5-2019

Auto-Calibration of WIM Using Traffic Stream Characteristics

Johnson Baker
University of Arkansas, Fayetteville

Auto-Calibration of WIM Using Traffic Stream Characteristics

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Civil Engineering

by

Johnson Baker
Clemson University
Bachelor of Science in Civil Engineering, 2012

May 2019
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

____
Sarah Hernandez, Ph.D.
Thesis Director

Clinton Wood, Ph.D. Committee Member	Kevin Hall, Ph.D. Committee Member

Abstract

This project evaluates the performance of Weigh-in-Motion (WIM) auto-calibration methods used by the Arkansas Department of Transportation (ARDOT). Typical auto-calibration algorithms compare the WIM measured weights of vehicles from the traffic stream to reference values, using five-axle tractor-trailer configured trucks for comparisons, e.g. Federal Highway Administration (FHWA) Class 9. Parameters of the existing algorithms including the Front Axle Weight (FAW) reference value, the sampling frequency required to update the calibration factor, and thresholds for Gross Vehicle Weight (GVW) bins were evaluated. The primary metric used to estimate algorithm performance was Mean Absolute Percent Error (MAPE) between the WIM and static scale GVW measurements.

Two altered auto-calibration algorithms based on methodologies utilized by ARDOT and the Minnesota DOT (MNDOT) were developed. Parameters for the algorithms are based on combinations that produced minimal MAPE at the study sites. WIM data from two sites (Lamar and Lonoke) and static scale data from one site (Alma) were collected along Eastbound I-40 in Arkansas during March 2018. The updated MNDOT auto-calibration algorithm reduced the MAPE by 2.5% compared to the baseline method at the Lamar site (n = 77 trucks) and by 1.1% for the Lonoke site (n = 44 trucks). The updated ARDOT algorithm reduced MAPE by 1.6% at the Lamar site and 0.6% at the Lonoke. Due to limitations of the field data collection methodology, the thresholds defining FAW reference values and the FAW reference values themselves were not able to be tested for spatial transferability, e.g. the samples of trucks at the Lonoke WIM site were a subsample of the trucks at the Lamar WIM site. Improvements in auto-calibration accuracy at

low volume sites but was not tested due to the small number of confirmed vehicle matches at a third WIM site (Bald Knob, n = 2 trucks).

Overall, site specific tuning of auto-calibration algorithms will improve the accuracy of WIM data which is used for pavement design, maintenance programming, and traffic forecasting. For example, improvements of 2.5% MAPE of WIM measured GVW results in a 39% difference in the estimated Equivalent Single Axle Load (ESAL) factors used for pavement design.

Acknowledgements

This research was possible with the guidance of Dr. Sarah Hernandez, Ph.D., and assistance of Fu Durandal, Taslima Akter, Magdalena Asborno, Joshua Vines, ARDOT Research Committee, and the Arkansas State Police.

I would also like to thank Karen Baker Minton, Quin Minton, Liz Cowle, and Clare Kane for their continuing support during this research.

Table of Contents

List of Figures

List of Tables

Abbreviations

AASHTO	American Association of State Highway and Transportation Officials
AIC	Akaike Information Criterion
ASTM	American Society for Testing and Materials, formerly
B-WIM	Bridge Weigh-in-Motion
BIC	Bayesian Information Criterion
EB	Eastbound
ESAL	Equivalent Single Axle Loads
FAW	Front Axle Weight
FHWA	Federal Highway Administration
GOF	Goodness of Fit
GMM	Gaussian Mixture Model
GVW	Gross Vehicle Weight
QC	Quality Control
TRB	Transportation Research Board
WIM	Weigh-in-Motion

1. Introduction

1.1. Overview

The goal of this study is to develop a protocol for the auto-calibration of weigh-in-motion (WIM) sites in Arkansas. WIM technology allows for the collection of large quantities of data without disturbing the traffic stream at the cost of lower accuracy when compared to the data produced by static scales. The degree of accuracy required of WIM data depends upon its intended use: enforcement, data collection, or both (Papagiannakis and Quinley 2008). Due to the limitations of static scale weigh stations, WIM sites are a primary tool used to gather traffic data and assist enforcement efforts. It is therefore essential to minimize error in the data and monitor the calibration of the weight sensor.

Errors can be reduced through effective calibration. WIM calibration requires test trucks to make multiple passes across the sensors at set speeds as prescribed in ASTM E1318. This procedure is time consuming and expensive due to the equipment and labor costs and as a result is not implemented in Arkansas. In fact, many states who do use test trucks to calibrate their WIM scales only follow the ASTM method partially (Papagiannakis and Quinley 2008). Arkansas instead utilizes a basic auto-calibration algorithm on WIM scales.

1.2. Auto-Calibration

Auto-calibration is an umbrella term that describes any algorithmic procedure which, once triggered, will periodically and automatically adjust data output. For WIM data, this criterion is usually the average front axle weight (FAW) and gross vehicle weight (GVW) compared against a reference value. These reference values are adjusted based on presumed or measured traffic stream characteristics. State-of-the-practice protocols differ in which specific vehicle element is analyzed. Florida DOT, Texas DOT and Caltrans monitor the average GVW, while Indiana DOT and Minnesota DOT monitor both the GVW and the FAW (Papagiannakis and Quinley 2008).

Standard practice in WIM calibration, as prescribed in ASTM E1318, is to use FHWA Class 9 trucks as test vehicles of known weight to perform the required passes over the WIM sensors. Vehicles are assigned to categories according to the classification scheme introduced by the Federal Highway Administration (FHWA) as shown in Figure 1. Vehicles are categorized according to their axle count and trailer configuration if a trailer is present. Commercial trucks are assigned to categories 5-13. FHWA Class 9 vehicles are composed of trucks that have one trailer and five axles. This is usually presented in the '3S2' configuration, which is a steering axle followed by tandem drive axles, and a trailer having tandem axles. However it is possible that the trailer axles are not in tandem, or even that a 2-axle truck may be pulling a 3-axle trailer. The use of FHWA Class 9 vehicles as the basis for WIM calibration was first proposed by Dahlin in 1983 (Davies and Sommerville 1987). This arrangement works well as FHWA Class 9 vehicles comprise 80% of all trucks (classes 5-13) (Fiorillo and Ghosn ; Nichols and Bullock) and are responsible for 70-90% of ESALs (Dahlin 1992). However, 3S2 trucks possess the largest variety of body configurations (Hyun et al. 2015) which affects the dynamic forces acting on the vehicle.

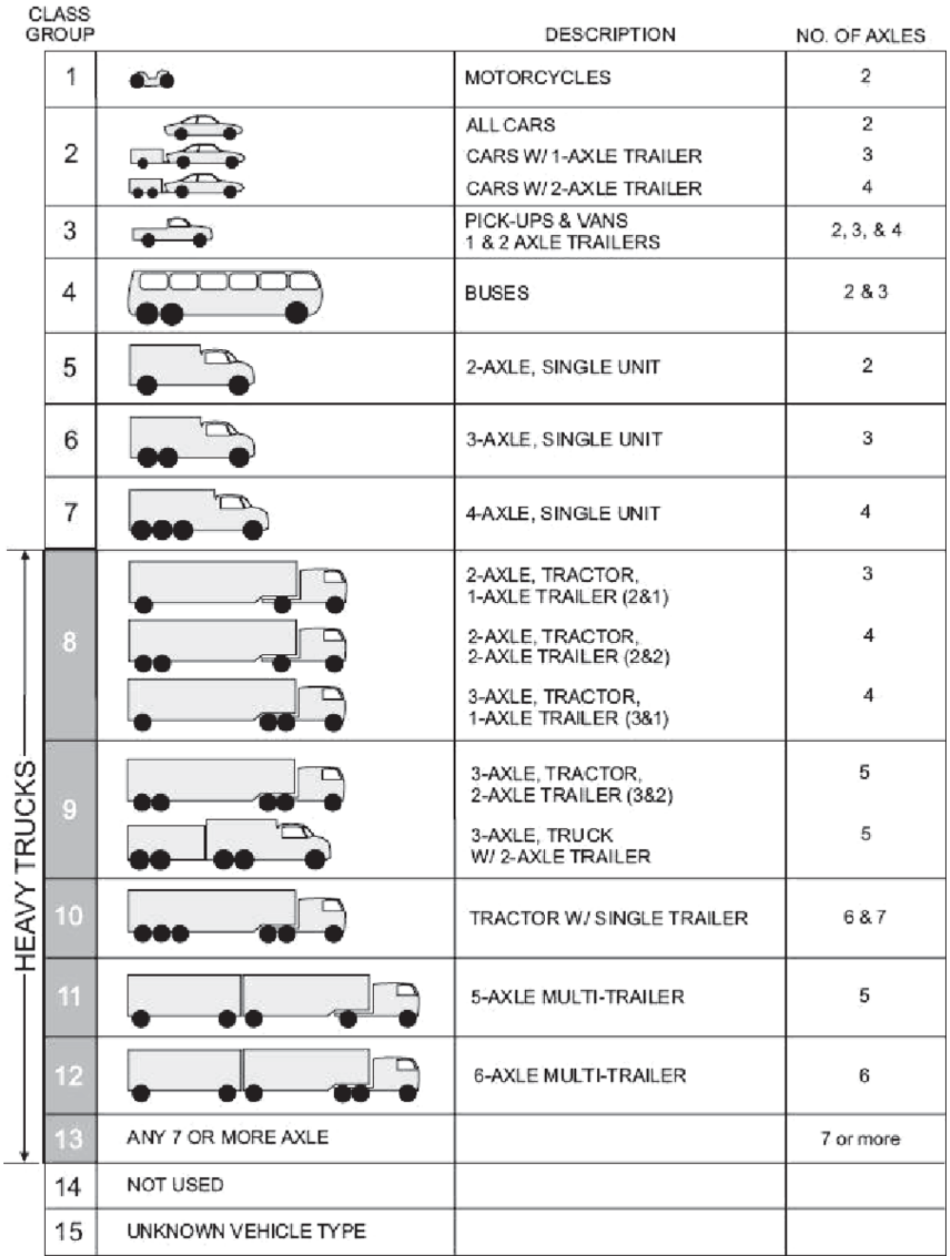


Figure 1. FHWA Vehicle Classification Scheme (Sayyady et al. 2013).

In Arkansas, the WIM auto-calibration algorithm computes a calibration factor by measuring 50 FHWA Class 9 trucks so that the product of the average measured weight and the calibration factor is equal to the expected average weight. All further measurements are then adjusted by the calibration factor. This method has several limitations. There are WIM sites throughout the state that experience low traffic volumes where it can take a long time to accumulate 50 measurements.

This, in turn, may not account for changing traffic stream characteristics or other time dependent factors. Error in data output can be caused by temperature changes in the ambient air and pavement, pavement roughness, and vehicle speed as shown in Figure 2. Only with a full understanding of WIM site characteristics can an effective auto-calibration algorithm be developed and employed.

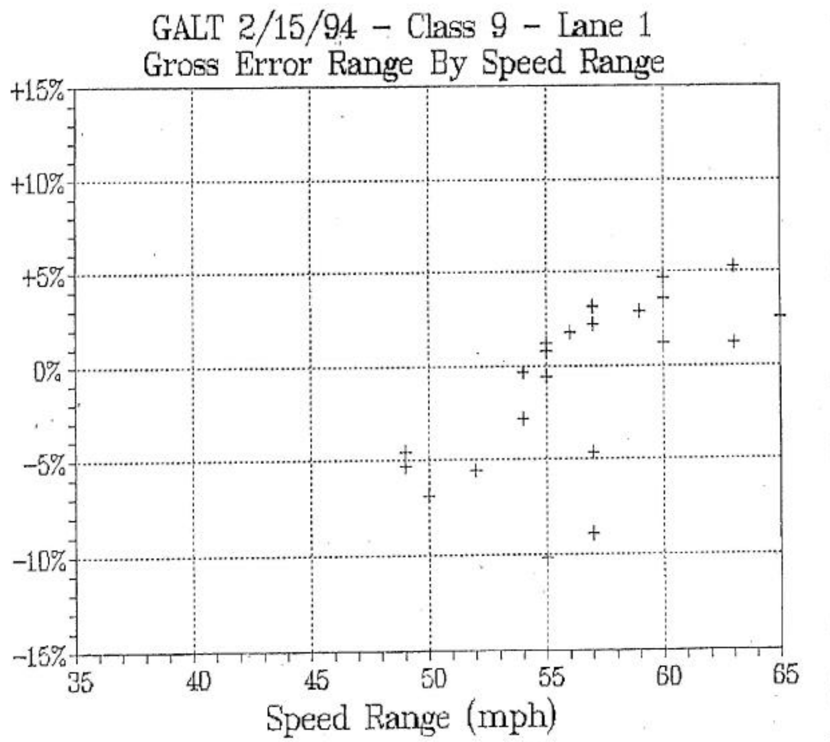


Figure 2. Effects of Speed on WIM Error (McCall and Vodrazka Jr 1997).

The proposed method to improve the auto-calibration algorithm in Arkansas will calibrate the WIM sensors by comparing an established reference value against measured data from traffic stream vehicles rather than a static scale measurement. This removes the requirement for static scale data in the regular calibration process. The reference value may be established during the initial site calibration or dictated based upon changing traffic stream characteristics. This method would be self-contained, meaning it does not require any additional weight data other than what is produced on site. This method offers cost savings over the ASTM standard procedure since there would be no human involvement required in regular operation. This research focused on the development of the improved auto-calibration algorithm and tuning algorithm parameters for application across Arkansas.

To develop and test the effectiveness of the algorithm, ground truth weight data was collected from the static scale enforcement site located in Alma, Arkansas and video and WIM data was collected at three Arkansas WIM sites located in Lamar, Lonoke, and Bald Knob. Using this data, a calibration analysis was performed to determine the effectiveness of the current algorithm ("baseline method"). Additionally, the WIM vendor suggested validating site characteristics to improve auto-calibration functions. Characteristics that were investigated include: site temperature, GVW distributions, average FAW values, and load status (fully loaded or not fully loaded).

1.3. Data Quality Control Checks

The auto-calibration process is triggered by data quality control (QC) checks. Data QC is an important part of the diagnostic process; it involves periodic analyses of WIM measurements to retain confidence that the sensing equipment is functioning properly and has not experienced calibration drift (Papagiannakis and Quinley 2008). In addition to identifying possible equipment problems, data QC will also assist the WIM user in ensuring data accuracy and precision (McCall and Vodrazka Jr 1997). An example of data QC is checking that measurements fall within a reasonable range, e.g. tandem axle spacing, FAW of FHWA Class 9 vehicles, and GVW distribution peak locations. It is recommended that auto-calibration not replace data QC, but rather act as a supplement, as data QC can alert users to issues more serious than calibration drift such as instrument malfunctions. Most state agencies recognize the importance of data QC and use it in their calibration practices (Papagiannakis and Quinley 2008).

2. Review of Relevant Literature

2.1. Weigh-in-Motion Technology

Weigh-in-Motion is the process of measuring the dynamic tire forces of a moving vehicle and estimating the corresponding tire loads of the static vehicle (Hallenbeck 1998). Development of WIM technology began in the United States in the early 1950's for use in collecting truck weight data (Al-Qadi et al. 2016). WIM usage expanded in the 1970's with the development of bridge WIM (B-WIM) technology (Al-Qadi et al. 2016) and advancements in computing technology (Bunnell et al. 2018). More recently, the American Association of State Highway and Transportation Officials (AASHTO) designated WIM as a "concept of focus technology" in 2004 (Al-Qadi et al. 2016).

A WIM system is comprised of a set of sensors and supporting instruments that process, display, store, and transmit the information collected from the sensors ("Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Methods" 2017). WIM scales can be installed in the mainline which enables measurements to be taken from vehicles without removing them from the traffic stream or reducing their speed. This contrasts with static scale technology employed at weigh stations which requires vehicles to leave the traffic stream and come to a complete stop to allow for weight measurements to be collected. WIM scales serve three primary applications: collecting statistical data, aiding enforcement, and enforcement (Papagiannakis and Quinley 2008). In addition to axle weights, a WIM system can collect data on a vehicle's number of axles, axle spacing, vehicle speed, and GVW as shown in Table 1. These characteristics allow for the calculation of equivalent single-axle loads (ESALs), which is a relative metric by which truck weights can be evaluated in terms of expected impact to the pavement structure. The WIM system software will also estimate the truck class by extrapolation from the axle configurations. This WIM data helps the states Departments of Transportation plan for pavement maintenance, inform pavement design, and better understand freight movement around the state. Additionally, WIM data is utilized in research regarding pavement and bridge design and helps fulfill federal reporting requirements. As shown in Figure 3, even small errors in axle weight measurements can result in large errors when estimating ESALs, and therefore in estimating expected pavement life (Hallenbeck 1998). This is due to the 4th power relationship between axle weight and ESALs (Dahlin 1992; Hallenbeck 1998).

Table 1. WIM Data Outputs ("Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Methods" 2017).

WIM Data output
Wheel Load
Axle Load
Axle-Group Load
Gross-Vehicle Weight
Speed
Center-to-Center Spacing Between Axles
Vehicle Class (via axle arrangement)
Site Identification Code
Lane and Direction of Travel
Date and Time of Passage
Sequential Vehicle Record Number
Wheelbase (front-most to rear-most axle)
Equivalent Single-Axle Loads (ESALs)
Violation Code

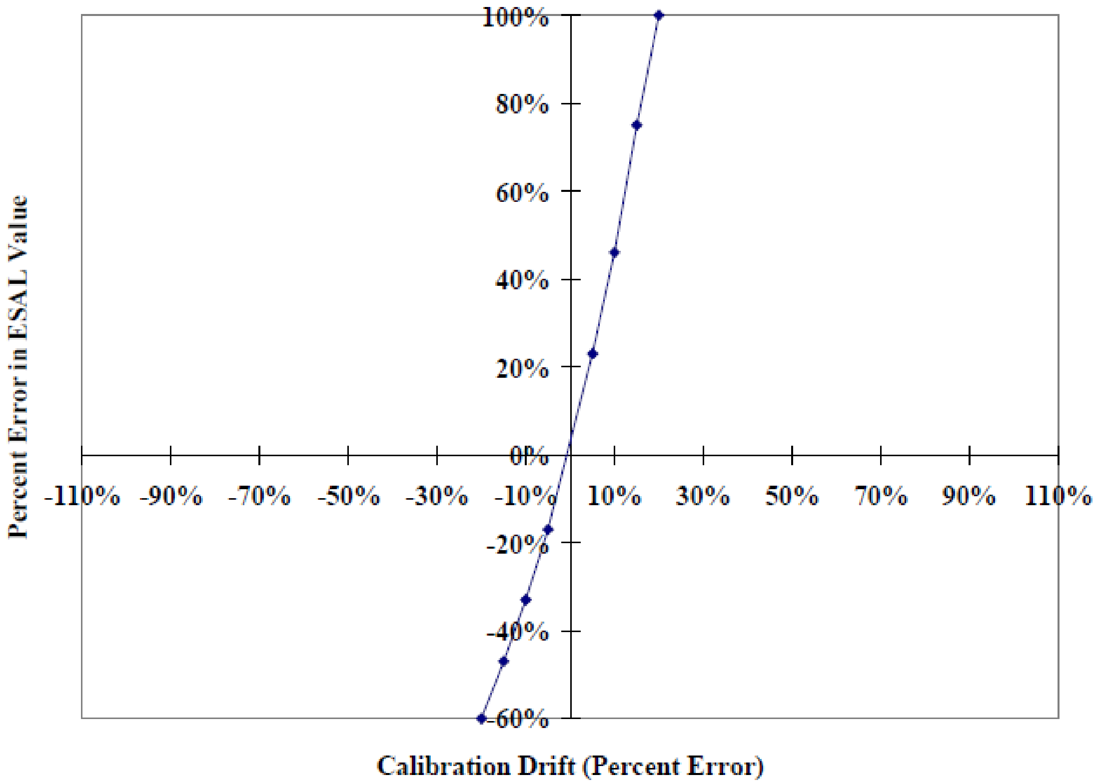


Figure 3. Relationship Between Calibration Drift and Error in ESAL Computations (Hallenbeck 1998).

2.2 WIM Components

The components of a WIM system include sensors used to detect and weigh vehicles, a system controller, communications hardware to transmit data, an equipment cabinet, and either a direct AC power connection or a battery bank and solar panels (Al-Qadi et al. 2016) The weight sensor is the fundamental component of any WIM system. It responds to the dynamic forces and outputs an electric signal corresponding to the magnitude of the force on the sensor. The process by which the sensor responds to a force varies depending on the sensor type. There are three main types of load sensors that are used in WIM systems: bending plate, load cell, and piezoelectric sensors. A fourth type of sensor, the fiber-optic sensor, is in development but is not yet commercially available (Al-Qadi et al. 2016).

2.2.1 Weight Sensors

Bending Plate sensors are comprised of strain gauges affixed to the bottom of high-strength, rectangular steel plates. When a wheel crosses a plate the strain gauges output a signal that corresponds with a strain measurement at a particular location. The strain gauges are arranged in such a manner so that software can be used to analyze the output from all sensors and interpret the force acting on the plate (Al-Qadi et al. 2016). The arrangement of the strain gauges allows for the measurement of small resistance changes which in turn allows for the measurement of minute strains. The bending plate system offers a middle ground in cost and quality compared to the load cell and piezoelectric sensors. It is more expensive and provides better quality data than the piezoelectric sensors (ceramic and polymer) but is less expensive and provides poorer quality data than the load cell sensor.

A load cell sensor is a transducer and operates in a similar fashion to a strain gauge. Typically, a load cell sensor is comprised of a machined shear beam with strain gauges mounted on both sides of the beam web (Al-Qadi et al. 2016). The operating principles are the same as with the bending plate, as both utilize strain gauges as the means of producing information and detecting small variances in signal strength. The load cells are typically supported by a steel frame and located underneath a steel loading plate. The load cell system is very durable and produces the highest quality data of any WIM sensor type (Al-Qadi et al. 2016). Likewise, the cost associated with its installation is the highest of the various WIM types because of the reinforced concrete vault required to support the system (Al-Qadi et al. 2016) and the higher cost of the sensor.

Piezoelectric sensors, like bending plate and load cell sensors, operate by converting an applied load into an electric signal of proportional magnitude. These sensors, made of a piezoelectric material, can be used for weight measurements and vehicle classification, or for vehicle classification only; however, the sensors that are used for weight measurements are subject to a stricter manufacturing process (Al-Qadi et al. 2016). Unlike a bending plate and load cell sensor, a piezoelectric sensor is only capable of measuring the magnitude of dynamically applied loads. This means that they are appropriate for high-speed WIM (HSWIM) applications but are less suited for slow moving vehicles and cannot produce useful information from static vehicles (Al-Qadi et al. 2016). Like the other sensor types, there are various configurations for which the sensors can be installed. Many systems use two lines of sensors per lane, called a double threshold, to increase the accuracy of the measurements and collect classification data (Al-Qadi et al. 2016). There are three common types of piezoelectric sensors, determined by the material used in the sensing element: ceramic, polymer, and quartz.

Piezoceramic sensors use powdered ceramic that surrounds a solid copper core and is contained within a copper sheath (Al-Qadi et al. 2016). The sensors have a small cross-sectional area and must be installed flush with the pavement surface. An illustration of a piezoceramic sensor product is shown in Figure 4. They provide data of average quality data and are appropriate for vehicle classification purposes (Al-Qadi et al. 2016). The benefits of piezoceramic sensors are that they are much cheaper than other sensor types (Table 2) and require less work to install. The ease of installation is countered by the detrimental effect the structural response of the roadway has on sensor output (Hallenbeck 1998).

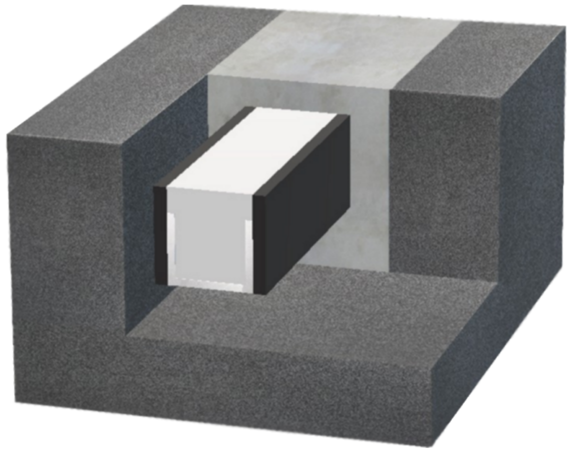


Figure 4. Cross-section of an Installed Piezoceramic Sensor ('Electronique Controle Mesure').

Table 2. Comparison of Weight Sensor Types (Al-Qadi et al. 2016).

Characteristic		Bending Plate	Single Load Cell	Piezoelectric Sensor	Quartz-Piezoelectric Sensor
Cost	Initial installation cost per lane (USD)	Medium (~$20,000)	High (~$50,000)	Low (~$9,000)	Medium (~$20,000)
	Annual maintenance and operation costs (USD)	Medium (~$6,000)	High (~$8,000)	Low (~$5,000)	High
Accuracy (GVW 95-percent confidence)		±10 percent	±6 percent	±15 percent	±10 percent
Sensitivity		Medium	Medium	High	None to temperature, but high to roughness
Expected life (years)		6	12	4	Expected > 15
Reliability		Medium	High	Low	Medium

Piezoquartz sensors are the newest type of piezoelectric sensor and offer several advantages over the ceramic and polymer variants. Piezoquartz sensors are not sensitive to temperature and offer "research-quality traffic load data" (Al-Qadi et al. 2016) The drawback of this type of sensor is that it is more expensive to purchase.

Typical configurations of piezoelectric sensors include one line of sensors with inductive loops on both upstream and downstream sides or a double threshold configuration with one or two inductive loops as shown in Figure 5 and Figure 6. Additionally, these sensors can be arranged in a multiple sensor array which allows for the repeated collection of vehicle data. This configuration yields higher data accuracy and allows for additional methods of calibration.

Figure 5. Double Threshold WIM System Configuration (Al-Qadi et al. 2016).

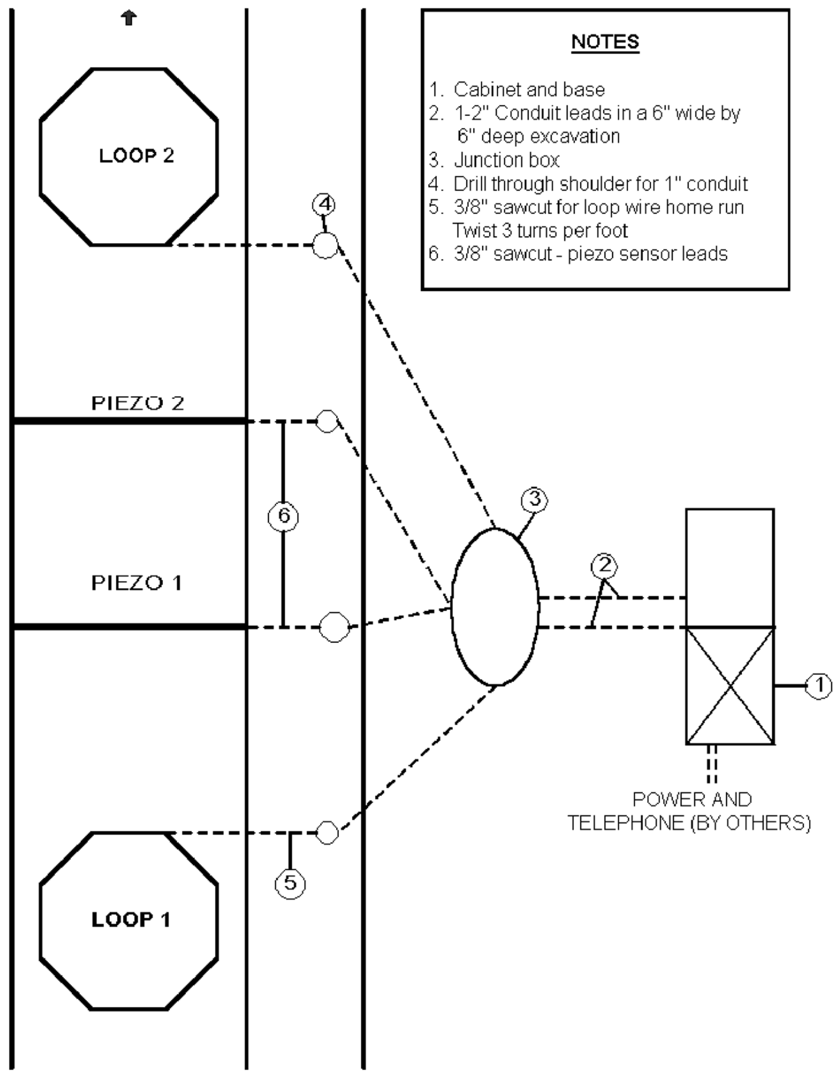


Figure 6. Example of a Piezoelectric System Configuration (McCall and Vodrazka Jr 1997).

Understanding how a weight sensor functions is essential to understanding why calibration is necessary for any sensor. A weight sensor is a precisely engineered instrument which utilizes the material properties of the sensing element to produce an electrical signal or a change in signal strength. The signal output is linearly proportional to the magnitude of the vertical component of the dynamic force applied to the weight sensor ("Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Methods" 2017). This raw signal is meaningless until it is correlated with the dynamic force that generated it. It is the function of calibration to assign a factor that is used in calculations for establishing this correlation between signal strength and magnitude of the vertical dynamic force received by the sensor. It is important to note that the signal output is determined by the force received by the sensing element, and not necessarily the force applied to the scale surface. This is because there are opportunities for outside influence and interruption of the force path from the scale/tire interface to the sensing element. The high degree of sensitivity required to make these measurements means that the measurements are affected by minute manufacturing defects, the method and materials of installation, and any change in environment after installation. Therefore, it is necessary to initialize the system after installation and provide routine calibrations over time to compensate for any changes that affect how the sensing element functions. It is during this system initialization that site reference values for FAW and GVW are established.

2.2.2 Inductive Loops

An inductive loop is comprised of a metal wire installed into the pavement in a closed-form shape. When electricity is supplied to the loop, the movement of metal objects passing over the loop induces a current which can be detected by the WIM controller. The primary purpose of inductive loop detectors is to signal the arrival and departure of a vehicle from the WIM site (Al-Qadi et al. 2016). They offer additional capabilities in the measurement of vehicle speed, vehicle class, and vehicle length depending on what components are present in the WIM system (Al-Qadi et al. 2016). While vehicle speed and vehicle class can be acquired using detector loops alone, it is advisable to use the weight sensors and axle detectors for this purpose as they provide higher quality data.

2.3 Data Accuracy

WIM load sensors are responsive to the dynamic forces of the traffic passing over it. The dynamic force of a vehicle is influenced by road surface roughness, vehicle acceleration, tire condition, tire pressure, wheel balance, vehicle suspension, vehicle aerodynamics, and wind ("Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Methods" 2017). On top of the dynamic forces, the sensor output can be affected by the number of axles (Gupta, Tang, and Yuan 2018) and environmental factors such as pavement temperature, pavement type, and installation materials and methods. Therefore, to interpret the sensor output accurately, the characteristics of the traffic vehicles and influencing factors present at the WIM site must be well understood. The vendor for the WIM controllers in Arkansas, Peek Traffic Corporation, has suggested that the site characteristics be verified to improve the sensor calibration. These characteristics are physical site characteristics and truck traffic characteristics. Site characteristics include pavement condition and profile, grade, traffic flow restrictions, and weather. Truck traffic characteristics include empty vs loaded trends, seasonal variations, enforcement effects, unique vehicles, and traffic operating characteristics (McCall and Vodrazka Jr 1997).

Data errors can be categorized as either systematic or random. Random error cannot be corrected by calibration (Mai, Turochy, and Timm 2013), therefore the scope of this research is limited to addressing systematic error. Systematic error, or calibration drift, is any consistent, non-random error in the data caused by some external influence that results in under- or overestimates (Hyun et al. 2015) as illustrated in Figure 7. While the exact cause of the error may not be known, it is nevertheless possible to correct for this error. Calibration drift can be corrected by computing and applying a calibration factor to the un-calibrated sensor output. Once applied, the controller software will update the existing calibration factor and all further output will be calibrated.

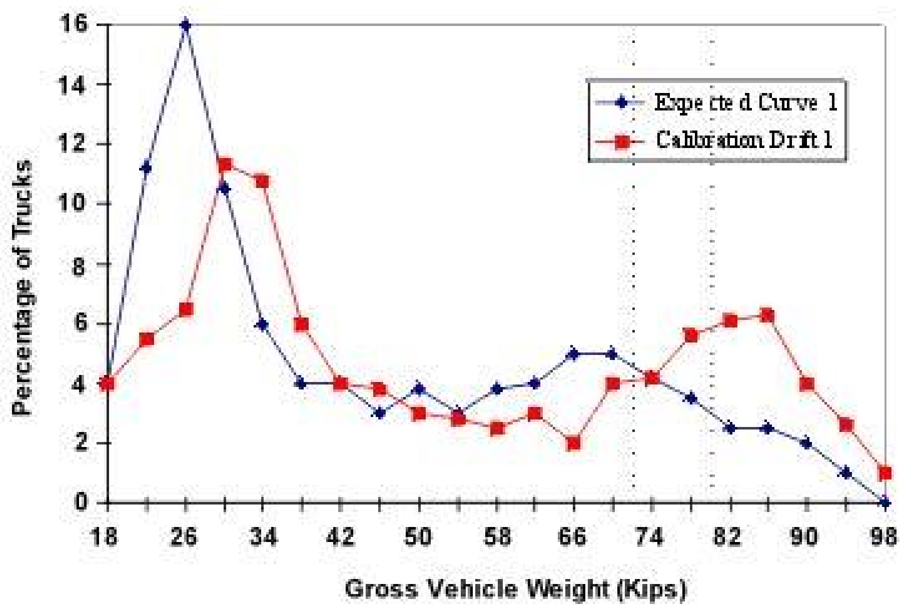


Figure 7. Illustration of Calibration Drift (Hallenbeck 1998).

WIM error is calculated from a WIM measurement and a vehicle's true static weight as shown in Equation 1 (Hashemi Vaziri et al. 2013; Quinley 2010). For the purpose of WIM calibration, the static scale measurements are assumed to be the true weight of the vehicle. The absolute percent error (APE) is calculated as shown in Equation 2. These measurements are obtained either from test trucks as prescribed in the ASTM E1318 standard or by using traffic stream vehicles. It is recommended to compute multiple calibration factors per site that represent different vehicle speeds (Papagiannakis and Quinley 2008). It will also be necessary to compute calibration factors for each lane. This is because piezoelectric sensors are prone to accuracy errors due to "lane edge effects" that influence the estimated weight (Jiang et al. 2009) as well as the differing traffic stream characteristics between lanes.

$$PE_i = \frac{WIM_i - GT_i}{GT_i} * 100$$ Equation 1

$$APE_i = |PE_i|$$ Equation 2

Where,

PE_i	= Percent Error for vehicle i, %
APE_i	= Absolute Percent Error for vehicle i, %
WIM_i	= dynamic weight of unique vehicle, i, converted to static weight recorded by the WIM sensor, usually in kips or lbs.
GT_i	= true static weight of unique vehicle, i, as measured by a static scale, with same units as WIM_i, e.g. Ground Truth

PE and APE can be used to evaluate WIM accuracy at a microscopic – per vehicle – level. Mean Absolute Percent Error (MAPE) is the appropriate metric if the accuracy of the WIM data across the sample population is desired. The MAPE is the average absolute percent error across the sample population as shown in Equation 3, and was used as the performance measure when comparing and tuning calibration algorithms.

$$MAPE = \sum_{i=1}^{N} \frac{APE_i}{N}$$ Equation 3

Where,

MAPE	= mean absolute percent error, %
N	= size of sample population

The calibration factor (CF) converts raw WIM sensor readings into estimated weights (Quinley 2010). The most common method of calculating calibration factors involves setting the mean error measurements to zero (Equation 4), however the measured element varies by WIM application and between agencies (Papagiannakis and Quinley 2008). For enforcement purposes it is more common to adjust the GVW measurement errors, while for data collection it is more common to adjust the FAW measurement errors (Papagiannakis and Quinley 2008).

$$CF = \frac{Reference\ Weight}{WIM\ Weight} \qquad \text{Equation 4}$$

Where

Reference Weight = average expected weight of specified vehicle element across local sample population, with same units as *WIM Weight*$_i$

WIM Weight = average dynamic weight of specified vehicle element across local sample population converted to static weight recorded by the WIM sensor, usually in kips or lbs

2.4 Calibration Standards

2.4.1 On-Site Calibration Standard Using Test Trucks

ASTM E1318 outlines the standard requirements for calibrating a WIM site using test trucks. The standard calls for the use of two test trucks that meet the following criteria ("Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Methods" 2017):

- Test trucks shall be FHWA Class 9
- Test trucks shall be weighed and measured at a static scale a minimum of three times before the calibration runs over the weight sensors

- Suspension type for all dual tired axles shall be air-type suspension, or of the type most representative of vehicles that operate at the site
- Vehicles shall be loaded to 90% of the registered maximum GVW
- Vehicles shall be loaded with non-shifting cargo in a symmetrical manner
- Tires are preferred to be mechanically balanced and shall be inflated to their recommended pressures and be in excellent condition

This list is not exhaustive, but meant to convey the extent of preparation and labor involved in conducting calibration according to the standard's methods. Before the test vehicles perform their calibration runs, the condition of the pavement must be evaluated for any deviations outside of the prescribed standards found in ASTM E1318. In a Transportation Research Board (TRB) survey it was found that most transportation agencies understand the importance of pavement roughness at WIM sites, yet most WIM sites do not meet the set standards for pavement smoothness (Al-Qadi et al. 2016). The test trucks will make multiple passes over each sensor in each lane at two speeds, differing by at least 20 miles per hour and not exceeding the speed limit without reasoning and permission from the local jurisdiction ("Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Methods" 2017). The two speeds must also be above and below the average speed of the vehicles operating at the WIM site.

The ASTM standard is strict and complex in practice. Surveys conducted by TRB reveal that most agencies do not follow the entire standard; many agencies use only one FHWA Class 9 truck, do not use the appropriate suspension type, and do not determine separate calibration factors for vehicles travelling at different speeds (Papagiannakis and Quinley 2008). Recommendations to improve this method include performing the test runs at more speeds, using more test trucks, and performing tests under varying environmental conditions (Papagiannakis and Quinley 2008).

2.4.2 Arkansas Department of Transportation Auto-calibration Algorithm

The ASTM E1318 standard describes auto-calibration as the ability for WIM software to automatically adjust the calibration factor ("Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Methods" 2017). The goal of auto-calibration is to correct possible systematic error present in the WIM data by adjusting the calibration factor based on the expected value or average value of specified traffic stream characteristics. The specific vehicle element being compared may vary depending upon the intended use of the WIM data. For example, sites which have the main purpose of enforcing maximum vehicle weights will typically compute the calibration factor using the measured GVW. Sites that are used mainly for traffic data collection will typically use the FAW instead.

Arkansas has 42 WIM systems installed throughout the state for data collection purposes (Figure 8). The current auto-calibration method employed in Arkansas compares the average FAW of 50 FHWA Class 9 vehicles against the expected average FAW and computes a calibration factor so that the average measured FAW is equal to the reference value (Figure 9). FAW measurements typically exhibit a smaller range and variation of errors than the corresponding GVW measurements due to the relative independence of the FAW from the truck's cargo and trailer weight. It is important to note that the FAW is not entirely independent from the cargo load, as the location of the truck/trailer connection, the "kingpin", is variable and is usually determined by factors such as driver preference, ride comfort, and fuel efficiency (Hallenbeck 1998; Dahlin 1992).

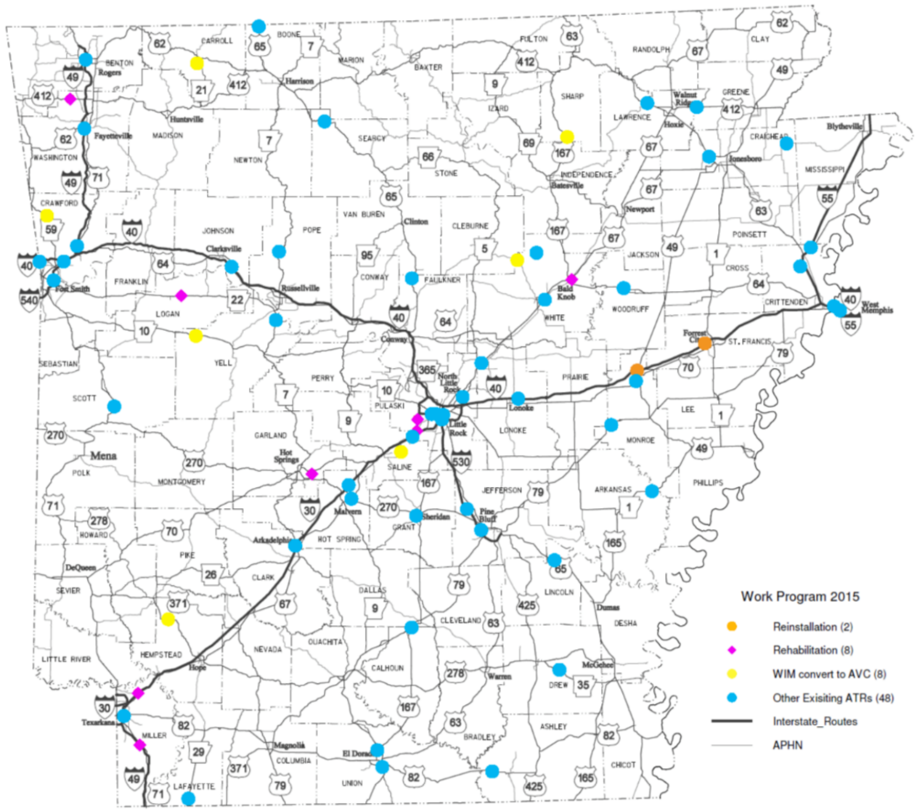


Figure 8. Data Collection WIM Site Locations in Arkansas (Open Source Image).

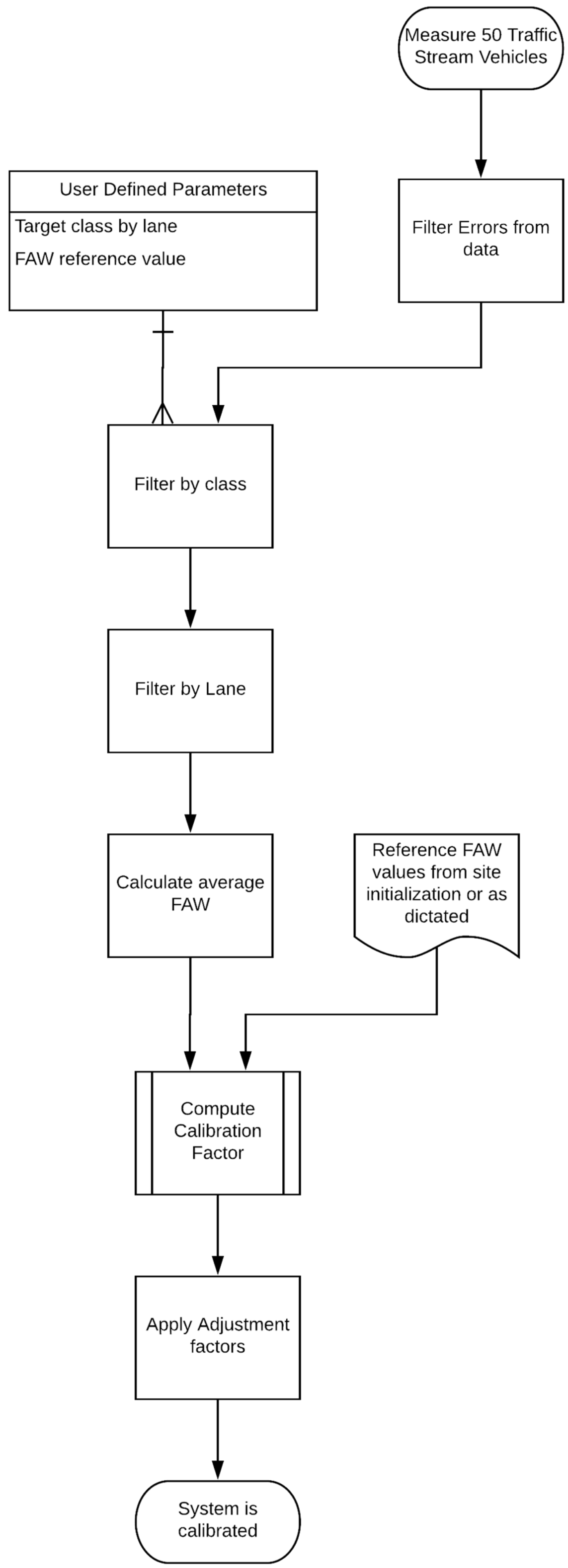
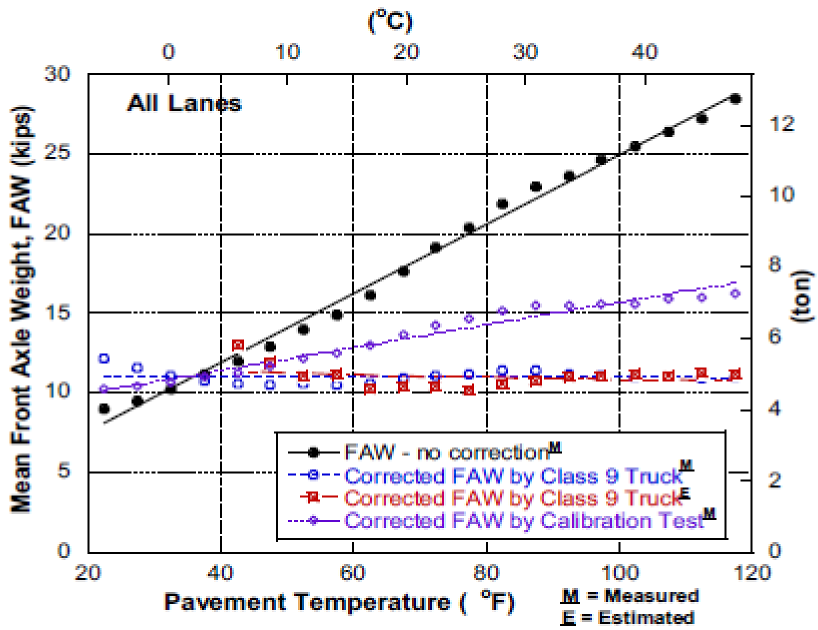


Figure 9. ARDOT Auto-Calibration Procedure.

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. It is possible to use vehicles other than FHWA Class 9 for calibration purposes, however the effectiveness of this choice needs to be investigated. During the time that it takes to accumulate fifty vehicles, the calibration of the WIM sensor will have drifted considerably due to temperature changes as shown in Figure 10 (Nassif et al. 2018). While the WIM software can utilize a temperature correction curve when computing a calibration factor, this feature is not used in Arkansas. Each WIM site is different, and each temperature sensor may be installed slightly differently in terms of sensor location, pavement type, and even the epoxy type used in the installation. Each of these factors could lead to errors in temperature readings and therefore errors in weight estimation, the possibility of sensor failure notwithstanding. Discussions with ARDOT staff have indicated that for these reasons, temperature data is not used in the calibration algorithm and as a result, temperature sensors are absent from Arkansas WIM sites. For the sites that experience high volume the calibration factor will be computed very frequently, so in theory the sensor will be calibrated as the temperature changes throughout the day. The issue of low volume site calibration is a problem that still needs to be investigated.



(a)

Figure 10. Effect of Temperature on Measured WIM Weights (Nassif et al. 2018).

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; Hashemi Vaziri et al. 2013). Although the FAW distributions may vary between WIM sites, each site's own distribution should remain consistent (Dahlin 1992). Using this method, the effectiveness of the calibration is only as good as the accuracy of the reference value. Lastly, it has been shown that the accuracy of piezoelectric sensors is heteroscedastic as a function of vehicle weight, illustrated in Figure 11 (Hashemi Vaziri et al. 2013). 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.

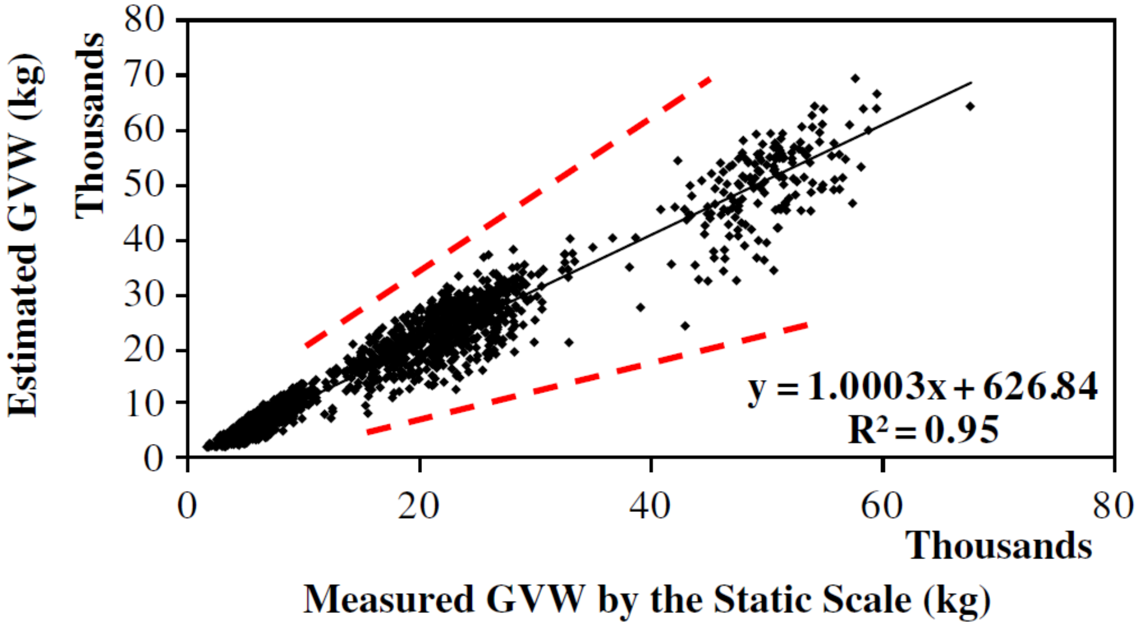


Figure 11. GVW Estimation of Ceramic Sensor Versus Static Measurement (Hashemi Vaziri et al. 2013).

2.4.3 Minnesota Department of Transportation Auto-calibration Algorithm

Because of these limitations, alternate methods of auto-calibration have been investigated. The most common methods of auto-calibrating WIM sites compare traffic stream measurements with reference values for those measurements. The system is deemed to require calibration when the traffic characteristics fall outside of a prescribed range from the reference values. By forcing the average traffic stream measurement to be equal to the reference value it is possible to instead be introducing systematic error if the sample population deviates from the expected characteristic values by a significant amount. A more complex analysis would provide the benefit of reducing the magnitude of any introduced error. This could be achieved by calculating several calibration factors, using more than one reference value to compute the calibration factor, or both.

Calibration factors are calculated differently depending on which auto-calibration algorithm is being used, which may vary by software and manufacturer. Best practice is to calculate different calibration factors that vary with site specific conditions and traffic stream characteristics. These conditions and characteristics can include GVW, FAW, temperature, and vehicle speed. This is the method used by the Minnesota Department of Transportation (MNDOT), outlined in Figure 12, which this research used as a foundation from which to build the new auto-calibration algorithm for Arkansas.

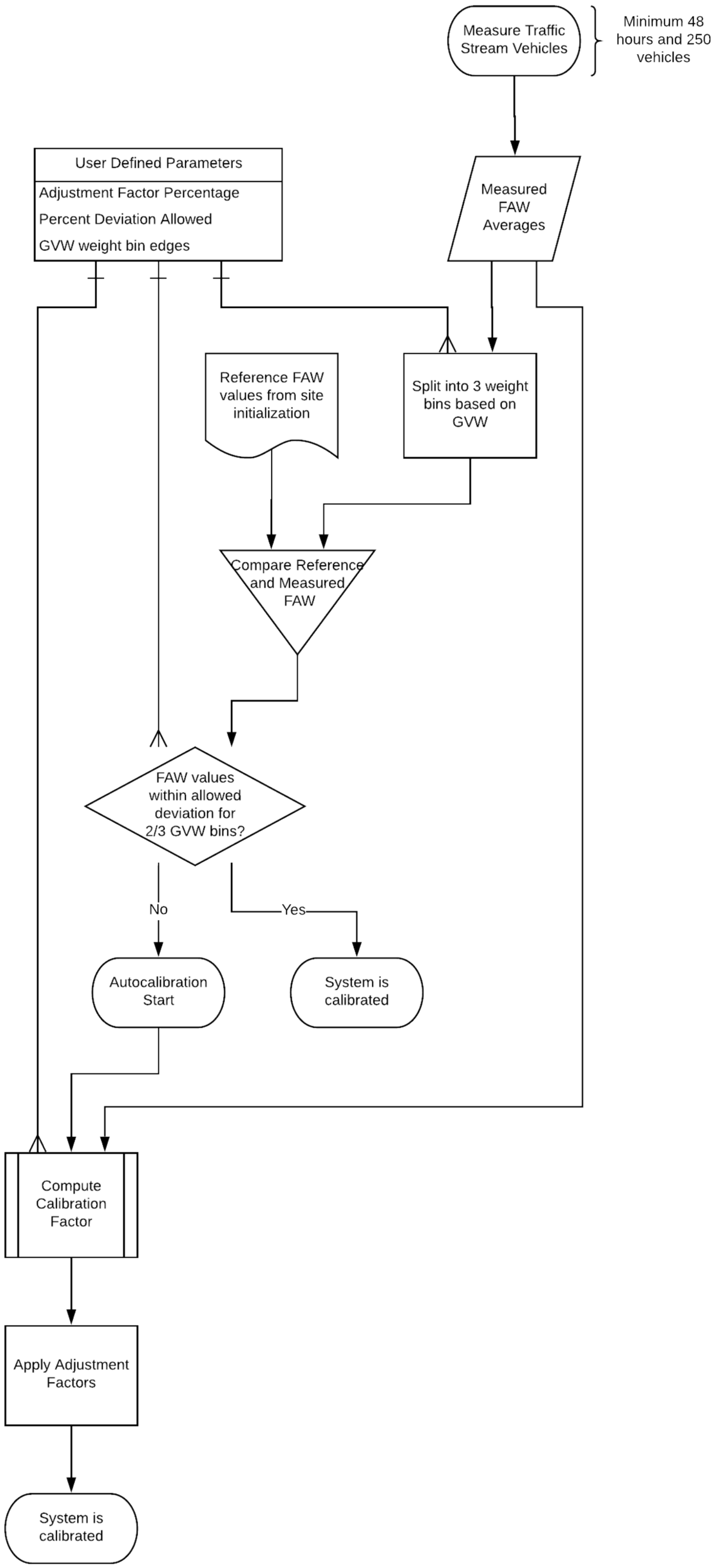


Figure 12. MNDOT Auto-Calibration Procedure.

The algorithm employed by the MDOT has several potential improvements over the ARDOT method. First, a calibration factor is computed for each GVW bins that represent unloaded trucks, partially loaded trucks, and fully loaded trucks as initially proposed by Dahlin (Dahlin 1992). This is beneficial for two reasons. First, estimation errors from polarized sensors are heteroscedastic by weight. Second, in a survey of nearly one thousand WIM site between 1974 – 1983 it was found that a bimodal distribution of loaded and unloaded vehicles is the most common weight distribution nationwide, being present at 80% of WIM sites (Hashemi Vaziri et al. 2013). The same survey showed a multimodal distribution having more than two peaks was present at 2% of sites and a unimodal distribution at 14% of sites (Hashemi Vaziri et al. 2013). While the proportions of these distribution types may have changed over the decades, it is likely that a WIM site exhibits more than one GVW mode. The calibration factor is computed using the average FAW of the vehicles in each weight bin. It is worth noting that individual body types may present their own distributions. For example, in some circumstances tanker trucks may present a bimodal distribution at a WIM site that is otherwise known as having a trimodal distribution (Hyun et al. 2015). This algorithm further considers the statistical importance of sample size and adjusts the weighted importance of each weight bin's calibration factor.

The MNDOT algorithm operates as follows. Periodically (e.g. hourly or every specified number of vehicles, n) a calibration check will be carried out. This calibration check shall be performed by taking measurements of GVW and FAW of an assigned number of FHWA Class 9 vehicles, specifically '3S2' vehicles. The MNDOT auto-calibration procedure has assigned weighted percentages to a range of traffic bin volumes as shown in Table 3. These FAW are categorized according to their corresponding GVW. The average FAWs are calculated for each GVW bin, and the deviation of the FAW from the specified reference values are calculated. If two of the three weight bins are outside acceptable deviation limits, then correction factors are calculated (McCall and Vodrazka Jr 1997).

Table 3 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

An example calibration factor calculation is shown in Table 4. The three GVW bins are shown in column 1 along with their bin limits. Column 2 shows the sample size of measured vehicles per GVW bin, which corresponds to the adjustment factors presented in Table 3. The average GVW per weight bin is shown in column 3. The percent deviation of the measured average weight from the desired average weight is displayed in column 4 and is calculated using Equation 5. The calibration factor for each GVW bin is then computed using Equation 6. Lastly the system calibration factor is calculated as the average of the three calibration factors produced in Equation 6 (one for each GVW bin).

Table 4. Example of Recalibration Results (McCall and Vodrazka Jr 1997).

Lane	Date		Time
#1	Fri. Mar.29, 1991		16:00:00
Gross Vehicle Weight Range [pounds (kg)]	Number of Vehicles Weighed	Average Recorded Weight [pounds (kg)]	Percent Deviation from Desired Weight
< 32,000 (14,500)	59	8,900 (4,000)	+ 4.7
32,000 - 70,000 (14,500-31,750)	112	9,700 (4,400)	+ 4.3
> 70,000 (31,750)	79	10,900 (4,950)	+ 4.8
Calibration Factor: 0.959		Sensor Weight Factor: 15.22	

$$percent\ deviation_i = \frac{\frac{\sum_{n=1}^{N(i)} WIM_{n(i)}}{N(i)} - Reference\ Value_i}{Reference\ Value_i} * 100$$

Equation 5

Where

Percent deviation$_i$ = The measure of how well the reference value represents the measure average FAW of the population, “N”, for GVW bin “i”

WIM$_n$ = The FAW recorded by the WIM system for unique vehicle “n” within GVW weight bin “i”

Reference value$_i$ = The desired or expected average FAW value for sample population “N” within GVW bin “i”

$$calibration\ factor_i = (1 - \frac{percent\ deviation_i}{100}) * \varphi_i$$

Equation 6

Where

Calibration $factor_i$	= the computed calibration factor for GVW bin “i”
Percent $deviation_i$	= the percent deviation for GVW bin as calculated from Equation 5
φ_i	= the sample size adjustment factor from Table 4 corresponding with the sample population, “N”, for GVW bin “i”

3. Data Collection

During the development of the auto-calibration protocol, it was necessary to validate the site-specific characteristics that are used in these algorithms. To accomplish this, WIM data and video were collected from three WIM sites on EB I-40 located at Lamar, Lonoke, and Bald Knob and static weight data and pictures were collected from the enforcement station in Alma, Arkansas also along EB I-40.

3.1 Site Selection

To validate algorithm parameters and evaluate the baseline procedure, individual FHWA Class 9 trucks were tracked across the static scale site and each of the three WIM sites. The MAPE of each trial algorithm was calculated by comparing the calibrated WIM data against ground truth data collected at the static weight station. This analysis was the primary method of judging an algorithms performance against the baseline procedure. As with any model validation, the larger the validation dataset, the more certain we can be that the conclusions drawn are correct.

Data collection sites were selected to maximize the number of trucks that we identify as crossing both the static scale and at least one of the WIM sites under surveillance (a "match"). Several heavily trafficked routes across the state were analyzed and the connectivity between WIM sites, shown in Figure 13, helped determine which routes have the highest number of vehicles that pass by a static scale and a WIM scale. This was done by processing GPS tracking data that was shared by the project partner Drivewyze. This tracking data consisted of unique vehicle identifiers, GPS coordinate location history, and temporal data. By using this data and setting up a geofence in GIS software, a table was generated containing WIM sites and the number of trucks that pass through them and also pass by a static scale. The GPS data is anonymous, and only contains information on trucks that subscribe to the Drivewyze service. The operating assumption of this analysis is that the number of trucks being tracked are proportional to the number of all trucks passing through these routes. The analysis produced heat maps that visually indicated how many trucks travel along a specific route between different static and WIM scales, an example of which is shown in Figure 14.

Figure 13. Connectivity of WIM Sites by Individual Trucks Subscribing to Drivewyze (Open Source Image).

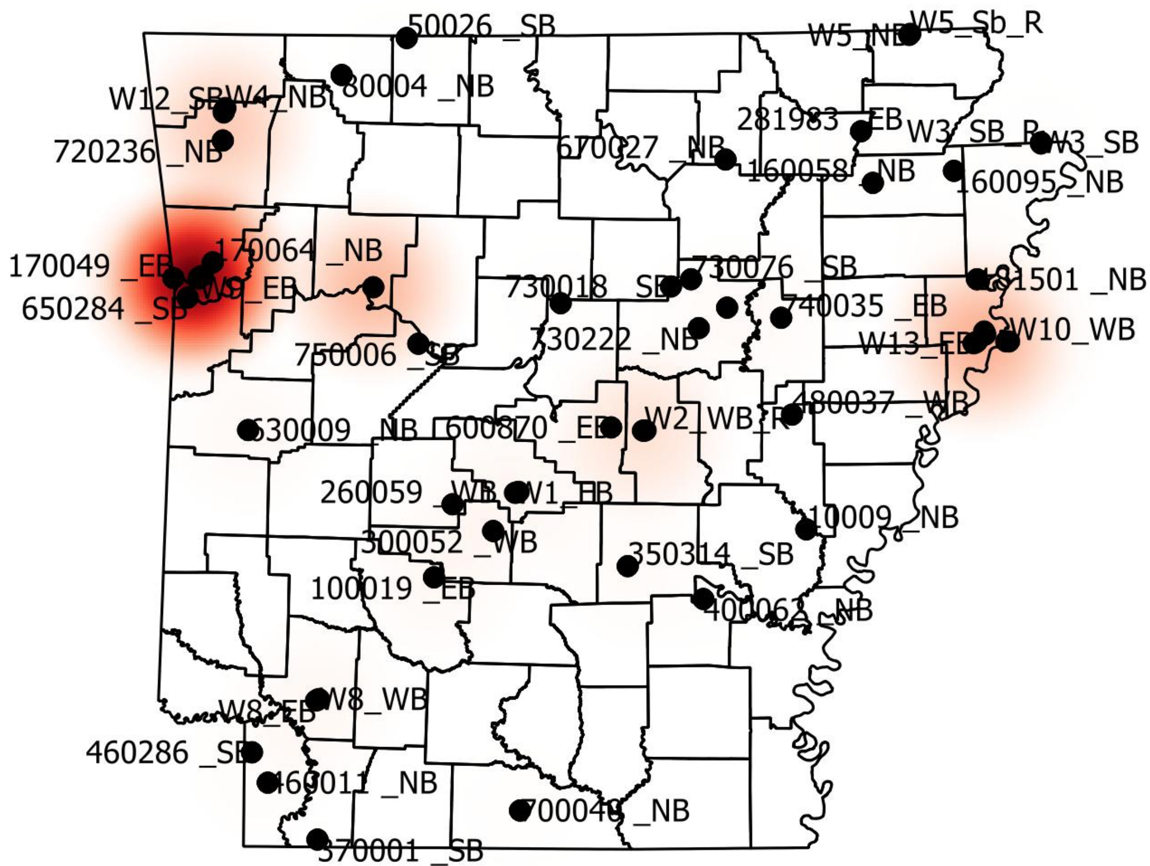


Figure 14. Heat Map of WIM Sites Indicating Relative Amount of Vehicles that Visit and also Pass Through the Alma Static Scale (Drivewyze Vehicles Only)(Open Source Image).

3.2 Truck Matching

Reviewing the video collected at each site allowed trucks that passed through both the static weigh station and a WIM site to be identified in a process that will hereto be referred to as ‘truck matching’. The truck matching process was a manual effort of reviewing all video and checking each truck that passes over the WIM sensor, shown in Figure 15, against the photographic record of trucks at the static scale, an example of which is shown in Figure 16.

The first step in this process was to create a database of all trucks that passed over the enforcement site static scale. This database contained the following information:

- Timestamp
- Static weights
- Truck class
- Visual identification information
- Ramp WIM ID number
- Photograph image filename

Figure 15. Video Snapshot of a Matched Truck Passing Through a WIM Site.

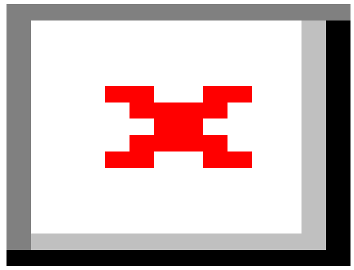


Figure 16. Photo of a Matched Truck Being Weighed at the Static Scale Enforcement Site.

Once the static scale database was compiled, the three videos recorded at WIM sites were analyzed. Every time a truck passed over the WIM sensor the video was paused and the video was compared to the photos of the static scale trucks to check if there is a match. Matches were identified by searching for any difference between the vehicle in the video and picture. If there was any doubt that a truck was a match then it was not included in the matched vehicle dataset. For example, Federal Express® trucks are very similar to one another and were not able to be distinguished for this purpose. To aid in this effort and greatly increase productivity, only trucks that could reasonably be expected to have been able to reach the WIM site at the time the video was recorded were considered for matching. The distance between the WIM and enforcement sites were determined, and the time it would take to travel this distance if the vehicle were going 80 mph (10 mph above the speed limit) was used as a cutoff threshold. This approach greatly decreased the time required to process the video, however the limitations of this strategy are that any truck that had an average speed of over 80 mph between the static and WIM scales might not be recorded as a match. By the same logic, any vehicles that travelled between sites too slowly had an increased chance of having performed a stop to offload or pickup cargo, which would change the vehicles weight, and therefore were also not considered for matching purposes. The threshold used for this was the time it would take a vehicle to drive between the two sites while travelling 10 mph below the speed limit, plus 30 minutes.

When a truck was confirmed as a match, the truck match log was updated with the following information:

- Video timestamp
- Remote site WIM ID
- Review comments such as traffic issues and trucks changing lanes over the sensors
- Full WIM record of the matched truck

Identifying the correct WIM record involved determining the time offset between the WIM record and the camera video. This was a trial and error procedure of finding a unique arrangement of traffic and vehicle headways over the WIM sensor that can be identified as a unique arrangement in the WIM record database with a high degree of certainty. At the end of this process the matched truck dataset contained all information needed to identify the truck from a picture and video image, as well as all data collected from the static weight receipts (ground truth data) and from the WIM sensor (WIM Data) for each matched vehicle. This allowed for the comparison of the WIM data against the known true static weight of the vehicle.

4. Data Analysis

Matched vehicle data allows for the identification of site-specific characteristics and the estimation of their values. In communications with the WIM vendor that supplies the ARDOT, Peek Traffic, it was recommended that the site-specific characteristics be validated to optimize calibration performance. This recommendation is in line with the recommendations of the literature which suggest that some traffic stream characteristics may vary from site to site (Dahlin 1992). This analysis was performed in two steps, ground truth data analysis and WIM data analysis.

4.1 Traffic Stream Characteristics

A Gaussian Mixture Model (GMM) of matched vehicles was created for the vehicles at each study site to help define thresholds for loaded and unloaded vehicles needed for the MNDOT auto-calibration method. The GMM represents the GVW distributions as multiple normal distributions with different means and standard deviation, combined via mixing proportions. After modelling the GVW distribution, the threshold separating loaded and unloaded trucks was identified. The location at which these normal distributions intersect, seen in Figure 18 and Figure 19, provide good starting values to tune for the calibration algorithm. GVW distributions and average FAWs of trucks crossing the static scale were used to determine initial values of parameter used in the ARDOT and MNDOT auto-calibration algorithms.

To “fit” a GMM to the GVW data, the number of components must be pre-determined. Previous research shows that a two or three component GMM typically fits a GVW distribution for FHWA Class 9 trucks. The lower distribution represents unloaded vehicles, the middle distribution represents partially loaded vehicles, and the upper distribution represents fully or overloaded vehicles. The best fit model is that with the lowest AIC and BIC values. AIC and BIC are goodness-of-fit measures that act as penalized-likelihood criteria (Sclove 1987). Mixtures of two to seven components were fit the GVW data and the model with the lowest AIC and BIC was chosen.

It was found that the traffic that passed through the static enforcement station located at Alma exhibited a tri-modal (e.g. three component) GVW distribution for FHWA class 9 vehicles (Figure 17), while the WIM sites at Lamar and Lonoke exhibited bimodal (e.g. two component) GVW distribution (Figure 18 and Figure 19). There were only two matched vehicles that passed through Bald Knob WIM site, therefore the distribution at Bald Knob was not fit using a GMM approach.

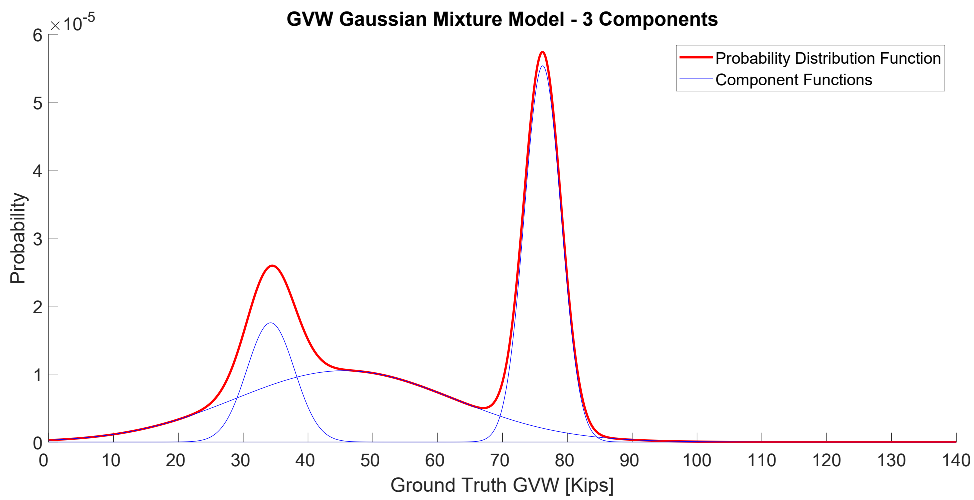


Figure 17. Alma GVW Distribution and Component Functions.

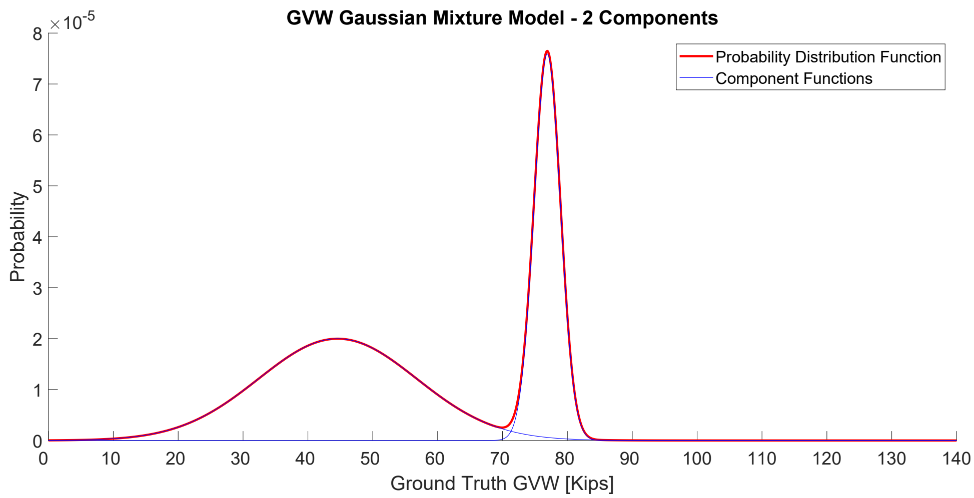


Figure 18. Lamar GVW Distribution and Component Functions.

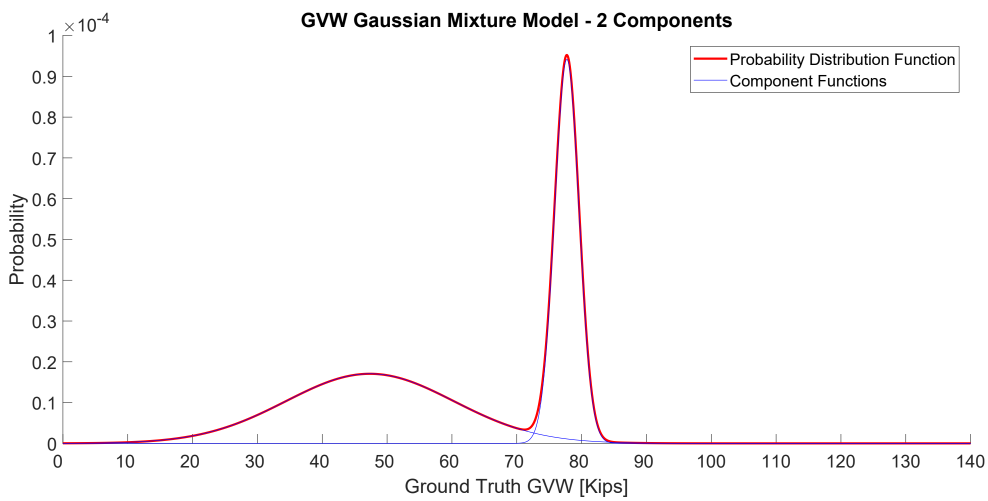


Figure 19. Lonoke GVW Distribution and Component Functions.

The shifts in peak location and amplitude of the GMMs indicates differing loading characteristics by location. The WIM sites at Lamar and Lonoke are located on EB I-40 but are on opposite sides of Little Rock and the I-30 interchange which leads to southwest Arkansas and Texas. The GVW distributions show a large proportion of loaded trucks traveling EB toward Memphis, TN, a major intermodal hub, or beyond. The GVW distribution at the Alma site likely differs due to the sample of vehicles measured rather than the location characteristics. The WIM scales measure all traffic, not just vehicles that are pulled from the traffic stream like with static scales. One possible explanation for the difference in distribution is that trucks that don't subscribe to static scale bypass programs like Drivewyze are loaded significantly differently.

FAW reference values were identified for each GVW bin, e.g. unloaded, loaded, etc. (Figure 20). Initial FAW references values were found using the Alma static scale weights by computing the average FAW per bin. The horizontal lines represent the reference FAW values in each GVW weight bin defined by the MNDOT calibration algorithm while points represent the actual GVW-FAW pairs of the trucks measured by the static scale.

The data was separated into vehicles that traveled to Lamar and Lonoke after stopping at the Alma static scales. Note that extreme outliers were removed by imparting a manual filter on minimum acceptable FAW and GVW pairs. As Figure 21 shows, the reference FAW values vary between WIM sites. However, this difference was not statistically significant. The hypothesis that the matched class 9 vehicles that passed through the WIM sites at Lonoke and Lamar were from the same continuous distribution was tested using a two-sample Kolmogorov-Smirnov test (Sheskin, 2011). The resulting statistics are shown in Table 5. The null hypothesis is that both sample populations are from the same continuous distribution. This test was performed for the FAW and GVW of the matched vehicles and for both tests the null hypothesis failed to be rejected at the 95% level of confidence.

Therefore, due to limitations arising from the field data collection methodology, we are unable to draw conclusions about the spatial transferability of FAW reference values. The trucks present at the Lonoke WIM site were a subset of the trucks at the Lamar WIM site, and both were subsets of the trucks at the Alma site. Thus, the Lamar and Lonoke samples were not independent samples. In future studies, it is necessary to select WIM sites that capture different, independent truck populations. For example, a site along I-30 in the South-west portion of the state and a site along I-40 in the North-west portion of the state would contain different truck populations.

Table 5. Two-Sample Kolmogorov-Smirnov Test Results

Null Hypothesis Population	P-value
Lamar and Lonoke Matched Vehicles; FAW	0.346
Lamar and Lonoke Matched Vehicles; GVW	0.055

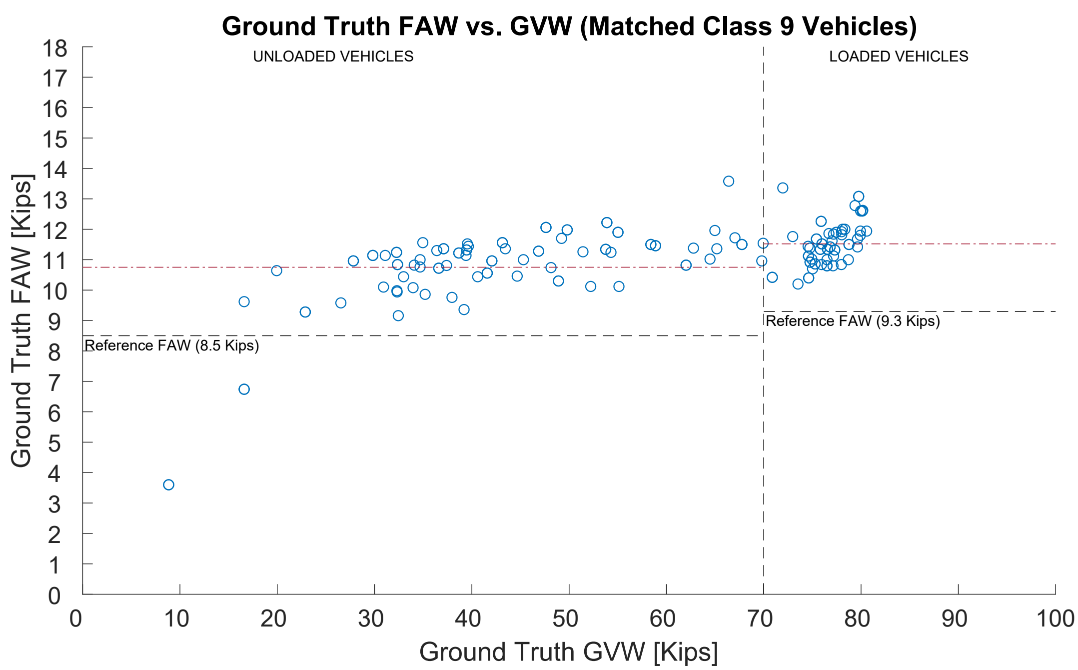


Figure 20. Ground Truth FAW vs GVW by GVW Bin for All FHWA Class 9 Vehicles.

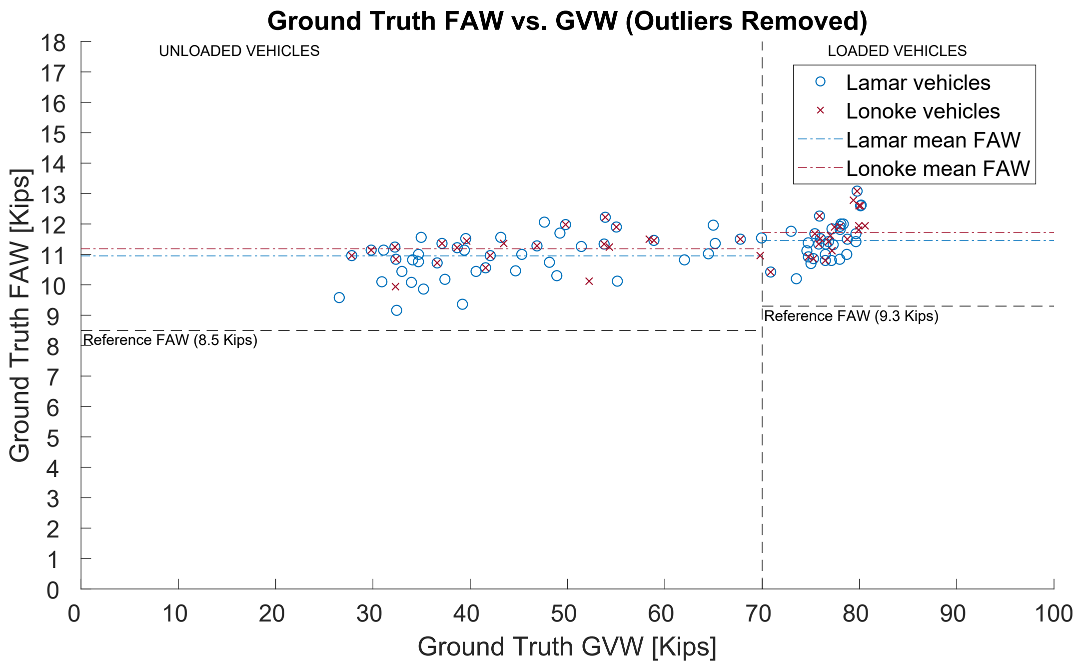


Figure 21. Ground Truth FAW vs GVW by GVW Bin Split by Matched WIM Site Location.

4.2 Initial Site Calibration

Trends in temperature and calibration drift were compared and FAW reference values were identified using the un-calibrated WIM data to illustrate the calibration drift that occurred throughout collection day. ARDOT turned off the auto-calibration function during the period of data collection. However, all initial calibration values (established when the site was originally installed) and auto-calibration values were both "turned off" and, as a result, the WIM data was severely shifted from the time that these functions were turned off (about 6 AM) until the end of the day (Figure 22). To re-create initial calibration settings, the first matched vehicle in each lane at each WIM site was used to estimate a calibration factor as shown in Equation 7. All data following the first vehicle match in each lane was adjusted by the initial calibration factor.

$$Calibration\ Factor = \frac{Matched\ Vehicle\ Ground\ Truth\ GVW\ Measurement}{Matched\ Vehicle\ WIM\ GVW\ Measurement}$$ Equation 7

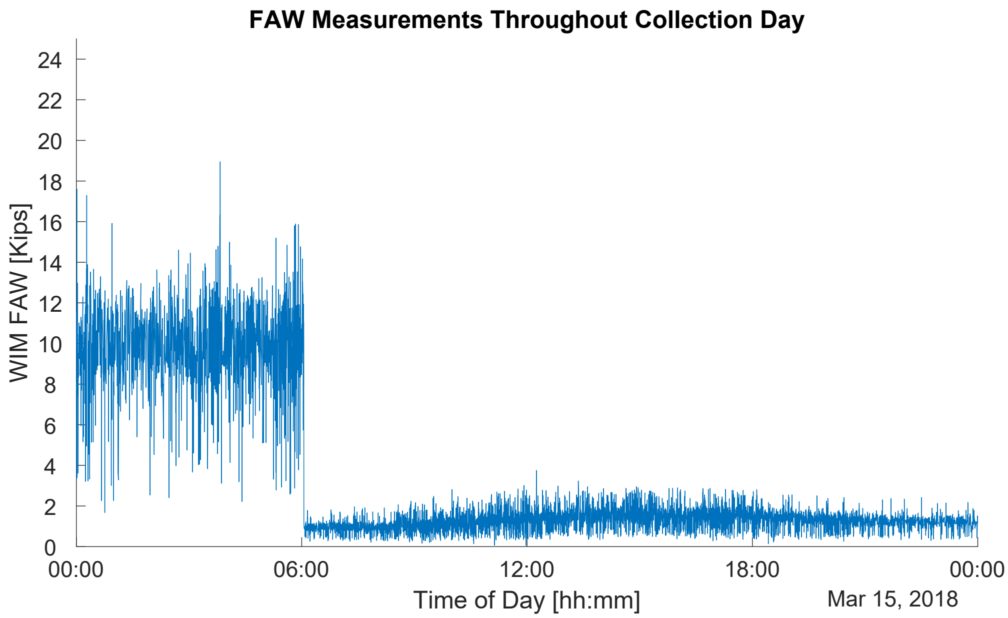


Figure 22. Unmodified FAW Measurements at Lonoke.

4.3 Site Characteristics

Air and pavement temperature were recorded at intervals of 15-30 minutes at each site during data collection. It is well understood that temperature greatly effects the calibration of weight sensors. A large contributor to this effect is the change in pavement stiffness which affects how the force is distributed from the wheel-pavement interface through the pavement and ultimately through the WIM weight sensor. As pavement temperature increases, pavement stiffness is reduced which results in an increased measurement by the sensor. The unmodified FAW, calibrated FAW, the FAW values for the inside and outside lane, and the corresponding air and pavement temperature are compared in Figures 23-25 for the un-calibrated data in Figures 26-28 for the initially calibrated data. The moving averages with a 50 vehicle window is shown in these figures. The inside lane and outside lane were calibrated independently. Note that there were no matched vehicles that passed through lane 2 at the Bald Knob WIM site, and therefore lane 2 was unable to be initially calibrated.

Overall, FAW increases as pavement temperature increases at all sites. However, air and pavement temperate are similar at Bald Knob and Lamar because the air temperature sensor was not shaded, while the sensor was shaded at Lonoke.

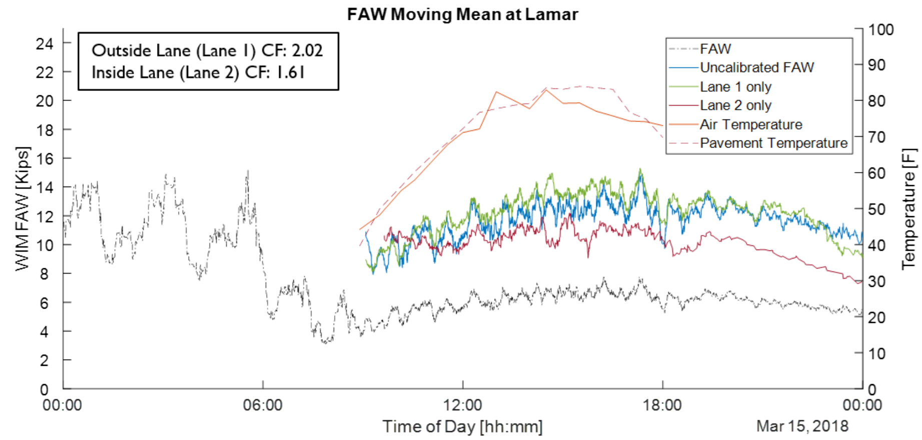


Figure 23. Initial Calibration of WIM Data from the Lamar WIM Site.

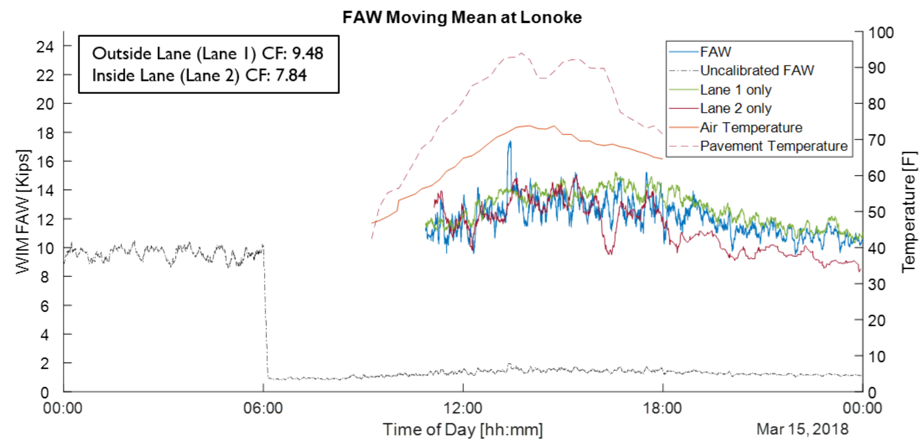


Figure 24. Initial Calibration of WIM Data from the Lonoke WIM Site.

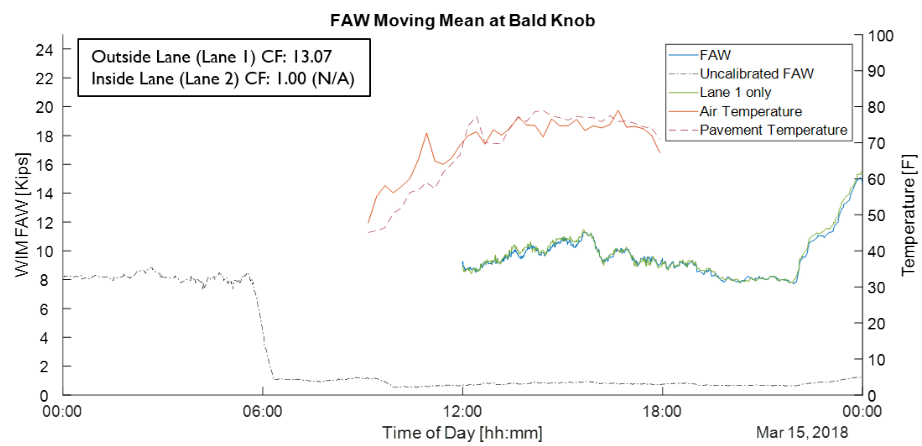


Figure 25. Initial Calibration of WIM Data from the Bald Knob WIM Site.

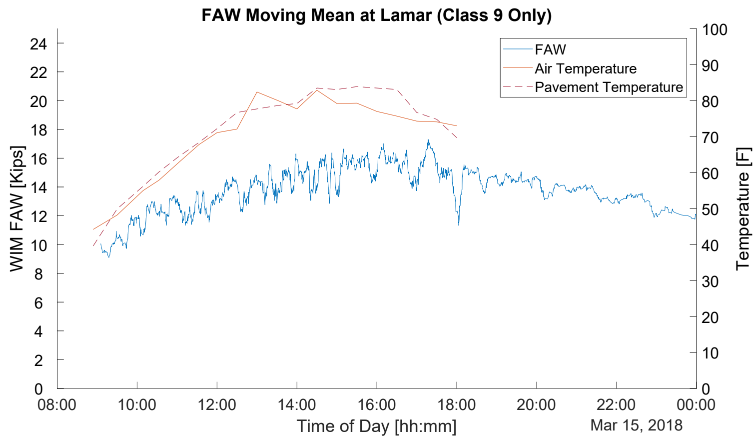


Figure 26. Calibration Drift at the Lamar WIM Site.

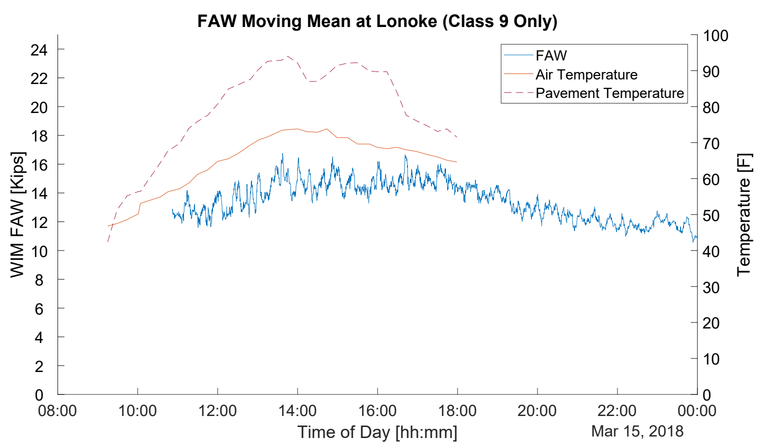


Figure 27. Calibration Drift at the Lonoke WIM Site.

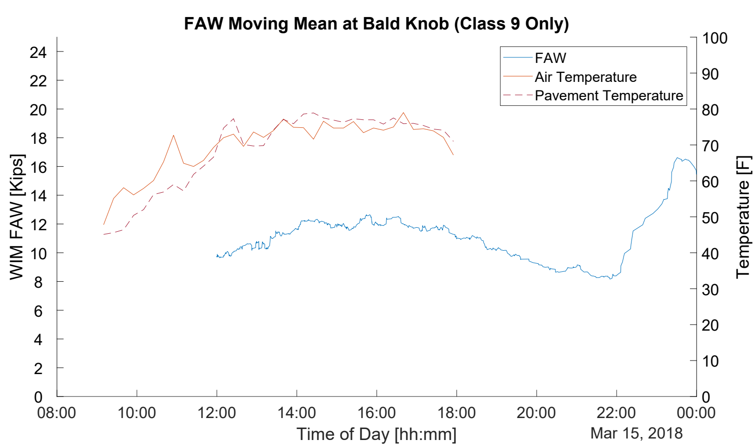


Figure 28. Calibration Drift at the Bald Knob WIM Site.

5. Baseline Auto-Calibration Results

5.1 ARDOT Baseline Performance

The auto-calibration algorithm that is currently used in Arkansas, the 'ARDOT method', served as a baseline procedure for evaluating the effectiveness of alternate procedures. The ARDOT method was applied to the WIM records of trucks that passed the static scales from each collection site to produce a "calibrated" dataset. The Percent Error (PE) of each vehicle record was calculated and the Mean Absolute Percent Error (MAPE) was used to summarize the performance of each auto-calibration algorithm. The MAPE of the ARDOT method was 12.9% at Lamar and 24.5% at Lonoke.

5.2 MNDOT Baseline Performance

The MNDOT calibration algorithm was applied to the WIM datasets from each collection site using PE and MAPE to summarize performance. The MAPE calculated using the original parameter values of the MNDOT procedure were 15.2% at Lamar and 26.5% at Lonoke. The MNDOT auto-calibration performance is worse than the baseline ARDOT procedure. This is to be expected as the initial parameter values for this procedure were not altered from the values used in Minnesota, and as already demonstrated, they are not a good match for Arkansas WIM sites.

6. Algorithm Parameter Tuning

The goal of parameter tuning is to minimize WIM measurement error by selecting site specific algorithm parameters for use at Arkansas WIM sites. Analysis included the following parameters:

- Calibration frequency
- GVW loading thresholds
- Reference FAW values

Each parameter had an effect on the calibration algorithm MAPE, however the parameters also influenced each other. To select a combination of parameter values that produced minimal error, an iterative analysis was performed (Figure 29). First each parameter was applied individually to measure change in MAPE. Once relative magnitudes of the parameter effects on MAPE were determined, the parameters were evaluated in order of increasing magnitude.

An update to the calibration frequency parameter was applied first. The trend of more frequent calibration yielding lower MAPE values was consistent across sites due to the need to compensate for temperature changes. An update to the loading threshold parameter was applied second. The starting values for loading threshold were derived from the GMMs in Figures 18 and 19. Third, the reference FAW values were adjusted. It was found that FAW parameters greatly affected the MAPE values and were highly dependent on the other parameter values. Two additional iterations of parameter turning were performed such that the resulting parameters from one iteration served as the starting point for the next iteration (Table 6). The final parameter values are as described in the remainder of this chapter.

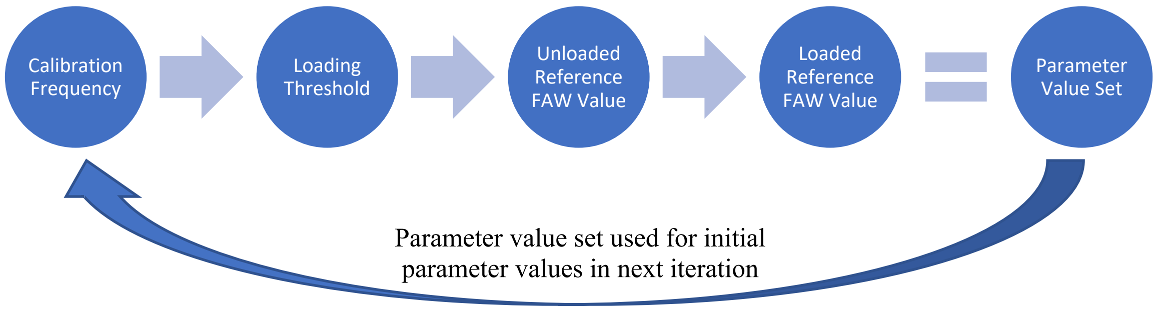


Figure 29. Steps of the Parameter Tuning Process

Table 6. Intermediate Parameter Values During Iterative Tuning

	Calibration Frequency [# of vehicles]		Loaded Threshold [kips]		Unloaded Reference FAW [kips]		Loaded Reference FAW [kips]	
Iteration	Lamar	Lonoke	Lamar	Lonoke	Lamar	Lonoke	Lamar	Lonoke
1	50	50	40.0*	20.0*	9.5	11.2*	9.0	10.5
2	50	50	71.0	72.5	10.5	11.5	9.3	10.2
3	50	50	71.0	72.5	10.2	11.00	9.3	10.2
* No local minimum established for MAPE within test range								

6.1 Calibration Frequency Parameter

Calibration frequency should ideally keep pace with temperature changes but should also maintain a reasonable number of sample vehicles in each calibration update. The ARDOT calibrates the sensors after every 50th 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.

The results of changing the calibration frequency of the ARDOT algorithm are shown in Figure 30. The more frequent calibrations yielded the lowest MAPE. MAPE continued to decrease with more frequent calibration, e.g. less than 50 vehicles per calibration cycle. Minimal improvement was achieved by reducing the frequency below 30 vehicles.

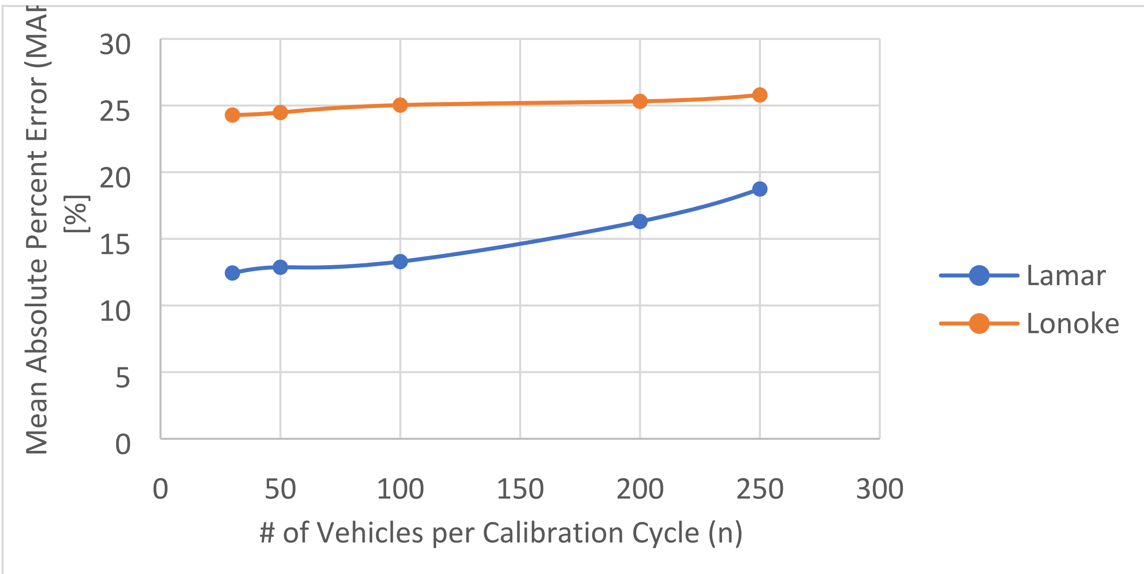


Figure 30. ARDOT Parameter Tuning for Calibration Frequency

In addition to a larger calibration update sample (250 trucks), the MNDOT method divides the sample into weight bin groups. For each weight bin group, a sample adjustment factor is applied (Table 3) and calculated into the calibration factor. Unlike the WIM sites in Arkansas, the MNDOT sites contain temperature sensors which helps account directly for temperature effects on WIM measurements. When no temperature sensors are installed, more frequent calibrations are needed to keep pace with temperature changes, and therefore smaller sample sizes are required (i.e. 50 vehicles). Figure 31 shows the cumulative vehicle count throughout the day at the Lonoke WIM site. The pavement temperature is overlaid on this figure to show the change in temperature throughout the day. The vertical lines represent every 250th FHWA class 9 vehicle to pass through the site. From this figure, it is apparent that the temperature can change quite drastically between calibrations if the site is only calibrated every 250 vehicles. A plot of MAPE vs calibration frequency was generated to identify an optimal value (Figure 32). Higher calibration frequency yielded better performance (lower MAPE) due to the need to compensate for calibration drift due to rapid changes in temperature.

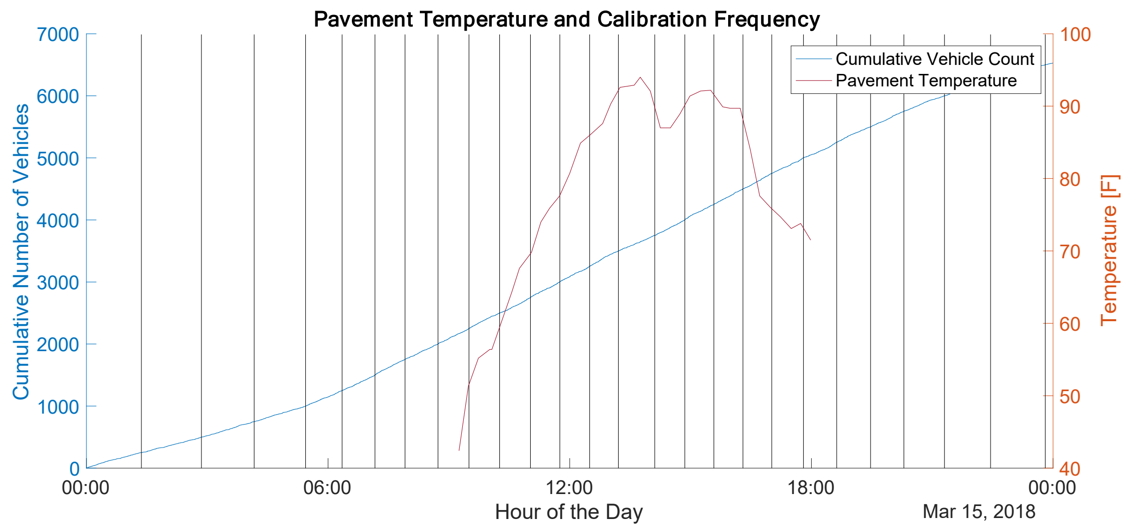


Figure 31. Effect of Pavement Temperature Change on Calibration Frequency

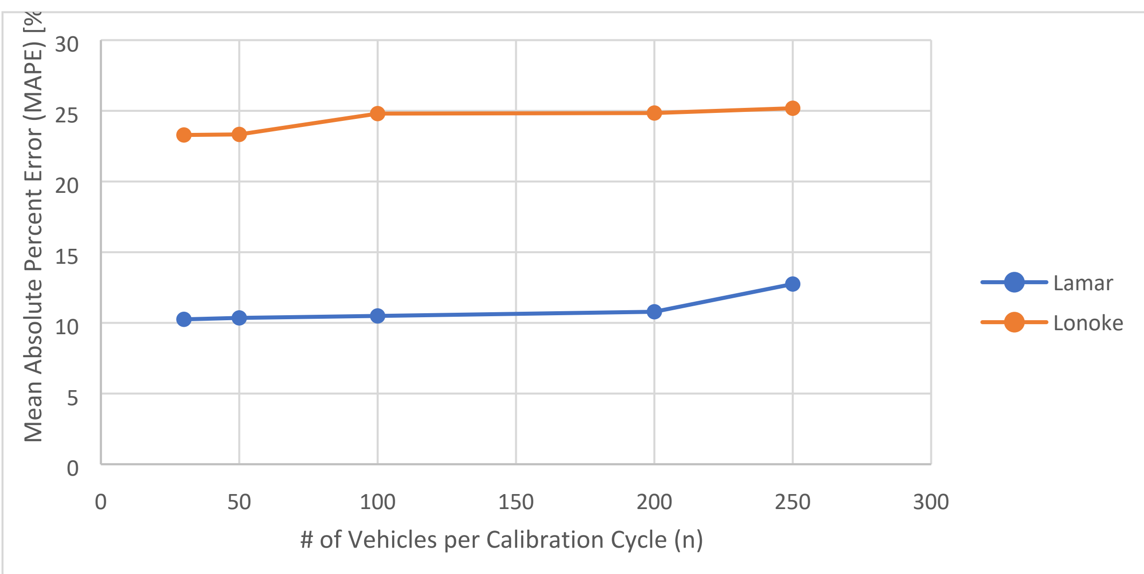


Figure 32. Calibration Frequency Tuning at Lamar & Lonoke WIM Sites.

6.2 GVW Weight Bin Threshold Parameters

Parameter tuning was performed on the GVW thresholds for each WIM site. The initial thresholds for loaded weight were defined by the GMM analysis for Lamar and Lonoke. Since the best fit GMMs were bimodal, only a single threshold between loaded and unloaded trucks was extracted, e.g. the "loaded" threshold. If the GMMs were found to be trimodal, two thresholds would be extracted, an unloaded and a loaded threshold. Initial loading thresholds of 71.0 kips for Lamar and 72.5 kips for Lonoke were used. Parameter values in 5 kip increments were evaluated ranging from 40 kips to 95 kips. MAPE comparison across GVW loaded bin parameter settings show the initial values defined by the GMM to produce the lowest MAPE (Figure 33). The maximum allowable, unpermitted GVW in Arkansas is 80 kips 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). This supports the finding that loaded thresholds are spatially transferrable between sites with similar traffic stream characteristics. This also supports the findings that the loaded threshold are located near the maximum permitted GVW; in the vicinity of 72 kips.

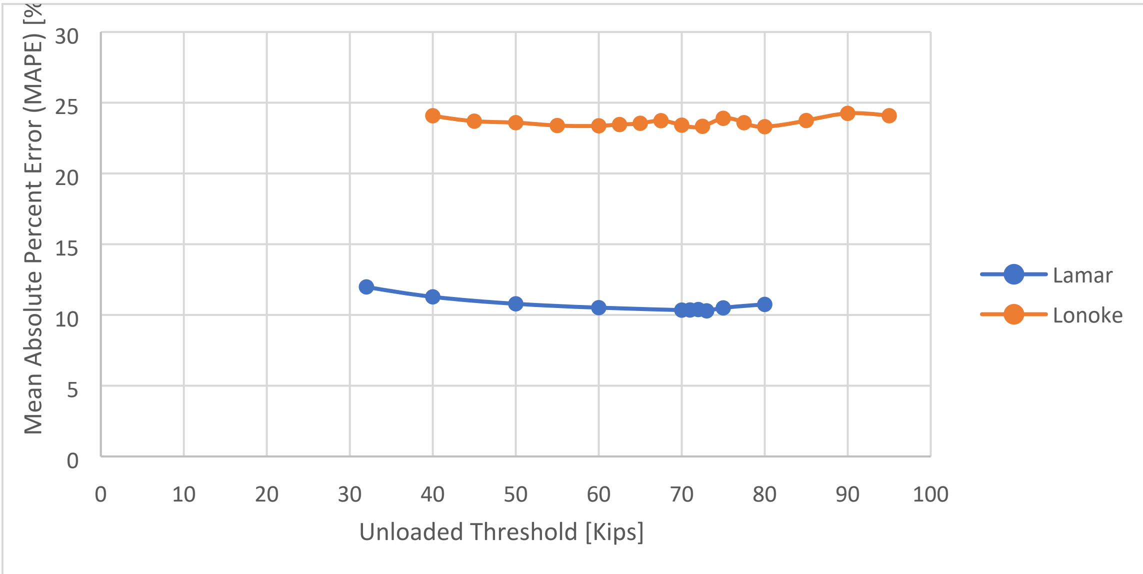


Figure 33. GVW Threshold Tuning at Lamar & Lonoke WIM Sites.

6.3 FAW Reference Parameters

For the ARDOT method the FAW parameters were adjusted in increments of 0.5 kips between 8 and 11 kips. The results of changing the FAW value on the ARDOT algorithm are shown in Figure 34. The reference FAW values differed by WIM site location and the values that produced the lowest MAPE were very close to the average of the bimodal parameter values found in the MNDOT algorithm tuning. The reference FAW values for Lamar and Lonoke were found to be 9.6 kips and 10.6 kips.

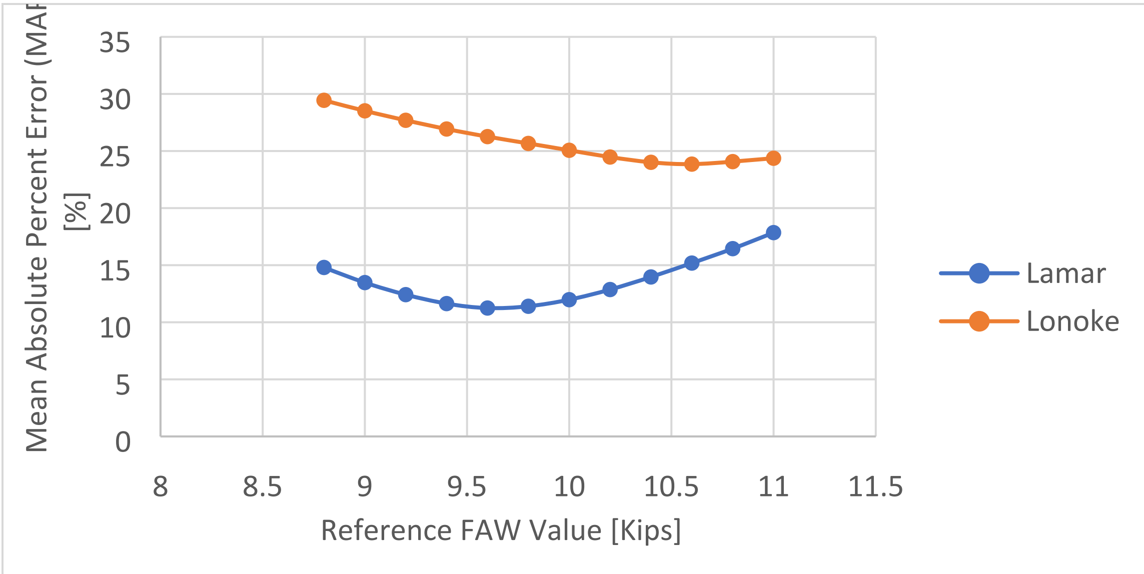


Figure 34. ARDOT Parameter Tuning for Reference FAW Value.

As described in Section 4.1, the initial reference FAW values at Lamar were 10.95 kips and 11.46 kips for the unloaded and fully loaded weight bins, respectively. These values at Lonoke were 11.18 kips and 11.72 kips. For the MNDOT method the FAW parameters were adjusted in increments of 0.5 kips over a range 7 to 13.5 kips. Figure 35 and Figure 36 show the results of the reference FAW value tuning. It was found that these reference values did vary between WIM sites. The resulting trend of reference value vs. error matched predictions of a significantly curved shape with a distinct minimum value. This trend is due to the close relationship between the reference values and calibration factors, as shown in Equations 5 and 6.

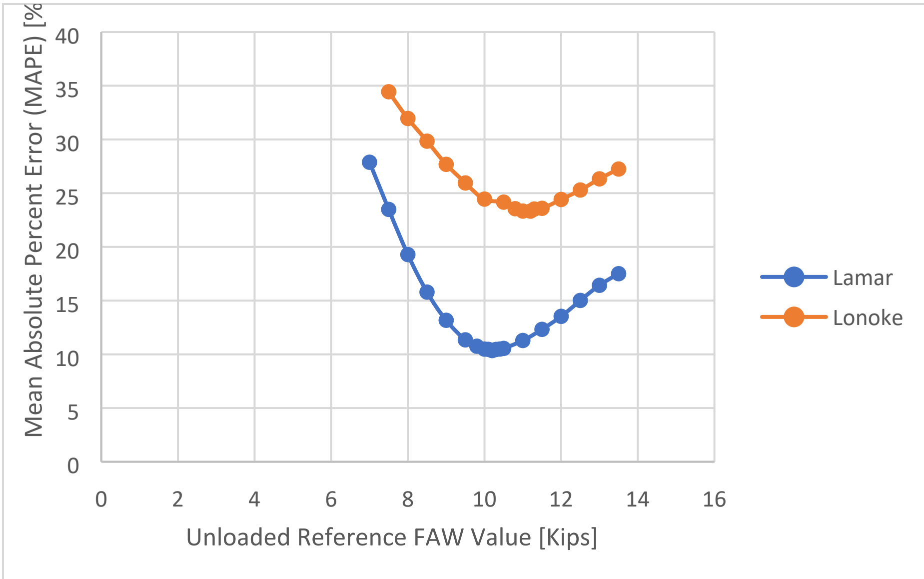


Figure 35. Parameter Tuning of the Unloaded FAW Reference Value for MNDOT.

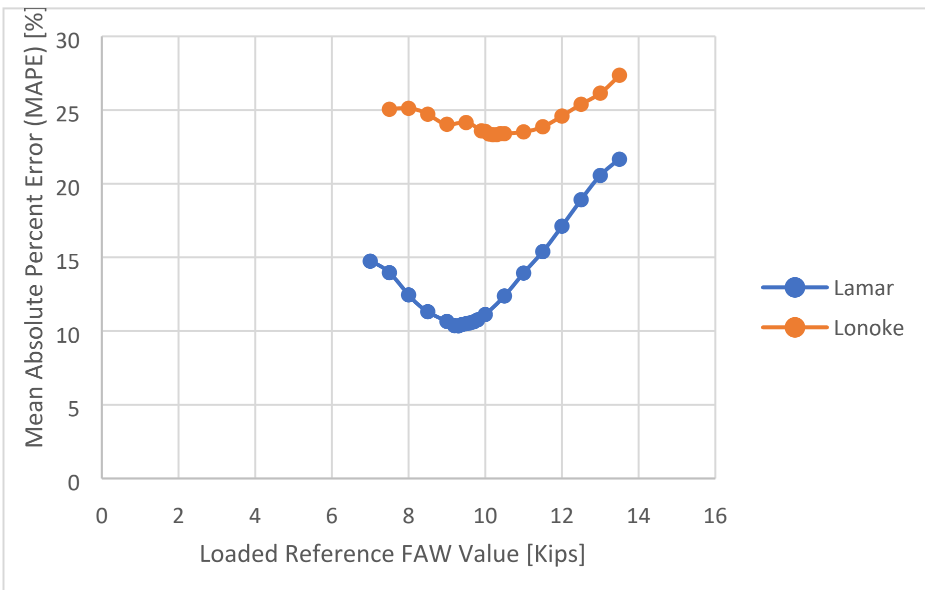


Figure 36. Parameter Tuning of the Loaded FAW Reference Value for MNDOT.

7. Discussion and Conclusions

7.1 Tuned Algorithm Performance

This project examined the effect of altering auto-calibration parameters to improve weight measurement accuracy. The parameters examined included: FAW reference values, GVW threshold values for loaded vehicles, and calibration frequency. A summary of the parameters that produced the lowest MAPE at each of the study sites in shown Table 7.

Table 7. Summary of Parameter Sensitivity Analysis.

	Site Characteristics								**Results**	
Site	Sample Size	GVW Distribution	% Class 9	Reference FAW per bin [kips]		Bimodal Threshold Values [kips]	Frequency		**Improvement Over Baseline Procedures [% difference]**	
Auto-Calibration Method				ARDOT	MNDOT	MNDOT	ARDOT	MNDOT	**ARDOT**	**MNDOT**
Baseline Parameter Values				10.2	8.5 9.3 10.4	72.0	50	250	**-**	**-**
Lamar	77	Bimodal	63.2	9.6	10.2 9.3	71.0	50	50	**2.5**	**4.8**
Lonoke	44	Bimodal	70.0	10.6	11.0 10.3	72.5	50	50	**1.1**	**3.1**
Bald Knob	2	Bimodal	61.6	Not evaluated						

The critical factors in improving auto-calibration performance is identifying and quantifying GVW distribution and the reference FAW value per GVW bin on a site-specific basis. Using GMM, the GVW distribution at the Alma static scale enforcement site was found to be tri-modal, while the WIM sites showed bimodal distributions. From the estimated GMM models, the threshold separating the fully loaded vehicles were found to be 71 kips and 72.5 kips for Lamar and Lonoke, respectively. From this, we conclude that the threshold between the heaviest vehicles and the rest of the traffic stream is spatially transferrable and approximately 72 kips as suggested in the literature. It is assumed that some sites may exhibit tri-modal GVW distributions. However, the WIM sites included in this analysis did not allow us to study the effect of changing the unloaded vehicle threshold which should be examined in future work.

This research indicates that site-specific data should be collected to establish reference parameter values for FAW and GVW shape. Site-specific data would need to include static scale and WIM measured weight comparisons for at least a 12 hour period (dawn to dusk) to understand the temperature effects on the sensor. It can be cost prohibitive and time consuming to collect site-specific data as described above (although it is less expensive that running test trucks over the site for calibration as per the ASTM guideline). To reduce labor costs of manually matching trucks across sites, Automatic Vehicle Identification (AVI) technology can be used.

Parameter tuning produced an improvement in MAPE of 2.5 % and 1.1 % at Lamar and Lonoke, respectively. Improved WIM data accuracy is beneficial for WIM data applications like pavement design, maintenance programming, and travel forecasting. For example, due to the 4^{th} power relationship between axle weights and estimated ESAL values, this improvement results in a difference in Equivalent Single Axle Load (ESAL) estimation of 39% and 1.4%, respectively, which has significant implications for pavement design.

7.2 Future Research

This research has evaluated the performance of two state-of-the-practice auto-calibration algorithms and investigated the limitations of each. Using data collected from two WIM sites and a static scale this research has shown 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. To capture spatial and temporal transferability, further data collection is required from WIM sites located somewhere other than I-40. The transferability of site characteristics between locations that exhibit a trimodal GVW distribution needs to be investigated. An analysis of GVW data from additional WIM sites would be sufficient to categorize a potential data collection site as having the desired GVW distribution. The FAW reference value has been shown to be specific to each site; the reference FAW value for each GVW weight bin should be determined for every WIM site. Moreover, testing at low volume sites, e.g. non-interstate WIM sites, is also needed. For low volume sites, calibration frequency performance in terms of time elapsed rather than vehicles elapsed should be evaluated. Due to the high volume of the WIM sites at Lamar and Lonoke, the effect of speed on WIM error was not evaluated due to the small range of measured vehicle speeds. At low volume sites, an analysis of the effect of speed on calibration error should be conducted. This analysis should include the calculation of additional calibration factors based upon vehicle speed. Lastly, the use of a temperature correction curve should be investigated, which would allow for larger sample sizes and less frequent calibration.

8. References

Al-Qadi, Imad, Hao Wang, Yanfeng Ouyang, Kirk Grimmelsman, and Jeffrey E Purdy. 2016. "LTBP Program's Literature Review on Weigh-in-Motion System." In.: United States. Federal Highway Administration.

Bunnell, Wayne, Howell Li, Michael Reed, Tim Wells, Dwayne Harris, Marc Antich, and Darcy M Bullock. 2018. "Weigh-in-Motion Sensor Calibration Using Outlier Analysis." In.

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

Davies, Peter, and Fraser Sommerville. 1987. *Calibration and accuracy testing of weigh-in-motion systems*.

'Electronique Controle Mesure'. Accessed November 30, 2018. http://www.easybib.com/guides/citation-guides/mla-8/cite-digital-image-mla-8/.

Fiorillo, Graziano, and Michel Ghosn. 2014. 'Procedure for statistical categorization of overweight vehicles in a WIM database', *Journal of Transportation Engineering*, 140: 04014011.

Gupta, Diwakar, Xiaxu Tang, and Lu Yuan. 2018. "Weigh-in-Motion Sensor and Controller Operation and Performance Comparison." In.: Minnesota. Dept. of Transportation. Research Services & Library.

Hallenbeck, Mark. 1998. 'Long term pavement performance program protocol for calibrating traffic data collection equipment'.

Hashemi Vaziri, Shahram, Carl T Haas, Leo Rothenburg, and Ralph C Haas. 2013. 'Investigation of the Effect of Weight Factor on Performance of Piezoelectric Weigh-in-Motion Sensors', *Journal of Transportation Engineering*, 139: 913-22.

Hyun, Kyung, Sarah Hernandez, Andre Tok, and Stephen G Ritchie. 2015. 'Truck Body Configuration Volume and Weight Distribution: Estimation by Using Weigh-in-Motion Data', *Transportation Research Record: Journal of the Transportation Research Board*: 103-12.

Jiang, Xiaohua, Shhram H Vaziri, Carl Haas, Leo Rothenburg, Gerhard Kennepohl, and R Haas. 2009. "Improvements in piezoelectric sensors and WIM data collection technology." In *Auunal Conference & Exhibition of Transport Association of Canada*.

Mai, Derong, Rod E Turochy, and David H Timm. 2013. 'Quality control of weigh-in-motion data incorporating threshold values and rational procedures', *Transportation Research Part C: Emerging Technologies*, 36: 116-24.

McCall, Bill, and Walter C Vodrazka Jr. 1997. 'States' successful practices weigh-in-motion handbook'.

Nassif, Hani, Chaekuk Na, Enson Portela, and Tulio Bittencourt. 2018. "An Approach to Minimize Weighing Error of WIM System Based on Ambient Temperature." In.

Nichols, Andrew P, and Darcy M Bullock. 2006. 'Automatic speed calibration methodology for traffic monitoring sites', *Journal of Transportation Engineering*, 132: 30-39.

Papagiannakis, A Thomas, and Rich Quinley. 2008. *High speed weigh-in-motion system calibration practices* (Transportation Research Board).

Quinley, Rich. 2010. "WIM data analyst's manual." In.: United States. Federal Highway Administration.

Sayyady, Fatemeh, Yahya Fathi, George F List, and John R Stone. 2013. 'Locating Traffic Sensors on a Highway Network: Models and Algorithms', *Transportation Research Record*, 2339: 30-38.

Sclove, Stanley L. 1987. 'Application of model-selection criteria to some problems in multivariate analysis', *Psychometrika*, 52: 333-43.

Sheskin, D., Handbook of Parametric and Nonparametric Statistical Procedures, Fifth Edition, 2011.

"Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Methods." In. 2017.