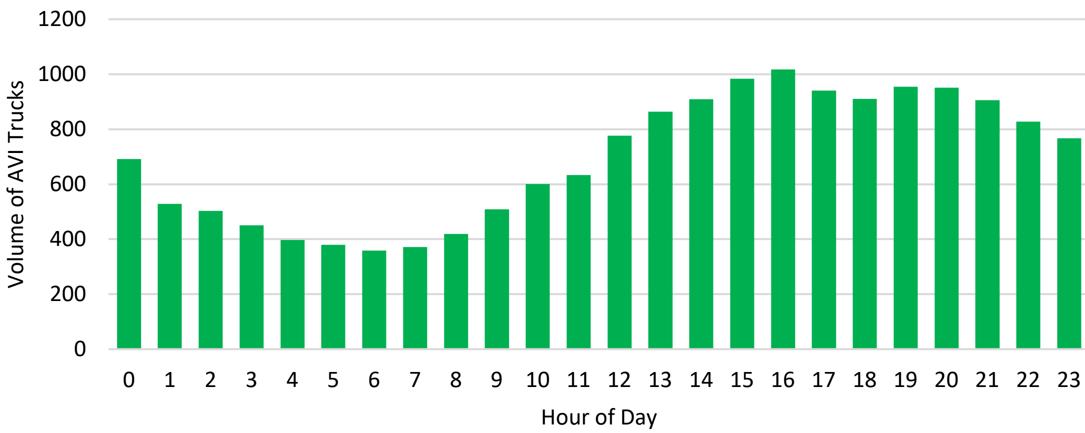


Figure 36. Comparison of Static and WIM Weights for Lamar and Lonoke

Second, the AVI-based auto-calibration algorithm produced lower FAW MAPE (increased accuracy) toward the end of the data collection time period (Figure 37). Correlation between time of day and MAPE could be attributed to the inability of the AVI-based method to keep pace with temperature changes earlier in the day when traffic volumes of GPS enabled trucks were lower. With less trucks available to compare across sites, the frequency and accuracy of calibration factor updates is reduced. Thus, we are likely to see temporal trends in FAW auto-calibration accuracy that is related to volumes of GPS tracked trucks. For the ARDOT and MnDOT methods, changes in accuracy do not trend with time of day.



(A) AVI-Based Auto-Calibration FAW MAPE Results by Time of Day

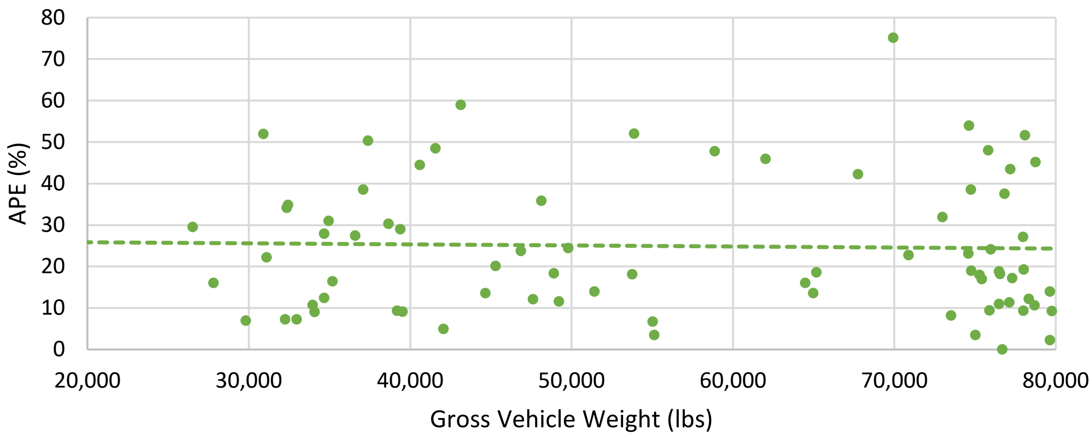


(B) Volume of AVI-Tracked Trucks by Time of Day for March 15, 2018 Data Collection

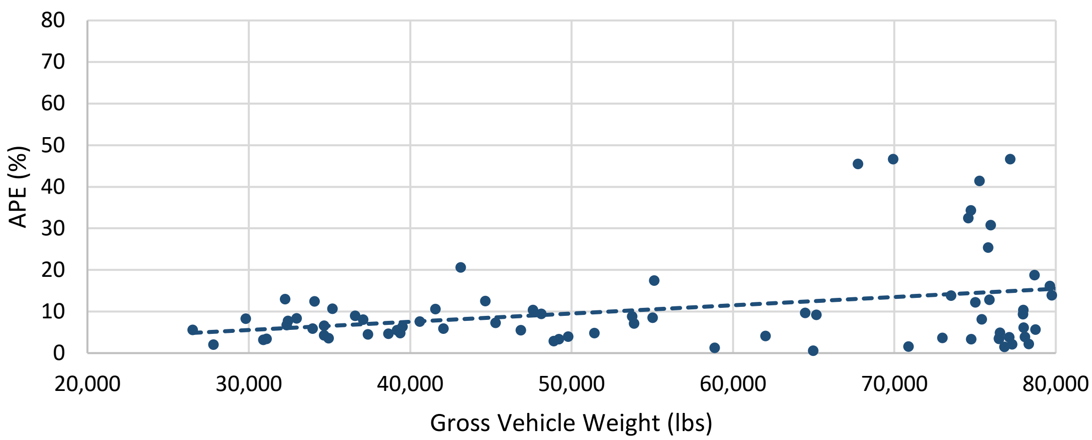
Figure 37. Comparison of FAW Estimation Error by Time of Day for AVI-based and ArDOT Methods for Lamar WIM Site

Third, the AVI-based method produced more consistent MAPE across GVW ranges than the ArDOT and MnDOT methods. The AVI-based method calculated calibration factors by comparing the same truck's FAW across WIM sites. This comparison takes into consideration the inherent relationship between FAW and GVW, e.g., with higher GVW slightly higher FAW is expected due to the location of the king-pin and loading pattern across the axles. Therefore we see no trend between GVW and MAPE for the AVI-based method but for the ArDOT method (Figure 38A), we observe an increase in MAPE as GVW increases (Figure 38B). The ability to consistently calibrate across GVW ranges can be interpreted as an advantage of the AVI-based approach.

Considering this, we assessed a variation of the AVI-based method in which we estimate *likely* GVW instead of *likely* FAW. To adapt the algorithm, we analyzed the FAW of each truck to determine which GVW was "correct", e.g., if the FAW of a truck was outside the tolerance of the reference FAW then we would not use that truck's GVW as the *likely* GVW. However, this approach resulted in less accurate results than the originally proposed AVI-based method. This can be attributed to the high variability in GVWs which makes it impossible to assume a reference GVW to compare the *likely* GVW when GVWs across sites are in disagreement.



(A) AVI-Based Method



(B) ARDOT Method

Figure 38. Comparison of FAW Estimation Error by GVW for AVI-based and ARDOT Methods for Lamar WIM Site

CHAPTER 7: CONCLUSIONS

SUMMARY OF FINDINGS

An AVI-based auto-calibration method was developed and compared to existing auto-calibration algorithms. The AVI-based method consisted of first, matching AVI-tracked trucks to WIM PVR records and second, applying a calibration procedure in which the measured weights of the same truck tracked by AVI across multiple WIM sites are used to generate a reference weight and calibration factor. The approach currently used by ARDOT generates calibration factors based on the measured Front Axle Weight (FAW) averaged for a sample of 50 five-axle tractor-trailers and compares it to a predefined reference weight. A more robust method, developed by MnDOT, expands on that approach by defining three FAW references based on Gross Vehicle Weight (GVW) bins and applying correction factors when sample sizes are small.

The proposed AVI-based approach was compared to the ARDOT and MnDOT approaches for a set of six WIM sites at Lamar, Lonoke, Bald Knob, Glen Rose, Arkadelphia, and Texarkana. During two data collections in March of 2018 and 2019, we collected WIM Per Vehicle Record (PVR) at each WIM site, AVI data from a national truck GPS data provider, and static weight recordings at Arkansas Highway Police weight enforcement sites at Alma and Hope. Auto-calibration at the Lamar and Lonoke WIM sites was disabled during data collection but was enabled for the Glen Rose, Arkadelphia, and Texarkana sites. An extensive data preprocessing methodology was developed and applied to provide data necessary for auto-calibration algorithm performance evaluation. Without auto-calibration, we observed FAW errors ranging from 24% to 85% with GVW error in the same range.

Site-specific tuning of the user-specified values required in the ARDOT and MnDOT auto-calibration algorithms resulted in errors for FAW between 9% and 11% and for GVW between 14% and 18%. Overall, through site-specific tuning of parameters like FAW reference values and GVW bin thresholds used in the ARDOT and MnDOT algorithms, we can potentially reduce measurement errors by 1% to 23%. Due to the fourth power relationship between measured weight and Equivalent Single Axle Load (ESAL) used in pavement design, an improvement in only 1% can result in 4% increase in ESAL. A limitation of developing site-specific user-specific values like FAW references is that it would require very detailed and time-consuming data collection efforts to gather necessary data. This is potentially expensive and time consuming. The AVI approach, on the other hand, alleviates some of the need to perform manual field data collection by leveraging AVI truck-tracking technologies such as GPS.

Comparison of FAW and GVW estimation accuracy across all three auto-calibration methods and study sites can be summarized as follows:

- The ARDOT method reduced errors to between 12% and 16% for FAW and 14% to 29% for GVW.
- The MnDOT method reduced errors to between 11% and 26% for FAW and 11% to 41% for GVW.
- The AVI-based method reduced errors to between 10% and 35% for FAW and 16% and 35% for GVW.

In general, the AVI-based method works well for sites with higher measurement error as seen at Lonoke and Glen Rose but maintains similar performance as the ARDOT and MnDOT algorithms in most other cases. Performance of the AVI-based algorithm was also found to correlate with the volume of trucks tracked by the AVI system, in this case a GPS tracking system. When more trucks are present, lower FAW and GVW estimation errors were observed. Using the ARDOT method, there is a correlation between higher GVW ranges and increased error. The AVI-based approach did not exhibit this same trend.

FUTURE IMPROVEMENTS

Using a single data provider for AVI data (in this case GPS data) could be considered a limitation of the current methodology since the data may not be representative of all truck industries and cargo types and may, in some time periods, represent small sample sizes. Although we did not note the cargo configurations of all trucks in the AVI and static weight sample, most trucks were van trailers. This means that calibration factors do not incorporate different trailer types that might have different loading patterns. For example, liquid bulk tanks, livestock, and logging trailers may have very different loading patterns that effect the FAW variation and resulting calibration factors calculated via our proposed auto-calibration algorithm. In future work, we would like to consider a broader spectrum of AVI data sources such as various GPS and Electronic Logging Device (ELD) providers or license plate matching technology installed at WIM sites. Another related issue was the size of the AVI data sample. Our sample represented only a very small proportion of the total truck volumes. With a larger sample, we could compute more accurate *likely weights* within the AVI-based algorithm which could potentially increase the accuracy of the calibration factors.

Through two separate data collection efforts we were able to gather a sample of around 500 trucks to evaluate the performance of the three auto-calibration methods. However, due to restrictions on collecting and tracking license plate data, considerable effort was required to produce “ground truthed” matches from side-fire video. This was a time-consuming process due to the low number of trucks that entered the weight enforcement station relative to the total number of trucks that crossed each WIM site during the data collection and the need to manually verify matches based on visual descriptions of trucks using video recordings. In future work, it would be highly beneficial to use license plate readers to automatically match trucks across sites during data collection.

Weight estimation errors as high as 85% were observed for WIM sites with auto-calibration disabled. Applying the currently used ARDOT auto-calibration algorithm reduced errors to as low as 14%. This discrepancy points to the major source of measurement inaccuracy, quality of the WIM sensors. Although they are maintained adequately, the piezoelectric sensor in the WIM system has a measurement tolerance of 30% (FHWA, 2018 Part 3). With no temperature sensors at the sites to adjust weight measurements in accordance with pavement and ambient temperature changes, it is difficult to produce accurate weight measurements, even with the proposed AVI auto-calibration algorithm. Our results show variation in CFs by time of day indicating the effect of temperature on sensor performance.

Further, piezoelectric sensors have short life spans (2-3 years) but it is likely infeasible due to budget restrictions to replace sensors this frequently. As the sensors degrade, they become more sensitive to weather and pavement conditions. Although budget prohibitive, a solution might be to transition higher volume WIM sites into higher quality scales such as strain gage or bending plate scales and possibly even relocate or drop sites that experience low volumes in order to shift funding priority to heavily-trafficked areas. Bending plates and load cells have a 6% to 10% error in GVW and 15% to 20% error in axle loads while piezo electric sensors have 15% error associated with GVW and up to 30% error in axle loads.

Moreover, the AVI-based algorithm which tracks and compares truck weights across multiple sensors can be used to prioritize WIM site sensor upgrades. For example, we can consider a set of “anchor” sites which are commonly crossed by all trucks. Then, as we track trucks from these sites to “satellite” WIM sites, e.g., those with lower volume or lower quality sensors, we can use the measured weight at the “anchor” site as a reference by which to calculate a calibration factor for the “satellite” location.

RETURN ON INVESTMENT

Effective auto-calibration provides a means to reduce costs associated with on-site calibration using test trucks, e.g., the practice recommended by ASTM E1318-09. This has both direct and induced impacts on costs. Direct impacts are those related to the costs of performing routine calibration. The induced impacts are those related to the increased accuracy of WIM data due to routine calibration. Since WIM data is used in bridge and pavement design, data that are more accurate can lead to more efficient design and thus possible reduced costs. The direct benefits are quantified here.

A comparison of the cost of performing on-site calibration using test-trucks to effective auto-calibration using either field data collection or AVI data was made to show the direct impacts of the proposed project. A summary of the comparison is provided in Table 12 and explained in the following paragraphs. All scenarios assume that Arkansas's WIM sites are calibrated once per year and that staff labor costs can be neglected since all approaches have nearly the same labor costs.

Table 12. Summary of Calibration Costs by Calibration Method

Calibration Method	Annual Cost	Cost Savings
Test Trucks (status-quo method)	$366,976	-
Auto-Calibration Using Field Data Collection	$150,000	59%
Auto-Calibration Using AVI-Based Data Collection	$0	100%
Auto-Calibration Using Field and AVI-Based Data Collection	$135,000	63%

States Successful Practices Weigh-in-Motion Handbook provides an example of costs related to WIM installation, operations, and maintenance for a WIM network (McCall and Vodrazka, 1997). The following assumptions were made in estimating the total calibration costs for Arkansas (all costs have been converted from 1997 dollars to 2017 dollars[1]):

1. Assume there are 50 WIM sites in Arkansas.
2. Each scale will be independently calibrated once per year.
3. 10% of the WIM sites (5 sites) are near static scales: The cost of a single-lane calibration effort when that effort is performed by pulling trucks from the passing traffic stream and weighing them both at the WIM scale and a nearby static scale is $5,715.20 ($3,800 in 1997 dollars) per calibration.
4. 90% of the WIM sites (45 sites) must be calibrated using test trucks: when test trucks are used, the calibration approach relies on two loaded test trucks to make multiple passes over the WIM site. The weight of the test trucks is known apriori. The cost per calibration using test trucks is $7,520.00 ($5,000 in 1997 dollars) per calibration.
5. Increases in staffing costs related to labor and travel associated with on-site calibration are not included in the cost estimate.

Thus, the total cost of calibration for the WIM sites in Arkansas is $366,976 per year[2]. This represents the minimum calibration effort with only one calibration of each site per year. Ideally, the sites should be calibrated quarterly (4 times per year) to offset the effect of seasonal weather changes. With quarterly calibration efforts at every site, the total cost would be $1,467,906 per year.

[1] **Source:** The Bureau of Labor Statistics' annual Consumer Price Index (CPI) was 160.500 in 1997 and 241.353 in 2017. The purchasing power of $100 in 1997 is $150.38 in 2017 (the CPI inflation between 1997 and 2017 is calculated as 241.353/160.500).

[2] **Calculation**: 5 sites x 1 calibration/year x $5,715.20/calibration + 45 sites x 1 calibration/year x $7,520.00/calibration = $366,976 per year

Effective auto-calibration procedures could eliminate the need to perform annual or quarterly on-site calibration thus removing the calibration costs of up to $1.5 million per year. Effective auto-calibration requires periodic updates of the calibration factors used in the update procedure. This can be carried out in two ways as proposed in this project: (A) on-site data collection and (B) usage of Automatic Vehicle Identification (AVI) data from truck pre-clearance programs.

For (A), field technicians would need to collect and process video data collected from trucks passing both the WIM and static scales. This would require two technicians, e.g., one at the static scale and one at the WIM site, to set-up cameras to match trucks that traverse both sites. Based on quotes from the WIM system vendor, the average rate for one day of field data collection is $1,500 per site per technician. Assuming again that each site is evaluated once per year, approximately $150,000 per year would be needed to update auto-calibration algorithms using field data[3].

Alternatively, if AVI data were to be used in-place of field data collection, then the field data collection costs would be eliminated. The costs associated with the use of AVI data for auto-calibration are expected to be minimal given that weigh station bypass programs like PrePass are willing to share data with the ARDOT at minimal or no cost in return for access to weight data. Thus for (B), the costs would be limited to staff labor hours which are assumed the same across all three approaches, e.g., on-site calibration with test trucks, on-site calibration using traffic stream vehicles, and AVI based approaches, and therefore not included in the comparative analysis.

Since it may not be realistic to assume that all WIM sites can be calibrated using AVI data, e.g. some WIM sites may not be along the same routes as the PrePass sites, a combination of (A) and (B) is more realistic. Assuming that 10% of the WIM sites (5 sites) can use AVI-based calibration and the remaining 90% (45 sites) require field data collection to update auto-calibration parameters, the total cost would be $135,000 per year.

[3] **Calculation**: $1,500/site/technician x 2 technicians/site x 50 sites/year = $150,000 per year

REFERENCES

1. Abbas, A. R., Frankhouser A., and Papagiannakis A. T., 2014. “Effect of Traffic Load Input Level on Mechanistic–Empirical Pavement Design”, Transportation Research Record No 2443.

2. Al-Qadi, I., Wang, H., Ouyang, Y., Grimmelsman, K., and Purdy, J.. 2016. "LTBP Program's Literature Review on Weigh-in-Motion System." Final report for the USDOT FHWA.

3. ASTM Standard E1318- 09. 2009. “Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Methods”, ASTM international, West Conshohocken, PA, www.astm.org

4. Bunnell, W., Li, H., Reed, M., Wells, T., Harris, D., Antich, M., and Bullock, D. 2018. "Weigh-in-Motion Sensor Calibration Using Outlier Analysis." Presented at the 2018 Meeting of the Transportation Research Board, Washington, D.C., January 7-11, 2019.

5. Chu, L. and Jeng, S., 2015. “Tracking Heavy Vehicles based on Weigh-In-Motion and Inductive Loop Signature Technologies.” IEEE Transactions on Intelligent Transportation Systems. Vol. 16, No. 2, April 2015.

6. Dahlin, C.. 1992. 'Proposed method for calibrating weigh-in-motion systems and for monitoring that calibration over time', Transportation Research Record: 161-61.

7. Davies, P. and Sommerville, F. 1987. “Calibration and accuracy testing of weigh-in-motion systems”. Transportation Research Record, Journal of the Transportation Research Board, Vol. 1123, pp 122-126.

8. Dixon, M. P., and L. R. Rilett. 2002. “Real-Time OD Estimation Using Automatic Vehicle Identification and Traffic Count Data.” Computer-Aided Civil and Infrastructure Engineering, Vol. 17, Issue 1, January 2002, pp. 7-21.

9. FHWA. 2016. “Traffic Monitoring Guide.” U.S. Department of Transportation/Federal Highway Administration, accessed online at https://www.fhwa.dot.gov/policyinformation/tmguide/tmg_2013/vehicle-types.cfm.

10. FHWA. 2017. “Verification, Refinement, and Applicability of Long-Term Pavement Performance Vehicle Classification Rules.” U.S. Department of Transportation/Federal Highway Administration, 31 Jan. 2017, www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/13091/002.cfm.

11. FHWA. 2018. “Weigh-In-Motion Pocket Guide Part 1 WIM Technology, Data Acquisition, and Procurement Guide.” FHWA Publication FHWA-PL-18-015.

12. FHWA. 2018. “Weigh-In-Motion Pocket Guide Part 3 WIM Calibration and Maintenance Guide”. FHWA Publication FHWA-PL-18-015.

13. Fiorillo, G. and Ghosn, M. 2014. “Procedure for statistical categorization of overweight vehicles in a WIM database.” Journal of Transportation Engineering, Vol. 140. Issue 5. May 2014

14. Gupta, D., Tang, X., and Yuan, L. 2018 "Weigh-in-Motion Sensor Controller Operation and Performance Comparison." *MnDOT Research and Services & Library*, January 2018. *University of Minnesota Department of Industrial Systems Engineering.*

15. Hallenbeck, M. 1998. "Long term pavement performance program protocol for calibrating traffic data collection equipment". Report for the Long Term Pavement Performance Program of the FHWA, accessed online at https://rosap.ntl.bts.gov/view/dot/4752

16. Hashemi V., Haas, C., Rothenburg, L., and Haas, R. 2013. "Investigation of the Effect of Weight Factor on Performance of Piezoelectric Weigh-in-Motion Sensors", Journal of Transportation Engineering, Vol. 139, Issue 9. September 2013.

17. Hyun, K., Hernandez, S., Tok, A., and Ritchie, R. 2015. "Truck Body Configuration Volume and Weight Distribution: Estimation by Using Weigh-in-Motion Data", Transportation Research Record: Journal of the Transportation Research Board. Vol. 2478. Issue 1.

18. Hyun, K., Tok, A., and Ritchie, S. 2017. "Long distance truck tracking from advanced point detectors using a selective weighted Bayesian model." Transportation Research Part C, Vol. 82, September 2017, pp. 24–42.

19. Jiang, X., Vaziri, S., Haas, C., Rothenburg, L., Kennepohl, G., and Haas, R. 2009. "Improvements in piezoelectric sensors and WIM data collection technology." Presented at the Annual Conference & Exhibition of Transport Association of Canada. Ontario, Canada.

20. Liao, C. 2012. "Traffic Data Quality Verification and Sensor Calibration for Weigh-In-Motion (WIM) Systems." Department of Civil Engineering University Minnesota, August 2012.

21. Mai, D., Turochy, R., and Timm, D. 2013. "Quality control of weigh-in-motion data incorporating threshold values and rational procedures". Transportation Research Part C: Emerging Technologies, Vol. 36. Pp. 116-24.

22. McCall, W. and Vodrazka, W.C. 1997. "States' Successful Practices Weigh-in-Motion Handbook", Iowa State University, Ames, Federal Highway Administration, Washington, D.C., December 1997.

23. Nassif, H., Na, C., Portela, E., and Bittencourt, T. 2018. "An Approach to Minimize Weighing Error of WIM System Based on Ambient Temperature." Presented at the 2018 Meeting of the Transportation Research Board, Washington D.C., January 7-11, 2018.

24. Nichols, A. P., and Cetin, M. "Use of Reidentified Vehicles to Evaluate Differential Calibration Accuracy Between Weigh-in-Motion Stations." Transportation Research Record: Journal of the Transportation Research Board, no.2527, 2015, pp. 37–48.

25. Nichols, A. P, and Darcy M Bullock. 2006. "Automatic speed calibration methodology for traffic monitoring sites", Journal of Transportation Engineering, Vol. 132, Issue 1. January 2006.

26. Ozbay, S. and Ercelibi, E. 2007. "Automatic Vehicle Identification by Plate Recognition." International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering. Vol. 1. No. 9.

27. Papagiannakis, A., and Quinley, R. 2008. “High speed weigh-in-motion system calibration practices”. National Cooperative Highway Research Program Synthesis of Practice. Transportation Research Board, 63p.

28. Quinley, R.. 2010. "WIM data analyst's manual." Technical Report for the FHWA. Report Number FHWA-IF-10-018.

29. Sayyady, F., Fathi, Y., List, G., and Stone, J. 2013. “Locating Traffic Sensors on a Highway Network: Models and Algorithms”, Transportation Research Record, Vol. 2339. Issue 1, 2013.

30. Tam, W. O. and Quintus, H. V. 2003.“Use of Long-Term Pavement Performance Data to Develop Traffic Defaults in Support of Mechanistic-Empirical Pavement Design Procedures”, Aviles Engineering, ERES Consultants, Transportation Research Record 1855, January 1, 2003.

APPENDIX A: AVI-BASED AUTO-CALIBRATION IMPLEMENTATION EXAMPLE

An illustrative example of the Truck Matching and Auto-calibration methods are provided here using WIM stations Lamar, Lonoke and Bald Knob.

TRUCK-MATCHING EXAMPLE

Vehicle inter axle spacing in feet were utilized in the example as the vehicle parameter to compare in order to find the WIM records for AVI trucks.

Step 1: Determine Time Offsets

The AVI to WIM time offsets at Lamar, Lonoke and Bald Knob are found to be 15s, 20s, and 12s, respectively.

Step 2: Identify Candidate WIM Records

An AVI truck with unique ID ‘123’ is identified crossing these sites within one day at the following times.

$[tuck_{123}]$: $Lamar\ @\ t_1 = 9{:}\,00{:}\,00\ AM$

$Lonoke\ @\ t_2 = 10{:}\,00{:}\,00\ AM$

$Bald\ Knob\ @\ t_3 = 10{:}\,30{:}\,00\ AM$

d_{123}:

Site 1	Time 1	Site 2	Time 2	Site 3	Time 3
$Lamar_{360009}$	$9{:}\,00{:}\,00\ a.m.$	$Lonoke_{430037}$	$10{:}\,00{:}\,00\ a.m.$	$Bald\ knob_{730068}$	$10{:}\,30{:}\,00\ a.m.$

Step 2: Assign WIM records to AVI records. A buffer of Δ = 5 minutes was used in this example to find sets of candidate WIM records at each site.

$Candidate\ WIM\ records\ for\ tuck_{123} 5\min around\ \ t_1 = 9{:}\,00{:}\,00\ AM\ at\ Lamar \rightarrow$

Record	Time	AVI Time	WB 1	WB 2	WB 3	WB 4
90	8:55:15 a.m.	8:55:00 a.m.	17.50	4.02	28.78	4.64
…	…	…	…	…	…	…
95	9:00:15 a.m.	9:00:00 a.m.	17.00	4.20	32.00	4.15
…	…	…	…	…	…	…
100	9:05:15 a.m.	9:05:00 a.m.	15.88	4.23	32.98	4.71

Where AW 1, 2… is the weight of the 1st, 2nd, etc. axle. Note that inter-axle spacing could also be included.

$Candidate\ WIM\ records\ for\ tuck_{123} 5\min around\ \ t_2 = 10{:}\,00{:}\,00\ AM\ at\ Lonoke \rightarrow$

Record	Time	AVI Time	WB 1	WB 2	WB 3	WB 4
200	9:55:20 a.m.	9:55:00 a.m.	16.72	4.21	32.93	4.00

…	…	…	…	…	…	…
220	10:00:20 a.m.	10:00:00 a.m.	17.00	4.20	32.00	4.15
…	…	…	…	…	…	…
240	10:05:20 a.m.	10:05:00 a.m.	15.99	4.13	30.84	4.02

$Candidate\ WIM\ records\ for\ tuck_{123} 5\min around\ \ t_2 = 10{:}30{:}00\ AM\ at\ Bald\ Knob \rightarrow$

Record	Time	AVI Time	WB 1	WB 2	WB 3	WB 4
172	10:25:12 a.m.	10:25:00 a.m.	19.50	4.28	31.50	3.96
…	…	…	…	…	…	…
175	10:30:12 a.m.	10:30:00 a.m.	17.00	4.20	32.00	4.15
…	…	…	…	…	…	…
180	10:35:12 a.m.	10:35:00 a.m.	17.69	5.72	29.56	4.32

The vehicle parameters differences were used to determine potential matches. The vehicle parameter used in this example was axle weights. For instance, WIM record 95 at Lamar would be compared to records 200 thru 240 at the Lonoke WIM site and to records 172 thru 180 at the Bald Knob site. Notice that records 95, 220 and 175 would have the least overall difference. The following are GVW differences:

A. Data vector for Lamar:

$$W_{95,Lamar} = [\ 17.0, 4.2, 32.0, 4.15]$$

B. Differences between truck 95 at Lamar and candidates at Lonoke:

$$W_{95,Lamar} - W_{220,Lonoke} = 0$$

$$W_{95,Lamar} - W_{200,Lonoke} = 1.47$$

$$W_{95,Lamar} - W_{240,Lonoke} = 2.37$$

Resulting match: $argmin\big([0, 1.47, 2.37], \{W_{s,i}, W_{s'\ \ ,j}\}\big) = (\ W_{95,Lamar}, W_{220,Lonoke})$

C. Differences between truck 95 at Lamar and candidates at Bald Knob:

$$W_{95,Lamar} - W_{175Bald\ Knob} = 0$$

$$W_{95,Lamar} - W_{172,Bald\ Knob} = 3.27$$

$$W_{95,Lamar} - W_{345,Bald\ Knob} = 4.82$$

Resulting match: $argmin\big([0, 3.27, 4.82], \{W_{s,i}, W_{s'\ \ ,j}\}\big) = (\ W_{95,Lamar}, W_{228,Bald\ Knob})$ as $W_{175Bald\ Knob}$ has a closer time stamp to the AVI given the time offset, and least overall difference in axle weights as well.

Records 95, 220 and 175 have the lowest overall difference therefore these records are assigned to truck 123 for auto-calibration. Thus, the final unique WIM pairings for AVI truck '123' are as follows:

Site	Record	Time	WB 1	WB 2	WB 3	WB 4
Lamar	95	9:00:12 a.m.	17.00	4.20	32.00	4.15
Lonoke	220	10:00:12 a.m.	17.00	4.20	32.00	4.15
Bald Knob	175	10:30:12 a.m.	17.00	4.20	32.00	4.15

The example above is suited for a WIM system network that does not have a lot of variability in weight recordings preferably using WIM systems with bending plates or load cells which are more accurate than piezoelectric sensors and WIM systems that are not Type II as they have a higher variability of 30% combined with the sensitivity to temperatures and pavement conditions of piezoelectric sensors (FHWA, 2018). For the case study, inter-axle spacing was selected as the comparable vehicle parameter as these are more consistent measurements across FHWA vehicle class and due to the wide range in weight variability across WIM sites that was experienced in the recorded data.

AUTO-CALIBRATION EXAMPLE

The following is an idealized example of the AVI auto-calibration method presented in Flow Chart 4. In this example a 1-hour sample of trucks taken from 9 a.m. to 10 a.m. at WIM Station A includes three trucks: truck IDs 101, 105, and 203. WIM records for the same AVI trucks found at WIM Station A within time window *T* of 3.5 hours included three additional stations: B, C, and F. The reference front axle weight, W_R, was set to 10 kips. The deviation among front axle weights among the sites (δ_S) was 10%. The deviation (δ_W) between the reference weight and *likely weight* was 10%. The volume of FHWA Class 9 trucks at each WIM site on the given day were: 300, 250, 190, and 50, for sites A, B, C, and F, respectively. The algorithm is as follows:

1. <u>Select samples:</u> Obtain a set of FHWA Class 9 trucks crossing WIM site A from the AVI data.

AVI Truck Sample at WIM A from 9 a.m. to 10 a.m.

Truck ID	Timestamp	FAW (kips)
105	9:03:00	10
101	9:05:0	12
203	9:30:00	9

2. <u>Find AVI trucks:</u> Find each of the AVI trucks from site A that traversed other WIM sites in the 3.5 travel time window and count the number of sites crossed by each truck: truck 105 crossed 4 sites, truck 101 crossed 3 sites, and truck 203 crossed three sites.

Trucks Cross WIM B:

Truck ID	Timestamp	FAW (kip)
105	10:03:00	9
101	10:05:00	11
203	10:30:00	10

Trucks crossing WIM C:

Truck ID	Timestamp	FAW (kip)
105	11:33:00	10

101	11:35:00	10
203	12:00:00	12

Trucks crossing WIM F:

Truck ID	Timestamp	FAW (kip)
105	13:03:00	8

2. Check Deviation: Check the deviation, δ_S, in *front axle weight* of each AVI truck recorded at each WIM site to see if the sites require calibration. For this example, δ_S was 10% so that if the difference in *front axle weights* (or FAW) recorded at two WIM sites for the same truck differed by more than 10%, we considered them to need calibration. The following table shows necessary calculations:

Deviation for sites A and B:

Truck ID	Timestamp A	FAW A	Timestamp B	FAW B	Difference |A-B|/A
105	9:03:00	10	10:03:00	9	10%
101	9:05:00	12	10:05:00	11	8%
203	9:30:00	9	10:30:00	10	11%

Deviation for sites B and C:

Truck ID	Timestamp B	FAW B	Timestamp C	FAW C	Difference |B-C|/B
105	10:03:00	9	11:33:00	10	11%
101	10:05:00	11	11:35:00	10	9%
203	10:30:00	10	12:00:00	12	20%

Deviation for C and F:

Truck ID	Timestamp C	FAW C	Timestamp F	FAW F	Difference |C-F|/C
105	11:33:00	10	13:03:00	8	20%

By computing the deviations above it may be observed that in most cases the weights for a an AVI truck are significantly different in all cases except for truck 101 where the weights recorded in sites A and B are below the deviation, therefore they are similar. In this case, the calibration factor resulting from truck 101 recorded at sites A and B is 1.0.

3. Find *Likely Weight*: To compute the *likely weights*, ω_a, we first differentiate between high and low volume sites based on historical AVI data such that a site with over 50 AVI trucks per day was considered to be high volume. This is referenced via a look up table. Clustering is used to find ω_a when there is more than one high volume site. Then the ω_a was compared to the reference FAW, W_R, of 10 kips to see if it is within a weight deviation, δ_W, of 10%. If deviation between ω_a and W_R exceeds δ_W then W_R was used to compute the calibration factor, otherwise ω_a was used. A calculation for Truck 105 was as follows:

Truck 105:

- FAWs were 10, 9, 10, and 8 for sites A, B, C, and F
- High-volume sites = A, B, C
- ω_a = 9.75 kips from clustering analysis (e.g., cluster with 10, 9, and 10 kip *front axle weights* and GVWs)
- Deviation to reference weight: (9.75-10.00)/10.00 x 100% = 2.5%
- Comparison to threshold: $\delta_W = 10\%$ > 2.5%, therefore use ω_a = 9.75 kips

4. Calculate Calibration Factors: The calibration factors were calculated as the ratio of the *likely weights*, ω_a, to the recorded FAW. An example for site A is:

Site A:

Truck ID	Timestamp A	FAW A	Likely Weight A	CF
105	9:03:00	10	9.75	9.75/10 = 0.975
101	9:05:00	12	11.20	11.20/12 = 0.933
203	9:30:00	9	10.40	10.40/9.0 = 1.15
Average	*9:00 to 10:00*	-	-	*1.02*

Since site F was a low-volume site, the *likely weight* of truck 105 determined from clustering FAWs from sites A, B, and C was used to calculate the calibration factor for Site F as follows.

Site F:

Truck ID	Timestamp F	FAW F	Likely Weight A	CF
105	13:03:00	8	9.75	9.75/8 = 1.22
Average	*13:00 to 14:00*	-	-	*1.22*

5. Calibrate Site: The resulting calibration factors generated from the AVI trucks were used to adjust the weights recorded by the WIM for all trucks by dividing each of the WIM measured weights by the calibration factor.

APPENDIX B: MANUAL IDENTIFICATION OF TRUCKS ACROSS WIM SITES

The following steps explain how trucks were identified and matched once the time offsets were found using the Lamar and Lonoke sites as an example:

1. Watch the video at Lamar of trucks near the Drivewyze timestamp considering the 17s offset between the video and DW records. Take screen shots (Figure B.1) and notes of the trucks.

(A)

(B)

(C)

Figure B.1. Example Truck Images from Video at Lamar WIM Site

2. Observe the video footage from Lonoke near the GPS timestamp considering the offset between the video and GPS records at Lonoke (1min 45s - 3min) to find if any of the trucks from the captured images at Lamar cross the Lonoke WIM site. Figure B.2 shows the truck found at Lonoke, which corresponds to a truck that previously crossed Lamar site in Figure 13.

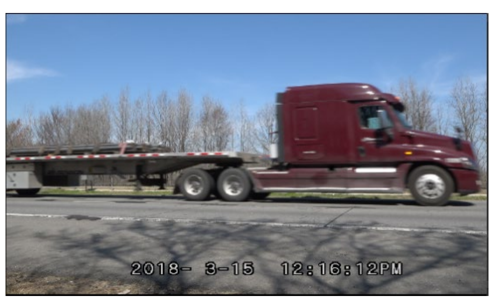


Figure B.2. Example of Truck Reidentified at Lonoke WIM Site

3. Record the "matched" truck (Figure B.3).

10:30:19	Lamar notes				match at Lonoke	
10:29:32	white truck	FEDEX trailer	12	lane 1		
10:30:05	red sleeper (darkred) white letters	carrying met:	9	lane 2	12:16:12	lane 2
10:30:17	red sleeper, white letters	white trailer	9	lane 2		

Figure B.3. Example of Notes Used to Match Trucks Across WIM Sites

This process resulted in a list of trucks and their WIM measurements for trucks that crossed Lamar and Lonoke (Figure B.4). A challenge was that the AVI to video time offset had some variability due to traffic flow at each site. GPS records were matched more precisely to WIM records by examining headways of WIM and video records. This was performed in order to obtain one-to-one matches at Lamar and Lonoke between the GPS and WIM vehicle records also looking at the truck lane and class sequence to find the exact match in order to develop the data to validate truck-matching and auto-calibration algorithms.

DW Timestamp		Video Timestamp		Video to DW Time Diff		WIM ID		WIM Timestamp	
Lamar	Lonoke	Lamar	Lonoke	Lamar	Lonoke	Lamar	Lonoke	Lamar	Lonoke
10:48:07	12:31:37	10:47:51	12:33:30	-0:00:16	0:01:53	**1852**	**5430**	10:48:05	12:34:28
10:48:36	12:26:52	10:48:19	12:28:34	-0:00:17	0:01:42	**1854**	**5371**	10:48:34	12:29:33
10:51:13	12:38:39	10:50:59	12:40:42	-0:00:14	0:02:03	**1864**	**5500**	10:51:14	12:41:40
10:49:58	13:18:12	10:49:44	13:20:15	-0:00:14	0:02:03	**1857**	**5884**	10:49:59	13:21:04

Figure B.4. Example of Finalized Matches Between Video, GPS, and WIM Records