WIM SENSORS ACCURACY, GUIDELINES FOR EQUIPMENT SELECTION AND CALIBRATION, AND TRAFFIC LOADING DATA APPLICATIONS

By

Muhammad Munum Masud

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Civil Engineering – Doctor of Philosophy

2022

ABSTRACT

Weigh-in-Motion (WIM) technology is one of the primary tools used for pavement management. It can provide essential and accurate truck traffic information, including vehicle class and speed, vehicle count, gross vehicle weight (GVW), single axle (SA) and tandem axle (TA) weights, axle spacing, and the date and time of the event. The State Departments of Transportation (DOTs) gather WIM data for various applications, including highway planning, pavement and bridge design, commercial vehicle weight enforcement, asset management, and freight planning and logistics. Because of the wide range of applications, the data obtained at WIM stations must be accurate, consistent, and reflect actual field conditions.

This study addressed four critical concerns related to WIM equipment performance, calibration needs, traffic loading data quality, and applications. Precisely, the current research advanced the state of the practice knowledge about (a) potential factors impacting WIM system accuracy, (b) accuracy and consistency of traffic loading data and calibration needs of WIM stations, (c) revised/modified guidelines for WIM equipment calibration, and (d) estimation of commercial freight tonnage from Gross Vehicle Weight (GVW) data. The research objectives were accomplished by synthesizing and analyzing the WIM performance and traffic loading data available in the Long Term Pavement Performance (LTPP) traffic database and data available through other state DOTs. The WIM sites analyzed in this study are from 30 states within the United States and 3 Canadian provinces.

Decision tree models were developed in this study to illustrate a potential for estimating the expected WIM measurement error range using information about the WIM site and sensor-related factors. The results show that the sensor array and sensor types are the most important predictors, followed by WIM controller functionality (speed points). The data analysis and

results also show that the climate can be important for some sensor types. One can integrate this information with equipment installation and life cycle costs to determine the most reliable and economical WIM equipment while also considering accuracy requirements by WIM data users.

One way to evaluate WIM measurement errors is by using the data collected immediately before and after equipment calibration. The limitation of this approach is that the data represent a snapshot in time and may not represent a long-term WIM site performance. Consequently, an alternative approach was needed to characterize temporal variations in WIM data consistency. This study presents a method to estimate WIM system accuracy based on axle load spectra attributes [Normalized Axle Load Spectra (NALS) shape factors]. This analysis's main objective is to determine WIM system errors based on axle loading without physically performing equipment calibration. Using NALS to estimate WIM system accuracy can save a significant amount of time and resources, usually spent on equipment calibrations yearly. Successful WIM equipment calibration can eliminate systematic weight, speed, and axle spacing errors. The suggested changes in current WIM calibration procedures related to truck type (loaded truck), number of truck runs, and truck speed (multiple speed points) can significantly reduce the time and resources needed for successful equipment calibration.

Accurate freight tonnage estimates and trends are essential due to their implications on economic, infrastructure development, and transportation policy decision-making. This study presents a practical application of WIM data to estimate freight tonnage and classify commodity types. The payloads computed for Class 9 trucks from GVW data strongly correlated with the average freight tonnage obtained from a commercial data source, i.e., Transearch from the IHS market. The user can independently verify the freight estimates from surveys at locations close to WIM sites.

I would like to dedicate my dissertation to my late grandfather Al-haj Muhammad Din, my beloved parents Masud Ahmed and Khurshid Bibi, my uncle Maqbool, Farooq, Mehboob, Sardar (late), and Iqbal (late), my dear sisters, and my elder brother Aamir. I also dedicate this dissertation to my wife Maryam, and our daughters Abeera and Zara who did not get the time and attention they needed from me during the last six and a half years because of my studies and had to participate in many school events, activities, and celebrations without me.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude and deep regards to my academic advisor Dr. Syed Waqar Haider, Associate Professor at the department of Civil and Environmental Engineering, for his constant motivation, monitoring, and guidance throughout my academic tenure here at Michigan State University. I will always remember his humble, soft and caring attitude, and the moral and emotional support that he provided over the past six and half years.

I would like to thank Professor Neeraj Buch, Professor Karim Chatti, Professor Emin kutay, and Professor Kirk Dolan for serving on my Ph.D. committee. I immensely benefitted from their intellect, valuable feedback, and academic discussions throughout my academic stay at MSU. I would also like to thank all other faculty members who taught me, especially, Professor Julie Winkler from the Department of Geography and Professor Robert Tempelman from the Department of Animal Science who helped me in attaining and improving my statistical and quantitative data analyses skills. I am also thankful to Dr. Jenahavive K. Morgan for providing mentoring and support during my teaching appointments.

I am thankful to my friends, family members, and well-wishers here at Michigan State University and in Pakistan who are always happy about my achievements and keep praying for my success. In particular, I would like to acknowledge, Mr. Jehangir Leghari, Mr. Naeem Tahir, and Mr. Hafiz Muhamamd Usman for their help and prayers.

I appreciate the consistent help and support provided by Laura Post, Laura Taylor, Baily Weber, and Joseph Nguyen, in solving my administrative problems and in making this academic stay a memorable chapter of my life. I cannot express in words how helpful and accommodating these individuals are for the graduate students.

I would like to thank Dr. Olga Selezneva, Principal Engineer at Applied Research Associates (ARA), and Mr. Dean Wolf, traffic services supervisor at ARA, for their positive critique and valuable feedback on research findings that significantly enhanced the quality of this work.

I would like to acknowledge the National Cooperative Highway Research Program (NCHRP) funding support to complete this study at Michigan State University. I would like to acknowledge the Long Term Pavement Performance (LTPP) and Michigan Department of Transportation (MDOT) management and staff for providing timely support in data collection.

Finally, I thank the Engineer-in-Chief Branch, Pakistan Army, Ministry of Defense, and Higher Education Commission of Pakistan for financially supporting my Master's and Ph.D. studies, and I am grateful to the staff members who are working day and night for making this Ph.D. program a success.

I am also thankful to my elder brother Aamir for continuously looking after my ailed parents in Pakistan over the past six and half years and never letting them feel my absence. Finally, I would like to extend my special gratitude to my wife Maryam Ali for providing unconditional support, motivation, and prayers for the successful attainment of my Ph.D. degree.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AADTT	Annual Average Daily Truck Traffic
AI	Artificial Intelligence
ALS	Axle Load Spectra
ANN	Artificial Neural Network
ARA	Applied Research Associates
ASTM	American Society for Testing and Materials
ATRI	American Transportation Research Institute
BP	Bending Plate
CART	Classification and Regression Trees
CFS	Commodity Flow Survey
CI	Confidence Interval
CL-9	Class 9
COV	Coefficient of Variation
DF	Dry Freeze
DNF	Dry No Freeze
DOT	Department of Transportation
ETG	Expert Task Group
FAF	Freight Analysis Framework
FHWA	Federal Highway Administration
FWD	Falling Weight Deflectometer
GMM	Gaussian Mixture Model
GVW	Gross Vehicle Weight

IRI	International Roughness Index
LC	Load Cells
LTPP	Long Term Pavement Performance
MDOT	Michigan Department of Transportation
MOE	Margin of Error
MPE	Maximum Permissible Error
MSE	Mean Squared Error
NALS	Normalized Axle Load Spectra
NCFRP	National Cooperative Freight Research Program
NCHRP	National Cooperative Highway Research Program
ODME	Origin Destination Matrix Estimation
PC	Piezo Cable
PCC	Portland Cement Concrete
PDF	Probability Density Function
PL	Peak Load
QA	Quality Assurance
QC	Quality Control
QGIS	Quantum Geographic Information System
QP	Quartz Piezo
RMSE	Root Mean Squared Error
RQD	Research Quality Data
SA	Single Axle
SD	Standard Deviation

SPS	Specific Pavement Sections
TA	Tandem Axle
TE	Total Error
TMAS	Travel Monitoring Analysis System
TPF	Transportation Pooled Fund
VIUS	Vehicle Inventory Use Survey
VMT	Vehicle Miles Traveled
WF	Wet Freeze
WIM	Weigh-in-Motion
WNF	Wet No Freeze
WRI	WIM-Roughness Index

CHAPTER 1 INTRODUCTION

1.1 BACKGROUND

Weigh-in-Motion (WIM) technology is one of the primary tools used for pavement management. It can provide accurate information about the traffic on road networks, including but not limited to vehicle class and speed, vehicle count, gross vehicle weight (GVW), wheel and axle weights, axle spacing, and date and time of the event [1]. The State Department of Transportations (DOTs) are required to collect and submit WIM data to the Federal Highway Administration (FHWA) as part of its traffic monitoring program. Apart from reporting WIM data to FHWA, agencies collect WIM data for many reasons, including highway planning, pavement and bridge design, commercial vehicle weight enforcement, asset management, and freight planning and logistics [2].

Overloaded trucks pose severe challenges to road transport operations. Compared to a truck loaded within legal weight limits, an overloaded truck is likely to cause more damage to the pavement and can lead to severe consequences if involved in a traffic accident. Law enforcement agencies divert potentially overloaded trucks to static scales and issue tickets based on the information collected at a WIM station [3]. Therefore, with so many potential uses, the data collected at WIM stations must be accurate and represent actual field loadings and conditions.

1.2 KNOWLEDGE GAPS AND RESEARCH NEEDS

Past studies pointed to the importance of various factors on WIM measurement accuracy [4-13]. However, no comprehensive study was found to quantify the relative importance of multiple factors and under what conditions these factors become critical for WIM measurement accuracy.

Most of the past studies were based on limited field data, which raised questions about the adequacy and broad applicability of the results.

While several WIM guidance documents are available to assist state highway agencies in collecting WIM data, the advice requires specialized knowledge to implement that traffic data collectors frequently don't have. There is a lack of practical tools to help agencies implement the available guidance. Based on the review of the current state of the practice, two research focus areas were identified in this study (a) advancing the state of knowledge in managing WIM data quality through an understanding of the effects of various site conditions and WIM equipment characteristics on WIM measurement accuracy and consistency, and (b) addressing the critical need for the practical tools that highway agencies can successfully implement to improve WIM data accuracy and support WIM data quality assurance functions.

1.3 RESEARCH OBJECTIVES

This study addresses multifaceted issues related to WIM systems performance, calibration procedures, and site and sensor-related factors that affect the accuracy of different sensors. The ultimate goal of the data analysis is to develop guidelines that state highway agencies can easily implement to collect accurate and reliable WIM data. The primary objectives of the research are to (a) describe statistical concepts in establishing WIM data accuracy, (b) draw comparisons between available WIM accuracy standards, (c) provide representative WIM measurement errors for different sensors (d) develop models for WIM equipment, and site selection (e) assess WIM data consistency and calibration needs based on axle load spectra (f) provide guidelines for successful WIM equipment calibration by quantifying the effect of sample size (truck runs), speed, temperature, and truck type on WIM errors, and (g) extend WIM loading data applications to estimate commercial freight from Class 9 trucks payloads. These objectives were accomplished by synthesizing and analyzing the WIM error data available in the LTPP database.

1.4 RESEARCH APPROACH

There is a need to understand the relative importance of various sources of error on WIM data accuracy and for methods that could help minimize the effect of external factors on WIM data quality. Several factors affecting WIM data quality have been identified through the literature review. A comprehensive study was designed to quantify the effect of multiple factors on WIM data accuracy and evaluate the relative significance of different factors on WIM performance. WIM calibration is an essential activity for maintaining WIM data accuracy. Statistical analysis and machine learning techniques were used to develop data-driven methods for identifying WIM calibration needs based on analysis of statistical attributes computed based on WIM data reported by the WIM system for FHWA Class 9 trucks. The models developed in this research investigate the use of axle load spectra attributes to assess the systematic changes (bias) in WIM measurements for gross vehicle weight (GVW), single axle (SA) load, and tandem (TA) load. This methodology can save significant time and resources required for field validation of WIM performance using test trucks when applied in practice. Additionally, depending on the extent of information related to the site, sensor, and calibration-related factors, the decision tree models developed in this study can help highway agencies to optimize WIM sensor type and array selection. This information can be integrated with WIM equipment installation costs and life cycle costs to determine the most reliable and economical equipment while also considering WIM data accuracy requirements received from WIM data users.

1.5 POTENTIAL BENEFITS OF THE STUDY

The details about representative WIM measurement errors by sensor type are presented in this report. These findings have an immediate practical application by providing highway agencies with the benchmark values demonstrating the practically achievable accuracy and variability of WIM measurements for different WIM sensor types after successful calibration.

Decision tree models were developed in this study to illustrate a potential for estimating the expected WIM measurement error range using information about the WIM site and sensor-related factors. One can integrate this information with equipment installation and life cycle costs to determine the most reliable and economical WIM equipment while also considering accuracy requirements by WIM data users.

Successful WIM equipment calibration can eliminate systematic weight, speed, and axle spacing errors. The suggested changes in current WIM calibration procedures related to truck type (loaded truck), number of truck runs, and truck speed (multiple speed points) can significantly reduce the time and resources needed for successful equipment calibration.

Accurate freight tonnage estimates and trends are essential due to their implications on economic, infrastructure development, and transportation policy decision-making. This study presents a practical application of WIM data to estimate freight tonnage and classify commodity types. The proposed method has good potential for application at WIM sites collecting loading data. Using WIM data is a different approach to traditional freight data collection methods like truck surveys, consumer reports, vehicle inventory & user surveys, commodity flow surveys, freight analyses framework, and other commercial data sources. The user can independently verify the freight estimates from surveys at locations close to WIM sites.

1.6 OUTLINE OF THE DISSERTATION

This dissertation contains eight chapters. Chapter 1 outlines the background, problem statement, research objectives, and potential benefits of this research and briefly describes the research approach. Chapter 2 documents a comprehensive literature review, including a description of WIM accuracy and system performance requirements, international WIM accuracy assessment standards, factors affecting WIM system accuracy, and issues with WIM system calibration. Chapter 3 describes the criteria for data selection, data sources, extent, and limitations. This chapter also discusses sources of various data types used in this study. The summary of available LTPP and other state-owned WIM sites considered for this analysis is also presented in chapter 3. Chapter 4 provides data analyses approach to evaluate and quantify the effect of site, sensor, and calibration-related factors on WIM measurement errors. This chapter presents a methodology that WIM data users and WIM data providers can use to estimate the expected WIM measurement accuracy for a given set of site conditions and WIM system design attributes. Chapter 5 provides a set of statistical procedures developed to identify and quantify changes in WIM measurement bias (calibration drift) based on analysis of changes in axle load spectra attributes for FHWA Class 9 vehicles (typically used as a calibration truck type) for WIM equipment calibration events. Chapter 6 addresses three core issues related to WIM systems accuracy and calibration procedures, i.e., how to; (1) perform successful calibration of a WIM system by quantifying the effect of sample size (truck runs), speed, temperature, and truck type on measurement errors, (2) model gross vehicle weight (GVW) WIM errors as a function of individual axle errors [(single axle (SA) and two tandem axles (TA), (drive and trailer)], and (3) estimate WIM measurement errors using the LTPP and the ASTM protocols. Chapter 7 demonstrates useful applications of axle load spectra to estimate commercial freight tonnage.

The presented methodology uses GVW loading data for Class 9 trucks to estimate vehicle payload and commodity type. Chapter 8 provides conclusions and highlights the most critical findings from the WIM data analysis, the significance of the results, and the potential benefits of the research outcomes for collecting high-quality WIM data. This chapter also makes recommendations for future data collection.

CHAPTER 2 LITERATURE REVIEW

2.1 BACKGROUND

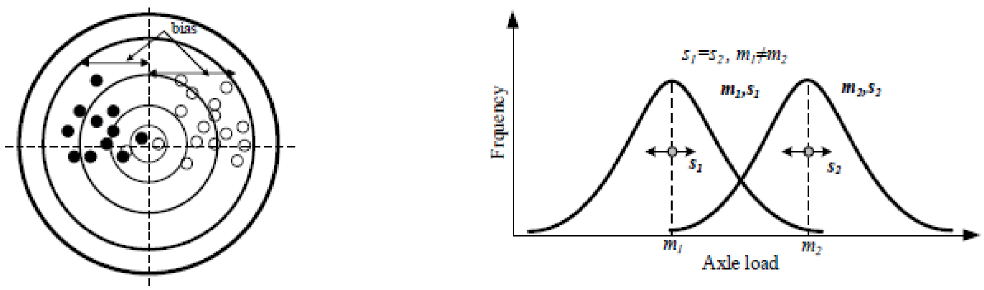
Weigh-in-motion (WIM) technology is one of the primary tools used for pavement management. It can provide accurate information about the traffic on road networks, including but not limited to vehicle class and speed, vehicle count, gross vehicle weight (GVW), wheel and axle weights, axle spacing, and date and time of the event [1]. The State Department of Transportation (DOTs) must collect and submit WIM data to the Federal Highway Administration (FHWA) as part of its traffic monitoring program. Apart from reporting WIM data to FHWA, agencies collect WIM data for many reasons, including highway planning, pavement and bridge design, commercial vehicle weight enforcement, asset management, and freight planning and logistics [2]. For the last few years, the traditional WIM stations amalgamated with advanced traffic monitoring technologies (e.g., image acquisition devices) collect additional vehicle and traffic information that Artificial Intelligence (AI) techniques can process. This data collection and processing approach at WIM stations has opened many innovative applications, including vehicle color identification, tire footprint information, missing/flat/mismatched tires, lane potion, out-of-lane detection, and load types detection [14]. Overloaded trucks pose severe challenges to road transport operations. Compared to a truck loaded within legal weight limits, an overloaded truck is likely to cause more damage to the pavement and can lead to severe consequences if involved in a traffic accident. Law enforcement agencies divert potentially overloaded trucks to static scales and also issue tickets based on the information collected at a WIM station [3]. Therefore, with so many potential uses, the data collected at WIM stations must be accurate and represent actual field loadings and conditions.

The accuracy of the WIM systems is a primary concern for its manufacturers and users. The accuracy of weighing results obtained from WIM systems largely influences the control of the overloaded vehicle on highways. Several WIM technologies exist to capture the applied forces and predict static weight. Because WIM technology estimates static weight for a moving vehicle, there are many potential sources of measurement error. Some errors are due to the variation in the forces transferred by the moving truck to the sensor; the others are because of WIM equipment type and site conditions.

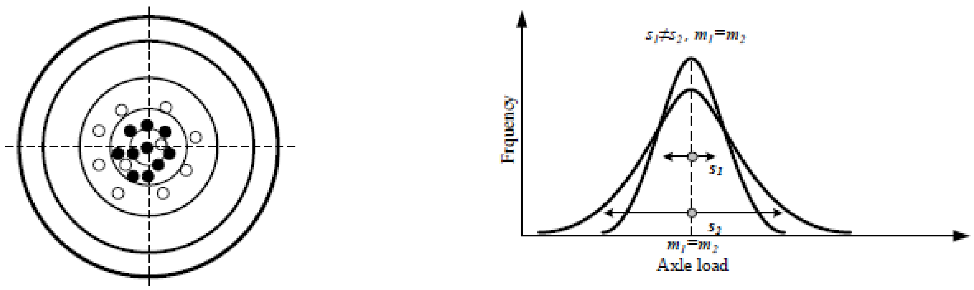
The long-term pavement performance (LTPP) traffic data module is one of its most significant components. The module provides data related to traffic inputs for pavement analyses, distributions to create AASHTOWare inputs, Pavement-ME tables, truck volumes, WIM calibration details, axle counts, vehicle classification, traffic summary statistics, and many more. The LTPP traffic data are the foundation for new pavement designs for years to come [11, 15]. Initially, the Pavement-ME traffic loading defaults were developed based on the data collected by the state agencies using the early generations of WIM sensors and submitted to LTPP. Since then, the LTPP has undertaken the specific pavement sections transportation pooled fund study 5(004) (SPS TPF) study. This program uses permanent WIM systems with more accurate and reliable sensors to collect high-quality axle loading data. The study was designed with the support of the Transportation Research Board Traffic Expert Task Group (ETG), and the data were collected by using (a) a centralized effort and (b) standardized data collection equipment and procedures [16].

2.2 WIM SYSTEM ACCURACY AND PERFORMANCE REQUIREMENTS

Establishing a baseline for assessing the impact of multiple factors on WIM data accuracy would require an understanding of measurement accuracy and consistency concepts. Figure 2-1 shows the target analogy to visualize the differences between accuracy and consistency. Accuracy is the conformity of results to the true value, i.e., the absence of bias. Bias is a tendency of an estimate to deviate in one direction from the true value. Consistency or precision is related to the repeatability of a process. The variability of repeat measurements can characterize precision under carefully controlled conditions. Figure 2-1 also illustrates that it is possible to be consistent (or precise, as applied to target shooting) without being accurate or accurate without being consistent (low precision). Ideally, we would like a measurement process to be accurate and consistent.



(a) Negative bias for the black data set and positive bias for the blank data set, similar precision for both data sets



(b) Good (black) and poor (blank) precision with no bias

Figure 2-1 Target analogy for understanding precision and bias.

The WIM system accuracy is measured in terms of the relative difference between WIM and static weights. The following equation can express the relative WIM error:

$$\varepsilon = \frac{WIM\ \ weight\ \ -\ Static\ weight}{Static\ weight} \times 100 \tag{2.1}$$

This relative error is commonly referred to as measurement error for a WIM scale. Further, this accuracy will vary for different types of WIM sensor technologies. For a well-calibrated WIM system, typical WIM measurement error follows a normal distribution with a zero mean (no bias) and a standard deviation [17], as shown in Equation 2.2:

$$\varepsilon = \frac{X' - X}{X} \sim N\left(0, \sigma_{\varepsilon}^{2}\right) \tag{2.2}$$

Where

$X' =$ load measured on a WIM scale for an axle configuration
$X =$ load measured on a static scale for the same axle configuration
$\sigma_{\varepsilon} =$ standard deviation (SD) characterizing the accuracy of the WIM scale

Several WIM accuracy assessment protocols are available these days. The American Society for Testing and Materials International Standard, ASTM E1318-09 [18] is mainly adopted in the US, and the European Road Specification COST-323 [19, 20] is used in European countries. In addition, the LTPP field operations guide also documents the procedures to evaluate the WIM system accuracy [21-25]. The following section presents a brief discussion of available WIM accuracy assessment protocols.

2.2.1 ASTM WIM Protocol ASTM E1318-09 (2017)

American Society for Testing and Materials International Standard, ASTM E1318-09 (updated in 2017), is a broadly recognized WIM measurement protocol in the United States. This specification classifies four types of WIM systems according to their application. Table 2-1 summarizes performance specifications for different WIM systems. The Types I and II systems can be installed at traffic data collection sites for vehicles moving at highway speeds (10 to 80 mph). Types III and IV are designed for weight-enforcement stations [18]. The ASTM Type I accuracy criterion was used to assess the WIM system performance for the SPS-TPF study. The static load (reference load) error limits as defined by ASTM WIM standard are ± 2 %, ± 3 %, ± 4

%, and ± 5 % for GVW, TA, SA, and wheel loads, respectively. WIM system performance is ascertained by comparing the reference and WIM weights for all the data items listed in Table 2-1. The following relationship is used in the specification to calculate the percent difference between the WIM system and the reference values, as shown in Equation 2.3:

$$D = C - R \tag{2.3}$$

The relative difference, d, in loads and weights (%) can be obtained by Equation 2.4:

$$d = \frac{C - R}{R} = \frac{D}{R} \tag{2.4}$$

where,

D = The difference in speed (mph), axle spacing (ft.), and wheelbase (ft.)

d = The difference in the value of the data item (wheel load, axle load, axle-group load, and gross vehicle weight) produced by the WIM system and the corresponding reference value is expressed as a percent of the reference value

C = Value of the data item produced by the WIM system

R = The corresponding reference value for the data item

Table 2-1 Functional performance requirements for WIM systems (ASTM).

Function	Tolerance for 95 % Compliance[a]				
	Type I	Type II	Type III	Type IV	
				value ≥ lbs.[b]	± lbs.
Wheel Load	± 25 %	-	± 20 %	5000	300
Axle Load	± 20 %	± 30 %	± 15 %	12000	500
Axle-Group Load	± 15 %	± 20 %	± 10 %	25000	1200
Gross-vehicle Weight	± 10%	± 15 %	± 6 %	60000	2500
Speed	± 1 mph				
Axle-Spacing and wheelbase	± 0.5ft				

[a] 95 % of data produced by the WIM must fall within tolerance. [b] Lower values are not a concern for enforcement.

2.2.2 COST-323 WIM Standard

This specification mainly addresses the issues associated with high-speed WIM systems, i.e., the WIM systems installed on one or more traffic lanes and operated under normal traffic conditions. According to this specification, under defined operating conditions (moving traffic, tire loads, etc.), the accuracy of a WIM system may only be defined statistically by a confidence interval of the relative error of a unit (an axle, an axle group, or a gross weight) defined as by Equation 2-1.

Such a confidence interval centered on the static load/weight is [-δ; +δ], where δ is the tolerance for a confidence level **π** (for example, 90 or 95%). A typical standardized table of δ values taken from European WIM specifications is shown in Table 2-2 [20].

Table 2-2 Accuracy classes definition, [value of δ, i.e., confidence interval width (%)].

Function	Accuracy classes						
	A(5)	B+ (7)	B (10)	C (15)	D+(20)	D (25)	E*
Gross Weight (GW)	5	7	10	15	20	25	> 25
Group of axle (AoG)	7	10	13	18	23	28	> 28
Single axle (SA)	8	11	15	20	25	30	> 30
Axle of group (GA)	10	15	20	25	30	35	> 35

* Class E is defined for the WIM systems which do not meet the class D (25) requirements.

2.2.2.1 Test Conditions and Confidence Levels ($\pi_\circ$)

This specification allows the user to set a test plan by selecting an appropriate combination of repeatability/reproducibility and environmental conditions. As per the specification, Table 2-3 provides the minimum levels of confidence (π_o) for different tests and environmental conditions. As compared to environmental repeatability (**I**), smaller π_o values are required for limited (**II**) and full (**III**) environmental reproducibility conditions.

Table 2-3 Minimum percentage levels of confidence πo of the centered confidence intervals case.

Test conditions		Sample Size (n)					
		10	20	30	60	120*	∞
Full repeatability	I	95.0	97.2	97.9	98.4	98.7	99.2
	II	93.3	96.2	97.0	97.8	98.2	98.9
	III	91.4	95.0	96.0	97.0	97.6	98.5
Extended repeatability	I	90.0	94.1	95.3	96.4	97.1	98.2
	II	87.5	92.5	93.9	95.3	96.1	97.5
	III	84.7	90.7	92.4	94.1	95.1	96.8
Limited reproducibility	I	85.0	90.8	92.5	94.2	95.2	97.0
	II	81.9	88.7	90.7	92.7	93.9	96.0
	III	78.6	86.4	88.7	91.1	92.5	95.0
Full reproducibility	I	80	87.4	89.6	91.8	93.1	95.4
	II	76.6	84.9	87.4	90.0	91.5	94.3
	III	73.0	82.3	85.1	88.1	89.9	93.1

* Sample sizes (*n*) not mentioned in the table may be interpolated.

2.2.2.2 Accuracy Assessment of the WIM System

The European WIM standard uses a pre-weighed or post-weighed vehicle to check the accuracy of a WIM system. The sample statistics, including mean (bias) ***m***, standard deviation ***s***, and the number values ***n*** are calculated and used as per the specification. "A lower bound ***π*** of the probability that an individual error falls within the specified interval [-δ; +δ] is calculated and compared to the specified **π**₀."

According to the statistics provided in the standard, an upper bound on the customer risk, **π,** for an α=0.05, is given by:

$$\pi = \Phi(u_1) - \Phi(u_2) \quad (2.5)$$

$$u_1 = \frac{\delta - m}{s} - \frac{t_{n-1,0.975}}{\sqrt{n}} \quad (2.6)$$

$$u_2 = \frac{-\delta - m}{s} + \frac{t_{n-1,0.975}}{\sqrt{n}} \quad (2.7)$$

The *function* Φ is the cumulative distribution function of a student variable and $t_{n-1,0.975}$ is a student variable with $(n-1)$ degrees of freedom. For a sample size greater than 60, the cumulative function Φ in the above equation can be approximated by the cumulative distribution of a standard normal variable. The following criteria are used for the acceptance of WIM systems:

- If $\pi \geq \pi_o$, the system is accepted in the accuracy class of tolerance for the criterion considered.
- If $\pi < \pi_o$, the system cannot be accepted in the proposed accuracy class, and the acceptance test is repeated with a lower accuracy class, i.e., a larger value of δ.

2.2.3 LTPP Field Operations Guide

The WIM equipment calibration or pre-validation is performed using a known truck weight on the static scales. As per the LTPP Field Operations Guide, the static weights are collected at the certified scales using the procedure documented in the ASTM WIM standard and remain constant during the WIM equipment calibration or pre-validation. However, the WIM weights may vary based on truck speed, temperature fluctuations, and other site factors. The WIM equipment pre-validation is a process of assessing the performance of a WIM system based on an earlier calibration event. The compensation (the process of altering the equipment calibration factors) does not apply during WIM equipment pre-validation [18, 26, 27]. The LTPP Field operations guide developed for SPS-TPF WIM sites describes a WIM site that can provide research quality loading data if it meets the ASTM Type I tolerance limits. The ASTM criterion of no more than 5% of the errors exceeding tolerance is not applied to determine the WIM site performance. The LTPP method does not apply to wheel loads. This guide presents a procedure for calculating WIM accuracy (total measurement error, abbreviated as TE in this paper) that uses measurement bias (mean error) and SD of errors based on sample size for multiple truck runs. As per this guide, the total WIM measurement error based on the test truck data obtained from a calibration event can be estimated using Equation 2.8. The equation is a combination of bias (mean error) and margin of error (MOE) with 95% CI as described in the LTPP Field Operations Guide for SPS WIM sites [26].

$$\text{Total Error} = \overline{X_{\varepsilon}} \pm t_{n-1, \alpha/2}\ \sigma \tag{2.8}$$

where,

$\overline{X_\varepsilon}$ = Mean error (bias) that can be reduced (to an extent) through successful equipment calibration.

t= t is the critical value (depending on the confidence level) of the student's t distribution based on the n-1 degree of freedoms

n= Sample size, 40 for the LTPP SPS-TPF WIM sites (20 each for fully loaded and partially loaded trucks)

σ= SD of the errors based on test truck data

α = The significance level, a 95% confidence level, its value is 0.05

2.2.4 WIM Technology Austroads (AP-R168)

Austroads (2000) defined WIM as a device that measures the dynamic axle mass of a moving vehicle to estimate the corresponding static axle mass. That is, the WIM device captures and records the axle or axle group mass and the gross vehicle weight as the vehicle is moving. WIM systems should not be confused with onboard vehicle weighing systems. Onboard weighing systems are mounted or attached to the vehicle, while WIM systems are independent of the vehicle being weighed. WIM system falls into two main groups concerning their speed; low-speed WIM (less than or equal to 15 Km/h) and high-speed WIM (greater than 15 Km/h). At the time of the Austroads report, there were 12 high speed and 5 low-speed WIM systems by different vendors and suppliers that were either being used or available in Australia (and New Zealand). Limited quantitative or field information exists on the performance and life span of mass sensors (i.e., the primary component of a WIM system) [10].

No standard Australian specification or test method is available to determine and report WIM system accuracy results. Generally, accuracy is specified in terms of 95% tolerance of the vehicle being weighed. For example, an accuracy result such as 95% of vehicles weighed was within 10% and 20% for gross vehicle mass and individual axle mass. Austroads (2000) recommended adopting or modifying any existing standard (ASTM 1994 and COST-323 1997) for the evaluation and accuracy of WIM systems in Australia.

2.2.4.1 Types of errors

Different types of errors can affect the accuracy of a WIM system. Austroads (2000) reported the following types of errors that can be associated with WIM system accuracy:

- Actual error: associated with the error in determining the true mass of the vehicle.
- Systematic error: associated with flaws in initial calibration or drift in existing calibration; quantified as the mean or average.
- Random error: associated with WIM system errors or vehicle characteristics quantified as the standard deviation.

2.2.4.2 Factors affecting WIM accuracy

Austroads (2000) also described the following factors which impact WIM system performance and accuracy:

- WIM location characteristics: Pavement-related factors like longitudinal and transverse profile, curvature, cross slope, pavement surface deflection, and pavement surface condition can be influential factors.
- Vehicular characteristics: vehicle speed, acceleration/deceleration, body and suspension type, and type condition can all affect the performance of the WIM system
- Environmental characteristics: Temperature, wind, and ice can significantly affect WIM system performance. Mostly, information related to these factors and their effect on the WIM system performance is well known to vendors, but most vendors don't disclose it.

2.2.5 The Dutch Metrology Institute (NMi) International WIM Standard

The Dutch Metrology Institute (NMi) International WIM standard was prepared in Europe by a group of international experts with specialized knowledge of metrology, standardization, and WIM technology. The NMi team believed the existing international standards (COST 323, ASTM, and OIML-R134) and specifications on WIM system performance have their areas of application with some pros and cons. They also found that none of the existing standards encompass all the applications and operating conditions for WIM systems, e.g., for direct enforcement of overloading under normal highway conditions. The NMi international standard was developed with specific intended characteristics, including ease of access, widely acceptable, objectiveness, and independence for technology or commercial bias. Some of the procedures related to accuracy and tolerance levels defined in the NMi standard are somewhat like ASTM 1318-09 and COST 323. The NMi standard provides the performance requirements of WIM systems and the minimum testing methods required to achieve desired performance. An advancement in NMi standard is its legal application specifications and test methods for WIM systems [28]. Essential features of NMi international WIM standard are discussed next.

2.2.5.1 Weighing specifications for statistical applications

For statistical applications, the NMi standard classified WIM systems according to their weighing performance into five accuracy classes using the capital letter 'S'. Accuracy levels for each statistical class are summarized in Table 2-4. Here, the accuracy level quantifies the maximum size of the two standard deviation interval [-2σ, $+2\sigma$] of the relative measurement error, and under the normal or Gaussian distribution, this interval includes 95% of all measurements.

Table 2-4 Statistical accuracy levels δ (%) per class.

Measured quantity	Accuracy classes				
	S (5)	S (7)	S (10)	S (15)	S (20)
Gross Vehicle Weight	5	7	10	15	20
Axle Group Load	8	11	15	20	25
Axle Loads	10	15	20	25	30

2.2.5.2 Weighing specifications for legal applications

For legal applications, the NMi standard classified WIM systems according to their weighing performance into four accuracy classes using the capital letter 'S'. Accuracy levels for each legal class are summarized in Table 2-5. For legal applications, the accuracy level quantifies the maximum size of the maximum permissible error (MPE) [-MPE, +MPE] of the relative measurement error. This interval includes 100% of all measurements. Table 2-6 summarizes NMi test specifications for statistical and legal applications of WIM systems.

Table 2-5 Legal accuracy levels MPE (%) per class.

Measured quantity	Accuracy classes			
	L (3)	L (5)	L (7)	S (10)
Gross Vehicle Weight	3	5	7	10
Axle Group Load	5	8	11	15
Axle Loads	7	10	15	20

2.2.5.3 Reference values (Static accuracy) and length measurements

According to this standard, the gross weight and the axle (group) loads shall be determined using a static weighbridge, portable scales, or low-speed WIM systems capable of weighing the complete vehicle at once with an error less than or equal to one-third (1/3) of the applicable error specified for statistical and legal applications.

Table 2-6 Summary of NMi standard specifications.

Application Type	Test Method	Test Method Description	Minimum number of test vehicles[a]	Number of runs	Acceptance criteria for accuracy
Statistical	Initial verification	Typically done after installation or major repairs affecting the sensors	2	20	95% measurements for relative error lie within ±δ[b]
Statistical	In-service verification	Used to verify if a system is still operating within specifications	1	10	95% measurements for relative error lie within ±δ
Legal	Type approval	First extensive performance test of a new system under full operating conditions	3	90	100% measurements for relative error should be less than 0.5MPE
Legal	Initial verification	Typically done after installation or major repairs affecting the sensors	2	60	100% measurements for relative error should be less than MPE
Legal	In-service verification	Used to verify if a system is still operating within specifications	1	10	100% measurements for relative error should be less than MPE

[a] Different types of test vehicles make multiple runs at maximum, minimum, and middle operating speed.
[b] Procedure for calculation of relative error measurements for each quantity and percentage of relative error measurements exceeding a specified criterion for each quantity is like ASTM 1318-09 [28].

2.3 FACTORS IMPACTING WIM SYSTEM PERFORMANCE

Vehicle, site, and sensor characteristics can influence WIM accuracy considerably. These factors have an individual and a combined effect on the WIM measurements [29].

Sensor type and array (number and spacing of sensors) are essential factors affecting WIM system accuracy. A recent synthesis of highway practice on WIM data reported findings based on survey data collected from 45 state DOTs within the US and six Canadian provinces. The results showed that 70%, 30%, 28%, 18%, and 20% of agencies use the quartz piezo (QP), bending plate (BP), piezo cable (PC), load cells (LC), and other WIM systems for data collection, respectively. Most agencies have more than one type of system and a few agencies with three or more kinds of WIM systems. Nearly 80% of the agencies were facing problems related to WIM sensors, and 60% indicated that they had problems with the WIM system going out of calibration. Some agencies faced more than one problem related to either WIM equipment or data [2]. The WIM vendors make recommendations about sensor configuration (number and spacing of sensors) that can be influenced by site limitations, road conditions, vehicle dynamics, and the expected speed. Intercomp (a renowned WIM sensor vendor); reported that the average WIM relative error could be reduced from 41% to 26% by using 4 to 6 sensors as compared to 2 to 4 sensors. More sensors result in improved WIM performance (low measurement errors), but the equipment cost increases. Currently, a single threshold (2 sensors) and a double threshold (4 sensors) are used for high-speed WIM stations. The triple (6 sensors) and tetra thresholds (8 sensors) are used primarily for low-speed WIM or static scales, which are more accurate [14]. More details related to the WIM sensor and array are given in the FHWA WIM Pocket Guide (Part-1) [30]. A study by Haider et al. reported that the multiple speed points functionality of the WIM controller has a /significant influence on WIM sensor precision. More speed points could significantly improve the WIM precision. The results were based on 35 LTPP WIM sites as part of the Specific Pavement Studies Traffic Pooled Fund Study (SPS TPF). The authors also reported that no consistent trends were observed between International Roughness Index (IRI) or WIM roughness index (WRI) and consistency in WIM measurements based on the available data [22].

The European road specification reports that WIM site characteristics influence vehicle motion behavior and may cause significant discrepancies between the impact forces and corresponding static loads [19, 20, 31, 32]. A recent study by Qin et al. presented a finite element model of a WIM system that allowed the WIM sensor to be placed anywhere in the pavement. The simulation results showed that multiple sensors embedded in the middle of the asphalt layer had improved their ability to capture dynamic responses [33]. Similar findings were reported by Darestani et al. [34]. Several other studies documented that regardless of the WIM system calibration, the WIM accuracy can deteriorate over time due to several factors, including temperature, pavement roughness, and fatigue of load sensors [4, 29, 35, 36].

Additionally, the vehicle suspension and oscillation can affect the WIM accuracy, resulting in the most significant possible errors in WIM systems. A multi-sensor WIM system may significantly reduce the influence of vehicle and axle oscillation. During the accuracy analysis of WIM systems using pre-weighed vehicles, Gajda et al. [7, 8, 37] reported that the dynamic component in the signal of the vehicle axle load exerted on the road surface is the primary cause of limited WIM system accuracy. The amplitude of the dynamic component depends on the pavement condition, vehicle speed, and suspension and may even amount to 40% of the static axle load values.

Table 2-7 summarizes potential factors affecting WIM system accuracy [4, 6, 9, 16, 22, 38, 39]. The WIM technology selection, site design, installation, maintenance, and calibration minimize WIM errors, while a portion is an inherent part of site characteristics. These errors lead to diminished WIM system performance, lower quality of WIM data, and a lack of users' confidence in the data. There is a need for practical tools to quantify the relative importance of various sources of error on WIM data accuracy for a given set of site-specific conditions. In addition, WIM data collectors need guidance and practical tools to improve WIM data quality through improved WIM site selection, technology selection, installation, calibration, maintenance, data analysis, and quality control/quality assurance (QC/QA) [22, 40, 41].

Table 2-7 Summary of factors influencing WIM system performance.

Potential factors	Description
Sensor, calibration, and traffic-related	Sensor type
	Number and spacing of sensors in the array
	WIM controller and speed points
	Calibration using heavy test trucks, truck dynamics
	Truck speed, acceleration, deceleration
	Traffic congestion and lane changes
	Drivers behavior, braking
Site conditions	Pavement type
	Pavement support under the sensor and
	Surface smoothness
	pavement surface distresses
	Roadway geometry (longitudinal grade and cross slope)
Installation and maintenance	Proper installation, including oversight
	Type of installation material (grout)
	Routine maintenance
	Calibration frequency
Environmental	Temperature
	Crosswind
	Precipitation
	Calibration season

2.4 CALIBRATION OF WIM SYSTEMS

The WIM systems go out of calibration, and their accuracy deteriorates over time because of many factors. These factors may include changes in measurement conditions (e.g., temperature and speed), pavement deflection, roughness caused by distresses, and fatigue of WIM sensors. The authors of the referenced studies also reported that regardless of the WIM system calibration, the WIM accuracy could deteriorate over time because of these factors [4, 5, 29, 36, 42, 43]. In another study in Arkansas, 10 out of 25 WIM sites yielded suitable loading data. The authors reported that the other sites exhibited WIM scale (sensor) failures and inconsistent loading data because of calibration concerns [44].

WIM equipment requires periodic calibrations to yield accurate and reliable loading data. Many agencies rely on a variety of auto-calibration techniques using different software-based algorithms to reduce the calibration cost. The most common auto-calibration methods offered by the WIM vendors include using the (a) average front axle weight of Federal Highway

Administration (FHWA) Class 9 trucks, (b) average weight of specific types of vehicles (often a loaded five-axle tractor semi-trailer). The auto-calibration techniques may be beneficial but have some limitations; for example, weight laws, truck characteristics, and front axle weights can vary among states. Therefore, these techniques could be implemented only after confirming the local WIM site conditions [45, 46]. The LTPP field operations guide uses multiple runs of a pre-weighed class-9 truck to perform equipment calibration. Figure 2-2 shows the FHWA vehicle Classifications.

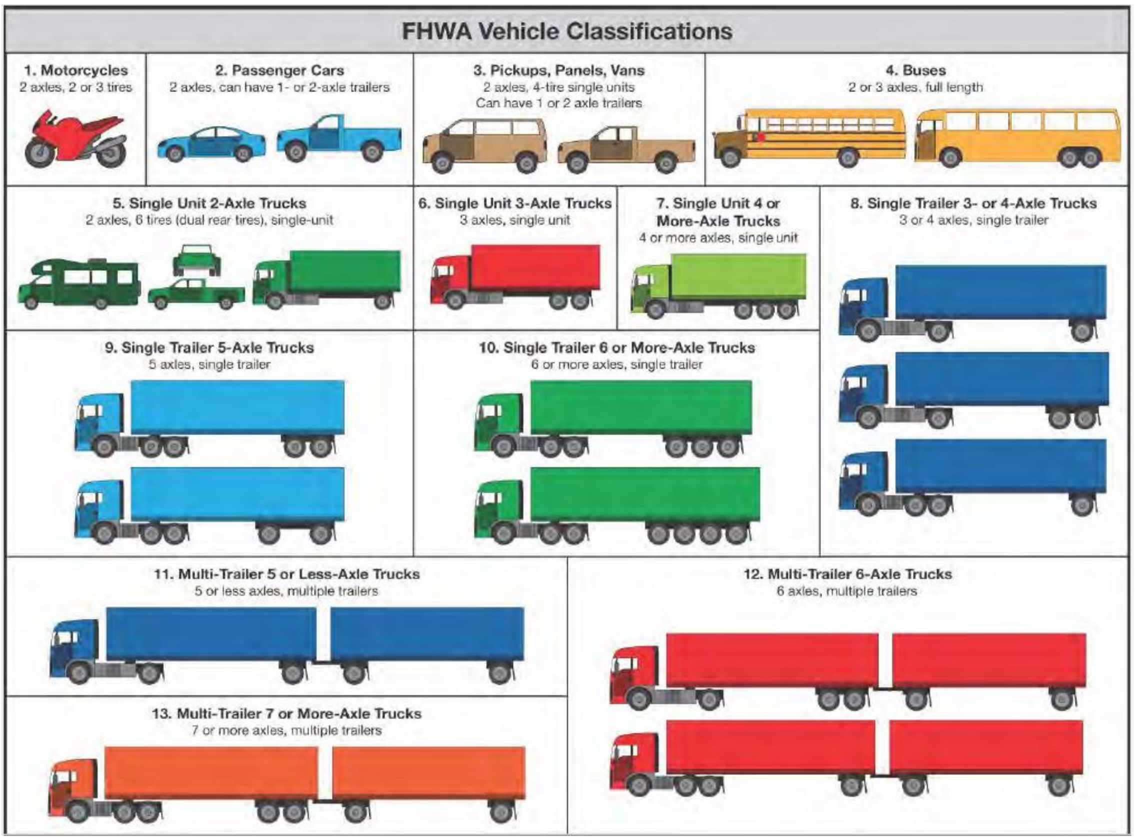


Figure 2-2 FHWA Vehicle Classifications [47-51].

2.5 CHAPTER SUMMARY

The types and sources of WIM error and the specific factors affecting WIM data quality and reliability were reviewed at length. Some WIM error sources are related to WIM site conditions, including road geometry, pavement roughness, pavement surface condition, pavement support under the sensor, and truck flow and composition. Other sources of WIM measurement errors are related to WIM system design (the sensor type, sensor array, sensor longevity, and WIM controller functions), quality of installation, calibration, and maintenance. Additional intermittent errors may result from temporal changes in pavement support under the sensor and changes in material properties of some sensors due to daily and seasonal temperature variations and environmental changes (softening of the support under pavement due to spring thaw, hardening of the support during a winter freeze, water penetration into the sensor). Some of the errors can be controlled through sensor selection and system design/configuration, QA of installation, routine maintenance, and calibration, while others are an inherent part of site characteristics that need to be understood and accounted for during the WIM site and WIM technology selection phase.

Most of the past studies were based on limited field data, which raised questions about the accuracy and adequacy of the results [4, 6-8, 12, 29, 52]. While many studies pointed to the importance of various factors, no comprehensive study was found to quantify the relative importance of multiple factors and under what conditions these factors become critical for WIM measurement accuracy. The factors mentioned above need to be evaluated further using quality WIM data to improve the desired accuracy of WIM systems.

CHAPTER 3 DATA ASSESSMENT AND EXTENT

3.1 PURPOSE

The literature review identified many factors affecting WIM data quality, including factors leading to systematic bias in WIM data and low precision (i.e., high WIM measurement error variability). These errors lead to diminished WIM system performance, lower quality of WIM data, and lack of users' confidence in the data. The purpose of the data assessment task is to investigate if sufficient data are available for the analyses to quantify the effect of different factors on WIM data accuracy and develop predictive models to infer the likely WIM data accuracy in the presence of certain site conditions and WIM operation practices.

3.2 FACTORS AFFECTING WIM BIAS AND VARIABILITY

Based on the literature review results, Table 3-1 presents the preliminary list of factors affecting WIM measurement bias and variability.

3.3 CRITERIA USED FOR DATA SELECTION

WIM data accuracy and consistency criteria and minimum calibration procedure requirements shown in Table 3-2 were used to identify WIM sites in the LTPP and other candidate databases suitable for analyses.

Table 3-1 List of potential factors affecting WIM measurement errors.

WIM Site Factors	WIM Equipment Factors
Pavement	**Sensor**
Pavement type	Sensor type
Pavement thickness	Sensor array
Pavement age	Sensor age
Pavement stiffness	**Controller function**
Surface condition	Additional Steering factor
Pavement Roughness	Number of speed points
Roadway Geometry	Temperature compensation
Grade	Auto-calibration
Curvature	**Calibration**
Slope	Number of test trucks
Traffic Flow	Test truck type
Truck dynamics	Number of test truck passes
Lane discipline	Test truck speed
Environment	Temperature during calibration
Average seasonal temperature	**Maintenance**
Average rainfall	Maintenance frequency
Climatic region	Corrective maintenance events
Wind force on the trucks	Installation quality assurance

Table 3-2 Criteria used for data selection.

Data Quality Category	Calibration		WIM Data accuracy criteria[a]			
	Total runs	Vehicle class	GVW bias	GVW total error	TA total error	SA total error
High Quality	≥10	9	≤ ± 5 %	≤ ± 10 %	≤ ± 15 %	≤ ± 20 %
Low Quality	≥10	9	≤ ± 15 %	≥ ± 10 %	≥ ± 15 %	≥ ± 20 %

Note: [a]= Must meet all 4 WIM data accuracy criteria if all 4 data attributes are available. Exception: If TA and SA errors are missing (not collected by the agency), only GVW bias and GVW total error values were used to qualify the data.

3.4 DATA SOURCES

The data needed for this study include information about WIM site performance encountered under different site conditions (pavement and road conditions and characteristics, traffic flow, and environment), WIM site designs, and WIM equipment installation and management practices. For this research, several potential data sources were considered. The primary data source was identified as LTPP program databases and supporting documentation. The LTPP databases contain WIM data from all the states within the US and the Canadian provinces. Because these data were collected as part of the Long-Term Pavement Performance program, in addition to WIM data, the databases also have extensive data about pavement, climate, traffic, and other site conditions.

3.4.1 Data Elements Identified in the LTPP Data Sources

Data elements associated with potential factors affecting WIM system efficiency were acquired from different LTPP database tables and ancillary data sources. Table 3-3 provides a summary of data elements identified in the LTPP databases, along with the name and description of the corresponding LTPP data tables containing the required data elements [21-25, 53-61].

In addition to the LTPP database mining, the paper reports and documentation collected by LTPP from the state agencies were reviewed, as well as the reports and documentation associated with the LTPP TPF 5(004) and SPS-10 WIM sites. These additional sources provided information about WIM installation, WIM calibration and/or validation, information about the WIM maintenance schedule, sensor type, array, and age, and information about pavement condition, road roughness, and road geometry. Table 3-4 summarizes the data elements identified in the LTPP documentation. The required data elements were obtained from the LTPP database standard release 32.0 using online Infopave® features. In addition to the LTPP WIM data identified for the study, the data were obtained from California, Wisconsin, Michigan, Pennsylvania, Indiana, New Jersey, and British Columbia for BP, LC, and QP sensors.

The data elements necessary for analyses included:

- WIM data, including (a) WIM measurements for calibration test trucks collected during field calibration and reference static truck weights obtained before and after each calibration, (b) daily axle load spectra computed based on WIM data,
- Information about the WIM sites, including (1) location, (2) roadway, (3) pavement characteristics, and (4) climatic data.

- Information about WIM equipment, including (1) sensor type, (2) sensor array, and (3) WIM controller functionality.
- Information about WIM calibration, including calibration dates, test truck characteristics, calibration speeds, number of test truck runs, and temperature data collected during calibration.

Table 3-3 The LTPP database tables and extracted data types.

Data type	Data fields	LTPP table name	Table alias	Table description	Class name
Pavement and Site Inventory	Pavement layer type and thickness	TST_L05B	Material Characterization and Thickness Data	Table containing layer descriptions for all constructions.	Representative Pavement Structure
	Construction no, experiment type, experiment no.	EXPERIMENT_SECTION	Experiment Section	Stores current experiment information that is driven by Maintenance and Rehabilitation activities.	Experiment Type and Improvement (M&R) History
	Site information	SHRP_INFO	LTPP Traffic Site Information	Data describing the traffic data relations and site conditions for a given SPS project or GPS Site.	Basic Information
WIM calibration and equipment	Calibration information	TRF_CALIBRATION_WIM	Weigh-in-Motion Equipment Calibration Data	Equipment calibration or calibration check information for WIM equipment used at LTPP test sites.	Equipment Calibration
	WIM sensor types	TRF_EQUIPMENT_MASTER	Traffic Equipment Master	Information on WIM and AVC equipment used at LTPP test sites collected from the calibration data sheet (Sheet 16).	Basic Equipment Information
Pavement Condition	Longitudinal profile	MON_HSS_PROFILE_SECTION	Longitudinal Profile Section Summary	High Speed Survey section level profile and computed statistics based on 150 mm interval data.	Section Level IRI
	Transverse profile	MON_T_PROF_INDEX_SECTION	Transverse Profile Index Section	Test section statistical summary of transverse pavement surface profile distortion indices.	Section Level Transverse Profile Distortion Indices (Rut)
	AC distresses	MON_DIS_AC_REV	AC Distress Survey Ratings	Distress survey ratings from manual field inspections of pavements with AC surfaces	Manual Distress
	JPCC distresses	MON_DIS_JPCC_REV	JPCC Distress Survey Ratings	Distress survey ratings from manual field inspections of pavements with jointed PCC surfaces	Manual Distress
	CRCP distresses	MON_DIS_CRCP_REV	CRCP Distress Survey Ratings	Distress survey ratings from manual field inspections of pavements with continuously reinforced PCC surfaces	Manual Distress
	JPCP faulting	MON_DIS_JPCC_FAULT_SECT	JPCC Faulting Section Data	Contains section faulting statistics from transverse joints and cracks using data from MON_DIS_JPCC_FAULT table	Section Level
	Back calculation (BC)	BAKCAL_MODULUS_SECTION_LAYER	Average BC modulus values	This table contains back calculated modulus values averaged for each FWD_PASS.	Section Level
Traffic	AADTT	TRF_TREND	AADT tend values	This table contains AADTT values for each section and each year they were in-study.	Estimated Traffic Data
Climate and Environment	Climatic regions, SN,D-Value(Effective slab thickness)	TRF_ESAL_INPUTS_SUMMARY	TRF ESAL Inputs Summary	Summary of ESAL equation inputs for a given section.	Computed ESAL Inputs
	Temperature	CLM_VWS_TEMP_MONTH	Virtual Weather Station Month Temperature	Virtual weather station monthly air temperature statistics.	Monthly
	Precipitation	CLM_VWS_PRECIP_MONTH	Virtual Weather Station Precipitation Month	Virtual weather station monthly precipitation statistics.	Monthly
WIM Data	Axle counts	MM_AX	Load Spectra	Axle counts by load bin by site, vehicle class, axle group, year, month, and DOW	Monthly

Table 3-4 The LTPP project documentation and reports and extracted data types.

Data type	Data fields	Report Name	Report/Document description
Pavement	Pavement type, age, and thickness	Phase II WIM Site Acceptability Report	Report provides information on pavement type, age or installation date, and construction.
Pavement Surface Condition	Longitudinal profile	LTPP ERD files	Profile data collected by the RSC and provided o WIM validation contractor for pavement profile roughness analysis
	Pavement distresses	LTPP Validation and Calibration Summary Report	Pavement discussion on the possible influence of pavement condition on WIM accuracy based on visual inspection.
	Average IRI	LTPP Validation and Calibration Summary Report	Report on average IRI values within WIM section and approach
	Maximum IRI	LTPP Validation and Calibration Summary Report	Report on maximum IRI value within WIM section and approach
Traffic	Truck Dynamics	LTPP Validation and Calibration Summary Report	Report provides truck dynamics in the WIM section and approach that may affect WIM accuracy
	Lane Discipline	LTPP Validation and Calibration Summary Report	Report provides information on whether trucks travel down the center of the lane.
Sensor	Sensor type	Phase II WIM Installation Report	Ancillary information provides site layout.
		Traffic Sheet 17	WIM Site inventory
	Sensor Array	Phase II WIM Installation Report	Ancillary information provides site layout.
	Sensor Age	LTPP Validation and Calibration Summary Report	Report provides the site installation date.
		Traffic Sheet 14	WIM Site installation information
WIM Controller	Steering factor, number of speed points, temperature compensation, auto-calibration	LTPP Validation and Calibration Summary Report	Report provides WIM controller information which is cross-referenced with vendor information.
WIM calibration and equipment	WIM calibration bias and standard deviation values	LTPP Validation and Calibration Summary Report	Report provides WIM calibration bias and standard deviation values
		Traffic Sheet 16	Information on WIM and AVC equipment used at LTPP test sites collected from the calibration data sheet (Sheet 16).
	Test truck data	Traffic Sheet 19	Provides information on test trucks used for validation
	Test truck speeds, number of passes	Traffic Sheet 20	Provides information on test truck runs, including speed
	Temperature during calibration	LTPP Validation and Calibration Summary Report	Temperature based analysis conducted on calibration results.
Maintenance	Maintenance frequency	LTPP Phase II Maintenance Reports	Provides information on semi-annual preventive maintenance and repairs

Figure 3-1 presents the distribution of sites with the most typical WIM sensor type for each State in the LTPP database. A majority of WIM sites have PC, followed by BP and QP sensors in North America. It should be noted that Figure 3-1 only shows the distribution of sites for states reporting WIM data to the LTPP. While it illustrates the WIM sites available in the LTPP database, it may not represent all WIM sites in the United States. Table 3-5 provides the distribution of available WIM sites and calibration records for different sensors. A record represents a single calibration event for which the bias and SD were calculated based on multiple runs of a class 9 truck. This dataset was used to study the potential factors that can impact WIM system performance.

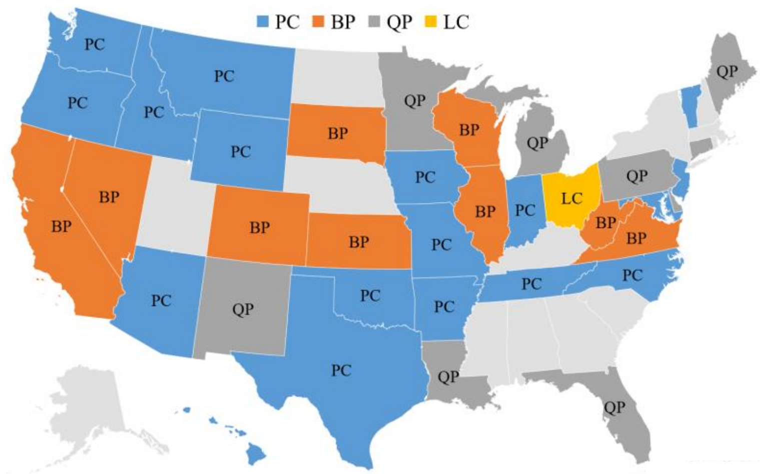


Figure 3-1 Distribution of LTPP WIM site location with different sensors in the US.

Table 3-5 Number of available WIM sites.

Data type	Sensor type				Total
	BP	LC	QP	PC	
Total sites	24	9	79	58	170
Total records	114	13	172	115	414

The daily axle load spectra (ALS) for class 9 trucks [single axle (SA) and tandem axle (TA)] were extracted from the LTPP database to assess the long-term performance of WIM systems between calibration events. This data set contained all four-sensor types, i.e., BP, LC, QP, and PC. Table 3-6 shows the distribution of the 51 sites by climate, pavement, and sensor types. The number of available WIM sites and calibration events for LC sensors was limited compared to the other three sensor types. The small sample (6 replicates for LC sensor) size can result in wider 95% confidence intervals (CI).

Table 3-6 Distribution of sites used to assess WIM consistency over time.

Pavement type	Sensor type	Climatic region				Total
		Dry freeze (DF)	Dry no freeze (DNF)	Wet freeze (WF)	Wet no freeze (WNF)	
Asphalt concrete (AC)	BP	-	-	-	1 (1)	1 (1)
	LC	-	-	-	-	-
	PC	-	-	8 (12)	2 (3)	10 (15)
	QP	-	3 (8)	6 (19)	7 (22)	16 (49)
Portland cement concrete (PCC)	BP	1 (8)	3 (8)	3 (8)	4 (12)	11 (36)
	LC	-	-	3 (6)	-	3 (6)
	PC	-	1 (1)	-	1 (2)	2 (3)
	QP	1 (2)	-	7 (16)	-	8 (18)
Total		2 (10)	7 (17)	27 (61)	15 (40)	51 (128)

Note: Numbers outside the parenthesis show available WIM sites, and numbers inside the parenthesis show number of available calibration records."-"indicates no data are available.

Table 3-7 presents the WIM data for BP and QP sensors before and after WIM equipment calibration. This dataset was used to provide guidelines for successful WIM equipment calibration. In total, 111 (53+58) and 62 (34+28) WIM records were available for pre-and post-calibration data, respectively. At least 40 test truck runs were used to obtain pre- and post-calibration data for these events.

Table 3-7 Distribution of WIM sites and records by sensor type and climate.

Data	Sensor type	Climatic regions				Total
		DF	DNF	WF	WNF	
Pre calibration	BP	-	3[a] (17 [b])	3 (18)	4 (18)	10 (53)
	QP	3 (9)	5 (16)	7 (18)	6 (15)	21 (58)
Post calibration	BP	-	3 (13)	3 (10)	4 (11)	10 (34)
	QP	2 (5)	3 (5)	3 (8)	6 (10)	14 (28)

[a] No of WIM sites, [b] No of WIM records

3.4.2 Description of Analysis Data Sets and Data Attributes

The following three categories of WIM sites were considered for analysis based on WIM data accuracy and consistency obtained from calibration records:

1. LTPP TPF 5(004) and SPS 10 research-quality WIM data (LTPP RQD): The WIM sites consistently meet the ASTM Type 1 performance requirements (i.e., GVW total error ≤ ± 10 % for ≥ 75% of the calibration events) were included in this data set. This data set consisted of 170 calibration records from 36 WIM sites that are part of the LTPP SPS TPF 5(004) and SPS-10 studies. These sites represent the highest quality WIM data set used for this study due to the stringent LTPP WIM calibration protocol and daily WIM data review implemented by the LTPP program. This subset contains WIM data for BP, QP, and LC sensors.
2. State-owned WIM sites providing high-quality WIM data (RQD Equivalent): This category included the state-owned WIM sites with the available WIM calibration data meeting or exceeding the criteria defined for LTPP RQD data accuracy standards. This data set included 164 calibration records from 94 WIM sites. This data set includes four sensor types: BP, QP, LC, and piezo-polymer cables (PC).
3. State-owned WIM sites providing data of lesser quality than LTPP RQD sites (Less than RQD): This category included the state-owned WIM sites that did not meet the LTPP RQD criteria based on the available calibration data. The subset includes 80 calibration records from 40 WIM sites. This subset contains WIM data for BP (two sites with one calibration record each) and PC sensors (predominantly PC data with 38 sites and 78 calibration records).

Tables 3-8 and 3-9 provide the distribution of available WIM sites and calibration records for different sensors and data categories, respectively. Note that based on the data collected for the analysis, all available WIM sites with LC, QP, and most BP sites had WIM data accuracy and consistency similar to the LTPP RQD sites. Only WIM sites with PC sensors had a significant number of sites with performance data lower than LTPP RQD. The low number of poor-performing WIM sites might be explained by the proactive actions by state highway agencies in correcting the problems or not sending WIM data to the LTPP if the WIM site was not meeting the required performance standards. Also, many WIM sites included in the LTPP database did not have calibration data, thus, reducing the pool of potential analysis sites.

Table 3-8 Number of available WIM sites.

Data type	Sensor type				Total
	BP	LC	QP	PC	
LTPP RQD	11	2	23	-	36
RQD Equivalent	11	7	56	20	94
Less than RQD	2	-	-	38	40
Total	24	9	79	58	170

Table 3-9 Number of available calibration records.

Data type	Sensor type				Total
	BP	LC	QP	PC	
LTPP RQD	84	5	81	-	170
RQD Equivalent	28	8	91	37	164
Less than RQD	2	-	-	78	80
Total	114	13	172	115	414

Table 3-10 provides the distribution of available WIM sites for different sensors and climates. The results show that several sensor-climate combinations in the matrix are missing or have a limited number of sites. For example, no LC sites are available in dry no-freeze (DNF) and wet no-freeze (WNF) climates. Also, only one BP site is available in the dry freeze (DF) climate. The majority of the QP sensor sites were available in the wet freeze (WF) climate. More PC sites were available in wet climates as compared to dry climates. The explanation for such distribution is that the WF climate is the most common in the highly populated regions of the US.

Table 3-11 shows the distribution of available records for multiple calibrations for sites presented in Table 3-10. There are 414 total records; however, the unbalanced design of this experiment matrix is apparent from the data availability. Because of the various missing or limited sensor-climate combinations, the experimental matrix considered is unbalanced. Therefore, it is challenging to conduct an overall ANOVA to isolate the influence of site factors on WIM data accuracy, considering multiple sensor types. The overall data extents show that majority of the data are available for QP, and PC sensors, followed by BP, while the LC sensor has only 13 WIM records.

Table 3-10 Distribution of WIM sites by sensor and climate.

Sensor type	Climatic region				Total
	DF	DNF	WF	WNF	
BP	1	8	8	7	24
LC	5	-	4	-	9
PC	2	9	26	21	58
QP	3	5	63	8	79
Total	11	22	101	36	170

Table 3-11 Distribution of calibration records by sensor and climate.

Sensor type	Climatic region				Total
	DF	DNF	WF	WNF	
BP	10	39	27	38	114
LC	5	-	8	-	13
PC	15	9	55	36	115
QP	9	12	123	28	172
Total	39	60	213	102	414

3.5 DATA SUMMARY

Table 3-12 summarizes data attributes for three data quality categories. The assessment of the individual data elements availability was required to analyze site conditions affecting WIM data accuracy and consistency (such as pavement smoothness, distresses, road geometry, pavement thickness, structural stiffness, traffic conditions, temperature, and speed during calibration, etc.). The data extent revealed that while the data elements to characterize WIM data accuracy and consistency were available, data elements for describing site conditions were missing for many candidate WIM sites. The absence of data elements characterizing site conditions resulted in the limited number of sites that could be used in the analysis of site factors affecting WIM data accuracy and consistency. The most complete data set was the LTPP RQD, with adequate data for most factors.

The site factors, including pavement type, thickness, longitudinal grade, curvature, and transverse slope, were analyzed for all the data categories wherever the data were available.

Table 3-12 Distribution of calibration records by sensor and climate.

Data attribute	Number of calibration records by data quality category			Total
	RQD	RQD equivalent	Less than RQD	
Climate	170	164	80	414
Calibration temperature	126	4	-	130
Pavement type	170	164	80	414
Pavement thickness	131	107	80	318
Longitudinal grade	170	39	18	227
Curvature	170	74	80	324
Cross slope	170	74	80	324
IRI	68	32	-	100
Deflection	-	8	-	8
Sensor array	170	164	80	414
Speed points	170	164	80	414
Calibration speed	158	84	20	262

3.6 CHALLENGES WITH THE AVAILABLE DATA

The data analysis approach depends on an adequate experimental design. The design of experiments uses power and sample size to examine the relationship between power, the number of replicates, and the maximum difference between the main effect means. Ideally, the experiment design precedes the data collection to ensure that the design has enough replicates to achieve adequate power.

For this study, the data analysis task used previously collected data from the LTPP and state highway agencies. The extent of the data was evaluated to determine its sufficiency and adequacy to support the analysis methods and objectives. There were several practical challenges in identifying enough WIM sites with documented high-quality WIM data and sufficient information about WIM site conditions. The LTPP technical support and state highway agencies were approached in an attempt to collect missing data. Unfortunately, since no experiment was designed to collect pavement data (smoothness, distress, stiffness, and thickness) and road geometry data at WIM site locations, the availability of these data elements beyond the LTPP SPS TPF and SPS 10 WIM sites was minimal. Pavement stiffness or other structural data were unavailable for any WIM site locations since no falling weight deflectometer (FWD) testing or pavement coring and testing were conducted near WIM sensor locations.

The FWD data at or closer to WIM sensors were collected by Indiana DOT specifically for this project at 8 non-LTPP WIM sites in Indiana. In summary, the following were some reasons for the unbalanced distribution of the WIM sites representing unique site conditions:

1. High-performing WIM sites were constructed under favorable site conditions and regularly maintained, thus, limiting the range of site conditions to be analyzed.
2. Sites were purposefully not installed under conditions likely to adversely affect WIM data accuracy and reliability, resulting in the absence of site entries for some site conditions.
3. Field data collection efforts at LTPP pavement experiments did not cover the exact WIM site locations, limiting the number of known factors at WIM site locations (except for pavement roughness data collected at LTPP TPF WIM sites).
4. The addition of state-owned WIM data for BP and QP sensors also resulted in an unbalanced design. Most of the BP and QP sensor's WIM data were provided by the states of California

and Michigan, located in dry and wet climates, respectively. These WIM sites were calibrated under similar site conditions for various factors, including climate, pavement type and thickness, sensor array, speed points, truck speed, and the number of truck runs.

5. The number of sites and calibration records for LC sensors was relatively small compared to the other three sensors, resulting in an unbalanced distribution of data for the different locations and sensor-related factors. Therefore, the data for LC sensors were analyzed separately.
6. The distribution of WIM data for PC sites was also not uniform for different factors, as most of these data are only available in wet climates.
7. The non-availability of continuous variables was another challenge in selecting the analysis approach because most of the variables available for the data analysis were categorical, i.e., climate, pavement, sensor, sensor array, and speed points.

3.7 CHAPTER SUMMARY

This chapter documents the criteria for data selection, sources, extent, and limitations. The purpose of the data assessment task is to investigate if sufficient data are available for the analyses to quantify the effect of different factors on WIM data accuracy and for building predictive models for inferring the likely WIM data accuracy in the presence of certain site conditions and WIM operation practices. The required data elements were obtained from the LTPP database standard release 32.0 using online Infopave® features. The data mainly comprised the LTPP research quality WIM stations. These sites are calibrated according to LTPP protocol with a complete set of supporting information about the WIM station and the pavement. These WIM sites are part of the Specific Pavement Studies Traffic Pooled Fund Study (SPS-TPF and SPS-10) and follow a more stringent LTPP WIM calibration protocol. In addition to the

LTPP WIM data identified for the study, the data were obtained from California, Wisconsin, Michigan, Pennsylvania, Indiana, New Jersey, and British Columbia for BP, LC, and QP sensors. In total, data from 170 WIM sites spread over 30 states within the United States and 3 Canadian provinces were analyzed.

The extent of the data was evaluated to determine its sufficiency and adequacy to support the analysis methods and objectives. There were several practical challenges in identifying enough WIM sites with documented high-quality WIM data and sufficient information about WIM site conditions. The LTPP technical support and state highway agencies were approached in an attempt to collect missing data. Unfortunately, since no experiment was designed to specifically collect pavement data (smoothness, distress, stiffness, and thickness) and road geometry data at WIM site locations, the availability of these data elements beyond the LTPP SPS TPF and SPS 10 WIM sites was minimal. Pavement stiffness or other structural data were unavailable for any WIM site locations since no FWD testing or pavement coring and testing was conducted near WIM sensor locations. The last section of this chapter presented the data limitations and potential challenges associated with the data analysis task.

CHAPTER 4 FACTORS IMPACTING WIM PERFORMANCE

4.1 PURPOSE

The relative influence of the factors presented in Table 2-7 on WIM measurement errors is not well understood or quantified. These factors contribute to poor WIM system performance and users' lack of confidence in the collected data. As a result, analytical techniques and models are needed to assess the relative significance of different sources of error on the accuracy of WIM data. WIM data collectors also require direction and practical tools to increase WIM data quality through improved procedures related to WIM site selection, technology selection, installation, calibration, maintenance, data processing, and quality control/quality assurance (QC/QA) [22, 40, 41].

4.2 INTRODUCTION

Vehicle, site, and sensor characteristics can influence WIM accuracy considerably. These factors have an individual as well as a combined effect on the WIM measurements [29]. Sensor type and array (number and spacing of sensors) are essential factors affecting WIM system accuracy. A recent synthesis of highway practice on WIM data reported findings based on survey data collected from 45 state DOTs within the US and six Canadian provinces. The results showed that 70%, 30%, 28%, 18%, and 20% of agencies use the quartz piezo, bending plate, piezo cable, load cells, and other WIM systems for data collection, respectively. Most agencies have more than one type of system, and a few have three or more kinds of WIM systems. Nearly 80% of the agencies were facing problems related to WIM sensors, and 60% indicated that they had issues with the WIM system going out of calibration. Some agencies faced more than one problem related to either WIM equipment or data [2]. The WIM vendors make recommendations about sensor configuration (number and spacing of sensors) that can be influenced by site limitations, road conditions, vehicle dynamics, and the expected speed. Intercomp (a renowned WIM sensor vendor); reported that the average WIM relative error could be reduced from 41% to 26% by using 4 to 6 sensors as compared to 2 to 4 sensors. More sensors improved WIM performance (low measurement errors), but the equipment cost increased. Currently, a single threshold (2 sensors) and a double threshold (4 sensors) are used for high-speed WIM stations. The triple (6 sensors) and tetra thresholds (8 sensors) are mainly used for low-speed WIM or static scales, which are more accurate [14]. More details related to the WIM sensor and array are given in the FHWA WIM Pocket Guide (Part-1) [30]. A study by Haider et al. reported that the multiple speed points functionality of the WIM controller has a /significant influence on WIM sensor precision. More speed points could significantly improve the WIM precision. The results were based on 35 LTPP WIM sites as part of the Specific Pavement Studies Traffic Pooled Fund Study (SPS TPF). The authors also reported that no consistent trends were observed between IRI or WRI and consistency in WIM measurements based on the available data [22].

The European road specification reports that WIM site characteristics influence vehicle motion behavior and may cause significant discrepancies between the impact forces and corresponding static loads [19, 20, 31, 32]. A recent study by Qin et al. presented a finite element model of a WIM system that allowed the WIM sensor to be placed anywhere in the pavement. The simulation results showed that multiple sensors embedded in the middle of the asphalt layer had improved their ability to capture dynamic responses [33]. Similar findings were reported by Darestani et al. [34]. Several other studies documented that regardless of the WIM system calibration, the WIM accuracy can deteriorate over time due to several factors, including temperature, pavement roughness, and fatigue of load sensors [4, 29, 35, 36].

Additionally, the vehicle suspension and oscillation can affect the WIM accuracy, resulting in the most significant possible errors in WIM systems. A multi-sensor WIM system may significantly reduce the influence of vehicle and axle oscillation. During the accuracy analysis of WIM systems using pre-weighed vehicles, Gajda et al. [7, 8, 37] reported that the dynamic component in the signal of the vehicle axle load exerted on the road surface is the primary cause of limited WIM system accuracy. The amplitude of the dynamic component depends on the pavement condition, vehicle speed, and suspension and may even amount to 40% of the static axle load values.

4.3 OBJECTIVES

This study addresses two core topics related to WIM technology, (a) representative WIM measurement errors for different sensor types and (b) factors affecting WIM data accuracy and consistency. The research outcomes presented in this chapter include (a) representative WIM measurement error ranges for different sensors after calibration and (b) statistical models (decision trees) to quantify the effect of site, sensor, and calibration-related factors on WIM data accuracy.

4.4 DATA ANALYSES APPROACH

Based on the data limitations discussed in Chapter 3 previously, the full factorial analysis was not a viable option for data analysis. Therefore, several statistical methods and strategies were adopted to address data limitations. Subsequently, wherever possible; the results obtained from different methods were compared and findings were reported based on the most accurate and easily interpretable prediction methods. The following data analysis methods were used to study the factors affecting WIM data accuracy:

1. The analysis of site factors' influence on WIM performance was focused on the evaluation of WIM data precision (as a measure of consistency and variability) and total error (computed as the mean error (or bias) +/- margin of error with 95% CI). The test truck data collected immediately after each WIM successful calibration event were used to ensure that the data were free of measurement bias (mean error) for quantifying the variability of WIM measurement error attributed to site and WIM equipment characteristics.
2. A full factorial design with two levels was considered for ANOVA if adequate WIM data were available.
3. Two-level partial factorial designs were used to analyze the factors with limited or unbalanced data. The ANOVA was conducted with partial factorial data to investigate the main effects of all the identified factors. Only two-way interactions were included in the model where adequate data were available.
4. One factor at a time analysis was considered to conduct a one-way ANOVA or paired t-test to compare means of different levels within a factor if WIM sites had limited and unbalanced data. However, no interaction between factors can be considered in this type of data analysis.
5. The interval plots were used to show the 95% confidence levels (CI) within the levels of various factors to show statistical differences (i.e., overlapping CI shows insignificant differences).
6. Multiple comparisons were conducted to compare multiple levels of factors where over two levels were available within a factor.
7. The non-parametric data analysis techniques were employed when the data did not meet the assumptions of parametric tests. The non-parametric tests are suitable and more robust for the data that do not meet normality assumptions.

8. Linear and multiple regression models were developed wherever adequate data were available. The artificial neural network (ANN) models were also developed, and the results were compared with multiple regression wherever possible.
9. The classification and regression trees (CART) regression techniques were also used in the data analysis. A CART regression is a predictive algorithm used in machine learning. The CART regression can be used for a continuous response variable with many categorical or continuous predictors. It illustrates important patterns and relationships between a continuous response and important predictors within highly complicated data without using parametric methods.

4.5 HOW TO QUANTIFY WIM ERROR

Class Compared to ground truth (static weights), WIM measurement errors can be divided into bias and precision. An estimate's tendency to stray in one direction from the true value is known as bias. Bias is typically characterized by mean error. The variability of repeated measurements under carefully controlled conditions can indicate measurement precision. It implies that lower random errors will result in higher precision. WIM data should result in low bias and high precision. WIM equipment requires calibration to yield highly accurate and precise data.

The relative difference between WIM and reference static weights is used to determine the accuracy of an individual WIM measurement. For the static weights to be used as a reference value, the ASTM E1318-09 section 7.1.3.4 requires the static weight limits to be within ± 4%, ± 3%, and ± 2% of the mean value for the SA, TA, and GVW measurements, respectively [62]. The true value for calibrating WIM systems is reference static weight measurements. The following equation can be used to express the relative WIM error [17, 25, 49, 63-65].

$$\varepsilon = \frac{\text{WIM weight} - \text{Static weight}}{\text{Static weight}} \times 100 \qquad (4.1)$$

Where: WIM weight = load, measured by a WIM sensor for an axle type, static weight = load measured on a static scale for the same axle type.

This relative error is commonly referred to as a WIM measurement error. Furthermore, the errors can differ depending on the type of WIM sensor technology used. Typical WIM measurement error for a calibrated WIM system follows a normal distribution [66] with a zero mean (no bias) and a standard deviation (SD), i.e.,

$$\varepsilon = \frac{X' - X}{X} \sim N\left(0, \sigma_{\varepsilon}^{2}\right) \qquad (4.2)$$

Where

$X' =$ measured load for an axle configuration on a WIM scale

$X =$ measured load for the same axle configuration on a static scale

$\sigma_{\varepsilon} =$ WIM measurement error standard deviation

The WIM equipment bias (mean measurement error) and σ_{ε} (SD of measurement errors) data were collected for all the calibration records. Another statistical attribute, the total WIM measurement error, was calculated using the equipment bias (X_{ε}) and SD (σ_{ε}) and the total number of runs for each calibration record. Equation 4.3 combines bias (mean error) and margin of error with 95% confidence, as described in the LTPP Field Operations Guide for SPS WIM sites [26]. A WIM site can be qualified as an ASTM Type I site if the gross vehicle weight (GVW) total measurement error after successful equipment calibration is ≤ ±10% [22]. There are additional requirements to qualify a site as ASTM Type I, i.e., error thresholds for wheel load, single and tandem axles, speed, axle spacing, and Wheelbase. More details can be found elsewhere [62].

$$\text{Total measurement error} = \left|\overline{x_\varepsilon}\right| + t_{\alpha/2}\ \sigma \qquad (4.3)$$

Where

$\overline{x_\varepsilon}$ = mean measurement error based on multiple test truck runs

$t_{\alpha/2}$ = student's t distribution with α=0.05

σ = WIM measurement error standard deviation

4.6 REPRESENTATIVE ERROR VALUES FOR WIM SENSORS

The three data sets described in the previous section (LTPP RQD, RQD Equivalent, and Less than RQD) were analyzed to determine typical WIM measurement accuracy and consistency achieved after equipment calibration. Ideally, a successful WIM equipment calibration should eliminate bias in all categories of weight measurements (GVW, group axle weights, and individual axle weights). However, in practice, it is nearly impossible to eliminate bias in all types of weight measurements simultaneously due to differences in dynamic forces present at the measurement time. Therefore, the data obtained after calibration still show some bias. Tables 4-1 to 4-3 present the representative values for GVW mean error (bias), the margin of error with 95% CI, and the total error for different sensors and data categories. The ASTM E1318-09 WIM protocol provides the threshold values for different classes of WIM systems. Type I WIM systems are typically used for highway traffic monitoring and pavement design using the AASHTOWare Pavement ME method. Type I WIM systems' gross vehicle weight error tolerance is ±10% [62, 67]. This value is compared with the total measurement error range (|bias|+ margin of error with 95% CI) for WIM performance evaluation.

WIM measurement errors and associated descriptive statistics for WIM performance attributes (i.e., mean error, the margin of error with 95% CI, and total error) were computed for each calibration event. Then, the averages were obtained to get representative values of all GVW WIM attributes for different sensor types and data sets.

Table 4-1 Representative values for GVW mean measurement errors (bias).

Data type	Sensor type			
	BP	LC	QP	PC
LTPP RQD	± 0.82%	± 1.60%	± 0.92%	-
RQD Equivalent	± 0.81%	± 1.00%	± 1.12%	± 1.50%
Less than RQD	-	-	-	± 4.51%
All except LTPP RQD	± 0.81%	± 1.00%	± 1.12%	± 3.01%
All combined	± 0.82%	± 1.30%	± 1.02%	± 3.01%

Table 4-2 Representative values for GVW margin of error with 95% CI.

Data type	Sensor type			
	BP	LC	QP	PC
LTPP RQD	3.65%	3.80%	4.86%	-
RQD Equivalent	3.20%	4.80%	4.22%	4.20%
Less than RQD	-	-	-	8.64%
All except LTPP RQD	3.20%	4.80%	4.22%	6.42%
All combined	3.43%	4.30%	4.54%	6.42%

Table 4-3 Representative values for GVW total errors.

Data type	Sensor type			
	BP	LC	QP	PC
LTPP RQD	± 4.47%	± 5.40%	± 5.78%	-
RQD Equivalent	± 4.01%	± 5.80%	± 5.34%	± 5.70%
Less than RQD	-	-	-	± 13.15%
All except LTPP RQD	± 4.01%	± 5.80%	± 5.34%	± 9.43%
All combined	± 4.25%	± 5.60%	± 5.56%	± 9.43%

The following observations can be made from the results in Tables 4-1 to 4-3:

- When the WIM systems were calibrated, the GVW mean errors (bias) were significantly reduced (within ± 1.60%) for all sensors available in LTPP RQD and RQD equivalent data categories. A considerably higher bias was observed for the PC sensor in less than the RQD data set (up to 4.5%, see Table 4-1).
- The average random errors due to GVW measurement variability (margin of error with 95% CI) did not expect to exceed ± 5.00% for all the sensors available in LTPP RQD and RQD equivalent data set. However, these ranges were higher for the PC sensor in the third (less than RQD) data set (up to 8.6%, see Table 4-2).

- The GVW total error for all available sensors in LTPP RQD and RQD equivalent data set was within ± 5.8%, well within ASTM Type I thresholds (± 10.0% for GVW total error) (see Table 4-3).
- Overall, BP sensors showed the least amount of error. The error ranges of LC, and QP sensors were similar for all three GVW attributes and within a 2% difference of the BP total errors.
- The PC sites part of the RQD equivalent data set showed low errors for all GVW error attributes compared to the sites part of less than the RQD category. Overall, the PC sites in less than the RQD data category showed the highest total error, where the average GVW total error values were within ± 13.15% (see Table 4-3).

The information presented in Tables 4-1 to 4-3 has an immediate practical application. It provides highway agencies with the benchmark values demonstrating the practically achievable accuracy and variability of WIM measurements for different WIM sensor types after successful calibration. Highway agencies could use this information to evaluate their WIM site performance against the benchmarks and establish realistic expectations for WIM measurement accuracy for different WIM sensors. However, the above benchmarks and findings are based on the WIM data obtained immediately after each calibration event. They may not reflect changes in the WIM sensor's performance between calibration events.

4.7 COMMON CHARACTERISTICS OF WELL AND POORLY-PERFORMING WIM SITES

4.7.1 BP Sensors

For the BP sensor, the WIM performance data and statistics computed after calibration events provided by the states of California and Wisconsin were found to be comparable or slightly more accurate than LTPP RQD data. The two BP sites with one calibration record each showed performance lower than LTPP RQD. These poorly performing BP sites were installed in Indiana with the staggered sensor array and calibrated using a single speed point with 10 test truck runs. One site showed unusually high bias (possibly due to calibration criteria used), and the other showed high SD. These sites were not considered for further analysis because there were only two records, and both records were statistically identified as outliers. Also, the number of records showing poor performance was significantly lower compared to LTPP RQD and RQD equivalent data categories. Table 4-4 indicates that except for the two outlier sites, the state-maintained BP sites had WIM performance data similar to LTPP RQD sites.

Table 4-4 Descriptive statistics for BP sensor by data categories.

Data category	Sites	Calibration events	Bias	SD	Total error
LTPP RQD	11	84	± 0.82%	1.83%	± 4.52%
RQD Equivalent	11	28	± 0.81%	1.60%	± 4.30%
Less than RQD	2	1 (site18-2009)	±0.50%	5.8%	± 13.60%
		1 (site 18-3031)	±6.80%	3.8%	± 15.40%

The effect of site, sensor, and calibration-related factors on BP sensors' WIM performance was analyzed by performing one-way ANOVA. Table 4-5 provides the descriptive statistics of the GVW performance data by different site factor levels computed for WIM sites with BP sensors in the RQD equivalent data category. The analysis showed no significant differences (i.e., no more than a 2 percent difference in the average total GVW errors) in the performance characteristics of BP WIM sites for the factors analyzed.

4.7.1.1 Common Characteristics of Well-Performing WIM Sites with BP Sensors

Some common characteristics of well-performing BP WIM sites were observed, as summarized below.

- Climate: The climate did not affect the performance of the BP sensor for the state-owned WIM sites. BP WIM sites perform well in all climates.

- Sensor array: The better quality data were obtained by using the BP sensors in a staggered configuration.
- Truck runs, and calibration speeds: The state-owned BP sites that were calibrated using 10 truck runs showed slightly lower variability in calibration data. However, the difference was small, with a 0.5% error difference.
- Speed points: All the BP sites part of these analyses were calibrated using multiple speed points and showed good performance with low variability in measurement error.
- Pavement: all high-performing BP plates were installed in 12-inch thick PCC pavements.

Table 4-5 Descriptive statistics for RQD equivalent BP data.

Factor	Levels	Number of calibration events	GVW SD	GVW total error
Climate	Dry	16	1.4%	± 4.1%
	Wet	12	1.8%	± 4.6%
Sensor array	BP in-line	24	1.6%	± 4.3%
	BP staggered	4	1.7%	± 4.1%
Truck runs	10	5	1.4%	± 3.8%
	>10	23	1.6%	± 4.4%
Speed points	Single	-	-	-
	Multiple	28	1.6%	± 4.3%
Pavement	AC	-	-	-
	PCC	28	1.6%	± 4.3%

4.7.2 QP Sensor

The WIM performance data for QP sensors were obtained for the WIM sites located in the states of Michigan, Pennsylvania, Connecticut, New Jersey, and Wisconsin. The computed WIM performance attributes for these sites were either comparable or slightly better than LTPP RQD data. The descriptive statistics of the GVW data are shown in Table 4-6. The results show insignificant differences between the GVW WIM performance data for these two data categories.

Table 4-6 Descriptive statistics QP sensor (GVW).

Data category	Sites	Calibration events	Bias	SD	Total error
LTPP RQD	23	81	± 0.9%	2.4%	± 5.9%
RQD Equivalent	56	89	± 1.1%	2.1%	± 5.7%

The effect of site, sensor, and calibration-related factors on QP WIM performance was analyzed for state-owned WIM sites in the LTPP equivalent category using one-way ANOVA. Table 4-7 provides descriptive statistics for various factors used in the analysis.

Table 4-7 Descriptive statistics for RQD equivalent QP sensor data.

Factor	Levels	Calibration events	GVW SD	GVW total error
Climate	Dry	-	-	-
	Wet	89	2.1%	± 5.7%
Sensor array	QP double staggered	6	1.8%	± 4.6%
	QP double in-line	16	2.7%	± 6.7%
	QP single staggered	67	2.0%	± 5.5%
Truck runs	10	14	2.1%	± 6.0%
	>10	75	2.1%	± 5.6%
Speed points	Single	16	2.18%	± 6.03%
	Multiple	73	2.13%	± 5.68%
Pavement	AC	40	2.3%	± 5.9%
	PCC	49	2.0%	± 5.5%

The analysis showed no significant differences (i.e., no more than a 2 percent difference in the average total GVW errors) in performance characteristics of QP WIM sites for the factors analyzed, with exception of the sensor array.

4.7.2.1 Common Characteristics of Well-Performing WIM Sites with QP Sensors

- Climate: All good-performing QP sites were located in a wet climate (no data in Dry climates were available for analysis).
- Pavement: QP WIM sites performed well both in AC and PCC pavements. WIM sites with QP sensors installed in PCC pavements showed slightly better but not statistically significant improvement in WIM performance. On average, less than 0.5% improvement was observed for QP WIM sites in PCC pavements.
- Truck runs: The state-owned QP WIM sites that were calibrated using more than 10 truck runs showed slightly lower variability in calibration data.

- Calibration speed points: Statistically, the GVW total error was not significantly different when multiple speed points were used to install calibration factors, as compared to a single speed point. The GVW average values for standard deviation and total error were slightly better for sites calibrated using multiple calibration speed points, as compared to the sites with single points.
- Sensor array: WIM sited with QP sensors in a double-staggered sensor array (i.e., a total of four half-lane sensors) showed the best performance. However, state-owned Michigan WIM sites with a single-staggered sensor array (i.e., a total of two half-lane sensors) also showed very good performance.

4.7.3 PC Sensor

The WIM performance data for PC sensors were obtained from the states of Alberta, Arizona, Indiana, Iowa, Manitoba, Maryland, Missouri, New Jersey, North Carolina, Pennsylvania, and Washington. The computed GVW WIM performance data for these sites were subdivided into two categories, (a) LTPP RQD equivalent and (b) less than LTPP RQD. The descriptive statistics of the GVW data attributes and comparisons between data types are shown in Tables 4-8 and Figure 4-1, respectively. The results show that GVW SD and total error are significantly lower for RQD equivalent data than the other category (see Figure 4-1 for details).

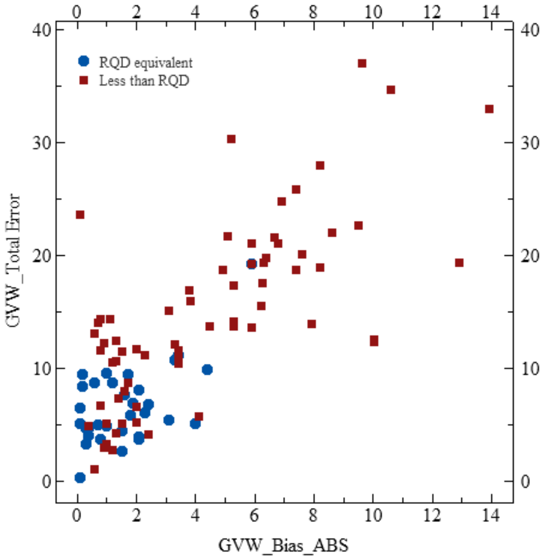
Table 4-8 Descriptive statistics PC sensor (GVW).

Data category	WIM sites	Calibration events	Bias	SD	Total error
RQD Equivalent	20	37	± 1.6%	2.2%	± 6.5%
Less than RQD	38	78	± 4.6%	4.4%	± 14.4%

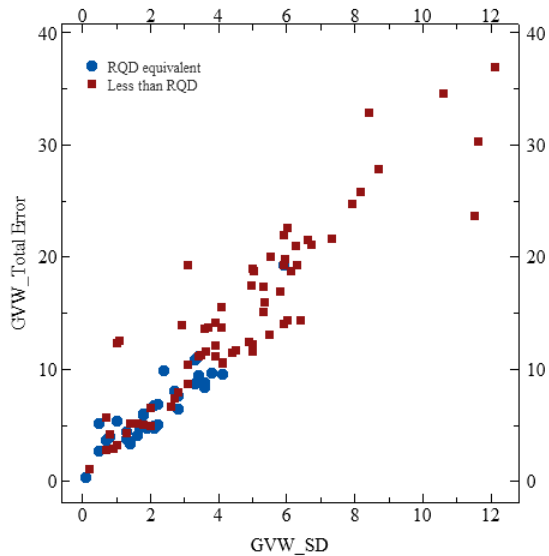
The analysis of different factors that may affect the performance of WIM systems with PC sensors was conducted separately for RQD equivalent and less than RQD data to identify different site factors associated with these two data sets. Table 4-9 provides descriptive statistics for RQD equivalent and less than RQD data sets. The distribution of the data for some of the factors is highly unbalanced and thus can lead to subjective outcomes. For example, the PC sensor part of the RQD equivalent data category had only two sites with one record, each located in a dry climate (Arizona). Similarly, the unbalanced distribution in the data matrix was observed for other factors. In general, statistically significant differences were observed between the two data categories, whereas the data were comparable for different factors within each category. An exception was the effect of climate, where PC sensor sites located in a dry climate showed significantly higher bias, SD, and total error as compared to the sites located in a wet climate. One explanation is that changes in hourly temperatures in dry climates are typically more rapid, and the effects of temperature changes on PC sensor performance during calibration are more apparent in dry climates (in WIM calibration practice, this phenomenon is termed as "chasing the error"), as compared to wet climates.

It was observed during the data analysis that the PC sensor showed significantly higher bias than other sensors even after the calibration (again, most likely attributed to the effect of changing temperature on WIM measurement accuracy). Therefore, the data analysis for GVW bias (using absolute values) was also conducted.

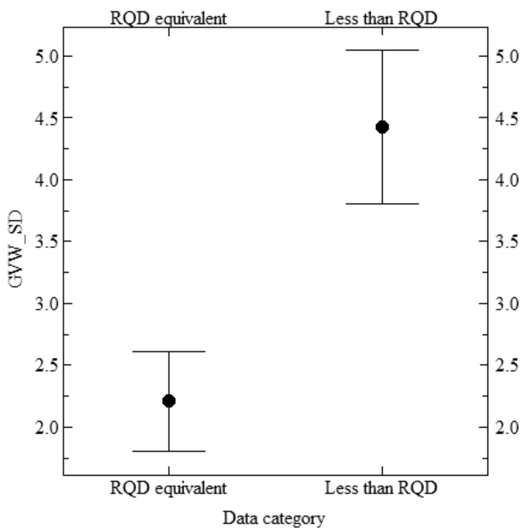
Statistically significant differences were observed in PC WIM error between LTPP RQD and RQD equivalent categories. However, the available data could not isolate the site characteristics for good and poorly performing PC sites.



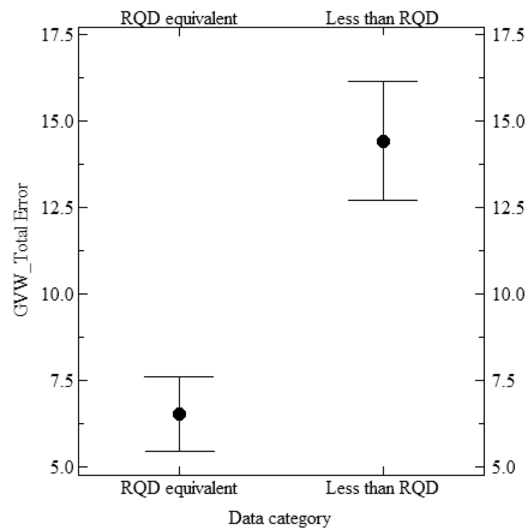
(a) GVW absolute bias and total error



(b) GVW SD and total error



(c) Effect of data type on SD (p-value <0.000)



(d) Effect of data type on total error (p-value <0.000)

Figure 4-1 Scatterplot and 95%CI interval plots for PC sensor by data categories.

Table 4-9 Descriptive statistics for RQD equivalent and Less than RQD data for PC sensor.

Factor	Data type	Levels	Calibration events	Bias (%)	GVW SD (%)	GVW total error (%)
Climate	RQD equivalent	Dry	2	± 2.35	2.0	± 6.4
		Wet	35	± 1.57	2.2	± 6.5
	Less than RQD	Dry	22	± 6.25	5.9	± 19.5
		Wet	56	± 3.91	4.37	± 13.58
Sensor array	RQD equivalent	PC half	4	± 0.56	2.5	± 6.2
		PC full	33	± 1.74	2.2	± 6.6
	Less than RQD	PC half	11	± 2.3	4.4	± 12.2
		PC full	67	± 4.9	4.8	± 15.76
Speed points	RQD equivalent	Single	28	± 1.62	2.2	± 6.6
		Multiple	9	± 1.52	2.3	± 6.4
	Less than RQD	Single	49	± 4.2	4.8	± 15.1
		Multiple	29	± 5.2	4.8	± 15.5
Truck runs	RQD equivalent	10	29	± 1.7	2.2	± 6.7
		>10	8	± 1.3	2.1	± 5.8
	Less than RQD	10	62	± 4.5	4.2	± 14.1
		>10	16	± 5.0	7.1	± 19.94
Pavement type	RQD equivalent	AC	26	± 1.75	2.1	± 6.4
		PCC	7	± 1.78	2.5	± 7.3
	Less than RQD	AC	41	± 5.05	5.3	± 16.9
		PCC	28	± 3.74	4.68	± 14.0

4.7.3.1 Common Characteristics of High and Poorly Performing PC WIM sites

- Climate: The PC sensor located in the wet climate showed significantly lower bias, SD, and total error as compared to the sites located in a dry climate in less than the RQD data category. The effect of climate on the PC sensor is similar but more significant to the one observed for the QP sensor. It is essential to highlight that most of the available PC sites in the analysis data set were located in wet climates.
- Pavement: On average, the PC sensor showed lower bias, SD, and total error in PCC pavements as compared to AC pavements in less than the RQD data category.
- Sensor array: For the available sensor array configurations, the effect of the sensor array was insignificant on GVW SD and total error in RQD equivalent data category. On average, higher values of GVW total error were observed for the PC full-lane sensor array in the Less than the RQD data category. This trend needs more careful evaluation as this is not usually observed in the field. The PC half-sensor array was only used in PC

sites located in Indiana (wet climate may have a more significant effect than the sensor array in this case). All other PC sites located in Alberta, Arizona, Iowa, Manitoba, Maryland, Missouri, New Jersey, North Carolina, Pennsylvania, and Washington were equipped with PC full-lane sensor array.

- Speed points: The PC sensor sites were calibrated using three or more speed points, and less than three-speed points showed insignificant differences in SD and total error within each data category (see Table 4-9). The GVW bias (absolute) values were higher for the events calibrated using 3 or more speed points. The probable reason for this trend is the extended duration of the calibration time needed for multiple speed points that may result in higher temperature fluctuations and ultimately result in higher differences between static and WIM weights.
- Calibration: several PC sites had bias over 2 percent event right after calibration, pointing to potential issues with maintaining consistency of WIM measurement for sites with PC sensors.

4.7.4 LC Sensor

The LC sample size (based on just 9 sites and 13 calibration records) was significantly small compared to the other three sensor types considered in the analyses. Table 4-10 provides the available number of GVW WIM calibration records and descriptive GVW error statistics for LC sensors.

Table 4-10 Descriptive statistics LC sensor (GVW).

Data category	Calibration events	Bias (%)	SD (%)	Total error (%)
LTPP RQD	5	± 1.64	1.86	± 5.40
RQD Equivalent	8	± 0.97	2.38	± 6.22

Based on the results, none of the factors available for analysis were found to have a statistically significant effect on the performance of WIM systems with LC sensors. The LC sensor data analysis results were very similar to the results obtained for the BP sensor. Similar to BP, all the LC sites were installed on PCC pavements. Table 4-11 provides the descriptive statistics of the LC sensor WIM performance data, stratified for different factors. Common Characteristics of High-Performing LC WIM Sites

4.7.4.1 Common Characteristics of High-Performing LC WIM sites

Based on the analysis results, WIM systems with LC sensors were found capable of collecting RQD data under the following conditions:

- Any climate (dry or wet)
- Any sensor array (in-line or staggered)
- PCC pavements(LC WIM technology requirement)
- Any controller compatible with LC sensors (single or multiple speed points)

4.8 MODELS TO DETERMINE EXPECTED WIM ERROR RANGES

This analysis aims to evaluate if effective statistical, machine learning, or logical modeling techniques could be used to quantify the effects of essential site, sensor, and calibration-related factors on the variability of WIM measurement error. In developing the model concepts, the practicality of the model implementation and how easy it would be for highway agency personnel to obtain the necessary model input factors were considered. The aim is to provide practical means to highway agencies to predict the expected level of error in WIM measurements based on local WIM site conditions, selected WIM equipment, and calibration efforts. The following independent variables were considered in the analyses based on the available data.

- Sensor type
- Sensor array
- Climate

- Calibration speed points
- Pavement type

Table 4-11 Descriptive statistics of LC sensor data by different factors.

Factor	Levels	Number of calibration events	GVW SD (%)	GVW total error (%)
Climate	Dry	5	2.1	± 5.5
	Wet	8	2.2	± 6.1
Sensor array	LC in-line	5	2.1	± 5.5
	LC staggered	8	2.2	± 6.1
Truck runs	10	7	2.2	± 6.0
	>10	6	2.1	± 5.8
Speed points	Single	8	2.4	± 6.2
	Multiple	5	1.9	± 5.4
Pavement	AC	-	-	-
	PCC	13	2.2	± 5.9

All the available independent variables were categorical and had limited replicates or were missing for several combinations in the data matrix. Due to limited data availability, the above list does not include several essential site factors (road geometry, pavement smoothness, and pavement strength information). Separate predictive models were developed for the PC, QP, and BP sensor types.

As mentioned earlier, the LC sensor data were available for a very limited number of sites compared to the other three sensors. There was no site or sensor-related factor that significantly affected the performance of the LC sensor. Therefore, the LC sensor data were not included in the model development. However, the models developed for BP could apply to LC, based on similarities of the effects of different site features on LC and BP sensor performances observed based on limited data and literature review findings. The following dependent variables were used in the analyses:

- GVW mean measurement errors (bias)
- GVW standard deviation of measurement errors (precision)
- GVW total measurement errors

4.9 DECISION TREE MODELS FOR WIM EQUIPMENT AND SITE SELECTION - NETWORK LEVEL

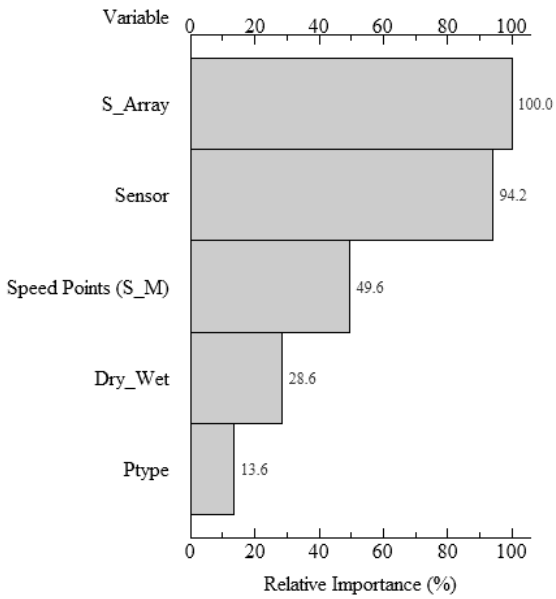
National-level models were developed based on all WIM data available for analysis. This analysis aimed to evaluate if effective statistical or logical models could be developed and used to quantify the effects of essential site, sensor, and calibration-related factors on the variability of WIM measurement error. In developing the model concepts, the practicality of the model implementation and how easy it would be for highway agency personnel to obtain the necessary model input factors were considered. The National-level models were developed for BP, QP, and PC sensors.

This analysis was conducted using the supervised machine-learning algorithm called classification and regression trees (CART®) available in Minitab. The CART Regression illustrates critical patterns and relationships between a continuous response and significant predictors within highly complex data without using parametric methods. Since CART regression uses non-parametric techniques, it is preferred for model development when data do not follow a particular distribution. The visual representation of the CART regression can make a complex predictive model much easier to interpret [68]. Moreover, this model development approach does not require variable transformations and is an effective alternative for categorical predictors. This study developed the decision tree models for all three WIM performance attributes, i.e., GVW bias, SD, and total error, to account for measurement precision and bias. Table 4-12 provides the model summary and description of terms related to model accuracy for all three attributes. Figure 4-2 presents the relative variable importance and model accuracy based on CART regression. The results show that the sensor array and types are the most important predictors, followed by controller functionality (speed points). The relative variable importance also shows that the climate (dry=DF and DNF vs. wet= WF and WNF) is important for predicting WIM measurement errors. However, the pavement type showed little significance.

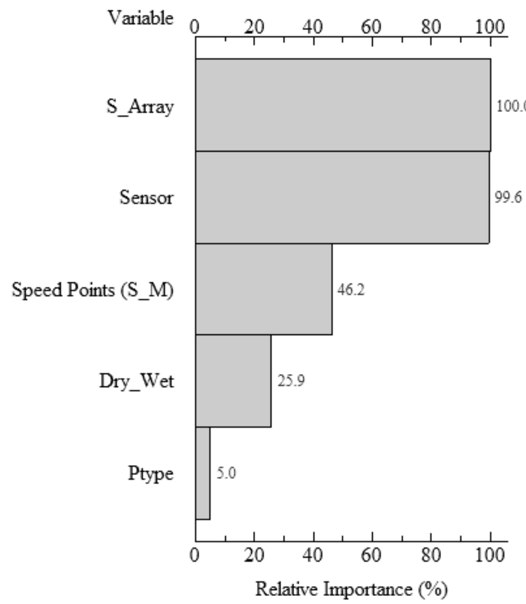
Table 4-12 CART regression (summary).

GVW data attribute	Total predictors	Important predictors	Number of terminal nodes	Minimum terminal node size	R2	
					(Trg[2])	(Tst[3])
Bias	5[1]	5[1]	5	13	0.45	0.40
SD	5[1]	5[1]	10	5	0.32	0.25
Total error	5[1]	5[1]	8	5	0.46	0.43

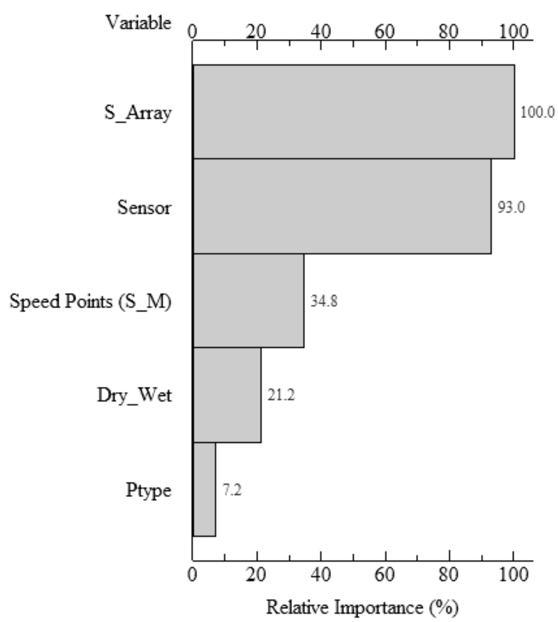
Note: **1**=sensor type, sensor array, speed points, climate, pavement type. **2**=Training, **3**=Testing



(a) GVW SD (Training R2 = 0.32)



(b) GVW total error (Training R2 = 0.46)



(c) GVW bias (absolute) (Training R2 = 0.45)

Figure 4-2 Relative variable importance CART regression.

4.9.1 Model for GVW Total Error and its Interpretation

Figure 4-3 presents the decision tree model developed for GVW TE. For simplification and decision-making for equipment selection based on: site, sensor, and calibration-related factors, the model can be interpreted with the help of Table 4-13. The first row shows that BP sensors in-line and staggered and QP sensors double staggered can yield the most accurate WIM data in any pavement and climate. Similarly, the last row shows that the PC sensor sites calibrated with a single-speed point and installed in a dry climate yielded the least accurate WIM data. It was observed that the BP sites used in this analysis were only installed on PCC pavements. The "-" symbol in Table 4-13 shows the factor is less/not important (insignificant effect on representative precision values based on available data). The "X" symbol in Table 4-13 shows the selection of a particular factor can lead to high or low WIM errors. "PC half" means 2 half-lane sensors in a staggered array providing a single threshold. "PC full" means 2 full-lane sensors, one on each side providing a double threshold.

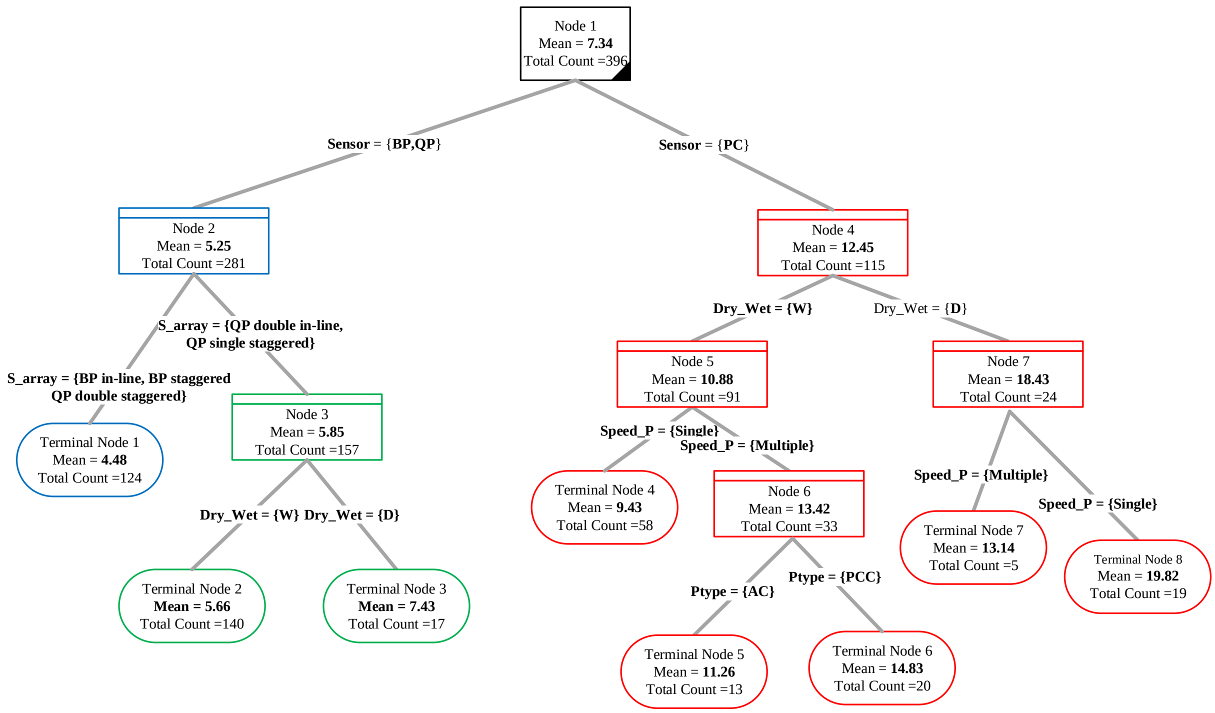


Figure 4-3 Decision tree model for GVW total error.

Table 4-13 Model interpretation for GVW total error.

Accuracy (Most to least accurate)	Factors resulting in high accuracy								GVW total error (%)
			Speed points		Climate		Pavement		
	Sensor	Sensor array	Single	Multiple	Dry	wet	AC	PCC	
1	QP, BP	BP in-line and Staggered, QP double staggered	-	-	-	-	-	-	± 4.48
2	QP	QP double in line, QP single staggered	-	-	-	X	-	-	± 5.66
3	QP	QP double in line, QP single staggered	-	-	X		-	-	± 7.43
4	PC	PC full, PC half	X	-	-	X	-	-	± 9.43
5	PC	PC full, PC half	-	X	-	X	X	-	± 11.26
6	PC	PC full, PC half	-	X	X	-	-	-	± 13.14
7	PC	PC full, PC half	-	X	-	X	-	X	± 14.83
8	PC	PC full, PC half	X	-	X	-	-	-	± 19.82

4.9.2 GVW Total Error Model Application

Utilizing the supervised machine learning decision tree models based on the CART® algorithm, the presented methodology shows good potential for estimating the WIM measurement error range using information about the WIM site and sensor-related factors. The decision tree model can be conveniently used for WIM equipment selection. Depending on the extent of information related to the site, sensor, and calibration-related factors, the decision tree model can help highway agencies choose the optimal WIM sensor type and sensor array by considering WIM errors. This information can be integrated with equipment procurement, installation, and life cycle costs to determine the most reliable and economical equipment while also considering WIM data accuracy requirements received from WIM data users.

4.9.3 GVW Total Error Model Application

The presented model is developed based on WIM error data that showed minor variations in site conditions (especially BP and QP). 98% of the QP sensor data (167 calibration records out of 171) and 99% of the BP sensor data (111 calibration records out of 112) used in the analysis were within ASTM Type I accuracy based on GVW total error, i.e., the total error was within ± 10%. Additionally, the results need careful interpretation because the data were split at different nodes based on the available replicates, inducing some bias in the results. For example, a PC sensor installed in dry climates and calibrated using multiple speed points had only five replicates.

4.9.4 Models for GVW SD and Measurement Bias

The GVW SD and measurement bias models are presented in Figures 4-4 and 4-5, respectively. Similar to GVW total error model, the first row shows that BP sensors in-line and staggered (single threshold, 2 half-lane sensors in total), and QP sensors staggered in double threshold configuration (4 half-lane sensors in total) can yield the most consistent and accurate WIM data in any climate. In contrast, PC sensor sites calibrated with a single-speed point and installed in dry climates resulted in the least accurate and most inconsistent WIM data (see Figure 4-4 and

Table 4-14). Table 4-15 shows the summary of the GVW bias model. Ideally, the measurement bias should be zero just after calibration. However, most PC sites show bias values above 2% (see Table 4-15). Therefore, practical limitations of representative measurement accuracy must be considered, and it is recommended that all three models should be consulted for the final equipment selection.

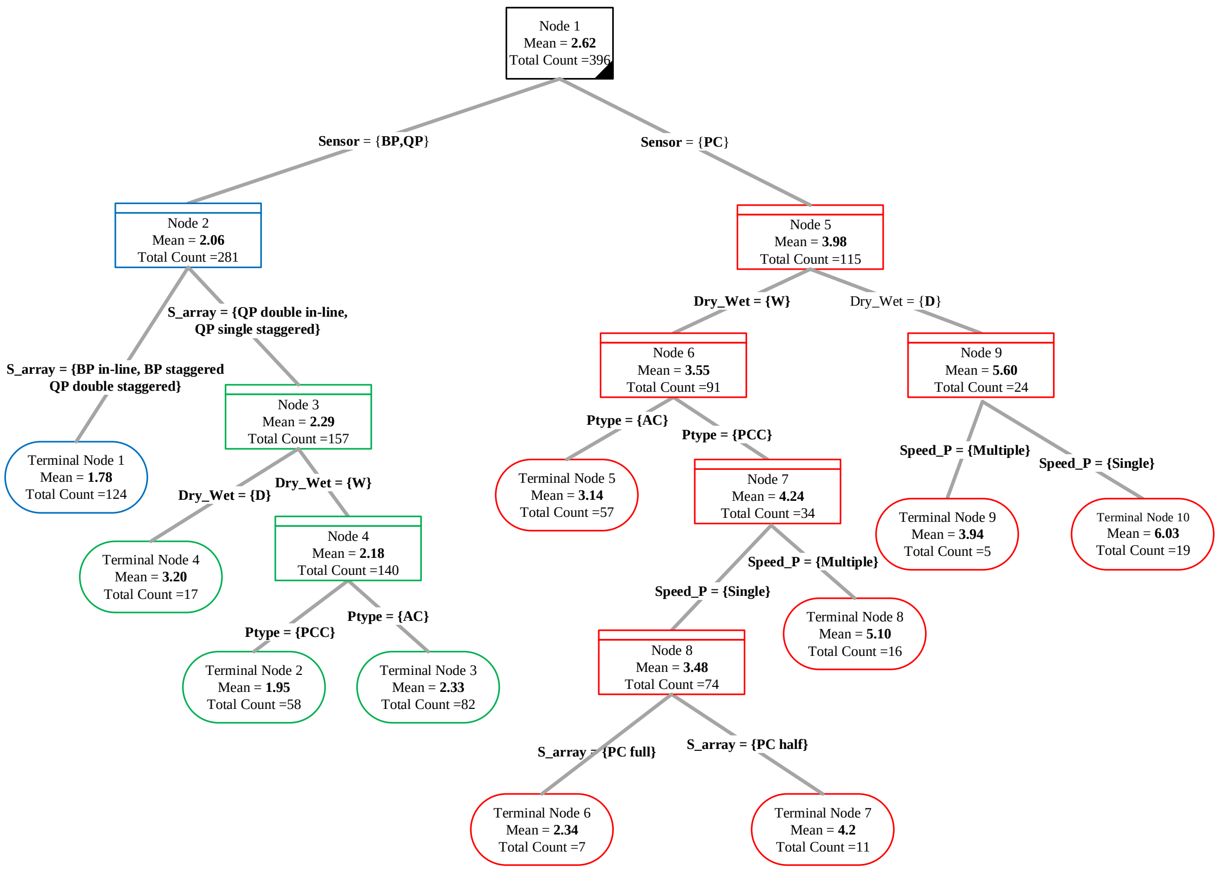


Figure 4-4 Decision tree model for GVW SD.

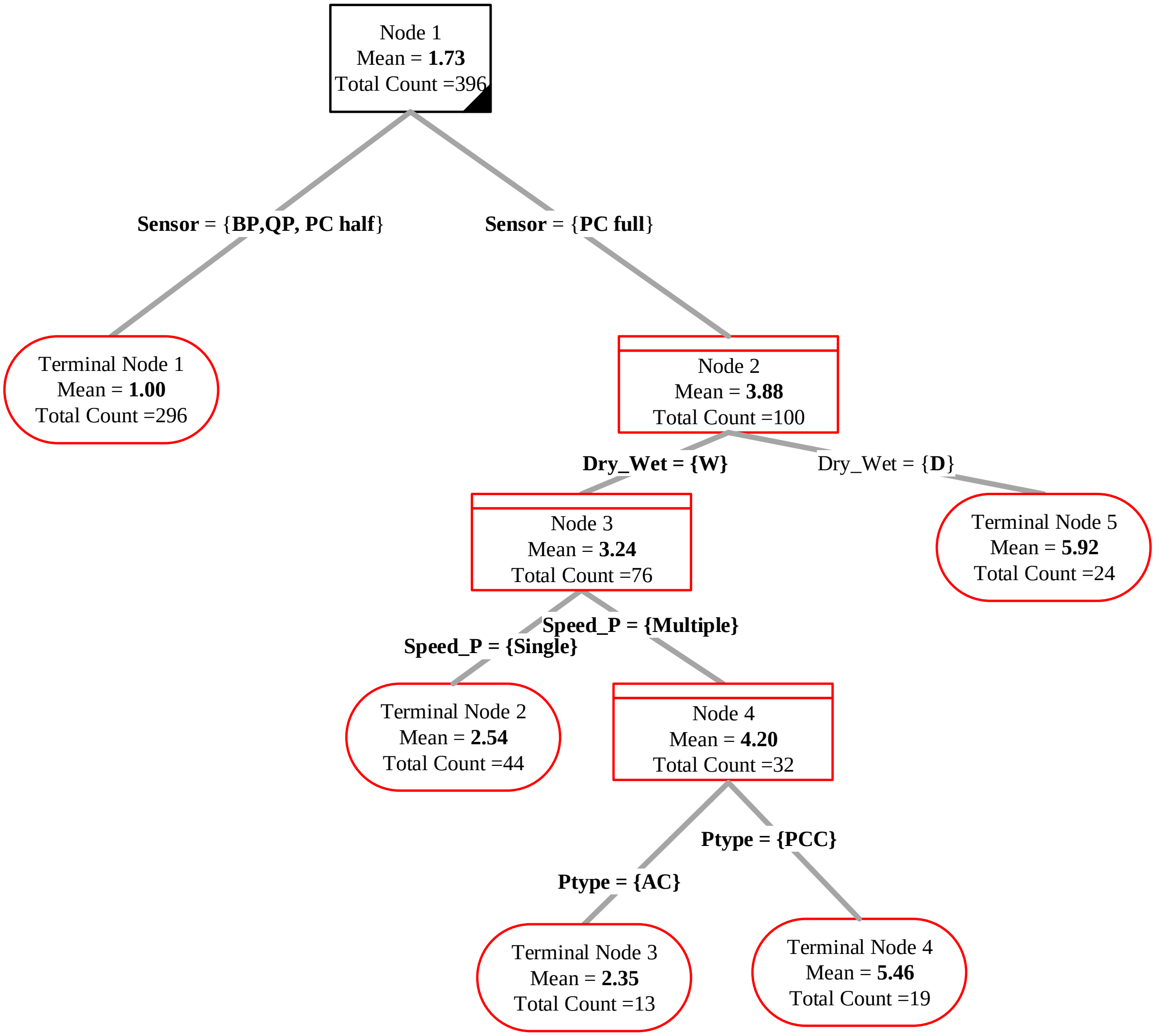


Figure 4-5 Decision tree model for GVW bias.

Table 4-14 Model interpretation for GVW SD.

Accuracy (Most to least accurate)	Factors resulting in high accuracy								GVW Precision (%)
	Sensor	Sensor array	Speed points		Climate		Pavement		
			Single	Multiple	Dry	Wet	AC	PCC	
1	QP BP	BP in-line and Staggered, QP double staggered	-	-	-	-	-	-	1.78
2	QP	QP double in line, QP single staggered	-	-	-	X	-	X	1.95
3	QP	QP double in line, QP single staggered	-	-	-	X	X	-	2.33
4	PC	PC full	X	-	-	X	-	X	2.34
5	PC	PC full, PC half	-	-	-	X	X	-	3.14
6	QP	QP double in line, QP single staggered	-	-	X	-	-	-	3.20
7	PC	PC full, PC half	-	X	X	-	-	-	3.94
8	PC	PC half	-			X		X	4.20
9	PC	PC full, PC half	-	X	-	X	-	X	5.10
10	PC	PC full, PC half	X	-	X	-	-	-	6.03

Table 4-15 Model interpretation for GVW bias.

Accuracy (Most to least accurate)	Factors resulting in high accuracy								GVW bias (%)
	Sensor	Sensor array	Speed points		Climate		Pavement		
			Single	Multiple	Dry	wet	AC	PCC	
1	QP, BP, PC	BP in-line and Staggered, QP double staggered, QP single staggered PC half	-	-	-	-	-	-	± 1.00
2	PC	PC full	-	X	-	X	X	-	± 2.35
3	PC	PC full	X	-	-	X	-	-	± 2.54
4	PC	PC full	-	X	-	X	-	X	± 5.46
5	PC	PC full	-	-	X	-	-	-	± 5.92

4.9.5 Analyses of Additional Factors Including Speed, Grade, IRI, Deflection

Highway agencies can use the information on representative WIM measurement errors presented in this study. They can compare their WIM site performance to industry standards and set reasonable expectations for WIM measurement accuracy for various sensor types and arrays installed in multiple climates.

A few additional factors were also studied where data were available, including traffic speed, longitudinal grade, IRI, and FWD deflections. The effect of climate and site-related factors on the performance of WIM systems with different sensors is summarized in Table 4-16. Table 4-17 summarizes the effects of WIM sensor type, array, calibration test truck speed, and WIM system features on WIM measurement errors.

Table 4-16 Effect of climate and pavement-related factors on the performance of WIM systems.

Factor	Sensor type	Statistical significance (Yes/No)	Comments
Climate	BP and LC	No	BP and LC errors are not affected by climate.
	QP, PC	Yes	Both sensors showed better precision in wet climates.
Pavement types	BP, LC	-	All BP and LC sensors were installed in PCC pavements.
	QP	No	Lower errors were observed in PCC pavements.
	PC	No	Lower errors were observed in AC pavements.
Longitudinal grade	BP, QP	Yes	Generally, flatter pavement (low grades, i.e., 1% or less) showed better precision.
IRI (pavement smoothness)	BP, PC, QP	No	There were no clear trends between IRI and consistency in WIM measurements based on the available data. (IN, NJ, and CA WIM sites).
FWD (pavement strength)	QP	No	Based on the available data for 8 WIM sites in Indiana, no consistent relationships were found between recorded deflection and consistency in WIM measurements.

Table 4-17 Effect of traffic speed and WIM system features on WIM errors

Factor	Sensor type	Statistical significance (Yes/No)	Comments
Sensor type	BP, LC, QP, PC	Yes	PC sensor accuracy and consistency were significantly different compared to other sensors.
Sensor array	BP, LC, QP, PC	Yes	Significant differences amongst sensor arrays were observed during the analysis. Sensor array design is a critical factor in achieving the desired WIM data accuracy.
Calibration Speed points	BP, LC, QP, PC	Yes	WIM controllers with multiple speed points could significantly improve WIM precision and reduce measurement bias. However, some inconsistencies were observed for the PC sensor.
Calibration speed	BP, LC, QP	No	A speed range between 5 to 10 mph at the time of calibration showed less variability in calibration data. The use of a narrow speed range may lead to incorrect computation of WIM measurement error for the sites with a wide range of operating speeds.

4.10 EFFECT OF GRADE, DEFLECTION, AND IRI AT STATE LEVEL

The possibility of developing a state-level model for estimating the WIM measurement accuracy was also evaluated based on a more extended set of the site and pavement-related factors available on a state level. The pavement-related factors include falling weight deflectometer (FWD) deflections, IRI, and longitudinal grade near WIM sensors. The state-level analyses were conducted using data obtained from the states of IN, NJ, and CA. Table 4-18 presents the descriptive statistics of the data obtained from all three states. The IRI data were available for 25 WIM sites. The detailed deflection data were available for 8 WIM sites located in IN. All the available WIM sites in CA sites followed ASTM specifications for the longitudinal grade, i.e., <2%. The average IRI values closer to WIM sensors were ≤ 84 inches/mile for all the sites considered for this analysis. The Indiana DOT provided Eclipse Resource Database (ERD) files with raw profiles and FWD deflections that were processed and synthesized to match WIM site locations. The average, maximum, and 95th percentiles were computed for the IRI (500 ft. segment) and deflection (300 ft. segment) data for IN WIM sites. The deflection data were not available for CA and NJ WIM sites. Information on roadway grades was also not available for NJ WIM sites. The states of CA and NJ provided the IRI values closer to the WIM site, and the ERD files with raw profiles were not available.

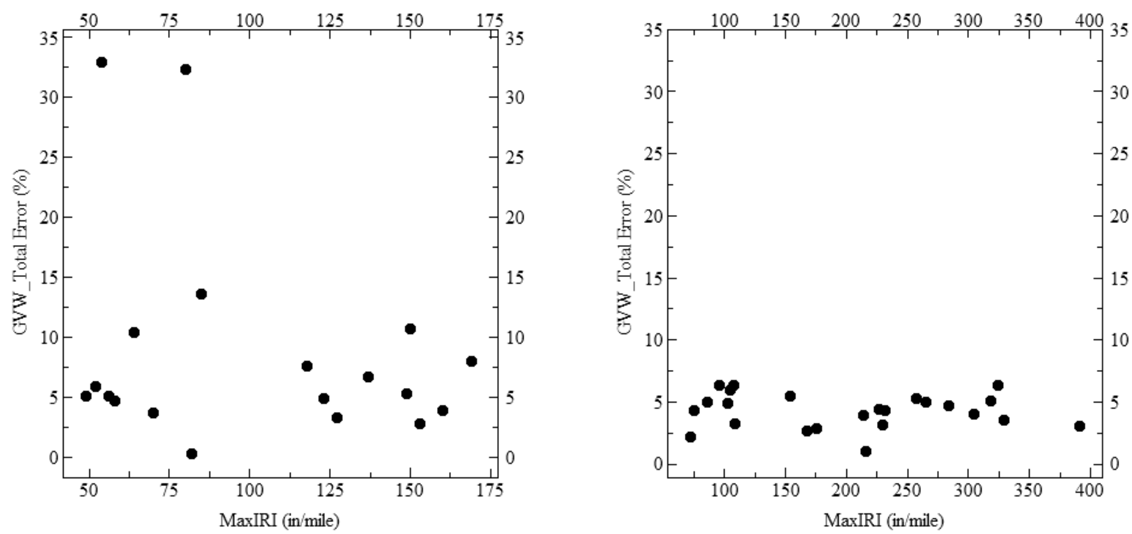
Table 4-18 Descriptive statistics of the IRI, deflection, and longitudinal grades (IN, CA, NJ).

State	Sites (records)	Average IRI (in/mile)	Maximum IRI (in/mile)	Average deflection (mils)	Maximum deflection (mils)	Average longitudinal grade
IN	8 (8)	61	149	3.3	5.3	0.64 %
CA	7 (24)	78.1	390.9	-	-	<2%
NJ	10 (19)	84.42	169.0	-	-	-

The scatter plots, boxplots, and correlations were used to assess the relationship between site factors and WIM measurement accuracy [see Figures 4-6 to 4-8]. Figures 4-6(a) and (b) show the IRI and GVW total error relationship for the WIM sites in NJ and CA, respectively. These WIM sites did not show a clear relationship (increase or decrease in WIM errors with increased or decreased IRI values) between the IRI and WIM measurement errors. All the NJ WIM sites considered for this analysis were equipped with PC sensors (2 full-lane, in-line sensors providing a double array threshold). Out of 19 calibration events for the NJ WIM sites, only 5 events (one each for five different sites) showed GVW errors greater than 10% [see Figure 4-6(a)]. All the CA WIM sites used BP sensors (2 half lane, in-line sensors providing a single array threshold). All the CA WIM sites showed highly accurate WIM data irrespective of the fluctuations in IRI values. The mean and maximum GVW total errors for the CA WIM sites were 4.3% and 6.3%, respectively [see Figure 4-6(b)].

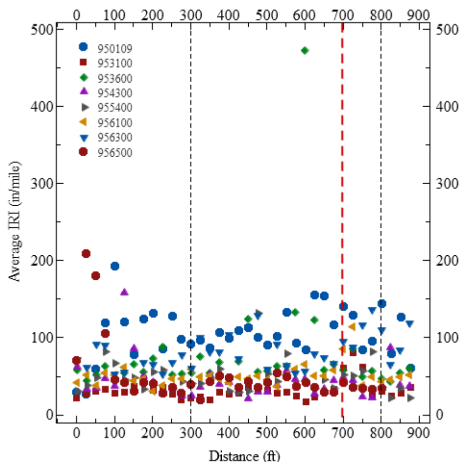
Figures 4-7(a) to (e) show the deflection, IRI, and longitudinal grade relationships with GVW total errors for the IN WIM sites. Figures 4-7(a) and (b) present the detailed IRI and deflection data for 8 WIM sites in Indiana. All the IN WIM sites were equipped with QP sensors (2 half-lane, staggered sensors providing a single array threshold). The mean and maximum

GVW total errors for the IN WIM sites were 5.7% and 9.6%, respectively [see Figure 4-8(b)]. No consistent trends were observed between the GVW errors and the deflection, IRI, and longitudinal grades for the IN WIM sites [see Figure 4-7(c) to (e)]. The IN WIM site 95-6100 with a significantly higher GVW total error (122.2%) was not included while calculating the GVW total error summary statistics. Figures 4-8(a) and (b) summarize IRI and GVW error values for the IN, CA, and NJ WIM sites. The key findings for this analysis based on Figures 4-6 to 4-8 are presented next.

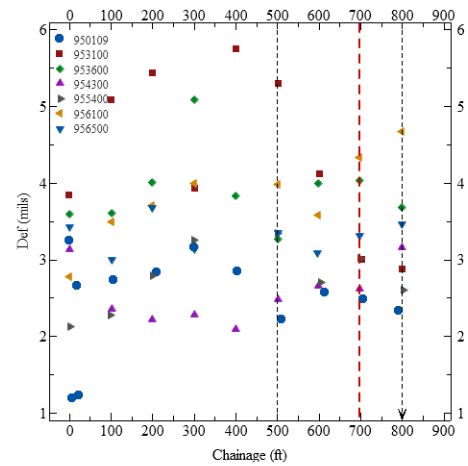


(a) Maximum IRI and GVW total error (NJ) (b) Maximum IRI and GVW total error (CA)

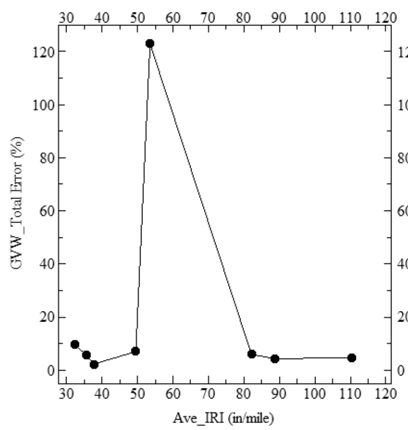
Figure 4-6 GVW errors and maximum IRI relationship - CA and NJ.



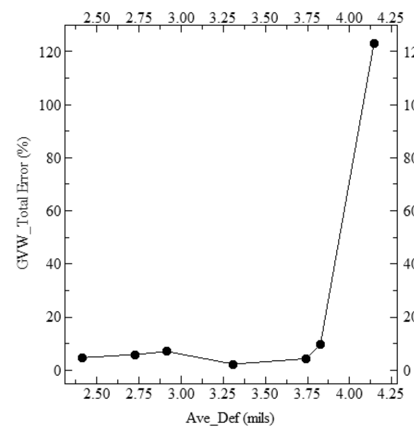
(a) Raw IRI data for IN



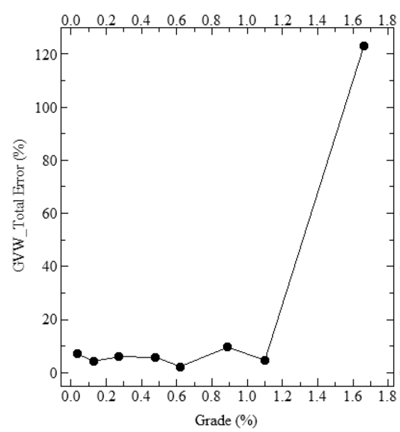
(b) Raw deflection data for IN



(c) Average IRI and GVW total error (IN)

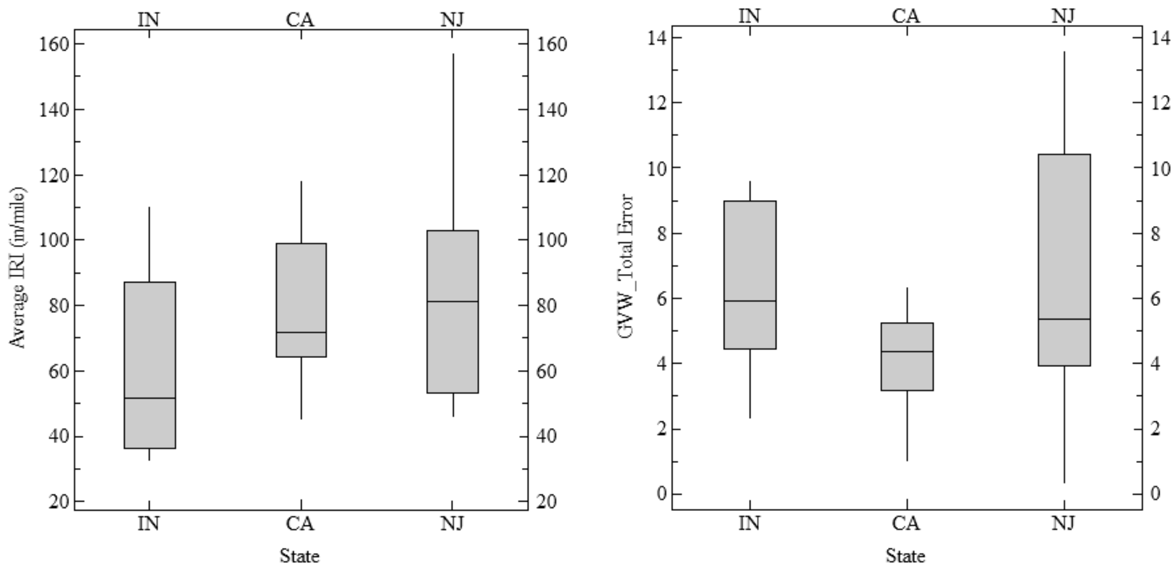


(d) Average deflection and GVW total error (IN)



(e) Longitudinal grade and GVW total error (IN)

Figure 4-7 GVW errors and IRI, and deflection relationship – IN WIM sites.



(a) Average IRI (b) GVW total error

Figure 4-8 GVW errors and IRI data extents –IN, NJ, and CA.

4.11 KEY FINDINGS FROM STATE-LEVEL DATA ANALYSES

- There were no clear trends between WIM measurement accuracy and IRI for all three states. Similar findings related to IRI data analysis were reported for the LTPP RQD dataset.
- The IN WIM site 956100 (AC pavement on I-64) showed significantly higher total error than all other sites used in the analysis. The same site showed the highest values for the longitudinal grade and the FWD pavement deflection average value. The higher values of deflection and grade could be the probable reasons for the WIM site 956100, resulting in unusually high errors. However, further investigation revealed that the leading cause for unusually high error was a bad sensor that was replaced.
- The mean deflection, IRI values, and grade observed for the WIM sites were within limits defined in the COST-323 (European) WIM standard for Class I (Excellent) WIM sites.

- Dynamic deflection (mean): ≤ 4 and 8 mils for rigid and flexible pavements, respectively.
- IRI: 0 to 82 inch/mile, 82 to 165, 165 to 250 for Class-I (Excellent), Class-II (Good), and Class-III (Acceptable) WIM sites, respectively.
- Grade: < 1%, < 2%, Class-I (Excellent), and all others sites, respectively.

4.12 CHAPTER SUMMARY

The details about representative WIM measurement errors by sensor type are presented in this chapter. These findings have an immediate practical application by providing highway agencies with the benchmark values demonstrating the practically achievable accuracy and variability of WIM measurements for different WIM sensor types after successful calibration.

The primary goal of this analysis was to evaluate if effective statistical or logical models could be developed to quantify the effects of essential site, sensor, and calibration-related factors on the variability of WIM measurement error. The purpose of such a model would be to help WIM data users and WIM data providers in estimating the expected WIM measurement accuracy for a given set of site conditions and WIM system design attributes. The presented methodology utilizing the decision tree models shows good potential for estimating the WIM measurement error range using information about the WIM site and WIM sensor-related factors. These decision tree models can support WIM equipment selection. Ideally, the WIM measurement bias should be zero just after calibration. However, the available WIM calibration data showed that some small bias was present even after calibration for most WIM sites. Therefore, practical limitations of the achievable measurement accuracy must be considered, and it is recommended that the model for predicting the GVW total measurement error, which accounts for both measurement bias and precision, should be used for practical implementation and support WIM equipment selection. Specifically, the decision tree model presented in this study can help highway agencies make an optimal selection of sensor type, sensor configuration, and controller functionality while considering achievable WIM errors and site conditions (climate and pavement type). This information, along with information about equipment longevity, length of data collection, and costs, can be used for equipment procurement, life cycle cost analysis and to assist WIM program managers in identifying the most reliable and economical equipment while also considering WIM data accuracy requirements specified by WIM data users.

CHAPTER 5 CONSISTENCY OF WIM DATA AND CALIBRATION NEEDS

5.1 PURPOSE

The analysis results reported in Chapter 4 focused on WIM measurement errors observed in the data collected immediately after equipment calibration. The limitation of this approach is that the data represented a snapshot in time and may not represent the long-term WIM site performance. Only a few WIM sites (SPS TPF sites) had WIM performance validation data available between calibrations and before the next calibration event. The available WIM performance data are not sufficient to analyze changes in WIM data over time following a calibration event. Consequently, an alternative approach is needed to characterize temporal changes in WIM data consistency. This chapter documents the procedures to analyze changes in WIM data over time during the year following the calibration event.

5.2 INTRODUCTION

This study investigated other ways of inferring WIM data accuracy and consistency over time. One approach is to relate errors in WIM data to the attributes of the normalized axle load spectra (NALS) for Class 9 vehicles. This approach can be employed to monitor and quantify temporal changes in WIM data consistency. There are several advantages to using axle weight data for Class 9 trucks. Class 9 is a recommended WIM calibration and validation class per ASTM E1318-09. Class 9 typically is the only vehicle class with supporting data for the computation of WIM precision and bias statistics because ASTM E1318 specifies this truck as a recommended calibration/validation test truck. Class 9 has a stable and well-understood gross vehicle weight (GVW) and axle weight distribution that helps identify and analyze WIM data changes over time. Class 9 is the most frequently observed heavy commercial vehicle type for most roads. The exceptions are load-restricted or secondary roads with a large percentage of small, lightweight service trucks. Typically, these are recreational, urban, or suburban roads with stop-and-go traffic not conducive to WIM measurements and thus do not represent recommended WIM site locations. The main objective of this analysis was to develop a methodology for assessing WIM measurement errors based on axle loading data analysis without physically performing WIM equipment validation in the field. The presented methodology can help highway agencies to monitor changes in WIM data and to select optimum timings for routine maintenance and calibration of WIM equipment without compromising data accuracy.

5.3 OBJECTIVES

This chapter addresses one core issue related to traffic loadings, i.e., getting accurate and consistent WIM data. Therefore, the primary objectives of the study are to provide (a) consistency of WIM data and recommendations for WIM equipment calibration frequency (b) WIM accuracy relationship with NALS shape factors, and (c) statistical analysis to develop a predictive model for WIM accuracy. These objectives were accomplished by synthesizing and analyzing the WIM and loading data in the LTPP database.

5.4 APPROACH FOR USING AXLE LOAD SPECTRA TO ASSESS CHANGES IN WIM SYSTEM PERFORMANCE

The following approach was followed to investigate how the WIM data collected from the uncontrolled traffic stream can be used to diagnose changes in WIM performance over time and to support decisions on whether to perform field WIM validation or calibration:

1. Use WIM data samples collected from the traffic stream (one month) to develop Normalized Axle Load Spectra (NALS) and assess NALS characteristics for FHWA Class 9 trucks.
2. Define statistical variables (shape factors) to monitor changes in NALS.
3. Analyze NALS before and after calibration of a WIM site and between two calibration events.

4. Assess WIM data consistency over time using the NALS shape factors developed for different periods after calibration (i.e., 1, 4, 8, and 12 months) for selected WIM sites.
5. Correlate changes in WIM accuracy and consistency over time using NALS statistics (shape factors) and WIM measurement error data collected for test trucks. Develop predictive models using changes in NALS statistics.
6. Develop a procedure for using NALS statistics (shape factors) to estimate the likely changes in WIM measurement errors and predict potential calibration drift.

5.5 DATASET DESCRIPTION

The data used for the analysis were obtained from LTPP research quality data (RQD) WIM sites installed with QP, LC, and BP sites. All available normalized axle load spectra (NALS) for Class 9 truck (single and tandem axles) data for TPS and SPS-10 sites were used in this analysis. As mentioned in Chapter 4, the LTPP research quality data only contains three sensor types (BP, LC, and QP); therefore, a few sites with piezo cables (PC) from the ASTM Type I dataset were added to this analysis. These sites will assist in quantifying the consistency of WIM measurements for sites with PC sensors and the other three sensor types. These sites represent the highest quality WIM data sets because of the more stringent LTPP WIM calibration protocol and daily WIM data review. The TPS and SPS-10 sites had detailed WIM measurement accuracy data collected before and after each calibration event that allowed the development of computational models to assess calibration drift. The additional data from the Michigan Department of Transportation (MDOT) for the QP sites were used for the model validations. Tables 5-1 and 5-2 present the summary of available WIM sites and records of axle load spectra data analyses. It should be noted that a record represents a single calibration event for which the bias and SD were calculated based on multiple class 9 truck runs (i.e., 25 to 40). It can be noted that the majority of the WIM accuracy data are available for the sites located in a wet climate.

Table 5-1 Distribution of sites for WIM data consistency analyses over time.

Pavement type	Sensor type	Climatic region				Total
		DF	DNF	WF	WNF	
AC	BP	-	-	-	1 (1)	1 (1)
	LC	-	-	-	-	-
	PC	-	-	8 (12)	2 (3)	10 (15)
	QP	-	3 (8)	6 (19)	7 (22)	16 (49)
PCC	BP	1 (8)	3 (8)	3 (8)	4 (12)	11 (36)
	LC	-	-	3 (6)	-	3 (6)
	PC	-	1 (1)	-	1 (2)	2 (3)
	QP	1 (2)	-	7 (16)	-	8 (18)
Total		2 (10)	7 (17)	27 (61)	15 (40)	51 (128)

Note: DF=dry freeze, DNF=dry no freeze, WF=wet freeze, WNF= wet no freeze, Numbers outside the parenthesis show available WIM sites, and numbers inside the parenthesis show number of available records."-"indicates no data are available.

Table 5-2 Distribution of WIM sites and records by the sensor, climate, and pavement type.

		Model development			Model validation		
Sensor	Pavement	Climate		Total	Climate		Total
		Dry	Wet		Dry	Wet	
QP	AC	2[a] (5 [b])	9 (25)	11 (30)	-	6 (9)	6 (9)
	PCC	1 (3)	1 (4)	2 (7)	-	10 (14)	10 (14)
	Total	3 (8)	10 (329)	13 (37)	-	16 (23)	16 (23)
BP	AC	-	-	-	-	-	-
	PCC	4 (11)	7 (22)	11 (33)	-	-	-
	Total	4 (11)	7 (22)	11 (33)	-	-	-

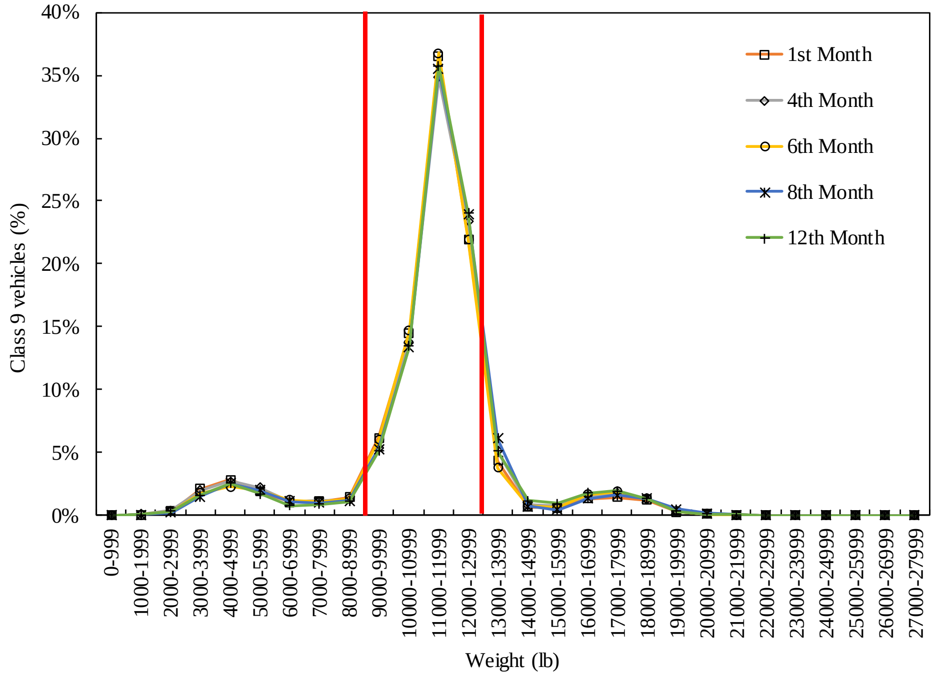
[a] No. of WIM sites, [b] No. of WIM records (one record each for pre and post-calibration)

5.6 MODELLING OF AXLE LOAD SPECTRA DATA

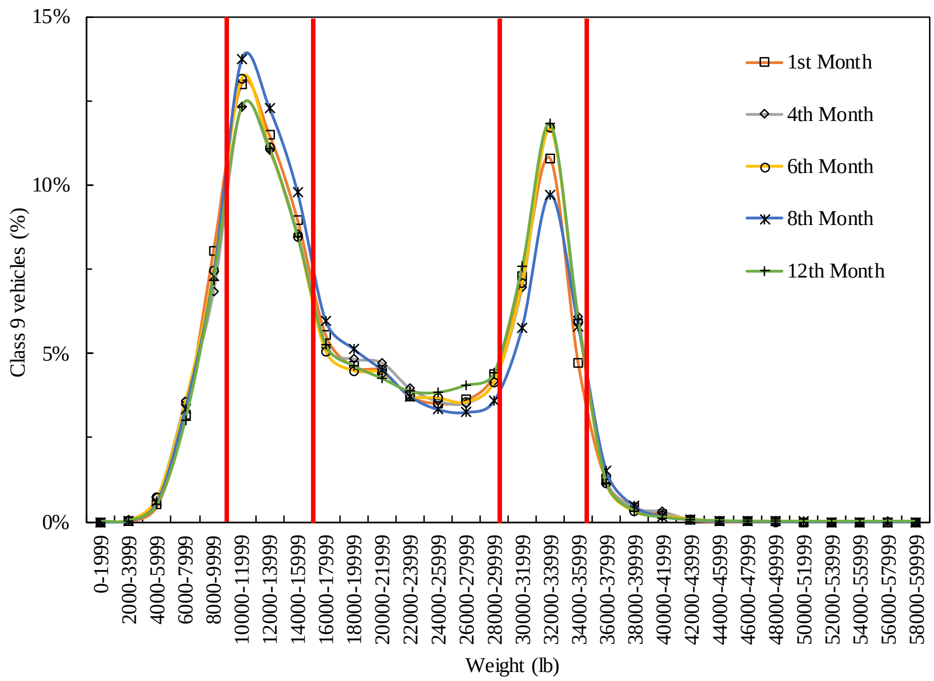
Class 9 single-axle (SA) NALS can be modeled as a single normal or log-normal distribution with a mean value corresponding to the NALS' peak load frequency value ("bell"-shaped distribution). The changes in the location of the peak of this distribution can be related to the changes in mean error. The spread of this distribution can be related to the changes in WIM measurement consistency or the precision of WIM measurements. Similarly, tandem axle (TA) NALS could be modeled by using a mixture of two normal distributions (i.e., the bi-modal distribution). The mean value of the first normal distribution corresponds to the unloaded (first) peak of tandem NALS, and the mean value of the second normal distribution corresponds to the loaded (second) peak of tandem NALS ("camelback"-shaped distribution).

Analysis of LTPP WIM data indicates more precise WIM data results for Class 9 NALS with well-defined high peaks (high mean) and skinny tails of the distribution (low standard deviation of normal distribution). Similarly, NALS based on the WIM data with low precision has low and poorly defined peaks of the distribution and fat tails of the distribution (corresponding to low mean and high standard deviation of normal distribution). NALS based on the data with a significant error due to bias has peaks of distribution shifted to the left or the right from the typical values. Consequently, such shifts in NALS may affect the pavement design thicknesses using mechanistic-empirical analysis and design procedures.

In this study, a sample of single and tandem axle NALS based on the data collected during 4 weeks immediately before or after calibration (i.e., based on the data that have well-documented measurement errors) is used to develop the approximating normal distributions and their descriptive statistics (height, mean, and standard deviation). These attributes and their combination are used to define axle load spectra shape factors. Figures 5-1(a) and 5-1(b) show two typical NALS for an SPS-2 WIM site in Colorado for single and tandem axles, respectively. This site was equipped with BP sensors. The bold vertical lines in the figures illustrate the typical ranges for peak loads.



(a) Single axle NALS



(b) Tandem axle NALS

Figure 5-1 Example of single and tandem axle NALS for site 8-0200 (BP).

The mean and variance of a single axle NALS were determined by using a discrete distribution. Equations 5.1 and 5.2 can be used to obtain the mean and variance of NALS with a single peak.

$$\mu_X = \sum_x x \times P(X = x) \tag{5.1}$$

$$\sigma_X^2 = \sum_x (x - \mu_X)^2 P(X = x) \tag{5.2}$$

For the tandem axle, typically, two peak loads are observed in a NALS. Figures 5-2 and 5-3 show examples of pre and post-calibration NALS data for four WIM sites with positive, negative, or negligible bias. In another study, a mixture of statistical distributions to characterize the predominantly bimodal axle load spectra were considered [69]. It was shown that two or more normal probability density functions (PDFs) could be added with appropriate weight factors to obtain the PDF of the combined distribution, as shown by Equation 5.3:

$$f^* = \sum_i^n p_i f_i \tag{5.3}$$

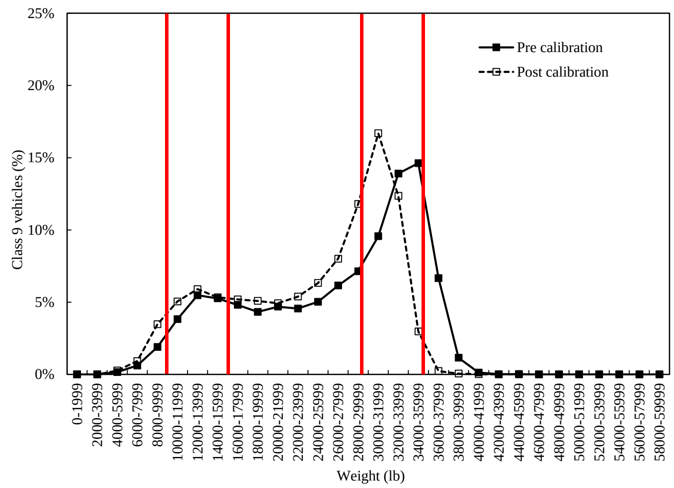
Where f^* = PDF of combined distribution, p_i= proportions (weight factors) for each normal PDF, and f_i= PDFs for each normal distribution.

For a bimodal mixed normal distribution containing two normal PDFs, the two-weight factors are complementary (i.e., $p_2 = 1 - p_1$), as shown in Figure 5-4. Haider and Harichandran determined that the bimodal shape of axle spectra could be effectively captured by using a combination of two normal distributions [17, 69-72]:

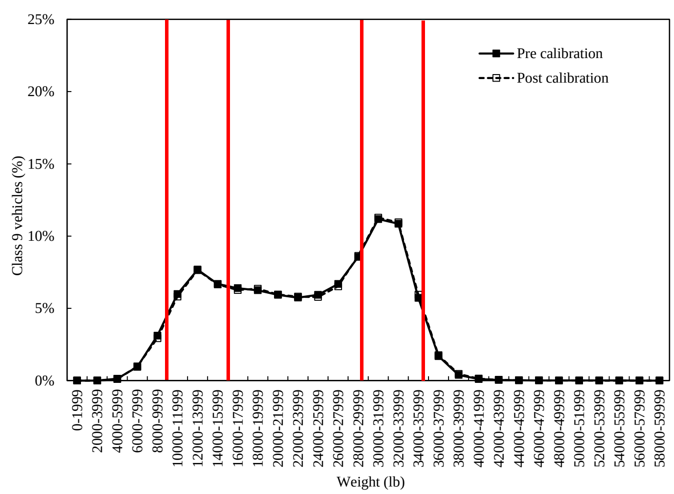
$$f^*(x;\mu_1,\sigma_1,\mu_2,\sigma_2,p_1) = \left(p_1 \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2 / 2\sigma_1^2} + p_2 \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x-\mu_2)^2 / 2\sigma_2^2} \right) \tag{5.4}$$

Where $\mu_1 =$ the average of empty or partially loaded axle loads, $\sigma_1 =$ the standard deviation of empty or partially loaded axle loads, $\mu_2 =$ the average of fully loaded axle loads, and $\sigma_2 =$ the standard deviation of fully loaded axle loads.

Figure 5-5 shows an example of the observed and fitted distribution for one of the tandem NALS in Minnesota. The vertical dotted lines show the typical range for the loaded and unloaded peaks for class-9 trucks.

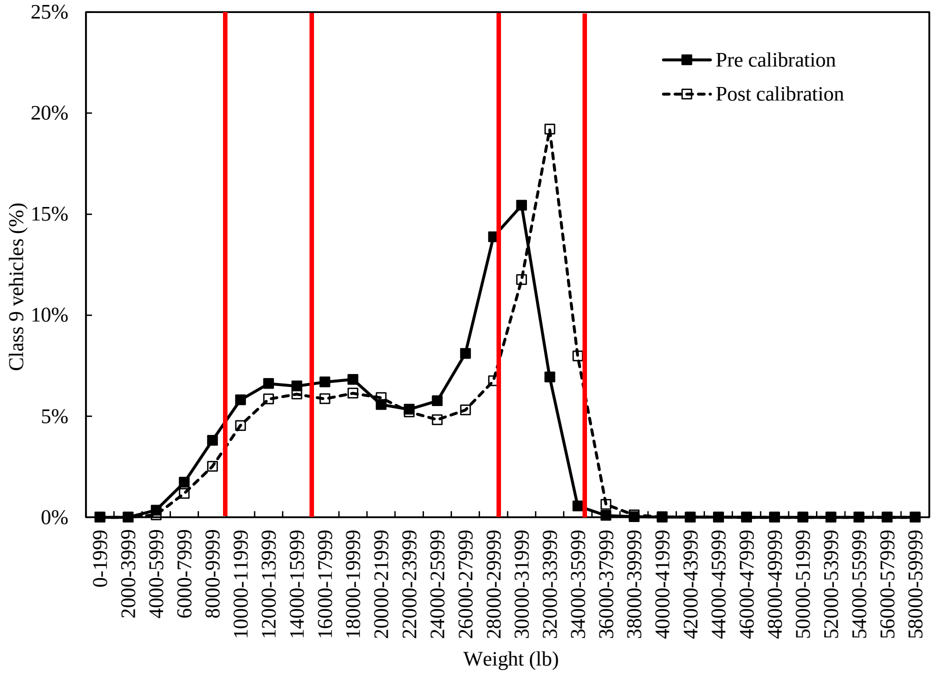


(a) QP sensor with 12.7% positive bias 53-0200 (2007)

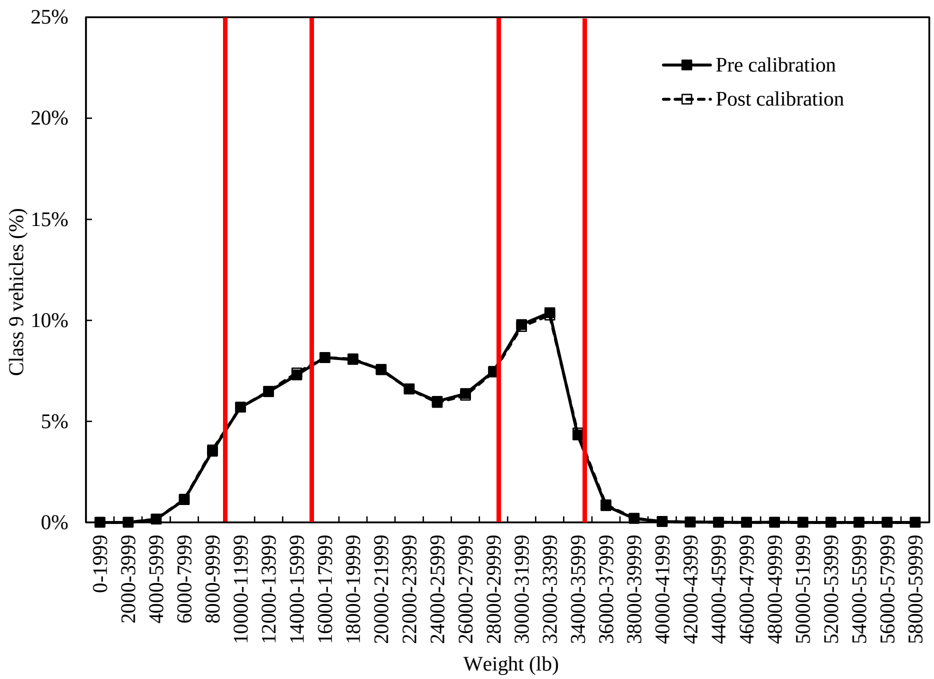


(b) QP sensor with 0.90% negative bias 42-0600 (2008)

Figure 5-2 Tandem axle load spectra example for QP sites.



(a) BP sensor with 6.3% negative bias 17-0600 (2014)



(b) BP sensor with 1.2% negative bias 20-0200 (2006)

Figure 5-3 Tandem axle load spectra example for BP WIM sites.

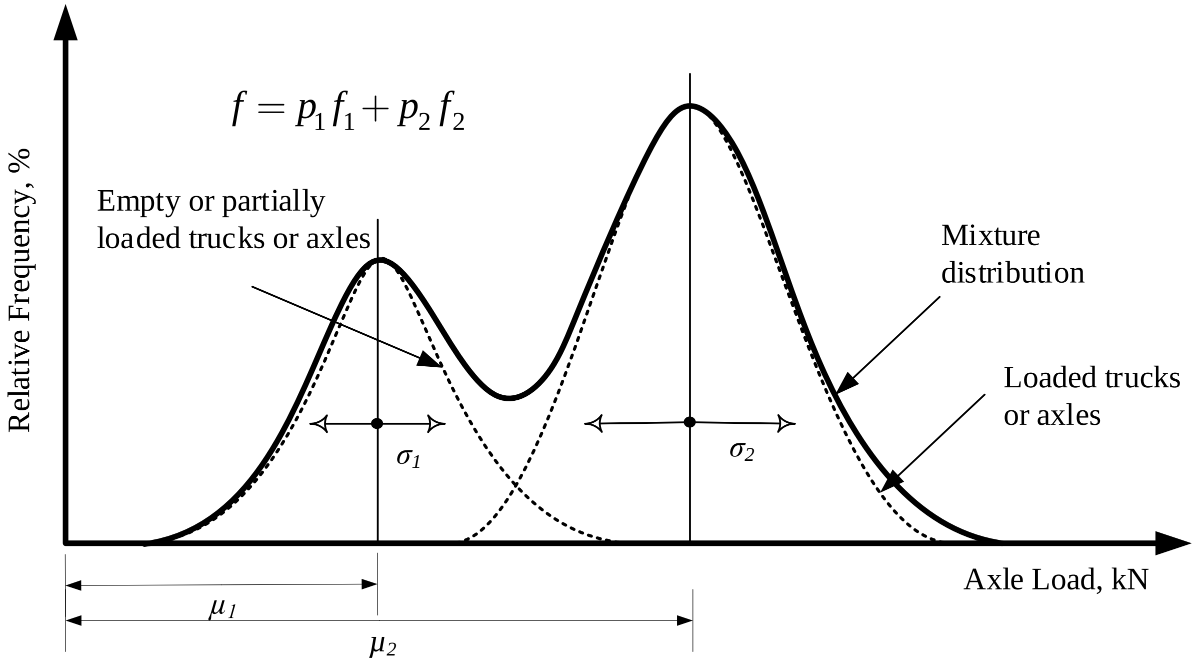


Figure 5-4 Tandem axle load spectra modeling using bimodal mixed normal distributions.

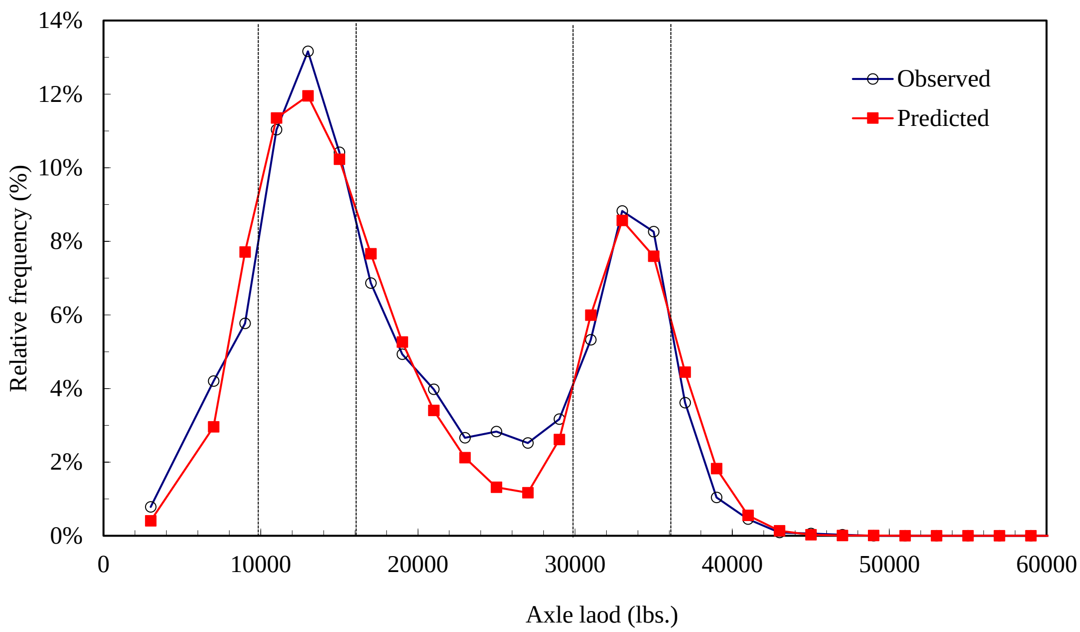


Figure 5-5 Example of a bimodal distribution fitting for TA NALS (27-0500 — Nov 2016).

5.7 CONSISTENCY OF WIM MEASUREMENT ERROR USING AXLE LOAD SPECTRA

For LTPP SPS TPF and SPS-10 WIM sites, measurement error data were available both before and after calibration. Pre-calibration WIM measurement errors were collected 1 to 3 days before every calibration event. Post-calibration WIM measurement errors were typically determined on the same day after a successful calibration event. The WIM measurement errors computed before and after each calibration event were analyzed to evaluate the effect of sensor calibration on the reduction in WIM measurement bias and variability. The pre and post-calibration data were only available for LTPP RQD sites. The consistency of WIM data was also evaluated using NALS shape factors. The NALS shape factors were obtained for 30 days, loading data collected instantly after calibration as a reference. The NALS were developed for 51 WIM sites using axle loading data. The daily data were used to compute SA and TA NALS for 1 month immediately after calibration and NALS (based on 1 month of data) at 4, 8, and 12 months after a calibration event. The analyses and comparisons of NALS over time were conducted separately for single and tandem axles of Class 9 trucks to assess the consistency of WIM data

5.7.1 Methodology for NALS Consistency Data Analyses

The NALS for single and tandem axles of Class 9 trucks were developed for the available WIM sites to analyze the consistency of WIM data over time. The axle load data for the following periods were considered:

- The NALS based on 30 days of WIM data collected after a successful calibration event.
- The NALS based on one entire calendar month of WIM data collected after a successful calibration event—4, 6, 8, and 12 months after a calibration event.

The NALS based on 14 and 30 days of data collected immediately after calibration are typically used by WIM practitioners for developing a comparison data set to evaluate consistency in WIM data over time. The 14 days of data are used for high truck volume sites, and 30 days (1 month) of data are used for low truck volume sites. Monthly NALS developed at different periods after the calibration event are useful for investigating changes in WIM data characteristics between calibration events. The process of obtaining NALS shape factors for single and tandem axles is presented in the previous section 5.6.

5.7.2 Single and Tandem Axle Shape Factors

The following statistical attributes were used to analyze differences in single axle SA and TA NALS over time:

5.7.2.1 Single Axle NALS Shape Factors

- The absolute differences in peak load (PL) values were computed to examine potential calibration drift or measurement bias overtime for the first 30 days after calibration, and the data collected at 4, 6, 8, and 12 months (i) After calibration were computed: $\Delta \mathrm{PL} = |\mathrm{PL}_{\mathrm{i}} - \mathrm{PL}_{30}|$
- The absolute differences in standard deviations (SD) of SA load values were calculated to analyze potential changes in measurement precision overtime for the first 30 days after calibration (as a reference), and data collected at 4, 6, 8, and 12 months (i) After calibration were computed: $\Delta \mathrm{SD} = |\mathrm{SD}_{\mathrm{i}} - \mathrm{SD}_{30}|$.

5.7.2.2 Tandem Axle NALS Shape Factors

- The absolute differences in peak load (PL_2) values of the loaded tandem axles (second peak in the TA load distribution) were computed to examine potential calibration drift or measurement bias over time using several time points: the first 30 days after calibration, and the data collected at 4, 6, 8, and 12 months (i) After calibration: $\Delta PL_2 = |PL_i - PL_{30}|$.

- The absolute differences in standard deviations (SD) of the loaded tandem axles distribution values were calculated to analyze potential changes in measurement precision overtime for the first 30 days after calibration (as a reference), and data collected at 4, 6, 8, and 12 months (i) After calibration were computed: $\Delta SD_2 = |SD_i - SD_{30}|$.
- The absolute differences in TA NALS mean of the loaded axles (axle weighing >26,000 lbs.) were calculated to analyze potential changes in measurement bias overtime for the first 30 days after calibration (as a reference), and data collected at 4, 6, 8, and 12 months after calibration were computed. $\Delta TAmean > 26{,}000 = |TAmean > 26{,}000_i - TAmean > 26{,}000_{30}|$.

5.7.3 Significant Differences Criteria for NALS Consistency

Changes in NALS distribution shape factors corresponding to the increase in the measurement error of 5 percent or more over time were considered as significant in this analysis:

1. If ΔPL >=5% for single NALS (or >=500 lb.), then there is a practical difference (measurement bias > 5%) between the peak loads for the reference month and i^{th} month.
2. If ΔPL_2>=5% for tandem NALS second peak or μ_2 (or >=1,500 lb.), then there is a practical difference (measurement bias > 5%) between the peak loads for the reference month and i^{th} month.

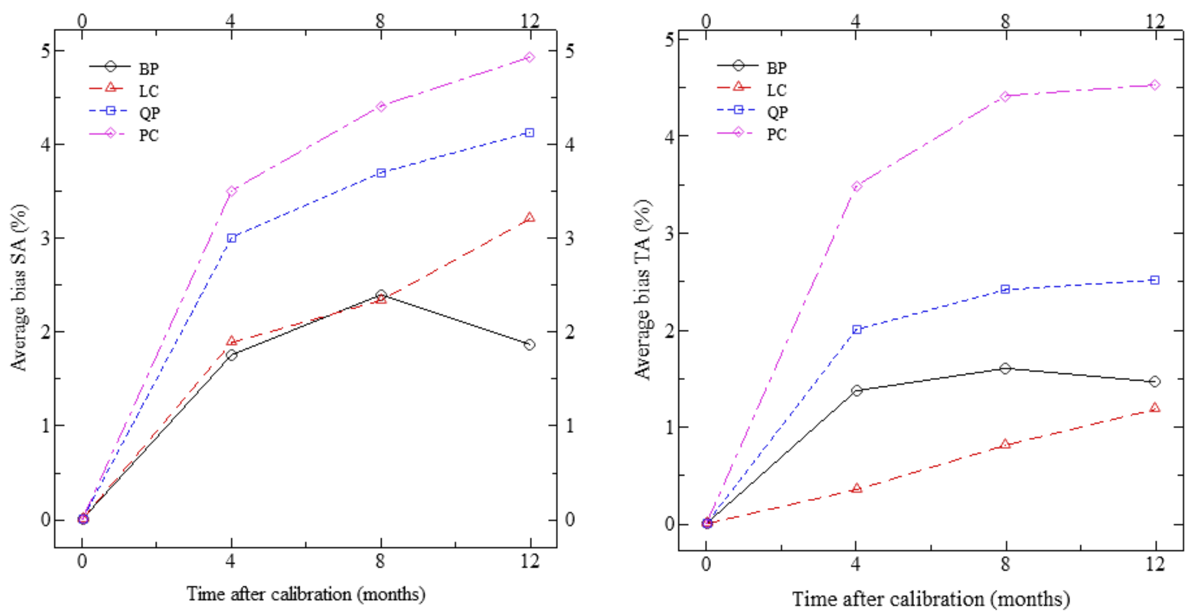
5.7.4 Key Findings Based on NALS Consistency Data Analyses

Table 5-3 and Figure 5-6 present the typical values for the percentage change in SA and TA NALS (calibration drift or bias) at 4, 8, and 12 months after calibration. The shape factors used for this analysis are SA mean load and TA mean of loaded bins (bins>26,000 lbs.). These shape factors could be used as surrogate measures of calibration drift. The available number of LC sites and records was limited (6 records for 3 sites) compared to the other three sensors considered for this analysis. Therefore, the results for the LC sensor were included only for completeness and may not represent the true performance of the sensor. The following are the key findings based on this analysis:

- The BP sensor showed the best performance with the lowest changes in SA and TA NALS one year after calibration. The changes for SA and TA NALS shape factors were less than 2 percent, indicating BP sensors can collect accurate data even one year after calibration. The results imply that calibration frequency longer than 1 year may be acceptable for the sites with BP sensors.
- The QP sensor relatively showed higher changes in NALS one year after calibration. The percentage changes in SA and TA NALS shape factors were 4.12 and 2.15, respectively. Therefore, at least an annual calibration frequency is recommended for sites with QP sensors.
- The PC sensor showed the highest changes in SA and TA NALS as compared to all other sensors. The sensor performance started deteriorating as early as four months after calibration (see Figure 5-6). The changes in PC NALS were even significant one year after calibration, with 4.92 and 4.52 percent changes for SA and TA, respectively. Due to significantly higher NALS inconsistencies, the sites with PC sensors may need multiple calibrations during the year.

Table 5-3 Percentage change in SA and TA NALS after calibration.

Sensor type	Number of sites	Calibration records	Time after calibration (months)	Average SA bias using NALS (%)	Average TA bias using NALS (%)
BP	12	36	4	± 1.75	± 1.37
			8	± 2.39	± 1.60
			12	± 1.86	± 1.46
LC	3	6	4	± 1.88	± 0.35
			8	± 2.33	± 0.81
			12	± 3.20	± 1.18
QP	23	60	4	± 3.00	± 2.00
			8	± 3.69	± 2.41
			12	± 4.12	± 2.51
PC	12	18	4	± 3.50	± 3.48
			8	± 4.40	± 4.41
			12	± 4.92	± 4.52



(a) SA bias over time

(b) TA bias over time

Figure 5-6 Percentage change in SA and TA NALS after calibration.

5.8 ASSESSING CALIBRATION DRIFT FROM AXLE LOAD SPECTRA

The current section presents an approach to estimate WIM system accuracy based on axle load spectra attributes (NALS shape factors). This approach can be employed to monitor and quantify temporal changes in WIM data consistency. The WIM measurement error computed before and after calibration was related to NALS shape factors for Class 9 vehicles. The main objective of this analysis was to develop a methodology for assessing WIM measurement errors based on axle loading data analysis without physically performing WIM equipment validation in the field. The presented method can help highway agencies to monitor changes in WIM data and to select optimum timings for routine maintenance and calibration of WIM equipment without compromising data accuracy. This section presents the procedure used to relate differences in WIM measurement errors, calculated based on pre and post-calibration data, with the differences in NALS shape factors. Table 5-4 presents the single and tandem axle NALS shape factors considered for analyses.

Table 5-4 SA and TA NALS shape factors.

Data	Single axle shape factors	Tandem axle shape factors
Based on 30 days of weight data collected before and after the calibration event	SA NALS mean (SAmean)	Unloaded peak (TAPL1)
	SA NALS SD (SASD)	Unloaded peak SD (TASD1)
	SA NALS mean of the distribution <16,000 lbs. (SAmean<16,000)	Loaded peak (TAPL2)
	Ratios (Pre/Post) of SA mean and standard deviation	Loaded peak SD (TASD2)
	-	The overall mean of the distribution (TAOAM)
	-	Overall SD of the distribution (TAOASD)
	-	TA NALS mean of the loaded axles (axle weighing >26,000 lbs.) (TAmean>26,000)
	-	Ratios (Pre/Post) of mean and SD for the first and second peaks for TA NALS.

The data selection, analyses, and model development process are explained with the help of a flow chart (see Figure 5-7).

- Step 1: is mainly data selection and syntheses
- Step 2: estimation of SA and TA NALS shape factors
- Step 3: differences of pre and post-calibration data (dependent and independent variables)
- Step 4: statistical modeling

The following variables (NALS shape factor differences) were obtained by taking differences of SA and TA NALS shape factors for pre and post-calibration loading data. These data attributes were used as independent variables for model development to assess changes in WIM errors over time.

5.8.1 Single Axle (SA) Shape Factors

Equations 5.5 to 5.7 were used to obtain the shape factors for SA. The ratios (pre/post) of SA mean and standard deviations were also obtained.

$$SAdiffMean = SAmean_{Pre} - SAmean_{Post} \tag{5.5}$$

where:

SAdiffMean=SA mean difference

$$SAdiffMean < 16,000 = SAmean < 16,000_{Pre} - SAmean < 16,000_{Post} \tag{5.6}$$

where:

SAdiffMean<16,000=SA mean difference<16,000 lbs.

$$SAdiffSD = SASD_{Pre} - SASD_{Post} \tag{5.7}$$

where:

SAdiffSD=SA SD difference

5.8.2 Tandem Axle (TA) Shape Factors

Equations 5.8 to 5.14 were used to obtain the shape factors for TA. The ratios (pre/post) of mean and SD for the first and second peaks were also obtained for TA NALS.

$$TAdiffM1 = TAPL_{1(Pre)} - TAPL_{1(Post)} \tag{5.8}$$

where:

TAdiffM1=TA unloaded peak difference

$$TAdiffSD1 = TASD_{1(Pre)} - TASD_{1(Post)} \tag{5.9}$$

where:

TAdiffSD1=TA unloaded peak SD difference

$$TAdiffM2 = TAPL_{2(Pre)} - TAPL_{2(Post)} \tag{5.10}$$

where:

TAdiffM2=TA loaded peak difference

$$TAdiffSD2 = TASD_{2(Pre)} - TASD_{2(Post)} \tag{5.11}$$

where:

TAdiffSD2=TA loaded peak SD difference

$$TAdiffMean > 26,000 = TAmean > 26,000_{Pre} - TAmean < 26,000_{Post} \tag{5.12}$$

where:

TAdiffMean>26,000=TA mean difference>26,000 lbs.

$$TAdiffOAM = TAOAM_{Pre} - TAOAM_{Post} \tag{5.13}$$

where:

TAdiffOAM=TA overall mean difference

$$TAdiffOSD = TAOASD_{Pre} - TAOASD_{Post} \tag{5.14}$$

where:

TAdiffOSD=TA overall SD difference

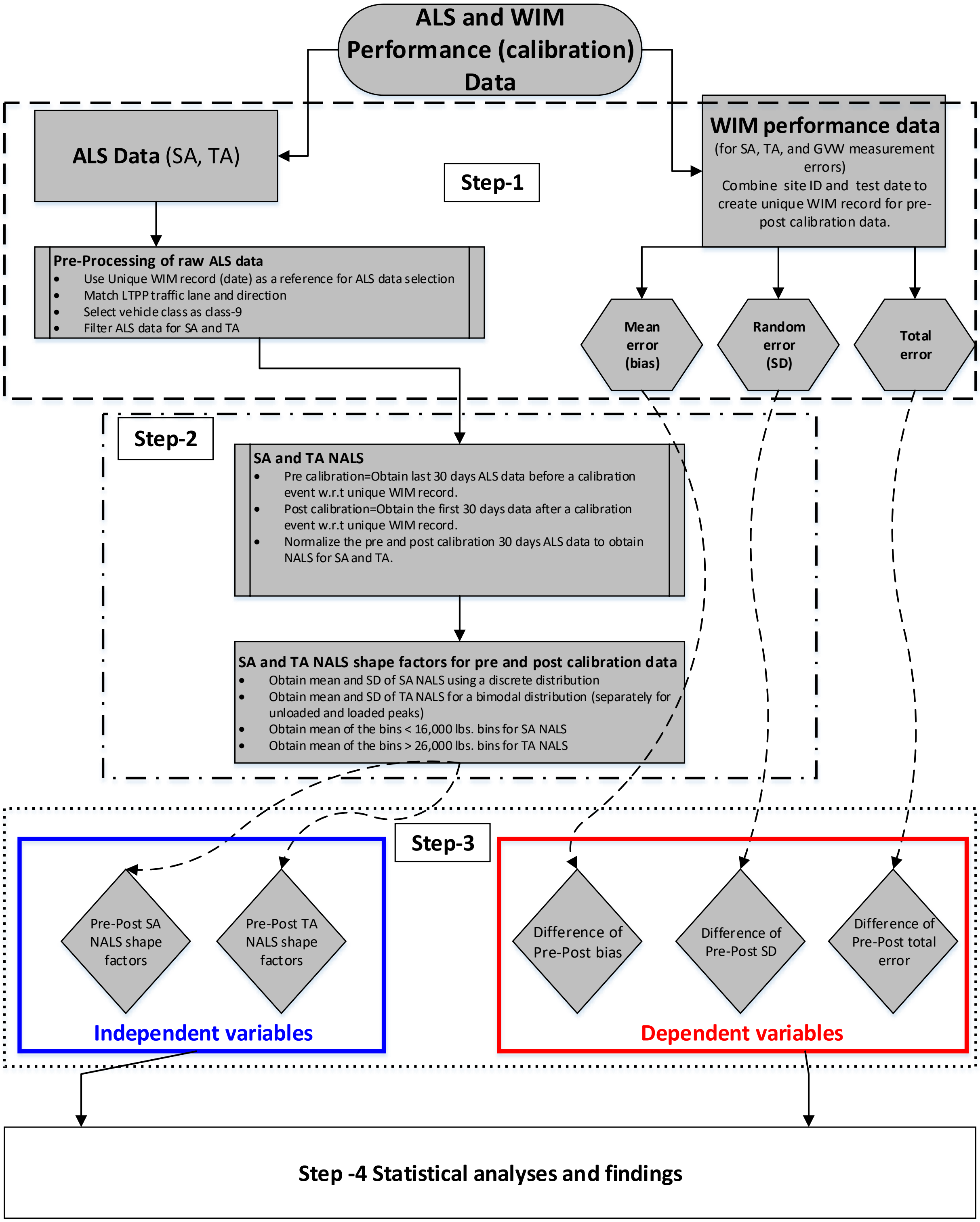


Figure 5-7 Flowchart for ALS and WIM performance data analyses.

5.8.3 Variables Obtained from WIM Calibration Data

The following variables (see Equations 5.15 to 5.23) were obtained by taking differences in SA, TA, and GVW WIM errors for pre and post-calibration data. These data attributes were used as dependent variables for model development to assess changes in WIM errors over time.

$$SABdiff = SAbias_{Pre} - SAbias_{Post} \tag{5.15}$$

where:

SABdiff=SA bias difference

$$SASDdiff = SASD_{Pre} - SASD_{Post} \tag{5.16}$$

where:

SASDdiff=SA SD difference

$$SATEdiff = SAtotalerror_{Pre} - SAtotalerror_{Post} \tag{5.17}$$

where:

SATEdiff=SA total error difference

$$TABdiff = TAbias_{Pre} - TAbias_{Post} \tag{5.18}$$

where:

TABdiff=TA bias difference

$$TA \text{ SD difference } (TASDdiff) = TASD_{Pre} - TASD_{Post} \tag{5.19}$$

where:

TASDdiff=TA SD difference

$$TATEdiff = TAtotalerror_{Pre} - TAtotalerror_{Post} \tag{5.20}$$

where:

TATEdiff=TA total error difference

$$GVWBdiff = GVWbias_{Pre} - GVWbias_{Post} \tag{5.21}$$

where; GVWBdiff=GVW bias difference

$$GVWSDdiff = GVWSD_{Pre} - GVWSD_{Post} \tag{5.22}$$

where; GVWSDdiff=GVW SD difference

$$GVWTEdiff = GVWtotalerror_{Pre} - GVWtotalerror_{Post} \tag{5.23}$$

where; GVWTEdiff=GVW total error difference

5.8.4 Statistical Analyses and Results

Data visualization is the first step before running any statistical analyses. A strong correlation was observed between TA shape factors and TA bias differences (see Table 5-5). The TA shape factors were also highly correlated with each other. This high correlation amongst TA shape factors could lead to the potential issue of multicollinearity. No clear relationship was observed between TA SD differences and TA NALS shape factors (see Table 5-6). A strong correlation was observed between SA shape factors and bias differences (see Table 5-7). The SA shape factors were also highly correlated with each other. This high correlation amongst SA shape factors could lead to the potential issue of multicollinearity.

Table 5-5 Correlation between TA bias and TA NALS shape factors

Variable	TAdiffM1	TAdiffM2	TAdiffM2 (Man)	TAM2 (Pre/Post)	TAdiffMean >26,000	TAdiffOAM	TABdiff
TAdiffM1	1						
TAdiffM2	0.22	1					
TAdiffM2 (Man)	0.10	0.86	1				
TAM2(Pre/Post)	0.21	1.00	0.85	1			
TAdiffMean>26,000	0.17	0.86	0.83	0.85	1		
TAdiffOAM	0.55	0.57	0.41	0.57	0.61	1	
TABdiff	0.22	0.75	0.72	0.74	0.89	0.60	1

Table 5-6 Correlation between TA SD and TA NALS shape factors.

Variable	TAdiffSD1	TAdiffSD2	TASD2(Pre/Post)	TAdiffOSD	TASDdiff
TAdiffSD1	1				
TAdiffSD2	-0.45	1			
TASD2(Pre/Post)	-0.32	0.63	1		
TAdiffOSD	0.065	0.23	-0.11	1	
TASDdiff	-0.11	0.17	0.21	-0.21	1

Table 5-7 Correlation between SA bias and SA NALS shape factors.

Variable	SAdiffMean	SAdiffMean<16,000	SAMean (Pre/Post)	SABdiff
SAdiffMean	1			
SAdiffMean<16,000	0.96	1		
SAMean(Pre/Post)	1.00	0.96	1	
SABdiff	0.88	0.86	0.89	1

The dependent and independent variables presented by Equations 5.5 to 5.23 were used to develop a model that would assess changes in WIM weight measurement errors over time. Different statistical techniques, including scatter plots, correlation, linear, non-linear, and multiple regression, were used to identify the most significant variables. The Next section presents the final models developed for SA, TA, and GVW bias estimation.

5.8.4.1 Model for Estimating Bias in TA Weight Measurement

Equation 5.24 shows the final model developed for QP and BP sensors. The sensor type was also considered an independent variable, but it was not significant. The coefficient of determination for the TA bias model is 0.8, showing that the independent variable can explain 80% of the variance in the dependent variable. Figure 5-8(a) shows the goodness-of-fit for the TA bias model. This graph compares the model-predicted and observed TA bias values for all the available data for the QP and BP sensors.

$$\begin{aligned} &TABdiff = 0.0041 * TAdiffMean > 26,000 \\ &R^2 = 0.80 \end{aligned} \tag{5.24}$$

The significant term, i.e., the difference between pre and post TA MEAN >26,000 (TAdiffMean>26,000), can be a good predictor for assessing and quantifying changes in TA bias in WIM systems. This shape factor represents the mean load of tandem axles weighing greater than 26,000 lb. For a bimodal tandem axle load distribution, it would be the loads in bins greater than 26,000 lb. The models can be improved further by adding more data in the future. The above model should be used in combination with the visual inspection of the shifts in the location of TA peak loads for the loaded peaks. This analytical approach can aid in estimation changes in WIM measurement accuracy and facilitate identifying the WIM calibration needs without performing the actual field validations of WIM equipment performance using calibration trucks. This methodology can save a significant amount of time and resources required for field validation using test trucks.

5.8.4.2 Validation of the Model for Estimating Bias in TA Weight Measurement

The WIM performance and axle loading data from the pre and post-calibration events were obtained from the MDOT and used for the model validation. Figure 5-8(b) shows the goodness of fit for the TA bias prediction model using the validation data. The TA bias predictions for the model validation data are reasonably accurate (R^2 = 0.82). These data were not used during the model development, and the prediction errors seem logical since both the data are subjected to different loading patterns and conditions. The TAdiffMean>26,000 data were simulated within the observed range to study the model's sensitivity. Figure 5-8(c) shows the sensitivity of the model to the independent variable. The model shows that when the pre and post-difference between TAdiffMean>26,000 for class 9 trucks exceeds almost 1250 lbs., the TA bias difference exceeds 5%, indicating equipment would require calibration.

5.8.4.3 TA Model Predictions

Table 5-8 provides the 95% confidence and prediction intervals based on the TA bias model as a function of TA shape factors. It can be noted that when pre and post-TA mean>26,000 difference exceeds 1250 lbs., the bias difference exceeds 5 percent.

Table 5-8 TA model predictions and 95% confidence and prediction intervals.

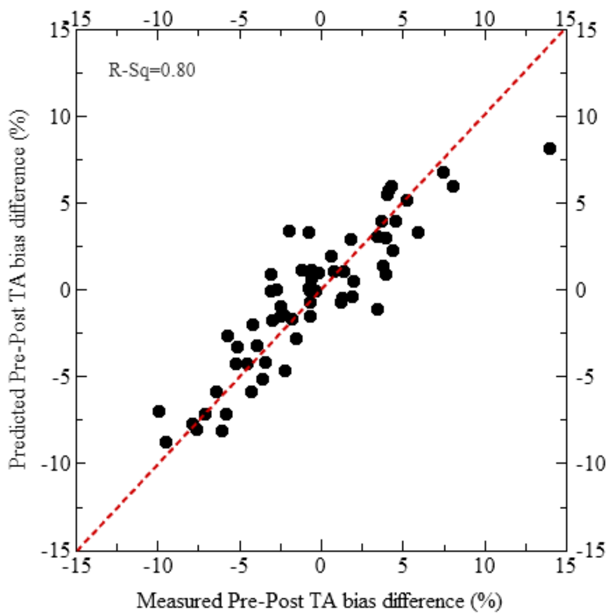
TAdiffMean>26,000 (lbs.)	TABdiff (%) (Predicted)	95 %CI		95 %PI	
		Lower	Upper	Lower	Upper
250	1.03	0.90	1.15	-3.03	5.08
500	2.05	1.79	2.31	-2.01	6.11
750	3.08	2.69	3.46	-1.00	7.15
1000	4.10	3.58	4.62	0.02	8.19
1250	5.13	4.48	5.77	1.02	9.23
1500	6.15	5.37	6.93	2.03	10.28
1750	7.18	6.27	8.08	3.02	11.33
2000	8.20	7.17	9.24	4.02	12.38

5.8.4.4 Model for Estimating Bias in SA Weight Measurement

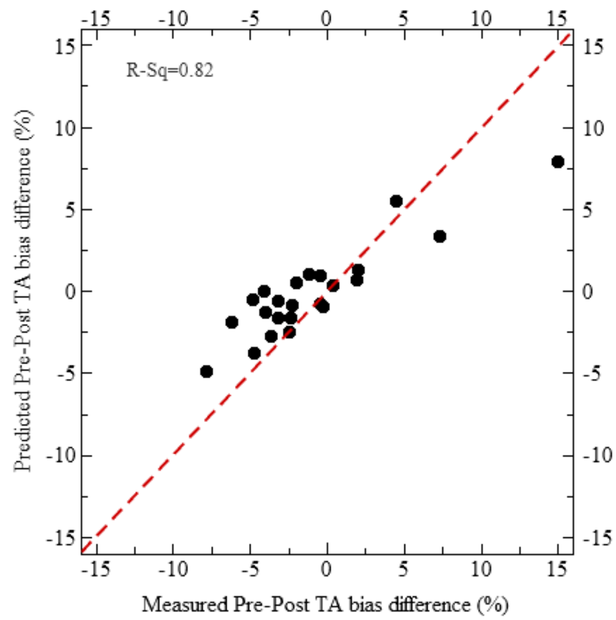
Equation 5.25 shows the final SA bias estimation model developed for QP and BP sensors. The coefficient of determination for the SA bias model is 0.78, showing that the independent variable SAdiffMean can explain 78% of the variance in the dependent variable SABdiff. Figure 5-9(a) shows the goodness-of-fit for the SA bias model.

$$\begin{aligned} SABdiff &= 0.008572 * SAdiffMean \\ R^2 &= 0.78 \end{aligned} \tag{5.25}$$

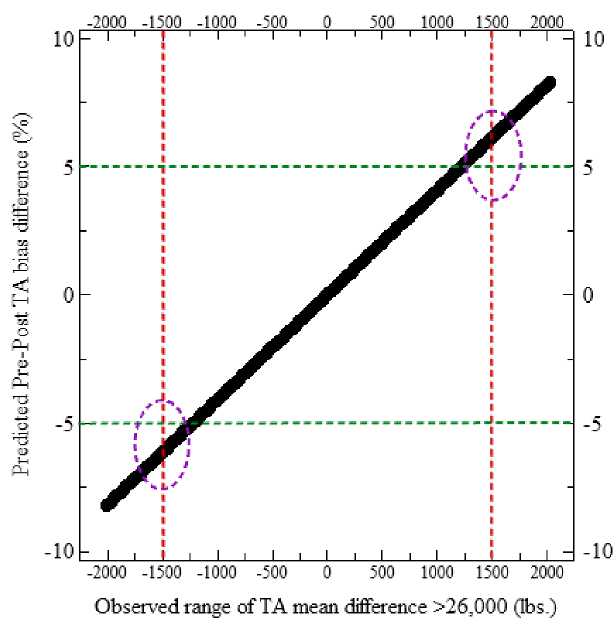
Overall, the SA model made predictions accurately (R^2=0.78). The significant term, i.e., the difference between Pre and Post SA MEAN (SAdiffMean) can be used as a good predictor for assessing and quantifying changes in SA bias in WIM systems. This shape factor represents the mean load value of the NALS for single axle load distribution. The models can be improved further by adding more data in the future. The above model should be used in combination with the visual inspection of SA peak load shifts in the NALS distributions.



(a) Goodness of fit



(b) Model validation



(c) Model simulations

Figure 5-8 Goodness-of-fit, validation, and simulations for the TA bias model.

5.8.4.5 Validation of the Model for Estimating Bias in SA Weight Measurement

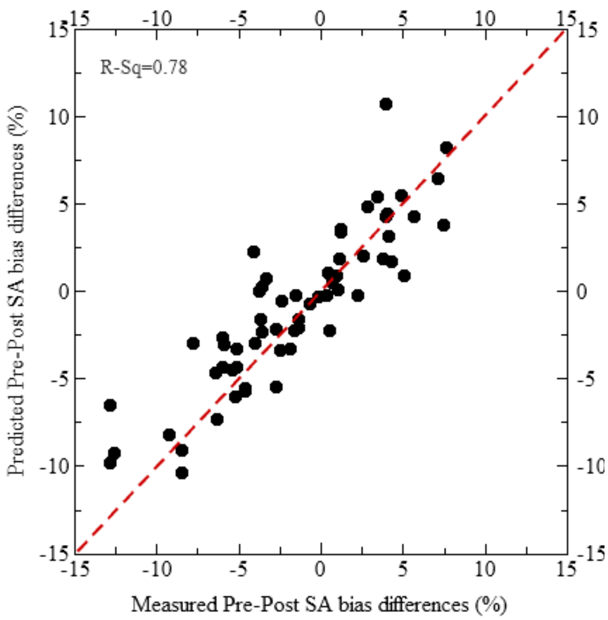
The pre and post-calibration WIM performance and axle loading data from the MDOT were used for the SA model validation. Figure 5-9(b) shows the goodness of fit for the SA bias model using the validation data. The SA bias predictions for the model validation data are reasonably accurate (R-Sq=0.68). The SAdiffMean data were simulated within the observed range to study the model's sensitivity. Figure 5-9(c) shows the sensitivity of the model to the independent variable. The model shows that when the pre and post-difference between SAdiffMean for class 9 trucks exceeds almost 500 lbs., the SA bias difference exceeds almost 4.5 to 5%, indicating equipment would calibration.

5.8.4.6 Model for Estimating Bias in GVW Weight Measurement

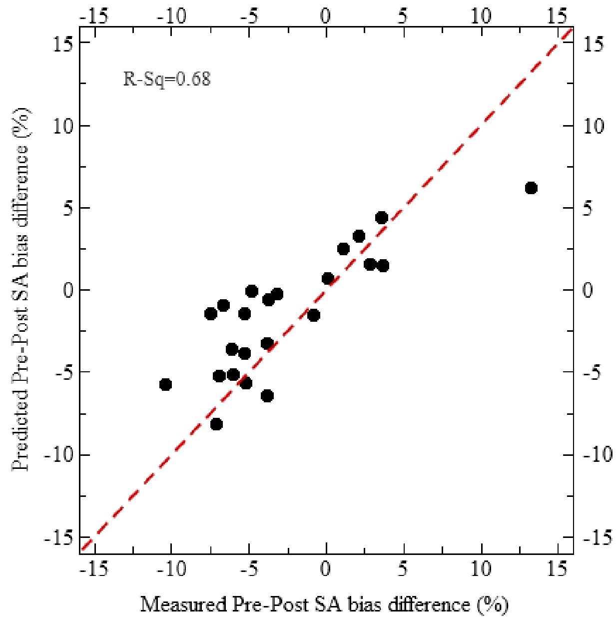
The SA and TA NALS shape factors were used as potential predictors to assess changes in GVW bias. The SA and TA shape factors were used in combination and separately to determine changes in GVW bias. Of all shape factors and combinations tested, TAdiffMean>26,000 was the best predictor for assessing GVW bias changes. Equation 5.26 shows the final GVW bias model developed for QP and BP sensors. The coefficient of determination for the GVW bias model is 0.75, indicating that the independent variable TAdiffMean>26,000 can explain 75% of the variance in the dependent variable GVWBdiff. Figure 5-10(a) shows the goodness-of-fit for the GVW bias model.

$$\begin{aligned} &GVWBdiff = 0.004030 * TAdiffMean > 26,000 \\ &R^2 = 0.75 \end{aligned} \tag{5.26}$$

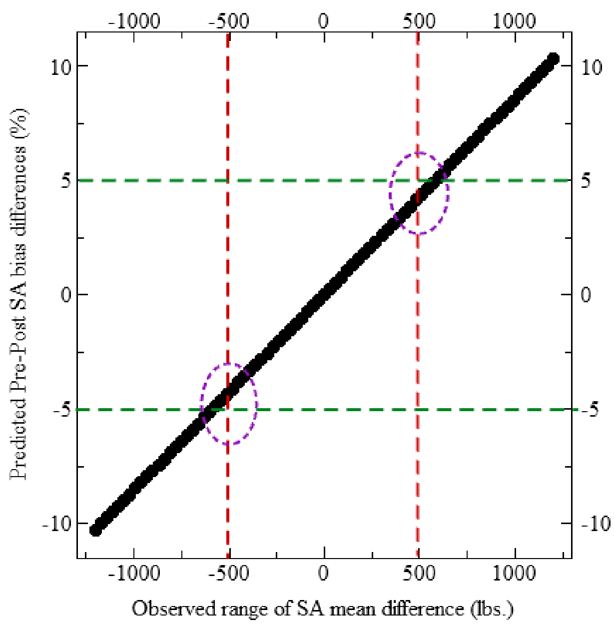
Overall, the GVW model made accurate predictions (R^2=0.75). The significant term, i.e., the difference between pre and post-mean weights of loaded TA (TAdiffMean>26,000), can be a good predictor for assessing and quantifying changes in GVW bias in WIM systems.



(a) Goodness of fit



(b) Model validation



(c) Model simulations

Figure 5-9 Goodness-of-fit, validation, and simulations for the SA bias model.

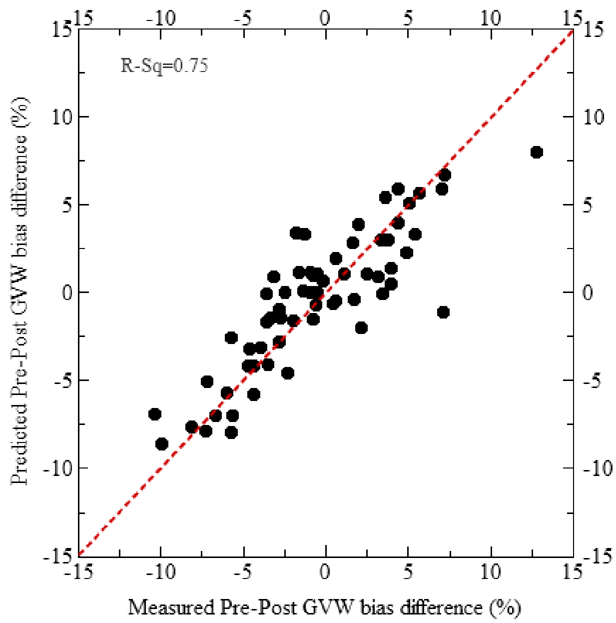
5.8.4.7 Validation of the Model for Estimating Bias in GVW Weight Measurement

Figure 5-10(b) shows the goodness of fit for the GVW bias model using the validation data obtained from MDOT. The TAdiffMean>26,000 data were simulated within the observed range to study the model's sensitivity. Figure 5-10(c) shows the sensitivity of the model to the independent variable. The model shows that when the Pre and Post difference between TAdiffMean>26,000 for class 9 trucks exceeds 1250 lbs., the GVW bias difference exceeds 5%.

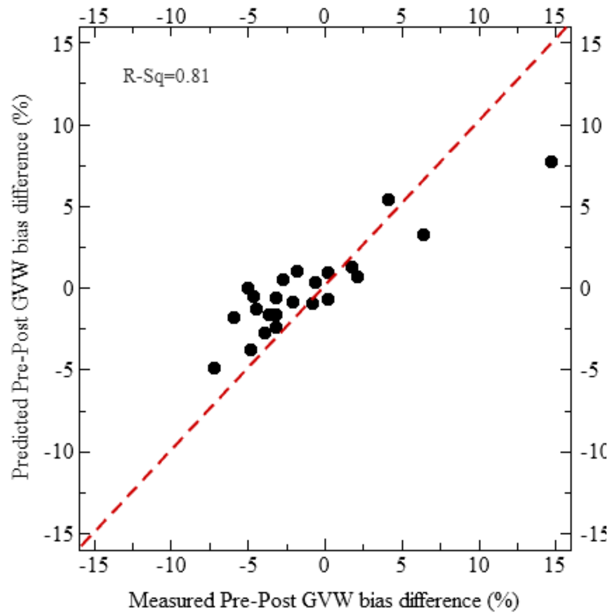
GVW bias models were also developed using a combination of SA (SAdiffMean) and TA (TAdiffMean>26,000) NALS shape factors. Also, the SA NALS shape factor, SAdiffMean; was independently used to estimate GVW bias differences. Equations 5.27 and 5.28 provide the models developed for GVW bias using both SA and TA shape factors combined and SA shape factor alone, respectively. Table 5-9 provides the coefficients for both models. The TA NALS shape factor TAdiffMean>26,000 showed up as the significant predictor when a linear combination of SA and TA NALS shape factors was used to estimate GVW bias differences. The model's accuracy is similar to the model presented in Equation 5.26 based on the TA NALS shape factor alone (R^2=0.75). The model accuracy significantly decreased (R^2=0.53) for the model based on the SA NALS shape factor. Figure 5-11 presents the goodness of fit for the model presented in Equation 5.28. Based on the results, it can be concluded that the TA NALS shape factors are better predictors to estimate GVW bias differences.

Table 5-9 GVW bias as a function of SA and TA NALS shape factors.

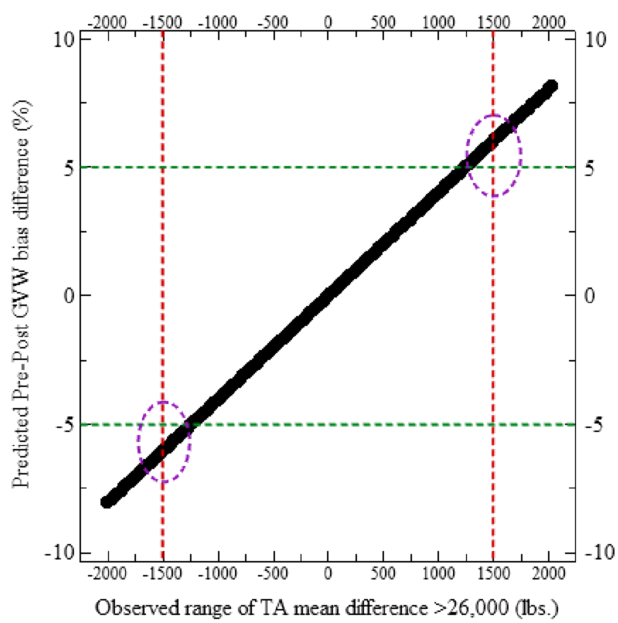
Term	Coef	SE Coef	T-Value	p-Value	VIF	Model type
TAdiffMean>26,000	0.003489	0.000476	7.32	0.000	2.54	Model-based on SA and TA NALS shape factors
SAdiffMean	0.001271	0.000877	1.45	0.152	2.54	
SAdiffMean	0.006275	0.000748	8.39	0.000	1.00	Model-based on SA shape factor alone



(a) Goodness of fit



(b) Model validation



(c) Model simulations

Figure 5-10 Goodness-of-fit, validation, and simulations for the GVW bias model.

$$GVWBdiff = 0.003489 * TAdiffMean > 26,000 + 0.001271 * SAdiffMean \qquad (5.27)$$
$$R^2 = 0.75$$

$$
\begin{aligned}
&GVWBdiff = 0.006275 * SAdiffMean \\
&R^2 = 0.53
\end{aligned}
\tag{5.28}
$$

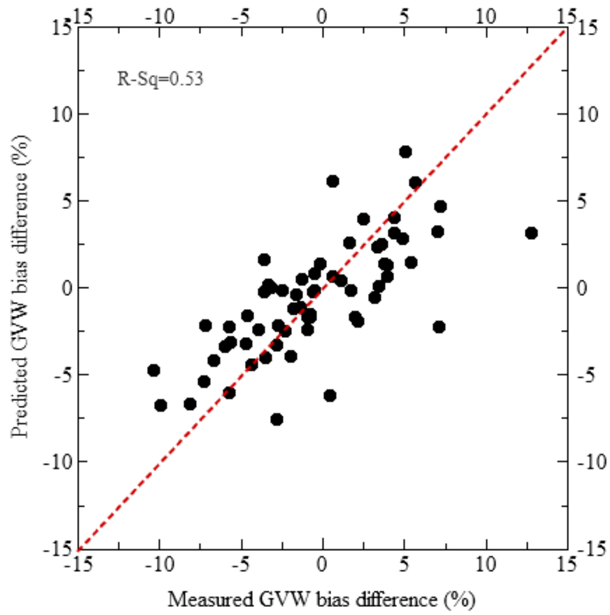


Figure 5-11 Goodness of fit for GVW model as a function of SA NALS shape factor.

5.8.4.8 Model for TA Measurement Bias Estimation for Different WIM Sensors

Equations 5.29 and 5.30 show the TA bias models developed separately for BP, and QP sensors. Although the sensor type is insignificant, these models can be used for making predictions for individual sensors. The model coefficients and accuracy are very much similar to the model that was developed by combining data for both sensors. Figure 5-12 shows the goodness of fit for the combined model.

For BP sensor:

$$
\begin{aligned}
&TABdiff = 0.0 + 0.004062 * TAdiffMean > 26,000 \\
&R^2 = 0.80
\end{aligned}
\tag{5.29}
$$

For QP sensor:

$$
\begin{aligned}
&TABdiff = -0.181 + 0.004062 * TAdiffMean > 26,000 \\
&R^2 = 0.80
\end{aligned}
\tag{5.30}
$$

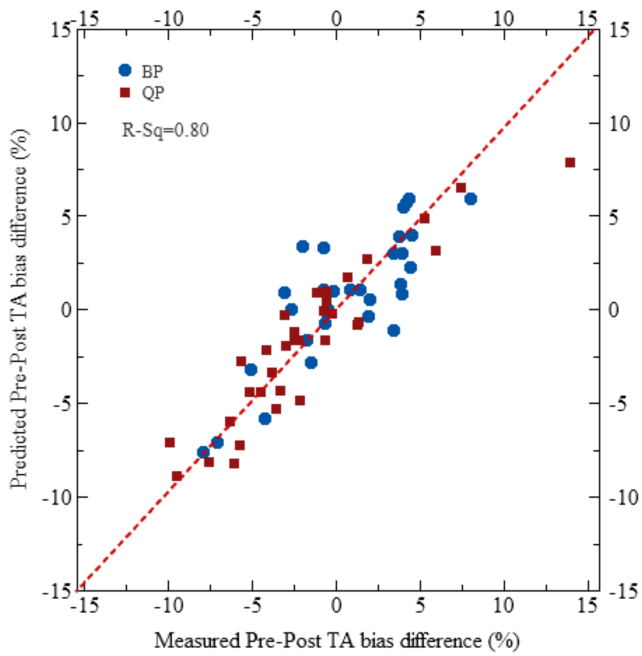


Figure 5-12 TA bias model for different sensors.

5.8.5 Application of Models – Case Study

This section presents the application of the TA bias model with the help of an example. The axle loading data were obtained from the SPS-10 WIM site located in Nevada (32AA00). This is a QP sensor WIM station installed in AC pavements. At this site, the calibration was performed on November 28, 2018, and the equipment showed negligible bias (-0.7% for TA). The next equipment calibration was scheduled for August 2019. The TA NALS data for one month after calibration (December 2018) and one month before the next scheduled calibration (July 2019) were obtained to study changes in WIM performance. Figure 5-13 shows the NALS for December 2018 and July 2019. The WIM site started overestimating TA loads within 7 months after calibration. The TA NALS shape factor (TAMean>26,000) was calculated for both datasets, and the TA bias was estimated using Equation 5.24. This shape factor can be calculated without fitting the bimodal distribution. The mean is the product sum of the midpoints and frequencies divided by the total of frequencies for the load bins greater than 26,000 lbs. The calculated values for TAMean>26,000 were 31810 and 34185 lbs. for December 2018 and July 2019, respectively. The TAdiffMean>26,000 value was 2375 lbs., and the estimated bias was 9.74 % for TA. The results show that the WIM system significantly overestimates weights and needs calibration. The WIM system field calibration and validation summary report also confirmed that the site is overestimating loads based on pre-validation results obtained using test truck runs on August 14, 2019. This example shows the application and significance of the TA bias model that can help identify the equipment calibration needs without physically making the test truck runs.

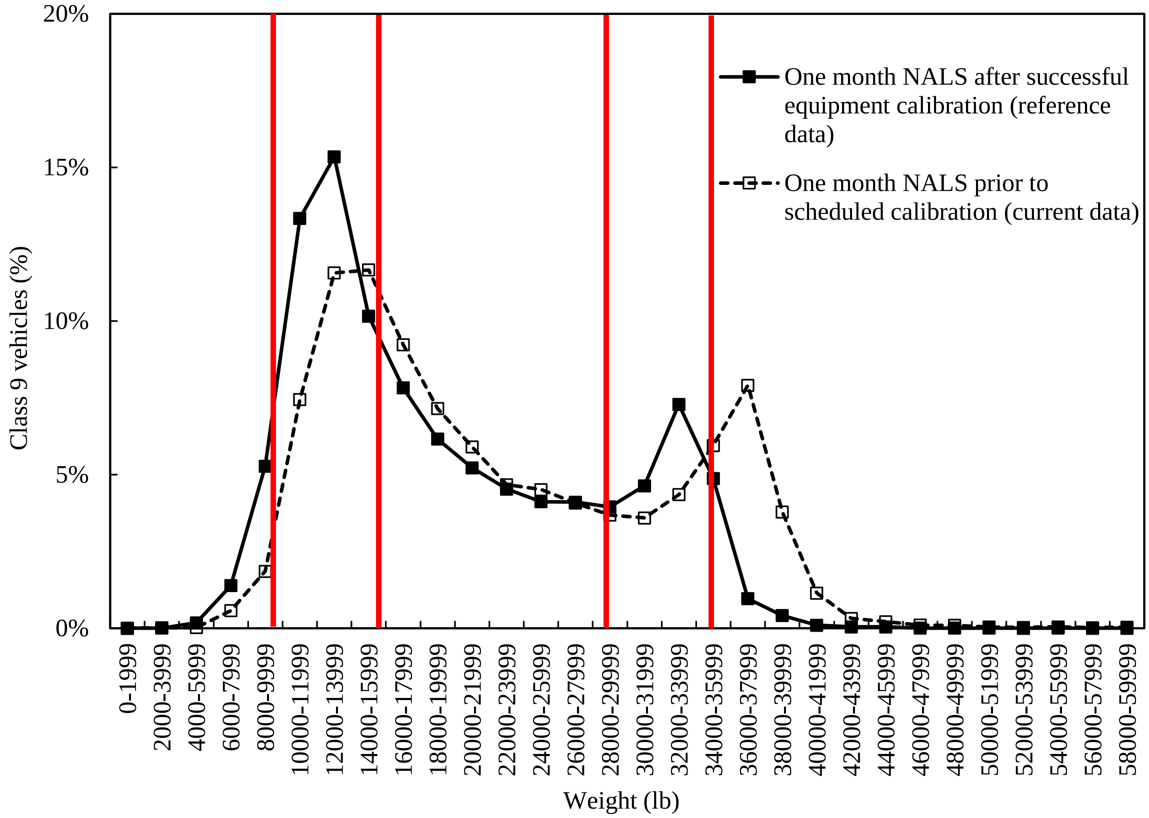


Figure 5-13 TA NALS for SPS-10 Nevada WIM site (32AA00).

5.9 KEY FINDINGS

The following are the key findings based on the analyses of NALS shape factors and WIM performance data:

- The NALS analyses show that a Calibration frequency longer than 1 year may be acceptable for the sites with BP sensors. A calibration frequency of at least 1 year is recommended for sites with QP sensors. Due to significantly higher NALS inconsistencies, the sites with PC sensors may need multiple calibrations in a yea
- No clear relationship was observed between the changes in SD values for SA and TA computed based on NALS and SD changes computed using pre- and post-calibration data based on test trucks data for SA and TA WIM SD.
- The pre and post-TA bias differences (TABdiff) can be accurately estimated using changes in TA mean value for the loaded (>26,000 lbs.) Class 9 trucks (TAdiffMean>26,000), obtained from pre and post-TA NALS. When the TADiffMean>26,000 difference exceeds 1250 lbs., the TA bias difference exceeds 5%, indicating the equipment requires calibration.
- The pre and post-SA bias differences (SABdiff) can be accurately estimated using differences in SA means (SAdiffMean) obtained from pre and post-SA NALS. When the SAdiffMean difference exceeds almost 500 lbs., the SA bias difference exceeds 4.5 to 5%, indicating the equipment requires calibration.
- A strong correlation exists between the GVW bias differences and TADiffMean>26,000 differences, indicating that TA WIM errors are significant contributors to GVW WIM errors.
- The pre and post-GVW bias differences (GVWBdiff) can be accurately estimated using pre and post-differences in TA mean>26,000 lbs. (TAdiffMean>26,000). When the TADiffMean>26,000 difference exceeds 1250 lbs., the GVW bias difference exceeds 5%, indicating the equipment requires calibration. The data results also showed that the

TA NALS shape factor (TADiffMean>26,000) is a better predictor (R^2=0.75) of GVW bias differences as compared to the SA NALS shape factor (R^2=0.53)

- The models presented should be combined with the visual inspection of SA and TA peak loads and the information about seasonal changes in traffic loading of Class 9 trucks due to land use activities (such as major agricultural harvests, if any).
- Using NALS to estimate the TA WIM accuracy can save a significant amount of time and resources, which are usually spent on equipment calibrations every year.

5.10 CHAPTER SUMMARY

A set of statistical procedures was developed to aid in identifying and quantifying changes in WIM measurement bias (calibration drift) based on analysis of changes in axle load spectra attributes for FHWA Class 9 vehicles (typically used as a calibration truck type) between WIM equipment calibration events. The results show that changes in single and tandem axle load spectra attributes, such as SA mean axle load and TA mean load for the loaded axles weighing over 26,000 lbs., can be effectively used to estimate the systematic changes (bias) in WIM measurements for GVW, SA, and TA.

WIM measurement accuracy estimation methodology through axle load spectra analysis can be used to identify WIM equipment calibration needs, saving a significant amount of time and resources required for field validation of WIM system performance using test trucks. The statistical models developed in this study for the prediction of WIM measurement bias for GVW and SA and TA loads could be fully automated and used to screen WIM data to identify data sets with significant deviations in key shape factors (SA mean axle load and TA mean load for the loaded axles weighing over 26,000 lbs.). Flagged WIM data sets could then be subjected to visual inspection of SA and TA load spectra, along with reviewing information about the expected seasonal changes in traffic loading due to land use (if any). These results could be used to decide if WIM equipment calibration is necessary.

CHAPTER 6 GUIDELINES FOR WIM EQUIPMENT CALIBRATION

6.1 PURPOSE

The relative influence of the factors presented in Table 2-7 on WIM measurement errors is not well understood or quantified. These factors contribute to poor WIM system performance and users' lack of confidence in the collected data. As a result, analytical techniques and models are needed to assess the relative significance of different sources of error on the accuracy of WIM data. WIM data collectors also require direction and practical tools to increase WIM data quality through improved procedures related to WIM site selection, technology selection, installation, calibration, maintenance, data processing, and quality control/quality assurance (QC/QA) [22, 40, 41].

6.2 INTRODUCTION

The WIM systems go out of calibration, and their accuracy deteriorates over time due to many factors. These factors may include changes in measurement conditions (e.g., temperature and speed), pavement deflection, roughness caused by distresses, and fatigue of WIM sensors. The authors of the referenced studies also reported that regardless of the WIM system calibration, the WIM accuracy could deteriorate over time due to these factors [4, 5, 29, 36, 42, 43]. In another study in the state of Arkansas, 10 out of 25 WIM sites yielded suitable loading data. The authors reported that the other sites exhibited evidence of WIM scale (sensor) failures and inconsistent loading data because of calibration concerns [44].

WIM equipment requires periodic calibrations to yield accurate and reliable loading data. To reduce the calibration cost, many agencies rely on various auto-calibration techniques using different software-based algorithms. The most common auto-calibration methods offered by the WIM vendors include using the (a) average front axle weight of Federal Highway

Administration (FHWA) Class 9 trucks, (b) average weight of specific types of vehicles (often a loaded five-axle tractor semi-trailer). The auto-calibration techniques may be beneficial but have some limitations; for example, weight laws, truck characteristics, and front axle weights can vary among states. Therefore, these techniques could be implemented only after confirming the local WIM site conditions [45, 73]. The LTPP field operations guide uses multiple runs of a pre-weighed class-9 truck for calibrating a WIM site.

6.3 OBJECTIVES

This study addresses three main issues related to WIM systems accuracy and calibration procedures; i.e., how to (1) perform successful calibration of a WIM system, (2) model gross vehicle weight (GVW) WIM errors as a function of individual axle errors [(single axle (SA) and two tandem axles (TA), drive and trailer tandem)], and (3) estimate WIM measurement errors using the LTPP and the ASTM protocols. Therefore, the primary objectives of the paper are to provide (a) a review of high-quality LTPP WIM data, (b) provide guidelines for successful WIM equipment calibration by quantifying the effect of sample size (truck runs), speed, temperature, and truck type on WIM errors, (c) develop models for GVW error predictions as a function of SA and TA, and (d) compare the ASTM and the LTPP WIM accuracy estimation methods using SA, TA, and GVW WIM errors. These objectives were accomplished by synthesizing and analyzing the WIM error data in the LTPP database for BP and QP sensors.

6.4 DATA EXTENTS

Table 6-1 presents the climate and sensor type distribution of WIM sites and associated records available in the LTPP database. It can be noted that the majority of the WIM sites are located in a wet climate. In total, 111 (53+58) and 62 (34+28) WIM records were available for pre-and post-calibration data, respectively. At least 40 test truck runs were used to obtain pre- and post-calibration data for these events.

Table 6-1 Distribution of WIM sites and records by sensor type and climate.

Data	Sensor type	Climatic regions				Total
		DF	DNF	WF	WNF	
Pre calibration	BP	-	3[a] (17 [b])	3 (18)	4 (18)	10 (53)
	QP	3 (9)	5 (16)	7 (18)	6 (15)	21 (58)
Post calibration	BP	-	3 (13)	3 (10)	4 (11)	10 (34)
	QP	2 (5)	3 (5)	3 (8)	6 (10)	14 (28)

[a] No of WIM sites, [b] No of WIM records

6.5 IMPROVED PROCEDURES FOR SUCCESSFUL WIM EQUIPMENT CALIBRATION

This section quantifies the effect of speed, temperature, number of runs (sample size), truck type (loaded vs. unloaded), and number of trucks on measured WIM errors.

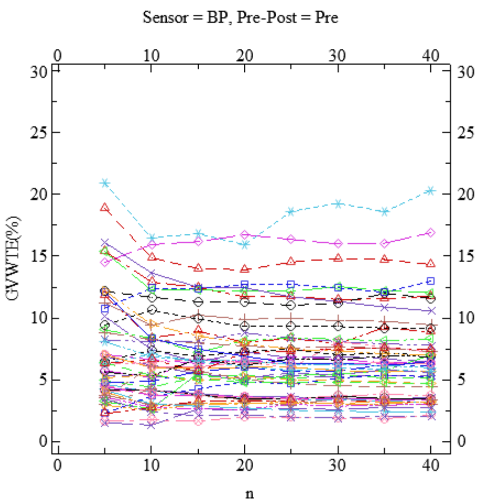
6.5.1 Desired Sample Size

The WIM equipment needs periodic calibrations, and the calibration frequency can vary from site to site for different sensor types [22]. The WIM equipment is calibrated using multiple runs of a test truck of known static weight, and the static weights are compared with the WIM weights. The truck type (fully and partially loaded Class 9 trucks) and the number of trucks (1 to 3) can vary for WIM equipment calibration. The number of runs per test truck can also vary from 10 (even fewer) to 60 depending on the calibration protocols in practice. The number of runs can be more than 60 if the truck data is used from the traffic stream [74]. The LTPP WIM protocol uses 40 test truck runs (20 each for two different trucks) to calibrate/validate a WIM site at varying speed levels. However, many other state DOTs use 10 truck runs or even fewer for a single test truck to calibrate/validate a WIM site. More runs can cover higher speed ranges and temperature fluctuations, consuming more time and resources. The sample size can influence the computed accuracy of WIM data and the reliability of the results.

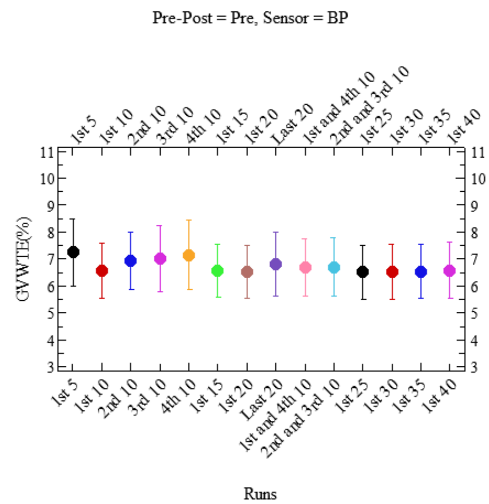
This section addresses an important question. i.e., what sample size is large enough to be considered representative for the mean (bias) and SD (consistency) error computations? Different sample sizes ranging from 5 to 40 were analyzed to evaluate their effect on WIM accuracy using pre and post-calibration WIM equipment data for BP and QP sensors. Different combinations (details below) of truck runs were used to impose randomness that can account for varying speed and temperature fluctuations. Figures 6-1 and 6-2 present the scatter plots and 95% CI interval plots of GVW total errors based on varying sample sizes (n) for BP and QP sensors, respectively. One horizontal line in the scatter plots represents a single pre- or post-calibration event. Figure 6-3 presents the line plots for GVW bias, SD, the margin of error at 95% confidence (MOE), and the GVW total error. The results show that the varying sample size has a statistically insignificant effect (mostly flat lines and overlapping 95% CI) on computed WIM errors, especially when n>=10, even when the errors were calculated for a different combination of truck runs [see Figures 6-1(a) to (d) and 6-2(a) to (d)]. The line plots suggest some differences (MOE and TE increase) when the sample size is extremely small, i.e., n<=5 [see Figures 6-3(a) to (d)]. It can be concluded based on data analyses that a WIM site can be successfully calibrated/validated using 10 or more runs.

The details of truck run combinations used in this analysis are shown below:

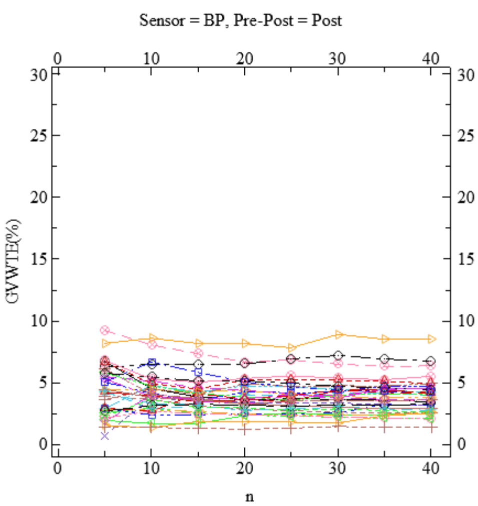
- 1st 5: 1 to 5
- 1st 10: 1 to 10
- 1st 15: 1 to 15
- 1st 20: 1 to 20
- 1st 25: 1 to 25
- 1st 30: 1 to 30
- 1st 35: 1 to 35
- 1st 40: 1 to 40
- 2nd 10: 11 to 20
- 3rd 10: 21 to 30
- 4th 10: 31 to 40
- Last 20: 21 to 40
- 1st and 4th 10: 1 to 10, and 31 to 40
- 2nd and 3rd 10: 11 to 20, and 21 to 30



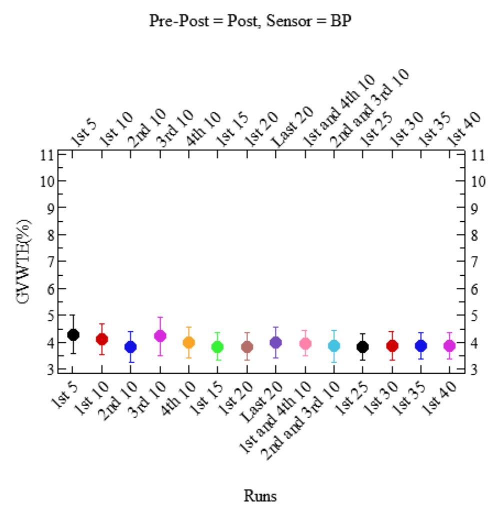
(a) BP-Pre (scatterplots with runs)



(c) BP-Pre (95%CI plots)

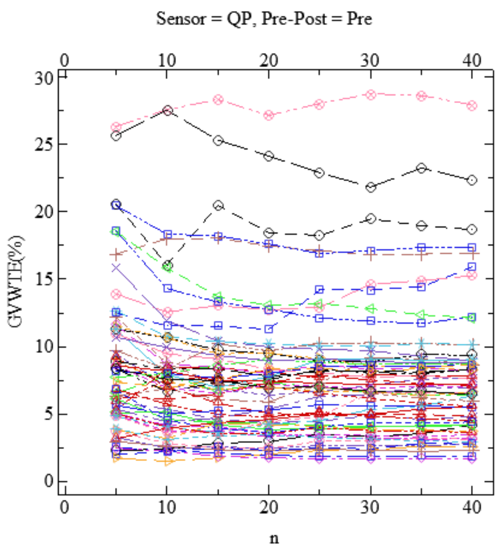


(b) BP-Post (scatterplots with runs)

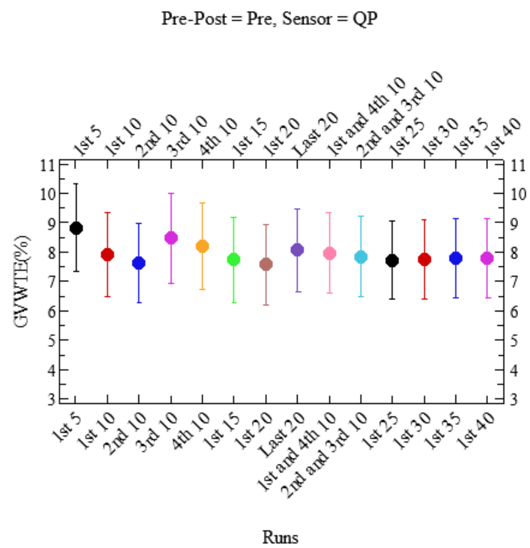


(d) BP-Post (95%CI plots)

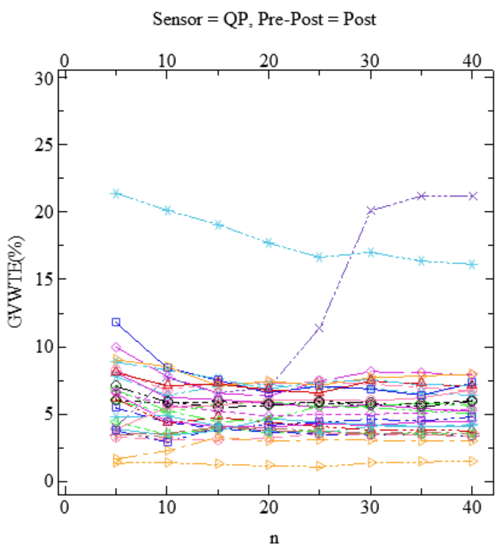
Figure 6-1 Scatter and 95% CI plots for varying sample size (BP sensor).



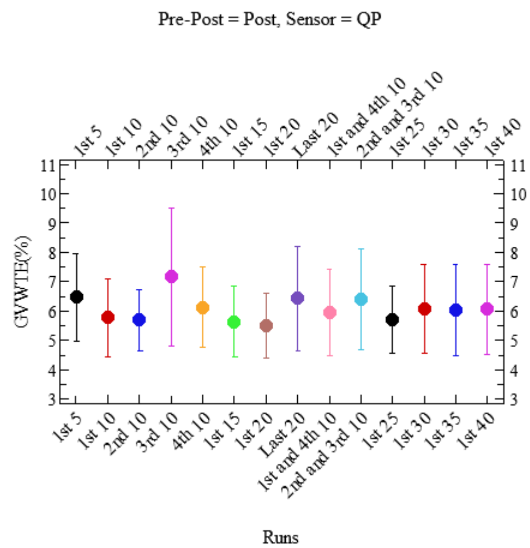
(a) QP-Pre (scatterplots with runs)



(c) QP-Pre (95%CI plots)

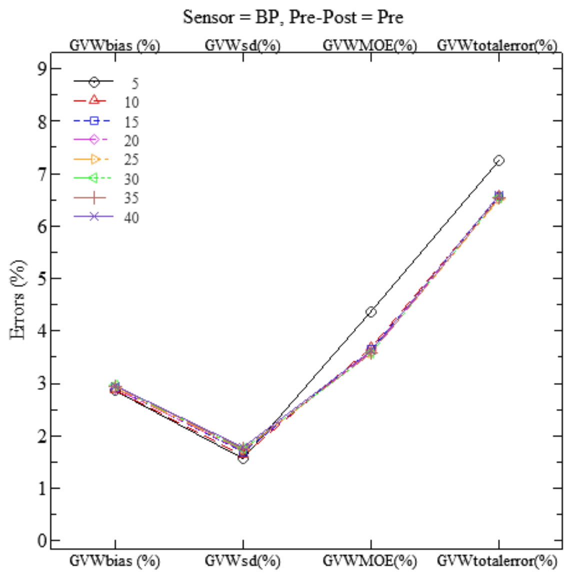


(b) QP-Post (scatterplots with runs)

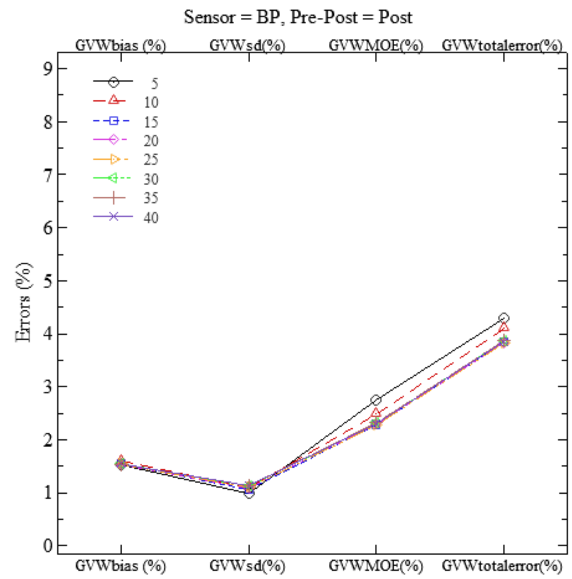


(d) QP-Post (95%CI plots)

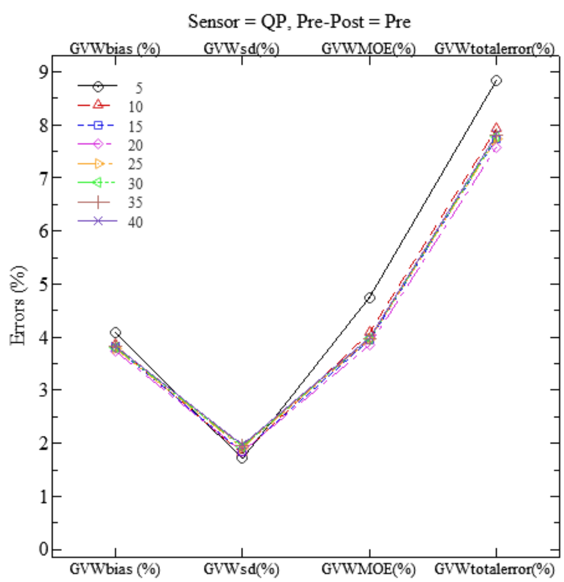
Figure 6-2 Scatter and 95% CI plots for varying sample size (QP sensor).



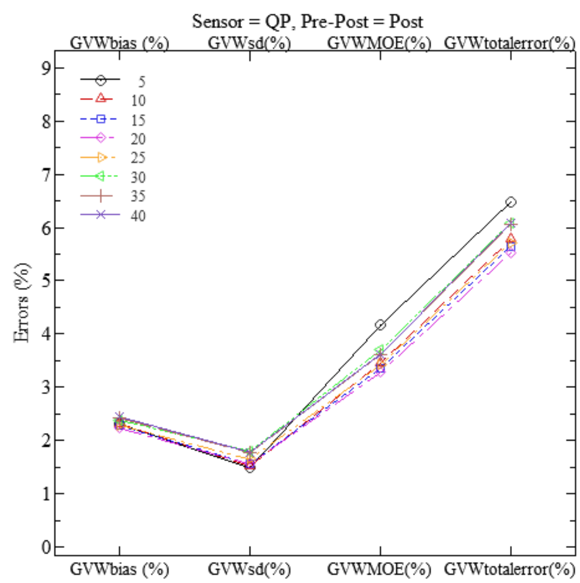
(a) BP-Pre



(c) BP-Post



(b) QP-Pre



(d) QP-Post

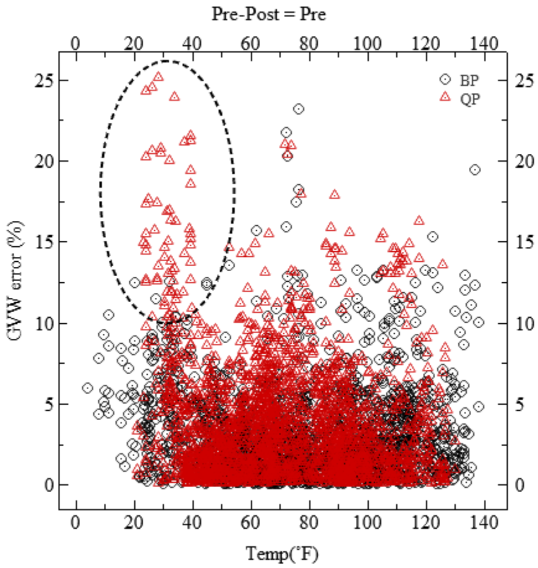
Figure 6-3 Impact of sample size on WIM error.

6.5.2 Effect of Temperature

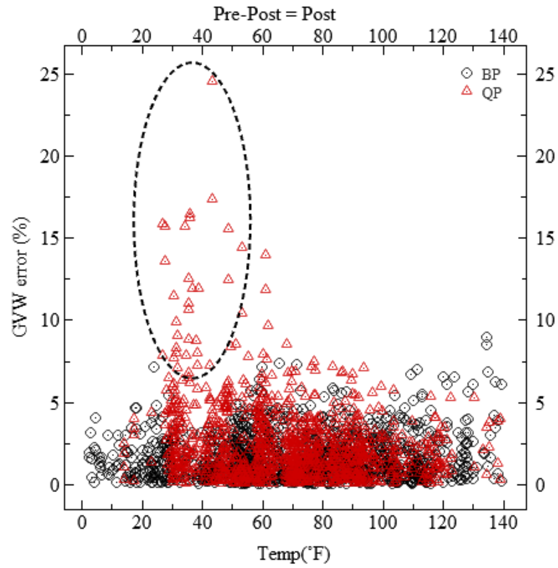
The LTPP Field operations guide for SPS WIM sites details the procedure for collecting pavement temperature data during equipment calibration [26]. The methodology is similar to the

LTPP FWD hand-held infrared temperature sensors. The guide recommends that the calibration be performed at a wide range of temperatures, 30˚F or more, and collecting data over more than 8 hours a day may be necessary. The protocol suggests that at least 12 runs (where possible) be performed for each temperature category.

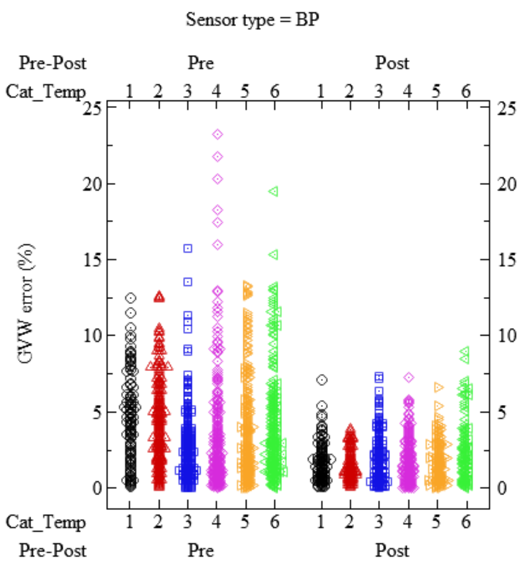
The calibration temperature data were categorized into 6 distinct categories, i.e., (<=30.0, 30.1-50.0, 50.1-70.0, 70.1-90.0, 90.1-110.0, >110.1) ˚F. Figures 6-4 and 6-5 present the results to evaluate the effect of temperature on GVW WIM errors. The scatter plots for temperature data show that mostly the effect is random except for QP sensor data below the freezing temperatures [see Figures 6-4(a) and (b)]. The individual value plots show that BP sensor GVW errors are very stable across different temperature categories [see Figure 6-4(c)]. However, some increase in GVW errors for QP sensors can be tied to low temperatures [see Figure 6-4(d)].
Further investigation revealed that all these data with high errors were collected at the QP site installed in PCC pavements located in Washington. Figures 6-5(a) to (c) clearly show that the GVW errors significantly increased as the temperature dropped below 40˚F at this site. This site's detailed calibration/validation report revealed that this WIM site also had issues with the PCC pavement conditions/support.



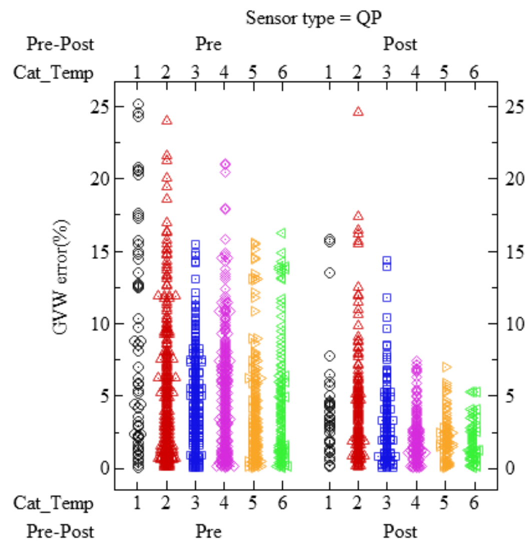
(a) Temperature-Pre



(b) Temperature-Post

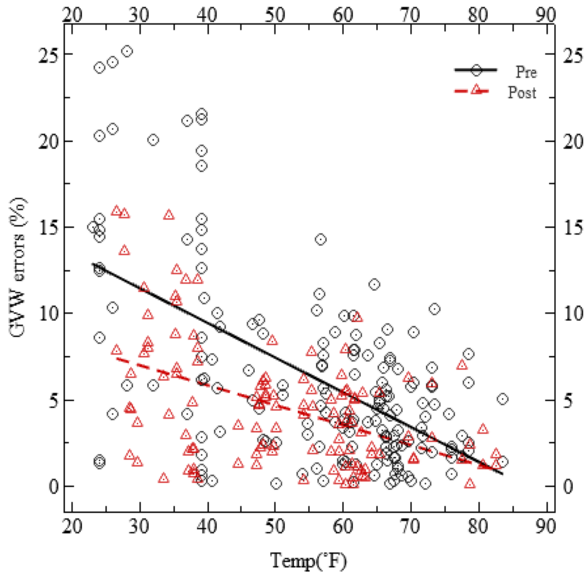


(c) Temperature-BP

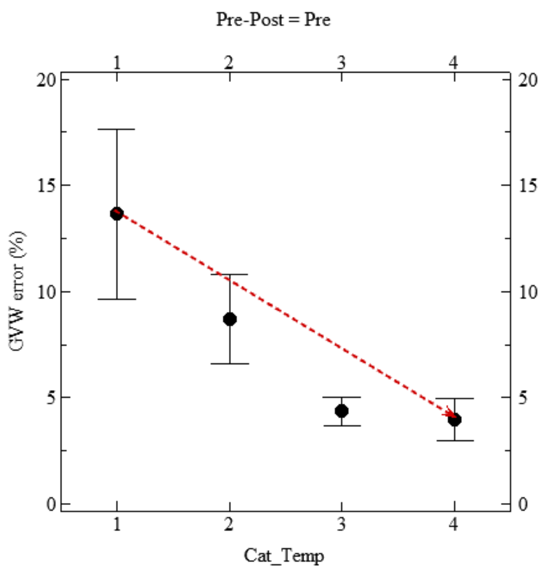


(d) Temperature-QP

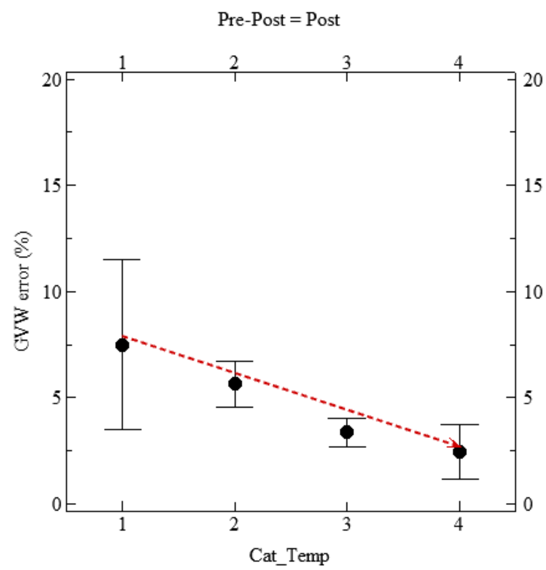
Figure 6-4 Effect of temperature on GVW errors



(a) QP-53-0200 (Washington)



(b) QP-53-0200 (Washington)-Pre

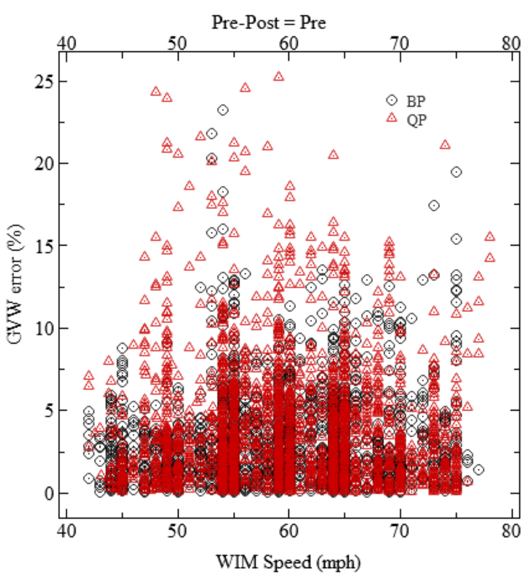


(c) QP-53-0200 (Washington)-Post

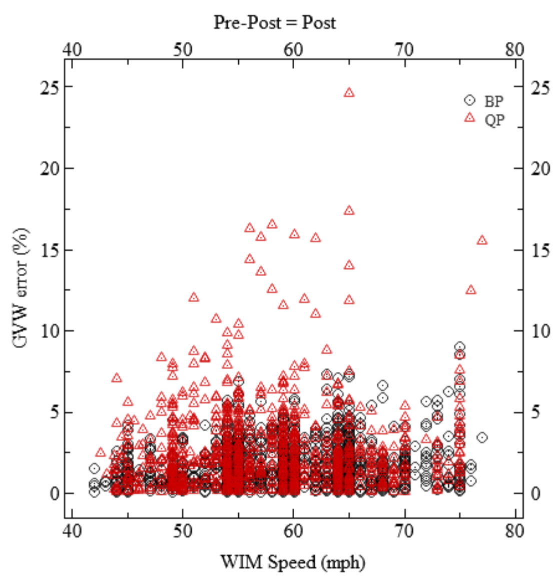
Figure 6-5 Effect of temperature on GVW errors

6.5.3 Effect of Truck Speed

The WIM error data were analyzed at different truck speeds. For this analysis, the calibration speed data were categorized into 7 distinct categories, i.e., (<=45.0, 45.1-50.0, 50.1-55.0, 55.1-60.0, 60.1-65.0, 65.1-70.0, >75.1) mph. The results are presented using scatter, individual values, and interval plots (see Figures 6-7 and 6-8). The scatter and interval plots do not show any clear relationship for the QP sensor, though a small increase in errors with increased speed was observed for BP sensors. Although the differences were statistically significant (using interval plots), the differences are very small (less than 1-2 %) and have no practical significance. The speed dependency (increase or decrease in errors with change in speed) was observed for individual sites/events; however, no clear effect was observed when all the data were combined. Therefore, applying the compensation based on different speed levels should continue for equipment calibration/validation.

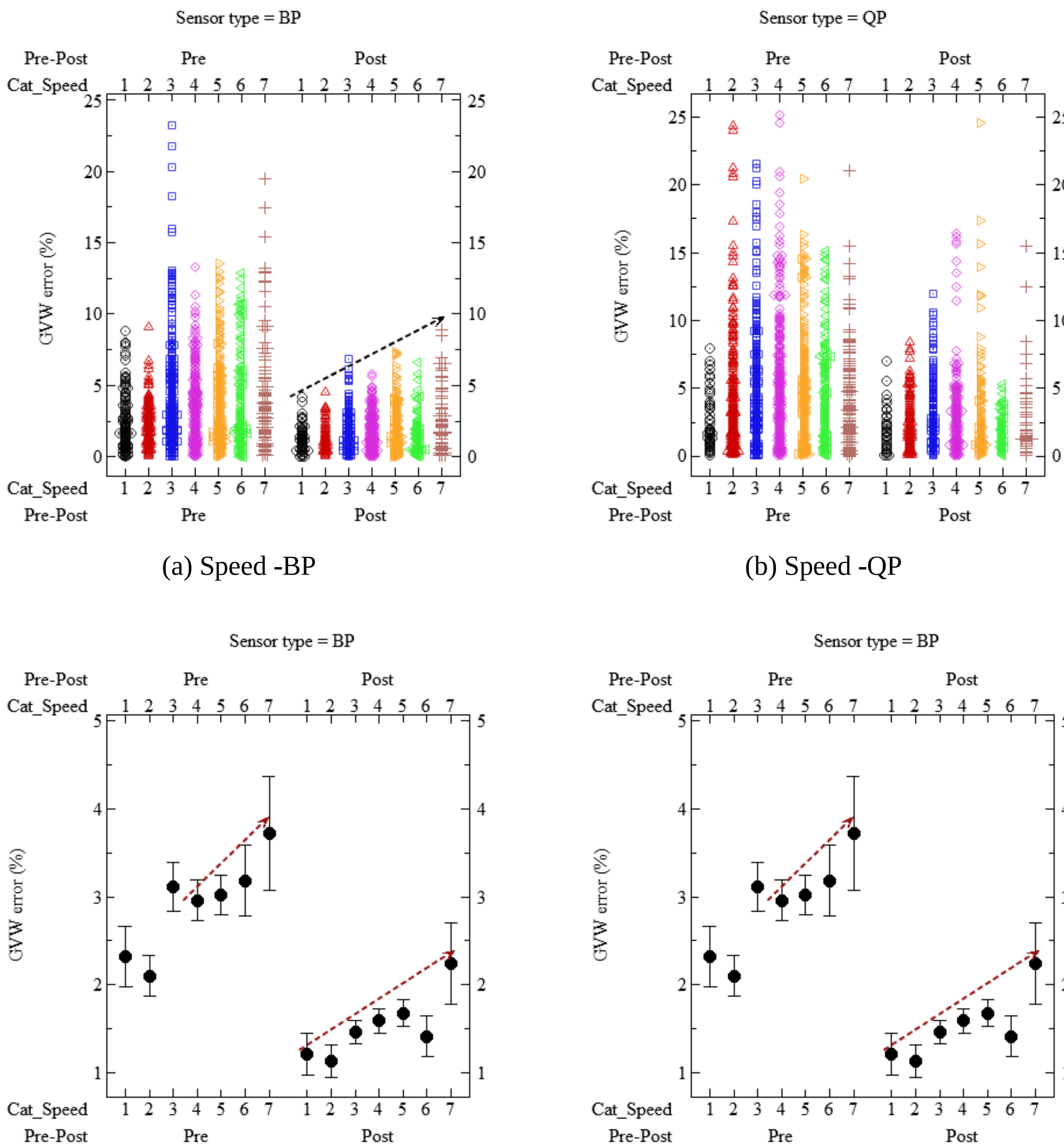


(a) Speed-Pre calibration data



(b) Speed-Post calibration data

Figure 6-6 Scatterplot-Effect of truck speed



(a) Speed -BP

(b) Speed -QP

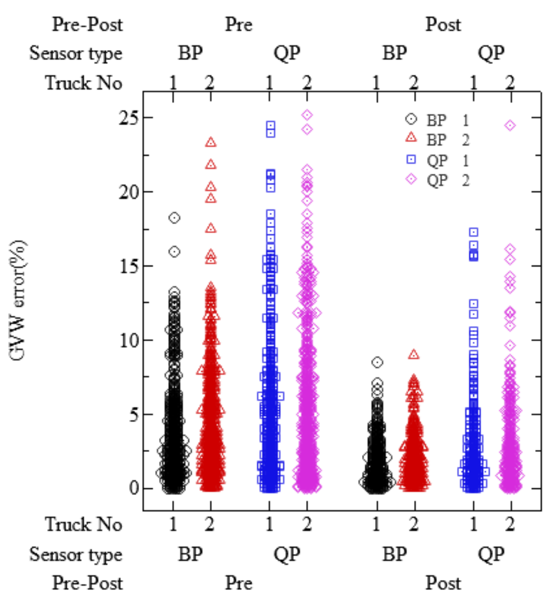
(c) Speed -BP

(d) Speed -BP

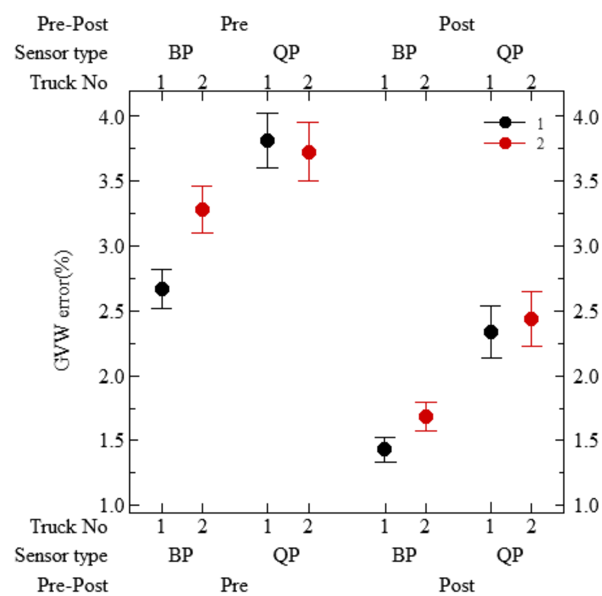
Figure 6-7 Effect of truck speed

6.5.4 Effect of Truck Type

The LTPP SPS WIM sites were calibrated and validated using two truck types, loaded Truck-1 and partially loaded Truck-2. Figure 6-8 and Table 6-2 present the GVW error results for trucks 1 and 2. The results show that the errors are significantly low for Truck-1 compared to Truck-2 for the BP sensor [see Figure 8 (b)]. The summary results for the BP sensor show that statistically significant (p-value <0.05) higher bias values were observed for Truck-2 as compared to Truck-1 in both pre and post-calibration data [see Table 6-2]. However, the magnitude of differences is very small and has limited practical implications, especially in post-calibration data (less than 0.25%). The truck type was not significant for QP sensor bias or SD values. It is pertinent to mention that both the truck types were used separately, using 20 test runs each to compute errors for individual calibration events.



(a) Individual value plots by truck type



(b) 95% CI plots by truck type

Figure 6-8 Effect of truck type (loaded vs unloaded)

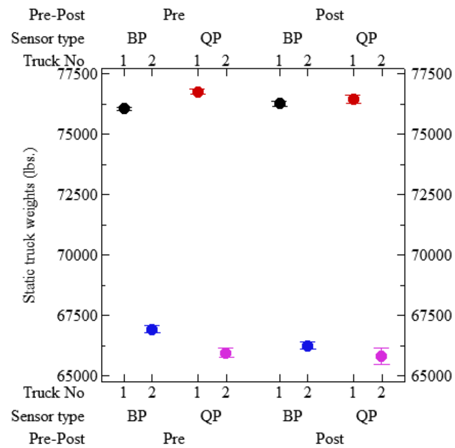
Table 6-2 Summary results with significance (loaded vs unloaded truck).

Sensor and calibration	Truck No.	Bias (%)	SD (%)	Significance (p-value)	
				Bias	SD
BP-Pre	Truck-1	2.66	1.57	0.003	0.055
	Truck-2	3.27	1.74		
BP-Post	Truck-1	1.44	1.10	0.03	0.054
	Truck-2	1.68	1.15		
QP-Pre	Truck-1	3.87	1.83	0.735	0.934
	Truck-2	3.78	1.82		
QP-Post	Truck-1	2.39	1.73	0.448	0.503
	Truck-2	2.50	1.79		

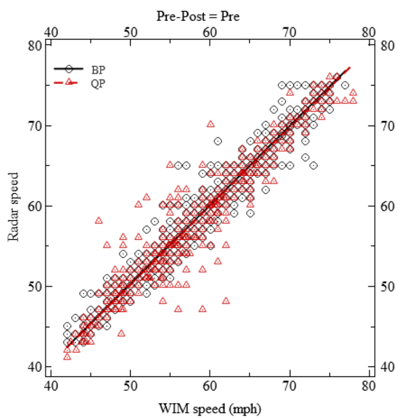
6.5.5 Static Weights, WIM Speed, and Overall vehicle Length

Finally, the truck speed and overall vehicle length estimated by the WIM system and the static truck weights were analyzed in this section. Figure 6-9 presents the results for static weights and WIM vs. radar speeds and WIM vs. static overall vehicle lengths. The results are based on the entire population, and it can be seen that the Truck-1 average static weight is 76,000 lbs. Truck-2 average static weight is around 66,000 lbs. for both sensor types in pre and post-calibration data. This variation in weights is imposed during calibration/validation procedures to account for the truck dynamics, adversely affecting WIM errors. Figures 6-9(b) and (c) show the comparisons of WIM and radar speed. The truck speeds collected by both; the WIM and the speed gun are generally in agreement. However, QP sensors underestimated WIM speeds for some pre-calibration records, and the issue was resolved in post-calibration data.

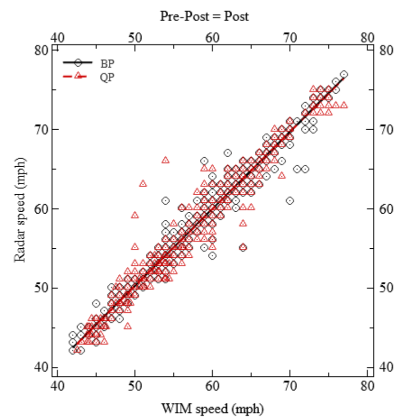
Similarly, the overall truck length estimates are more accurate for BP sensors [see Figures 6-9(d) and (e)]. The error in overall length or axle spacing can lead to vehicle misclassification. The data showed that the issue of under or over-estimation of vehicle length was eliminated in post-calibration data.



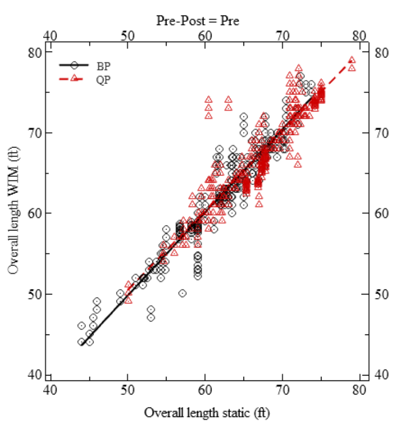
(a) Static weights for Truck 1 and 2



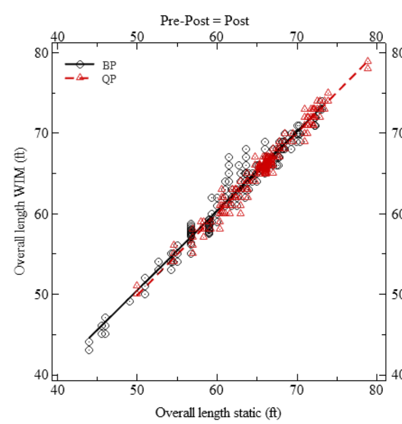
(b) WIM vs radar speed (pre-calibration)



(c) WIM vs radar speed (post-calibration)



(d) WIM vs static vehicle length (pre-calibration)



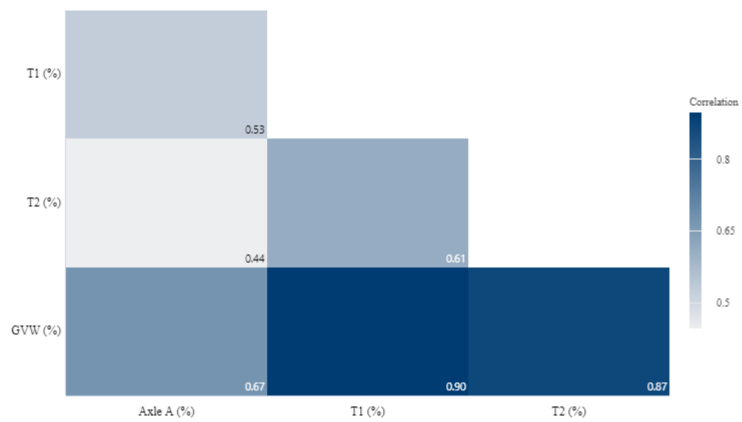
(e) WIM vs static vehicle length (post-calibration)

Figure 6-9 Static truck weights, speed, and vehicle length results

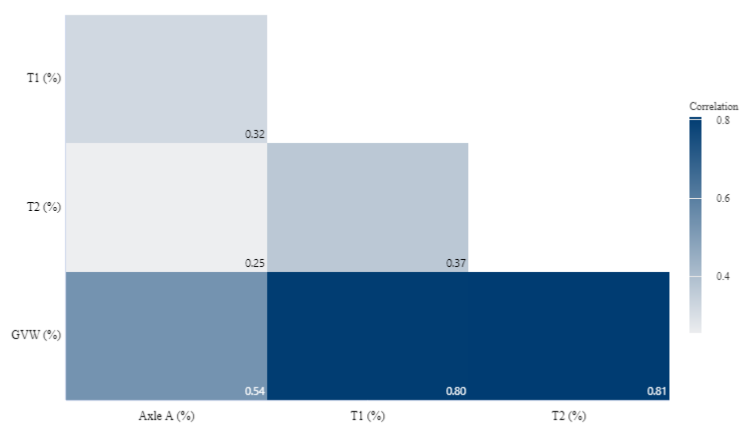
6.6 GVW ERRORS AS A FUNCTION OF SA, TA1, AND TA2

A WIM site can be categorized as ASTM Type I, if SA, TA, and GVW errors are within ±20%, ±15%, and ± 10 %, respectively. Due to complex truck configurations and dynamics, the static and dynamic weights collected for SA and two TA can vary substantially because individual wheel weights (left and right) are added to obtain a single axle (SA) and TA weights. The SA is a front axle in a Class 9 truck. During the WIM equipment calibration, the calibration factors (also known as compensation) are generally applied based on GVW errors obtained as a function of SA and two TA (drive tandem and trailer tandem). The LTPP field operation guide suggests a combination of the front axle (FA) and GVW can apply for compensation if a WIM site is equipped with such technology. This section presents the modeling of GVW errors as a function of FA and two TA. Before the model development, the pre and post-calibration data correlograms were generated to see the correlation between dependent and independent variables. Figures 6-10(a) and (b) show the results for Pre and post-calibration data, respectively. Strong correlations (0.80 to 0.90) were observed between GVW errors and drive tandem (T1) and trailer tandem (T2).

The Multiple Linear Regression (MLR) technique is used when one dependent variable is affected by more than one factor, assuming a linear relationship. Equation 6.1 shows the general form of MLR, where response y (independent variable) is predicted using inputs (dependent variables) x1, x2, and xi, βo is the intercept (constant term), and βi is the coefficient of the predictor xi. The multiple linear regression models developed for pre and post-calibration data are shown in Equations 6.2 and 6.3, respectively. All the independent terms (FA, T1, and T2) were significant (p-value <<0.05)) in both models.



(a) Pre-calibration WIM errors



(b) Post-calibration WIM errors

Figure 6-10 Correlogram for pre and post-calibration WIM errors

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_i x_i \tag{6.1}$$

$$GVW - \Pr e(\%) = -0.02536 + 0.162758 * SA(\%) + 0.41884 * T1(\%) + 0.408246 * T2(\%) \tag{6.2}$$

$$GVW - Post(\%) = 0.01003 + 0.16140 * SA(\%) + 0.40427 * T1(\%) + 0.41215 * T2(\%) \tag{6.3}$$

Table 6-3 presents the summary of pre and post-calibration models developed for GVW errors. The model goodness of fit shows that the GVW errors for pre and post-calibration data can be accurately estimated using the front axle and two tandem axles [see Figures 6-11(a) and (b)]. The results show that the models are sensitive to T1 and T2 errors (higher slope of predicted vs. the measured fit line), followed by SA errors [see Figures 6-11(c) and (d)]. The sensitivity analysis was performed by changing one variable at a time while keeping constant values for the other two variables. The average values and ranges were estimated for all the variables based on the actual WIM errors for pre and post-calibration data.

Table 6-3 Summary- pre and post-calibration GVW models.

Data	Significant terms	Variable importance (higher to lower)	MSE [a]	RMSE [b]	R-Sq (Trg)	R-Sq (Test)
Pre	3	SA (45.48% c), T1 (40.53%), T2 (13.64%)	0.077	0.277	99.65	99.64
Post	3	T1 (44.58%), T2 (25.79%), SA (28.67%)	0.080	0.283	99.05	99.04

[a] Mean squared error, [b] Root mean squared error, [b] Higher percentages indicate that the source (variable) accounts for more of the variation in the response.

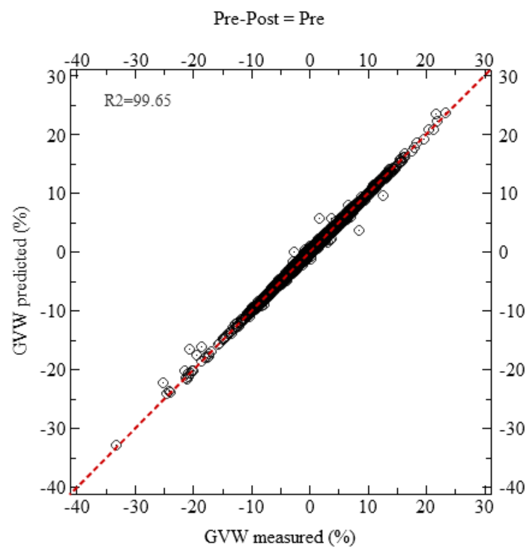
The SA load remains relatively stable after WIM equipment calibration because it mainly carries the engine's weight and is not affected by other truck payloads. Therefore the main contributors to the GVW errors are the two TAs. The distribution of load differences can cause a shift or change in axle weights observed by the WIM. The LTPP field guide suggests the use of steel plates or concrete blocks or beams securely attached [26]. At some TPF SPS sites, the test trucks loaded with crane counterweights were used for calibration. The developed models show that SA and two TA can accurately predict GVW errors. Therefore, the equipment calibration factors can be applied considering the calibration drift (positive or negative bias) in GVW errors. This information has huge potential for immediate application by highway agencies to optimize calibration procedures. This information validates agencies' practice of calibrating WIM sites based on GVW errors. Finally, agencies calibrating the WIM sites using SA only may revisit their technique and compare the results with the approach suggested in this study.

6.7 COMPARISONS OF WIM ERROR ESTIMATION METHODS

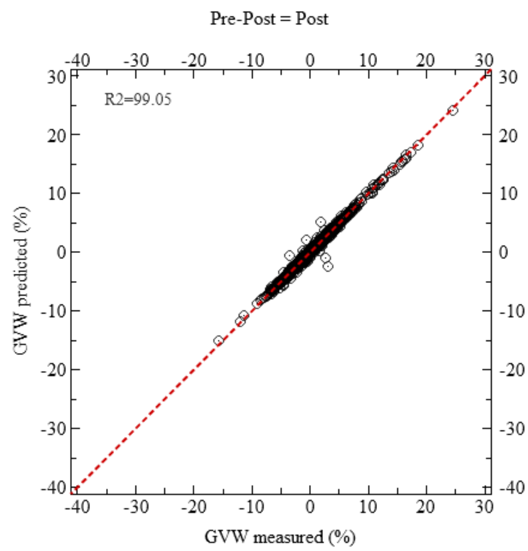
This section provides the results to accomplish the third and final objective of the paper. The three accuracy estimation methods described above are compared based on pre-and post-calibration data for varying sample sizes. Figures 6-12(a) and (b) show the number of passing records based on SA (<=20%), TA (<=15), and GVW (<=10) tolerance limits. Figures 6-13(a) and (b) show the number of failure events based on SA, TA, and GVW total errors computed for pre and post-calibration data. In general, all three methods are in agreement with each other. The LTPP accuracy estimation method is the most conservative of all three, especially with a smaller sample size (10 or fewer runs). Both the techniques using ASTM methods with slightly different interpretations are comparable.

Any pre- or post-calibration event that qualified as a passing event based on GVW error also passed SA and TA accuracy checks. However, there are events in pre-calibration data (compared to the LTPP method with 40 runs) that qualified as passing events based on SA (19 events) or TA (10 events) tolerance but did not pass the tolerance threshold based on the GVW errors (21 failure events). This analysis further augments the findings from the last section that if a WIM system is calibrated/validated using GVW errors, there is enough data-driven evidence that the system will also meet the SA and TA tolerance threshold. Some effect of sample size is also seen in the failure events part of post-calibration data. Only two events truly failed the equipment calibration from the WIM calibration reports. The two events were identified for the QP sensor, each for SPS WIM sites 350500 (New Mexico) and 530200 (Washington). Both these events were also declared as failure events using all three methods. Reviewing the detailed calibration reports revealed that the New Mexico site had a bad sensor that needed replacement. In contrast, the Washington site reported issues with pavement conditions/support. In addition, the LTPP accuracy computation approach resulted in three passing events as failures based on GVW data for a smaller sample size (10 runs). The three additional events [(total 5 in Figure 6-13(b)] were declared as failures with tolerance marginally crossing the thresholds. Due to a few outliers, these additional events had issues with the normality assumption considered in the

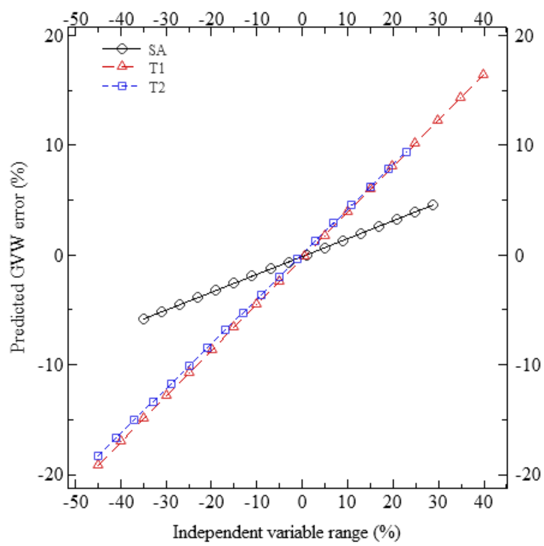
LTPP estimation approach. Based on this analysis, it can be concluded that for a smaller sample size (10 or less), all three methods should be compared to characterize a calibration event as pass or fail. Overall, the differences among all three methods were negligible, especially in post-calibration data. The results of the pre-calibration data also supported this analysis. However, at that time, the WIM sites were assessed based on a previous calibration performed 8 to 12 months (sometimes even more) before the data collection.



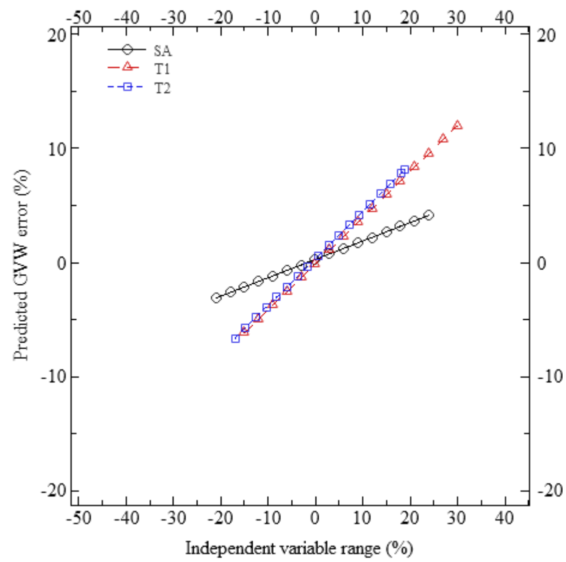
(a) Goodness of fit (pre)



(b) Goodness of fit (post)

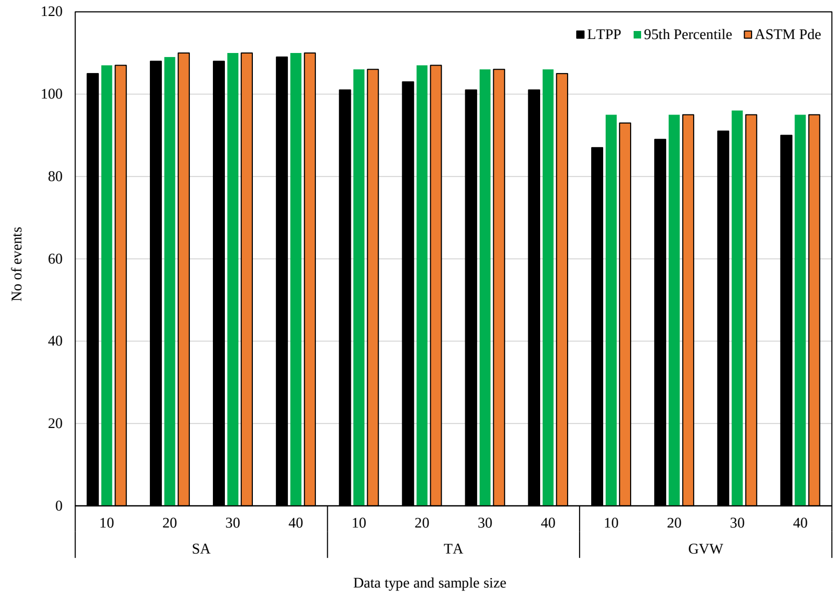


(c) Model sensitivity (pre)

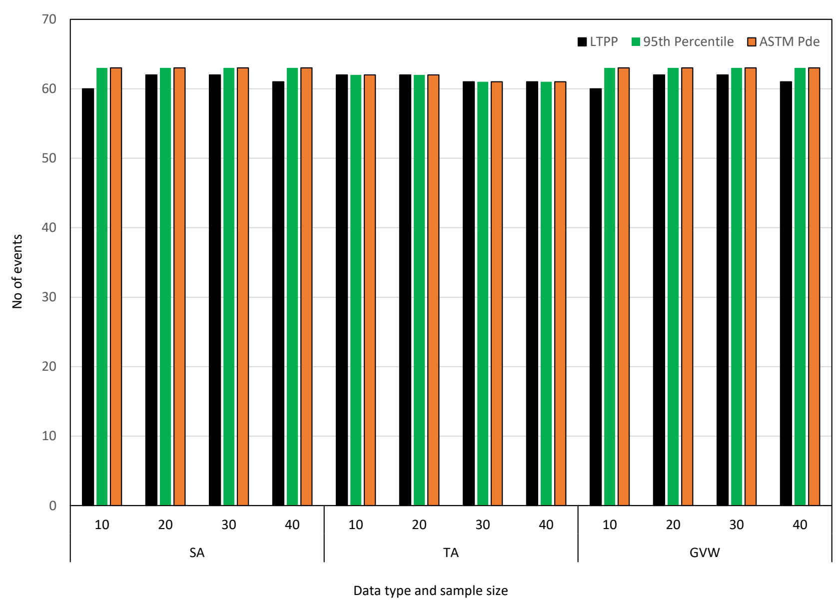


(d) Model sensitivity (post)

Figure 6-11 Comparison of models and sensitivity

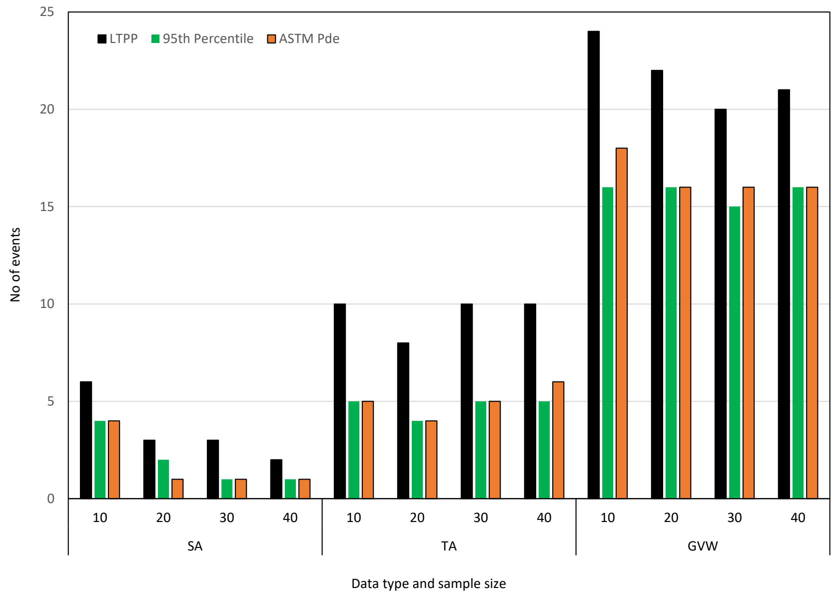


(a) Passing events- pre-calibration

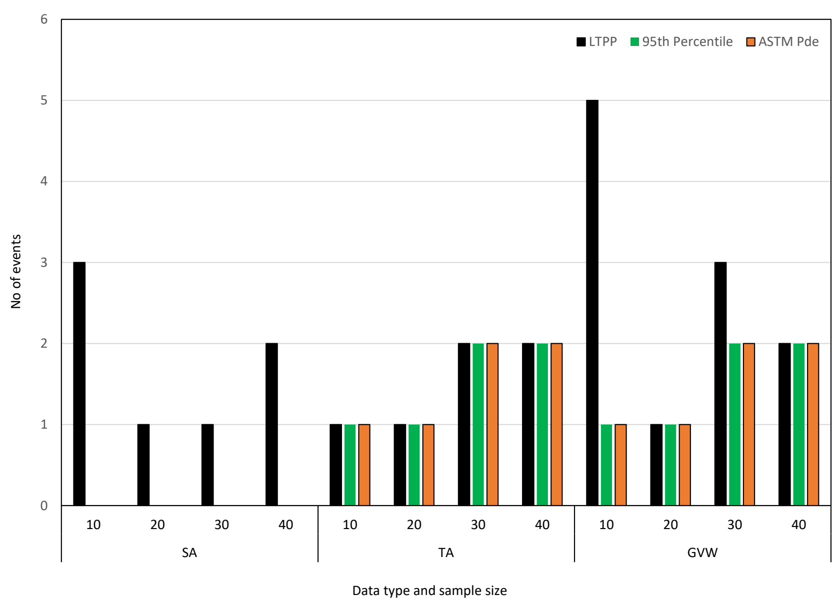


(c) Passing events- post-calibration

Figure 6-12 Comparisons of different accuracy estimation methods



(a) Failure events- pre-calibration



(b) Failure events- post-calibration

Figure 6-13 Comparisons of different accuracy estimation methods

6.8 KEY FINDINGS

Successful WIM equipment calibration can eliminate weight, speed, and axle spacing errors.

Following are the conclusions and recommendations based on data analyses.

- The results show that the effect of sample size on WIM errors was negligible, especially when the sample size is sufficiently large (n>=10) for QP and BP sensors. faltered
- The WIM site calibration can be performed using one test truck to achieve a representative range of BP and QP sensor errors. A single test truck with 12 runs (4 at each speed point) can be used for equipment calibration.
- The current LTPP filed operation guide recommendations of calibrating a WIM site at different speed levels should continue, preferably at three-speed points 50, 60, and 70 mph or as per the recommendations of the posted speed limits.
- Pre and post-calibration data can be collected on the same day for BP sensors, as no apparent effect of temperature was observed for BP WIM sites. If possible, the pre and post-calibration data can be collected for an extended period for QP sensors to account for higher temperature fluctuations.
- The representative post-calibration data can be collected accurately using one test truck with 12 passes at 3-speed points for QP and BP sensors. If a site shows higher speed dependency, the number of test truck runs may be increased to 20.
- The ASTM and the LTPP accuracy estimation methods are generally in agreement; however, the methods should be compared when the sample size is small and in the presence of potential outliers.
- The developed models showed that the GVW errors could be accurately predicted using SA and two TAs.
- The results show that if GVW errors are within ASTM Typ1 I tolerance, the SA and TA errors will also be within acceptable limits. Therefore, the practice of calibrating a WIM site using GVW errors should continue.

- The suggested changes in current WIM procedures can significantly reduce time and resources for successful equipment calibration.

6.9 CHAPTER SUMMARY

This chapter addresses three core issues related to WIM systems accuracy and calibration procedures, i.e., how to; (1) perform successful calibration of a WIM system by quantifying the effect of sample size (truck runs), speed, temperature, and truck type on measurement errors, (2) model gross vehicle weight (GVW) WIM errors as a function of individual axle errors [(single axle (SA) and two tandem axles (TA), (drive and trailer)], and (3) estimate WIM measurement errors using the LTPP and the ASTM protocols. The research objectives were accomplished by synthesizing and analyzing the WIM error data available in the LTPP database for bending plate (BP) and quartz piezo (QP) sensors.

Successful WIM equipment calibration can eliminate systematic weights, speed, and axle spacing errors. The ASTM and the LTPP accuracy estimation methods agree; however, the methods should be compared when the sample size is small (10 or fewer truck runs). The representative pre and post-calibration data can be collected accurately using one test truck with 12 or more runs at multiple speed points for QP and BP sensors. The developed models showed that the GVW errors could be accurately predicted using SA and two TAs. The results also show that if GVW errors are within ASTM Type I tolerance, the SA and TA errors will likely be within acceptable limits. Therefore, calibrating a WIM site using GVW errors should continue. The suggested changes in current WIM procedures can significantly reduce time and resources for successful equipment calibration. The preliminary models developed in this study can be validated in the field and improved further by adding more data in the future.

CHAPTER 7 ESTIMATION OF VEHICLE PAYLOAD FROM GVW DATA

7.1 INTRODUCTION

The freight transportation system in the United States contributes significantly to the country's economy, security, and quality of life. Strategic, operational, and investment decisions by governments at all levels will be necessary to maintain freight system performance and requires sound technical guidance based on research. The National Cooperative Freight Research Program (NCFRP) highlighted that the quality and extents of freight data are important for freight demand models to support public sector decision-making [75]. A region's economy substantially benefits from increased intra, and inter-regional freight flows between trading partners and intermodal centers. Freight generation and movement patterns are not well understood by planners and policymakers tasked with making complex strategic land use and transport planning decisions [76-78]. A study by Hwang et al. reported that the commonly used inputs for freight and regional travel demand and emission models include vehicle miles traveled (VMT), payload by commodity type, and vehicle loading. The study also documented that according to the commodity flow survey (CFS) of 2012, 71% and 73% of total goods by weights and values are transported by trucks compared to other mode shares (rail, water, air, pipeline, etc.). Also, FHWA Class 9 (5-axles semis) trucks accounted for 54% of the truck volumes collected in the 2017 Travel Monitoring Analysis System (TMAS). This study also presented an approach to calculating the average payload of loaded trucks by subtracting the empty trucks' estimated average GVW from the loaded trucks' average GVW. This report had a significant limitation: no data source was available to compare and validate the estimated payloads. A Gaussian mixture model (GMM) procedure is adopted to compare the weights of empty and loaded trucks and average payloads [79, 80]. This study computed average payloads by subtracting the average empty weight from the loaded weights using data from 4 WIM stations. The findings reported differences in estimated weight values (loaded, empty, and payload) from WIM data and the values obtained from the National and California Vehicle Inventory Use Survey (VIUS). Luis et al. applied a similar technique to Tandem axle distributions to improve the characterization of the axle load spectra for pavement design [81].

The State Departments of Transportation (DOTs) use several methods and data sources for freight tonnage estimation. A study in Florida thoroughly reviewed truck tonnage estimation methodologies and data sources [82]. The report documented several methods to estimate freight, including but not limited to freight analysis framework (FAF), commodity flow survey (CFS), truck traffic and counts, WIM data, and Origin-Destination Matrix Estimation (ODME). The authors also provided a list of data sources to estimate freight, including Annual Average Daily Truck Traffic (AADTT), American Transportation Research Institute (ATRI), and commercial data sources like Transearch. This study also proposed a new methodology to estimate freight tonnage based on WIM data and compared it with FAF tonnage estimates. The average aggregated tonnage estimated by WIM based method was 22 % and 23% higher than FAF-based methods for the years 2012 and 2017, respectively. Naveen et al. presented a freight data fusion approach by combining FAF and Transearch data to study commodity flow in the spatial domain [83]. The results also covered empty truck flow generation using WIM data and the origin-destination matrix. A study in Manitoba evaluated the use of portable WIM systems to collect axle weights for several applications, including pavement design and traffic patterns [84]. This study estimated and validated the weights of empty, partially loaded, and loaded trucks based on GVW WIM data. However, no discussion was available related to freight tonnage. Daniel et al. successfully demonstrated using WIM data to study temporal analyses of freight in Southern

California. The authors presented models to estimate the average GVW per year in metric tons based on data from 22 WIM sites [85]. However, the estimates were not validated against actual freight at these locations. Finally, a study completed in Florida used Class 9 GVW average values for partially loaded, full trucks, and empty trucks to investigate truck empty backhaul issues [86]. The empirical analyses resulted in strategies to help address the empty backhaul issue and improve Florida freight mobility and trade plan.

Most studies discussed above used the WIM data to get valuable information about freight movement. However, the freight tonnage estimates were not validated by other data sources, except in a few studies [80, 82, 83]. The main reason is the non-availability of adequate data sources in the public domain: additional costs and labor limit freight data monitoring, recording, and reporting regularly. Therefore, there is a need to develop a cost-effective and easily implementable approach to get general freight trends on highways and other state routes. The Long Term Pavement Performance (LTPP) database traffic module contains detailed gross vehicle weight (GVW) data for different FHWA truck classes. The data summaries are available by truck class daily, monthly, and yearly [73, 87]. This data can be used to identify empty, partially loaded, and fully loaded trucks based on the GVW reference weight ranges. Subsequently, this information can be validated with the freight and commodity survey data gathered by the state DOTs. The results can be incorporated into freight demand models to make informed transportation policy decisions. An effort is made to formulate a procedure to get freight information from WIM data, considering the freight data limitations. This study estimated freight tonnage (vehicle payload) from WIM data and validated it using Transearch data from IHS Markit. In addition, this research evaluates the feasibility of using the LTPP WIM data to estimate freight tonnage over time. The analysis presents freight estimates using the LTPP WIM data from the states of Michigan, Ohio, and Washington ranging from 1997 to 2020 (23 years).

7.2 OBJECTIVES

This chapter further extends applications of WIM data to address an important issue related to freight data, i.e., how to estimate freight tonnage and classify commodities based on GVW WIM data. The methodology uses GVW loading data to estimate vehicle payload and commodity type. The primary objectives of the research are to provide (a) a review of Michigan freight and GVW data, (b) an estimation of freight tonnage from GVW data, (c) a methodology to classify freight commodities based on GVW data, and (d) feasibility of potential application using the LTPP case studies. These objectives were accomplished by synthesizing and analyzing the freight and GVW loading data from the Michigan Department of Transportation (MDOT). Further, the models' adequacy and potential applications were assessed using the GVW WIM data for three LTPP sites.

7.3 DATA USED FOR ANALYSES

The authors obtained the GVW loading and freight data from the Michigan Department of Transportation (MDOT). The MDOT acquired freight and commodity type from the Transearch data from IHS Markit. The shape files containing freight and location information were processed in Quantum Geographic Information System (QGIS) software to obtain total tonnage and tonnage per commodity for a year. All the available sites were assigned a unique ID by combining county and route IDs. In addition, MDOT also provided GVW distributions for all truck classes based on available WIM stations on the same route for the same year as freight data. The county and road IDs were matched to correlate the GVW and freight information.

Table 7-1 and Figure 7-1 show the distribution of sites for GVW and freight data. In summary, 35 sites were available to analyze freight and GVW data.

Table 7-1 Detail of available WIM sites.

Station for GVW	County	Located on route	Functional class	Transearch freight data
03-7319	Allegan	I-196	Interstate	Available
09-6429	Bay	I-75	Interstate	Available
11-7189	Berrien	I-94	Interstate	Not available
12-7269	Branch	I-69	Interstate	Available
13-7159	Calhoun	I-94	Interstate	Not available
13-7169	Calhoun	I-94	Interstate	Available
19-5019	Clinton	US-127	Freeway and Expressway	Available
19-5319	Clinton	I-96	Interstate	Available
21-1459	Delta	US-2	Other Principal Arterial	Available
21-2229	Delta	US-2	Other Principal Arterial	Available
22-1199	Dickinson	M-95	Other Principal Arterial	Available
23-8869	Eaton	I-69	Interstate	Available
25-6119	Genesee	I-75	Interstate	Available
25-6449	Genesee	I-69	Interstate	Available
30-8129	Hillsdale	US-12	Other Principal Arterial	Available
33-8029	Ingham	US-127	Freeway and Expressway	Available
38-7029	Jackson	I-94	Interstate	Available
38-7049	Jackson	US-127	Other Principal Arterial	Not available
40-3069	Kalkaska	US-131	Other Principal Arterial	Available
41-9759	Kent	M-6	Freeway and Expressway	Available
47-8049	Livingston	I-96	Interstate	Available
49-2029	Mackinac	US-2	Other Principal Arterial	Available
58-8729	Monroe	US-23	Freeway and Expressway	Available
61-5289	Muskegon	US-31	Freeway and Expressway	Available
69-4049	Otsego	I-75	Interstate	Available
70-5059	Ottawa	I-196	Interstate	Available
70-5099	Ottawa	I-196	Interstate	Available
72-4129	Roscommon	US-127	Freeway and Expressway	Not available
72-4149	Roscommon	I-75	Interstate	Available
75-2199	Schoolcraft	M-28	Other Principal Arterial	Available
77-6369	Saint Clair	I-69	Interstate	Available
77-6469	Saint Clair	I-94	Interstate	Available
78-7119	Saint Joseph	US-131	Other Principal Arterial	Available
80-7219	Van Buren	I-94	Interstate	Available
81-8239	Washtenaw	US-23	Freeway and Expressway	Available
82-8839	Wayne	I-94	Interstate	Available
82-9189	Wayne	I-275	Interstate	Available
82-9699	Wayne	I-75	Interstate	Available

7.3.1 Overview of Freight Data

The total freight tonnage for different commodities was visualized first. Most of the sites contained multiple records for freight tonnage. Therefore, the representative freight statistics, including minimum, average, and maximum freight values, were calculated for each location.

The freight list contained information on 32 different commodities. However, the heat map and pie charts in Figures 7-2 and 7-3 show the details of the top 5 commodities for each site. Each county and road exhibited a unique distribution of freight. The predominating commodities in different counties were food and farm products, ores and minerals, petroleum products, logs and lumbers, chemical products, transportation equipment, and waste materials. Overall, the available data had farm products, food products, and nonmetallic ores and minerals as the top 3 commodities [see Figures 7-2 and 7-3(a)]. The trends varied for individual sites; for example, nonmetallic ores and logs/lumber products represent the maximum tonnage on M-95, Dickinson, and US-2, Mackinac Counties, respectively [see Figures 7-3(c) and (d)].

Figure 7-4 shows the relationship between different freight statistics computed for each site. The average freight shows a strong correlation (R^2 >0.86) with minimum and maximum freight values [see Figures 7-4(a) and (b)]. The sites' maximum and minimum freight values show a weaker relationship [see Figure 7-4(c). Therefore, this research assesses the GVW relationship with average freight values only. Figure 7-4(d) presents different routes' average and maximum freight information. The results show that the maximum freight travels on interstates I-75 and I-94 within Michigan, whereas; US-2, US-12, and US-131 carry minimum cargo.

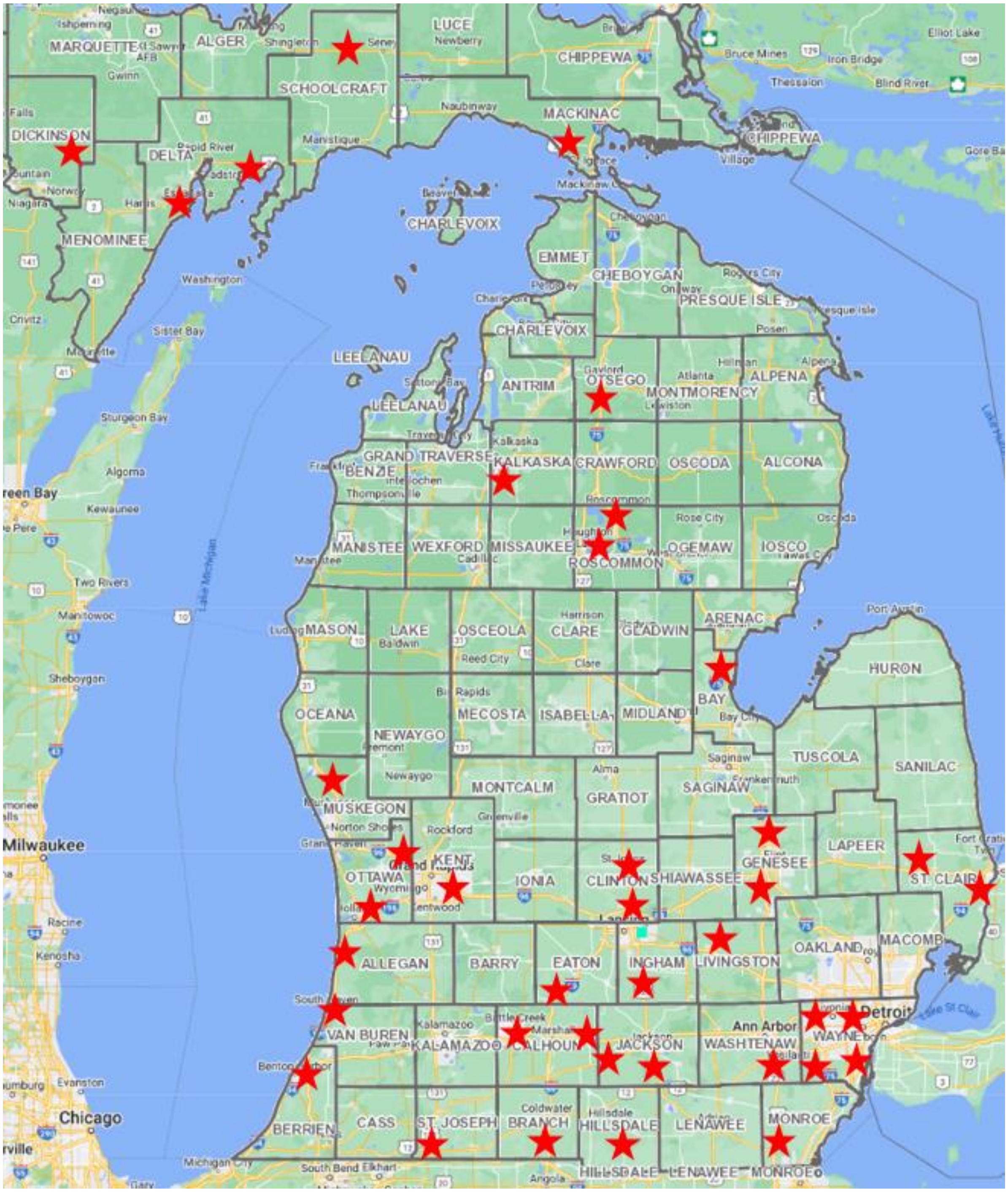


Figure 7-1 Location of available WIM sites.

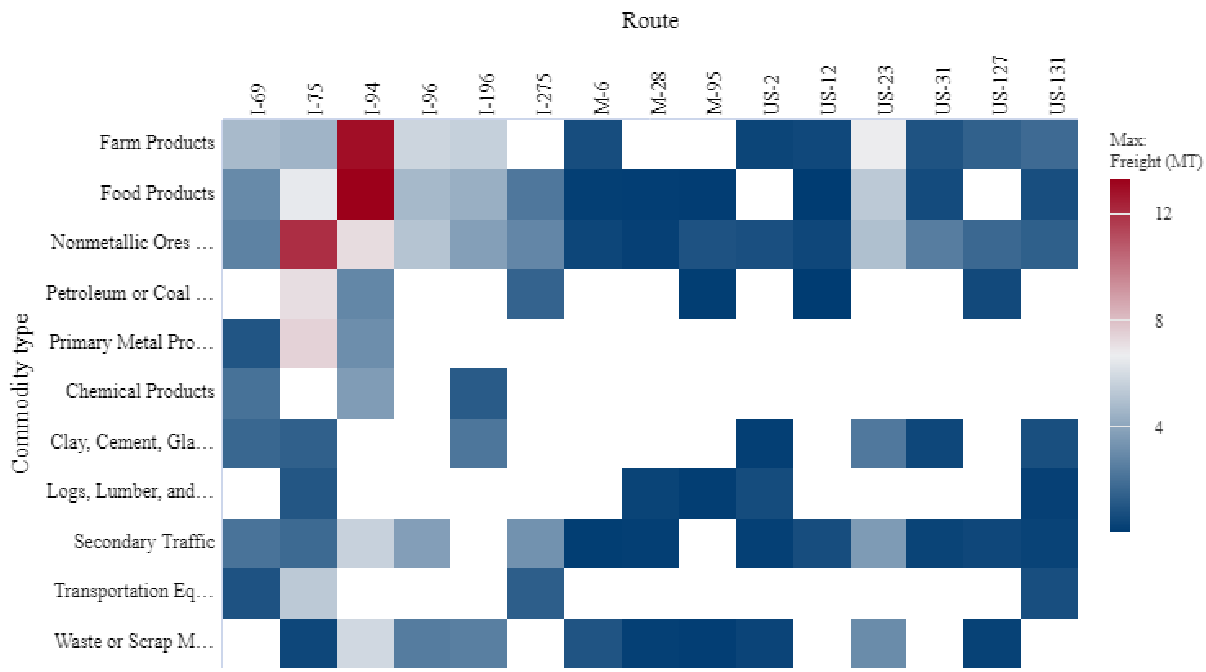
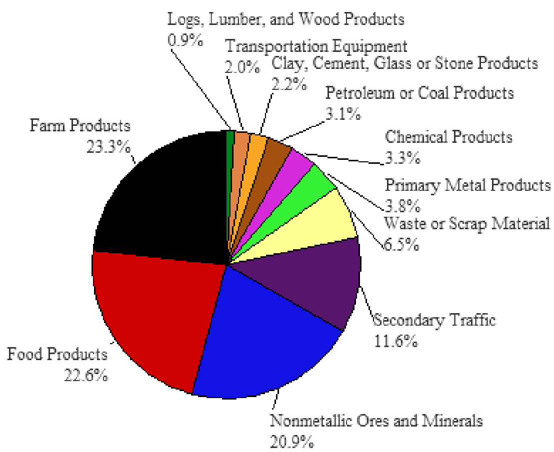
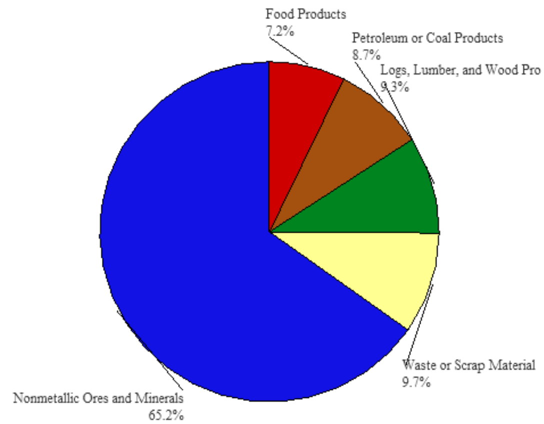


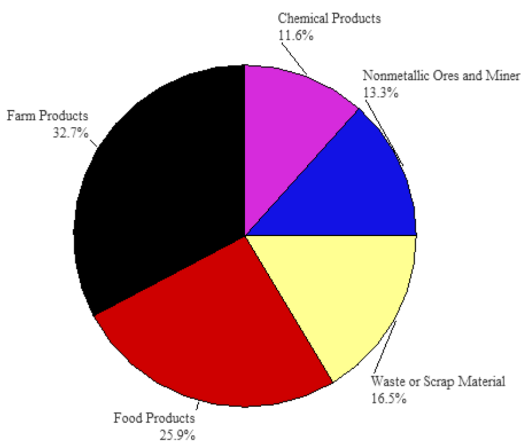
Figure 7-2 Heat map for freight data by commodity type and route.



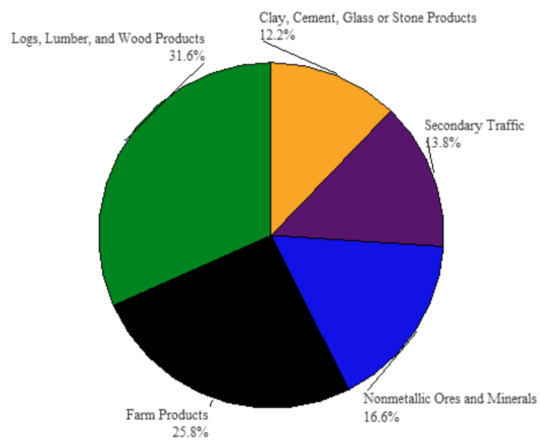
(a) Overall (based on 35 sites)



(c) Location 22-95

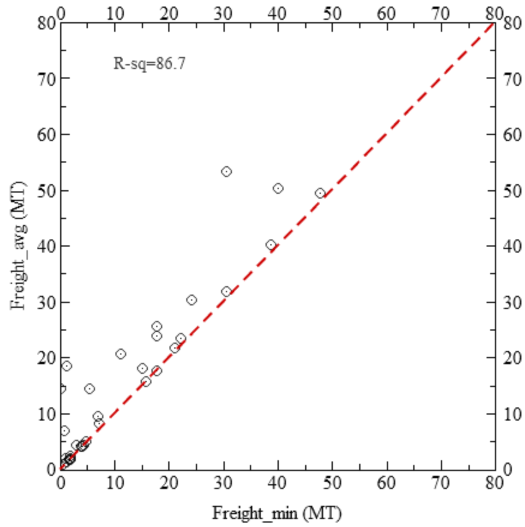


(b) Location 11-94

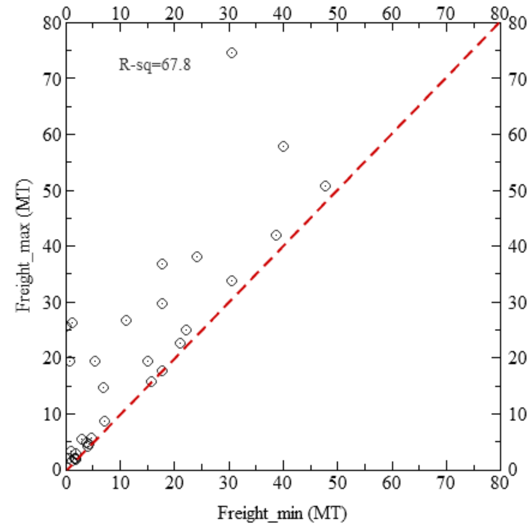


(d) Location 49-2

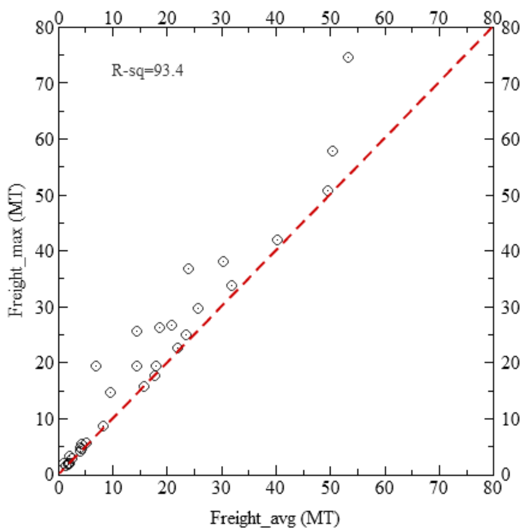
Figure 7-3 Freight data for different predominating commodities.



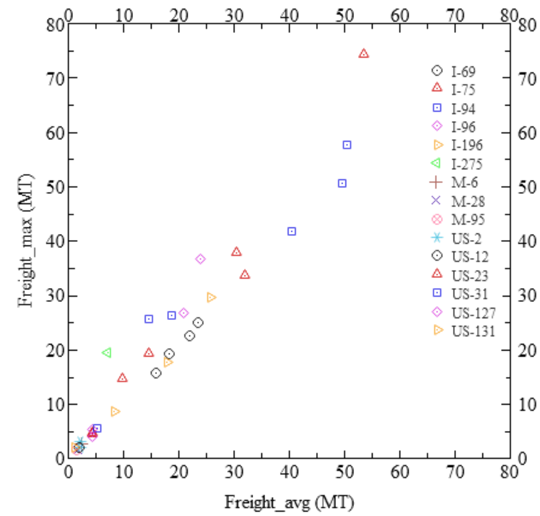
(a) Minimum freight vs. average freight relationship



(c) Minimum freight vs. maximum freight relationship



(b) Average freight vs. maximum freight relationship



(d) Average vs. maximum freight relationship (by route)

Figure 7-4 Relationship between freight data statistics.

7.3.2 Axle Loading Data for Gross Vehicle Weight

The GVW data were available on a monthly and yearly basis. The annual data were analyzed because the freight data were available yearly. The investigation used one year (i.e., 2018) of GVW data from 35 WIM stations. Each datasheet contained information on WIM ID truck class, direction, route functional class, city, and county. The GVW data were quarried separately for

Class 9 trucks and all other truck classes (4 to 13 excluding Class 9). The GVW data contained 41 bins at 3 kip intervals ranging from the smallest and the largest bins of 0-3 kip, and 120+ kip, respectively. Figures 7-5(a) and (b) show the GVW data for Class 9 trucks and all other truck classes. The GVW data for Class 9 trucks show two prominent peaks. The first and second peaks occur approximately from 24 to 36 Kip, and 68 to 80 Kip, respectively.

7.4 MODELLING OF GVW DISTRIBUTIONS

The GVW distributions were modeled using a set of three distinct distributions, i.e., empty, partially loaded, and fully loaded. Typically, two peak loads are observed in the GVW data. A mixture of statistical distributions was considered to characterize the predominantly bimodal axle load spectra [69]. It was shown that two or more normal probability density functions (PDFs) could be added with appropriate weight factors to obtain the PDF of the combined distribution, as shown by Equation 7.1:

$$f^* = \sum_i^n p_i f_i \tag{7.1}$$

Where f^* = PDF of combined distribution, pi= proportions (weight factors) for each normal PDF, and fi= PDFs for each normal distribution.

For a mixture distribution containing three normal PDFs, the three-weight factors are complementary (i.e., p1+ p2+p3 = 1). Haider and Harichandran determined that the shape factors of axle load spectra could be effectively captured by using a combination of the normal distributions:

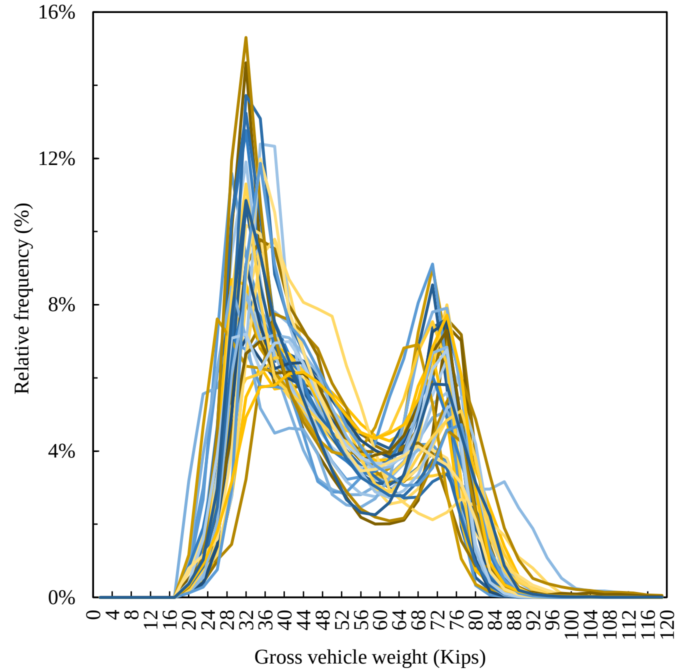
$$f^*\left(x;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,p_1,p_2,p_3\right)=\left(\frac{p_1}{\sigma_1\sqrt{2\pi}}e^{-(x-\mu_1)^2/2\sigma_1^2}+\frac{p_2}{\sigma_2\sqrt{2\pi}}e^{-(x-\mu_2)^2/2\sigma_2^2}+\frac{(1-p_1-p_2)}{\sigma_3\sqrt{2\pi}}e^{-(x-\mu_3)^2/2\sigma_3^2}\right) \tag{7.2}$$

Where $\mu_1 =$ the average for GVW of empty trucks, $\sigma_1 =$ the standard deviation for GVW of empty trucks, $\mu_2 =$ the average for GVW of partially loaded trucks, $\sigma_2 =$ the standard deviation for GVW of partially loaded trucks, $\mu_3 =$ the average for GVW of fully loaded trucks, and $\sigma_3 =$ the standard deviation for GVW of fully loaded trucks, $p_1, p_2, p_3 =$ weights of the three probability distributions.

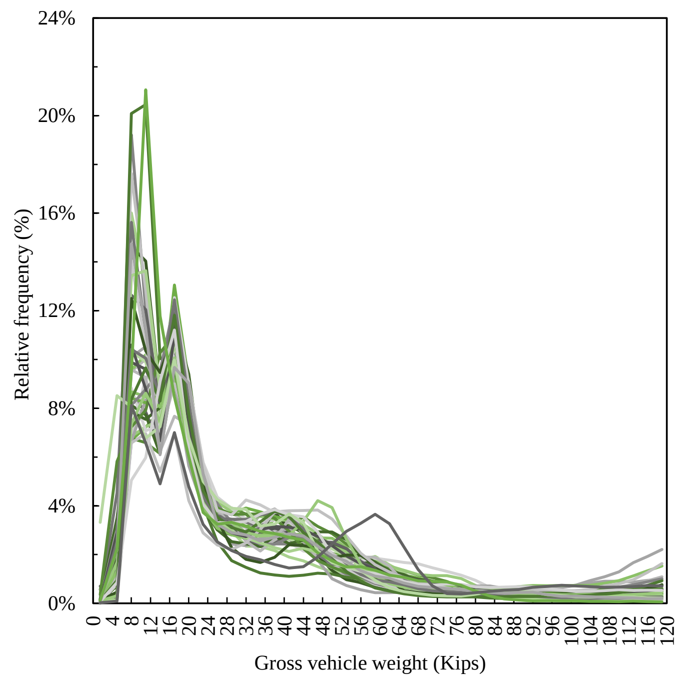
Figures 7-6(a) and (b) show an example of the observed and fitted GVW distribution and individual distributions for one WIM station in Clinton County.

7.5 PROCEDURE FOR RELATING WIM-BASED GVW PAYLOAD WITH TRANSERACH FREIGHT

This section presents the procedure to estimate freight tonnage based on GVW data for Class 9 and other truck classes. The data selection, analyses, and model development process are explained with the help of a flow chart (see Figure 7-7). Different statistical techniques were used to identify the most significant variables, including scatter plots, correlation, linear, non-linear, and multiple linear regression. The final models developed to estimate freight based on GVW data are presented next.

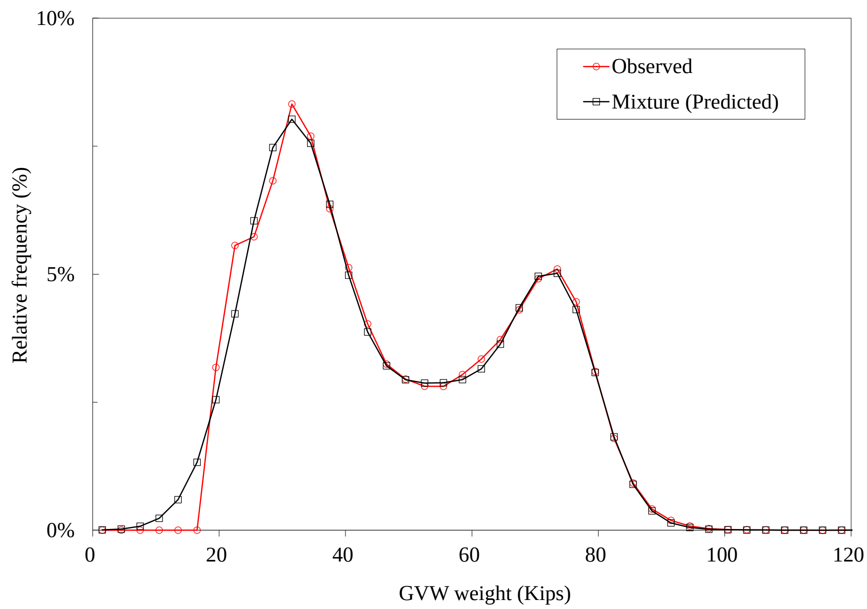


(a) GVW weights for Class-9 trucks

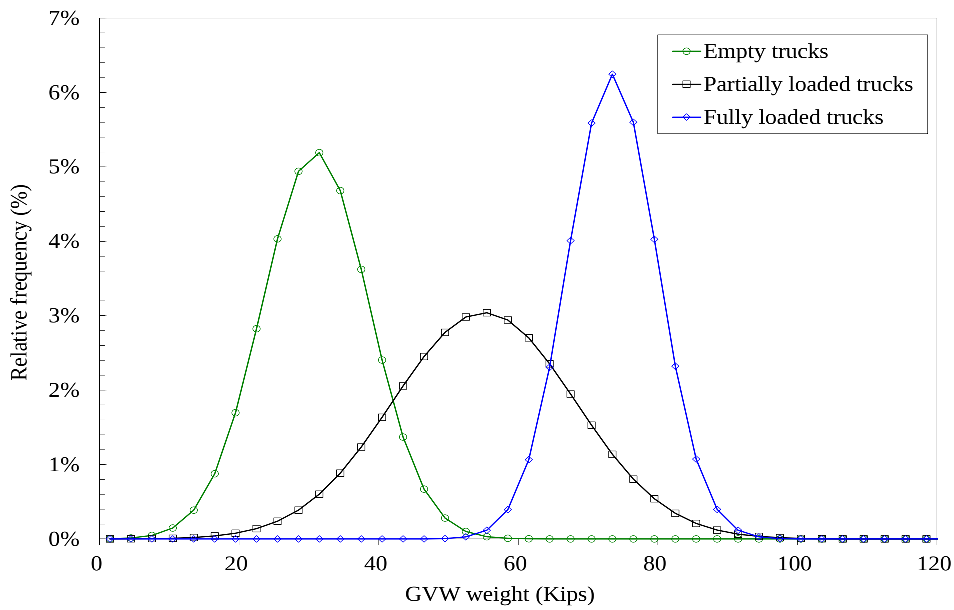


(b) GVW weights for all other trucks (Class 4 to 13 excluding Class 9)

Figure 7-5 GVW weight data from different Class 9 and other trucks (2018, Michigan).



(a) Example of GVW distribution fitting (Station 19-5019, Clinton County, US-127)



(b) Individual distributions (Station 19-5019, Clinton County, US-127)

Figure 7-6 GVW weight data - example of GVW distribution fitting.

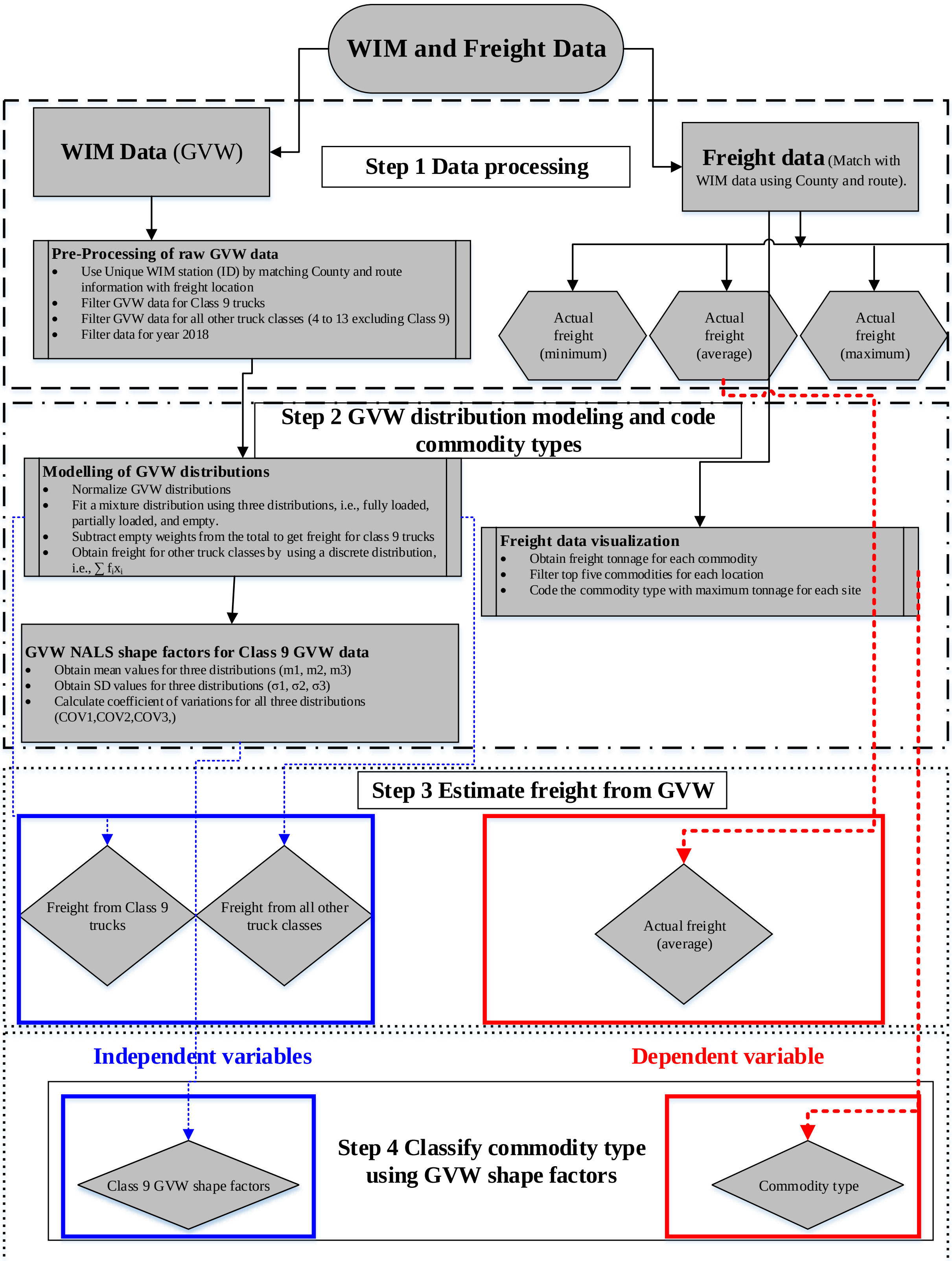


Figure 7-7 Flowchart for GVW and freight data analyses.

Equations 7.3 to 7.6 were used to estimate vehicle payload from GVW data for Class 9 and other truck classes.

$$Total\ GVW\ load = \sum f_i \times x_i \times N \tag{7.3}$$

where:

Total GVW load = Total load for GVW mixture distribution

f_i = Noramalized frequency for GVW mixture distribution

x_i = mid point of *i*th bin

N = Total counts for Class 9 trucks

$$Empty\ GVW\ load = \sum f_i \times x_i \times N \tag{7.4}$$

where:

Empty GVW load = GVW load carried by empty trucks

f_i = Noramalized frequency for GVW empty truck distribution

x_i = mid point of *i*th bin

N = Total counts for Class 9 trucks

$$Pay\ load_{Class9} = Total\ GVW\ load - Empty\ GVW\ load \tag{7.5}$$

where:

Pay $\text{load}_{\text{Class 9}}$ = Freight carried by Class 9 trucks

$$Pay\ load_{Others} = \sum f_i \times x_i \tag{7.6}$$

where:

Pay $\text{load}_{\text{others}}$ = Freight carried by other trucks

f_i = Frequency for GVW distribution (other trucks)

x_i = mid point of *i*th bin

The model to estimate freight tonnage uses dependent and independent variables presented in Step 3 of Figure 7-7. The presented model estimates freight as a function of Class 9 GVW payload and GVW load for other truck classes. The tonnage computed based on the Class 9 truck's GVW data strongly correlated with actual freight average values [see Table 7-2]. Equation 7.7 shows the model based on the payload computed from Class 9 trucks. The review of model diagnostics highlighted one unusual observation that was deleted. This point showed a significantly large residual as compared to other data points. Equation 7.8 shows the freight estimation model that contains payloads for both, i.e., Class 9 and all other truck classes. Although the second term was significant (p-value <0.05), its contribution to explaining model variance was negligible. Equation 7.9 shows the model developed as a function of GVW freight for other truck classes only. This model is not a very good fit for estimating freight. Figures 7-8(a) to (c) show the goodness of fit for the freight estimation models (Equations 7.7 to 7.9). The deletion of unusual observations slightly improved the model goodness of fit with approximately similar regression coefficients.

Further, no issue was found with regression assumptions [see Figures 7-9(a) to (c)]. The developed methodology has the potential for immediate application. This can be applied to estimate freight at any route, provided GVW WIM data are available.

Table 7-2 Correlation between GVW and average freight.

Variable	Pearson correlation	Spearman correlation	p-value
GVW freight_CL9 vs Freight_avg	0.879	0.868	<<0.005
GVW freight_others vs Freight_avg	0.446	0.583	<<0.005
GVW freight_CL9 vs GVW freight_others	0.694	0.806	<<0.005

$$Freight\_avg(MT) = 1.937 + 1.055\ GVW\ freight\_CL9$$
$$(R^2 = 84.57, N = 34)$$

where: (7.7)

Freight_avg(MT)=Average freight in Mega ton

GVW freight_CL9=Freight (paylaod) calculated from GVW distributions for class 9 trucks

$$Freight\_avg(MT) = 4.30 + 1.244\ GVW\ freight\_CL9 - 0.608\ GVW\ freight\_others$$
$$(R^2 = 86.91, N = 34)$$

where:

Freight_avg(MT)=Average freight in Mega ton

GVW freight_CL9=Freight (payload) calculated from GVW distributions for class 9 trucks

GVW freight_others=Freight (paylaod) calculated from GVW distributions for other truck classes

(7.8)

$$Freight\_avg(MT) = 3.43 + 1.541\, GVW\ freight\_others$$
$$(R^2 = 32.58, N = 34)$$

where:

Freight_avg(MT)=Average freight in Mega ton

GVW freight_others=Freight (payload) calculated from GVW distributions for other truck classes

(7.9)

7.6 CLASSIFYING FREIGHT COMMODITIES FROM GVW DATA

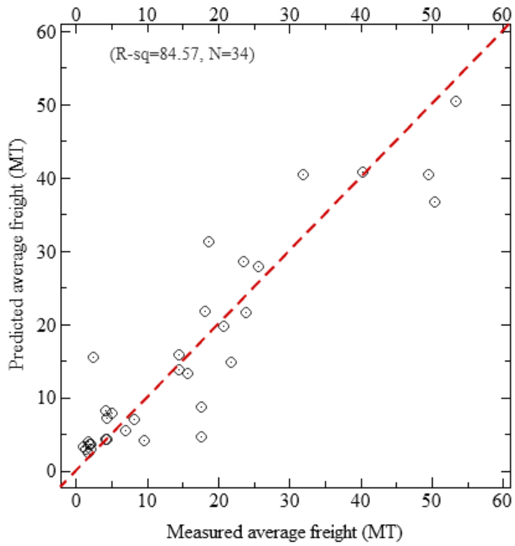
This section presents the procedure to classify freight commodities based on normalized GVW shape factors for Class 9 trucks. Because it was not possible to model their GVW distributions for other trucks, those were not considered. Table 7-3 presents the normalized GVW shape factors considered for analyses. The partially loaded trucks SD is the largest among the three groups. In this analysis, an attempt is made to classify freight commodity types based on shape factors as predictors. The freight commodities with maximum tonnage for each site were grouped into four classes. The four classes contain the following freight commodities:

Class 1: farm and food products (counts: 19)

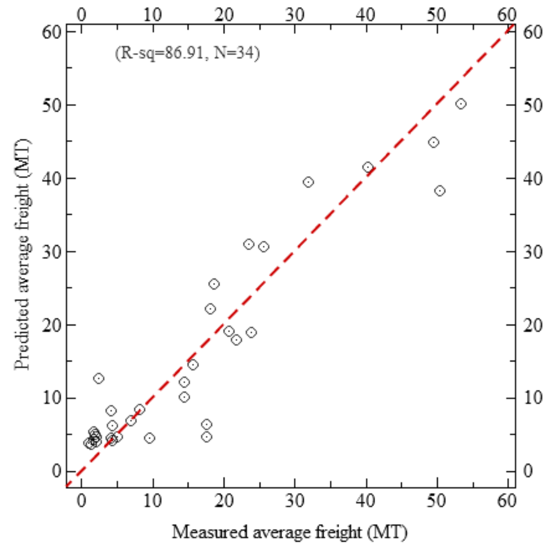
Class 2: nonmetallic ores and minerals and waste or scrap material (counts: 8)

Class 3: logs, lumber, and wood products (counts: 2)

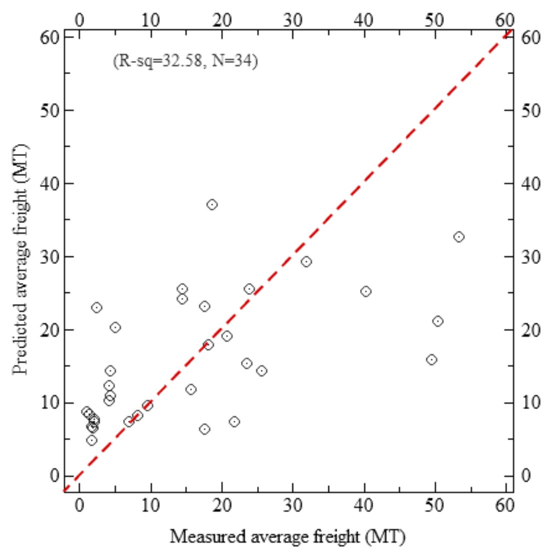
Class 4: chemical products and secondary traffic (counts: 6)



(a) Goodness of fit (Class 9 only)

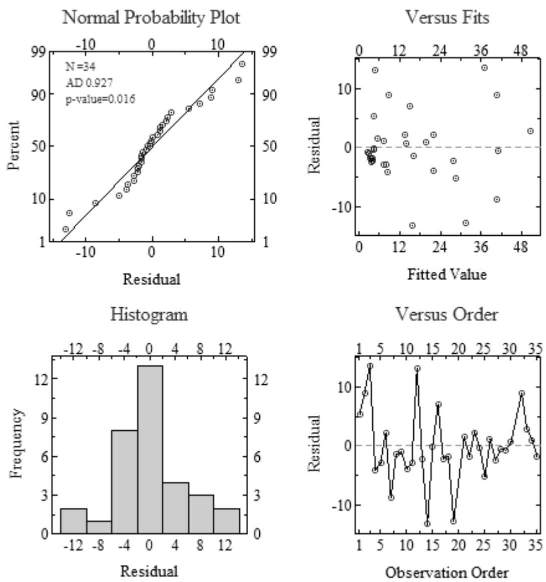


(b) Goodness of fit (Class 9 and others)

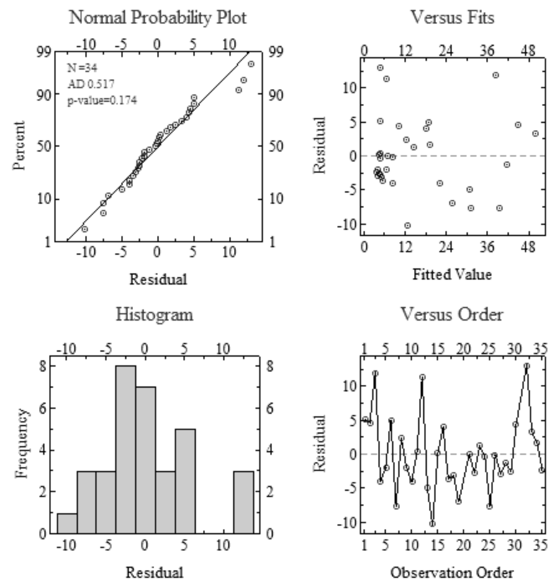


(c) Goodness of fit (Other truck classes only)

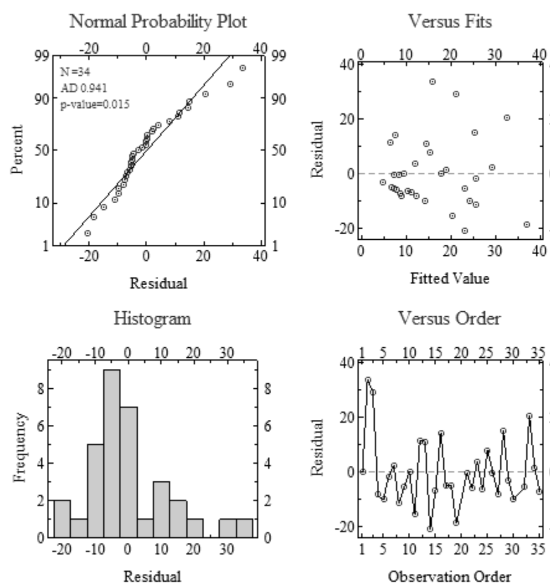
Figure 7-8 Goodness of fit for freight estimation models.



(a) Residuals for model (Equation 7.7)



(b) Residuals for model (Equation 7.8)



(c) Residuals for model (Equation 7.9)

Figure 7-9 Diagnostics for freight estimation models.

Table 7-3 GVW shape factors for Class 9 trucks.

GVW shape factors	Symbol	Min.	Max.	Mean	Average values from the literature				Units
					[88]	[89]	[86]	[13]	
Empty trucks distribution	m1	26.1	35.7	31.6	34.2	33	<40	32.2	kips
Partially loaded trucks distribution	m2	37.3	55.1	43.6	-	46	40 - 60		
Fully loaded trucks distribution	m3	62.2	74.9	70.8	-	67	>60	82.7	
Empty truck loads	σ1	2.4	7.7	3.7	-	-	-	1.8	
Partially loaded trucks distribution	σ2	7.6	13.1	10.4	-	-	-	-	
Fully loaded trucks distribution	σ3	3.9	16.5	6.9	-	-	-	4.0	
Empty trucks distribution	COV1	0.08	0.25	0.11	-	-	-	0.06	No units
Partially loaded trucks distribution	COV2	0.18	0.27	0.23	-	-	-	-	
Fully loaded trucks distribution	COV3	0.06	0.25	0.09	-	-	-	0.05	

This analysis was conducted using the supervised machine-learning algorithm called classification and regression trees (CART®) Classification. The CART Classification illustrates critical patterns and relationships between a categorical response and continuous or categorical predictors within highly complex data without using parametric methods. The visual representation of the CART regression can make a complex predictive model much easier to interpret [68]. Table 7-4 provides a set of logical rules for classifying freight commodities. These rules can help classify a commodity type by analyzing the available information on GVW shape factors. Figures 7-10(a) and (b) present the relative variable importance and model accuracy based on the CART Classification model. The optimal tree with 7 terminal nodes has a relative misclassification cost of 0.31, i.e., the model can correctly classify 69% of the total events (24/35). The results show that the average of fully and partially loaded trucks are the most important predictors, followed by the average of empty trucks. Other variables listed in Table 7-4 showed an insignificant effect. Figure 7-11 presents the CART decision tree model of this data. The results show that 11 out of 19 (57.9%) events were correctly classified as Class 1. The percentage of correctly classified events for Classes 1 and 2 was 100% (10 out of 10). In contrast, 50% (3 out of 6) events were correctly marked in Class 4. Figure 7-12 presents the receiver operating characteristics (ROC) curves for all four classes. The area under the ROC curve (AUC) is a measure of discrimination; a model with a high area under the ROC curve suggests that the model can accurately predict observation value [90]. All the plots show an AUC > 0.9, indicating that the commodity type can be classified using GVW shape factor information. This analysis considers one class at a time as an event like a binary (event, no event) response. The model also incorrectly classified 31% of the total events. In Class 1, 8 events were misclassified; 3 and 5 events were marked as Classes 2 and 3, respectively. In Class 4, 3 events were misclassified: 1 and 2 as Class-1 and 2, respectively. These results are based on a very small dataset and need careful interpretation. The variability in independent variables was limited because the GVW distributions were similar for most available locations. The number of cases in each sub-class was also limited, especially in Class 3. Additionally, although the top commodity carries a maximum share of freight at a particular location, it does not explain the entire freight pattern. This top commodity is just one out of a list of 32 other commodities part of total freight tonnage. The process presented here is a way of getting valuable information from WIM data. More data can augment the existing findings in the future.

Table 7-4 Decision tree model rules for the commodity classification model.

Variable	Terminal node	m1 (empty) kip		m2 (partially loaded) kip		m3 (loaded) kip	
Class 1	6	-	-	> 45.42		<= 73.28	
	2	<= 31.93		39.50 <	<= 43.77	<= 70.20	
Class 2	3	< = 31.92		39.50 <	<= 43.77	> 70.20	
	7	-	-	> 45.42		> 73.28	
Class 3	5	-	-	43.77 <	<= 45.42	-	-
	1	-	-	< = 39.50		-	-
Class 4	4	> 31.93		39.50 <	<= 43.77	-	-

7.7 POTENTIAL APPLICATIONS TO LTPP WIM DATA - CASE STUDIES

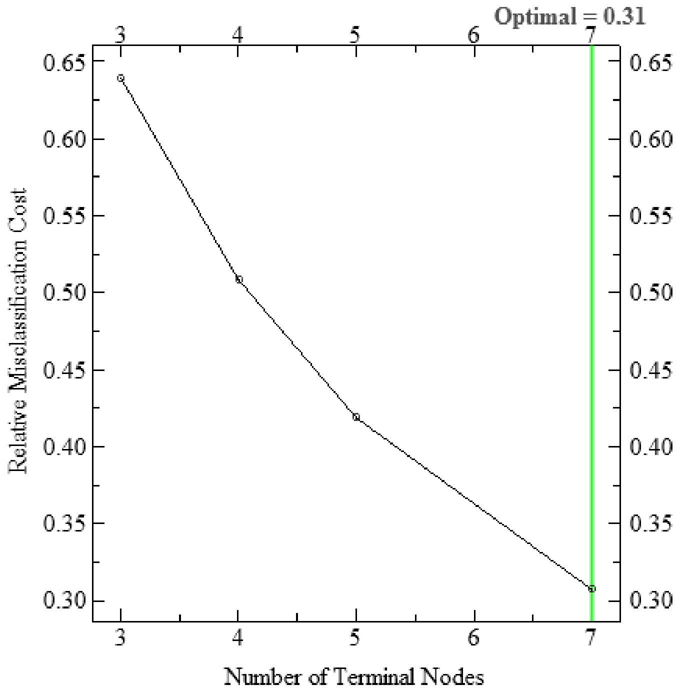
This section presents the feasibility of applying the proposed method to three LTPP WIM sites. Table 7-5 details the relevant LTPP tables used for data extraction. The site details are given below:

- Michigan SPS-1 site located on US-27 (South) Rural Principal Arterial – Other lanes all), Clinton County (26-0113).
- Ohio SPS-1 site located on US-23 (South) Rural Principal Arterial – Other all lane (1), Delaware County (39-0101).
- Washington GPS-6A site located on State Route 167 Urban Other Principal Arterial lanes (all), King County (53-6049).

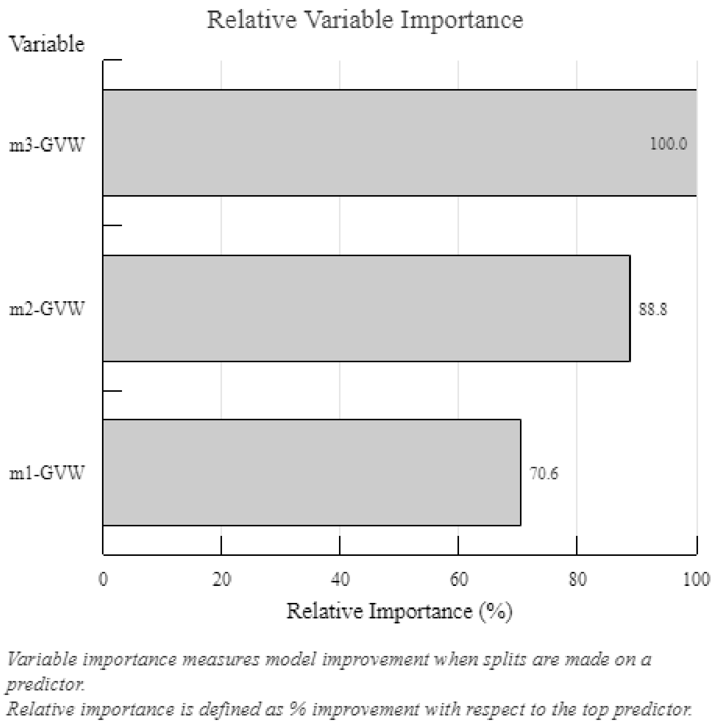
The Michigan site was selected because the Transearch freight tonnage was available closer to this section. The other two sites were chosen because of varying traffic levels and patterns.

Table 7-5 LTPP database tables used to extract data elements.

Type of data	Data elements	Relevant LTPP tables	Table description
General information	LTPP section inventory	EXPERIMENT_SECTI ON	The three key fields that define a unique record in this table are STATE_CODE, SHRP_ID, and CONSTRUCTION_NO.
General information	LTPP Traffic Site Information	SHRP_INFO	This table contains combined data from INV_ID, INV_GENERAL, SPS_ID, SPS_GENERAL, and SPS_PROJECT_STATIONS.
GVW counts	Yearly Aggregate Of GVW	YY_GVW	Gross Vehicle weights are aggregated by vehicle class yearly by day of the week.



(a) Model performance



(b) Relative variable importance

Figure 7-10 CART Classification model performance and variable importance.

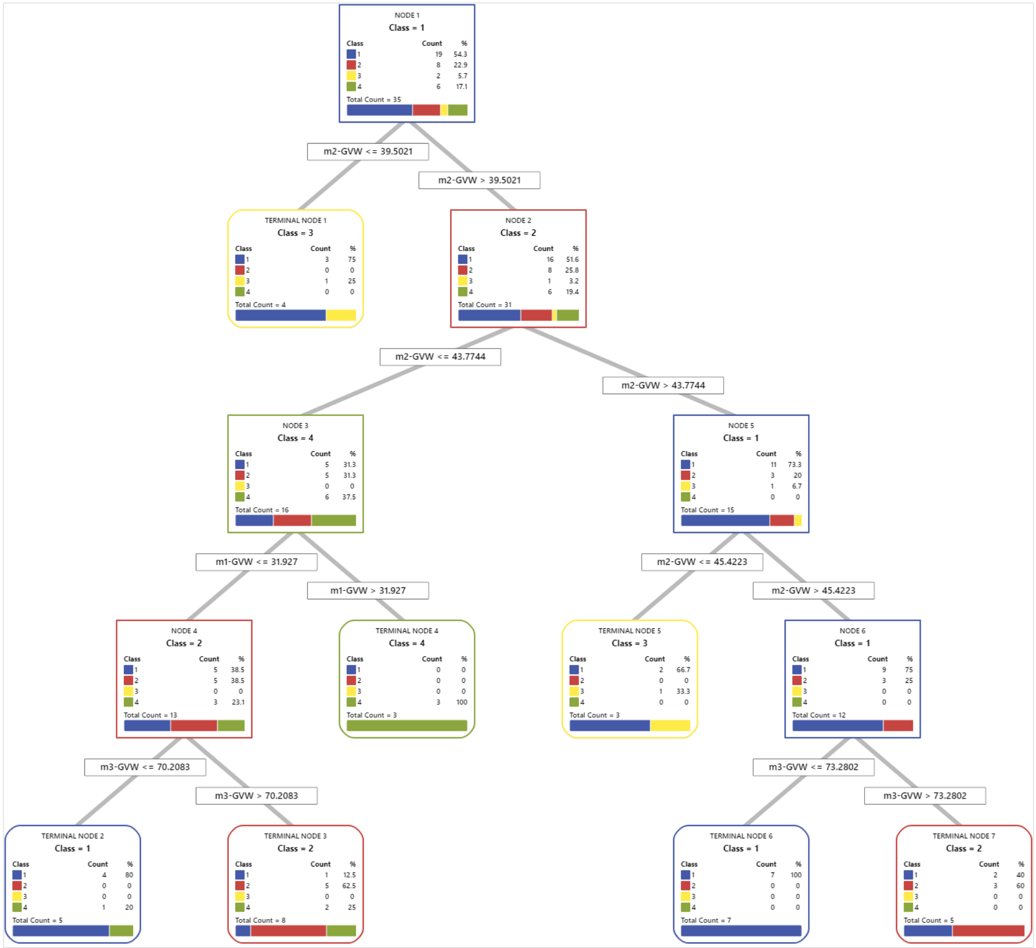
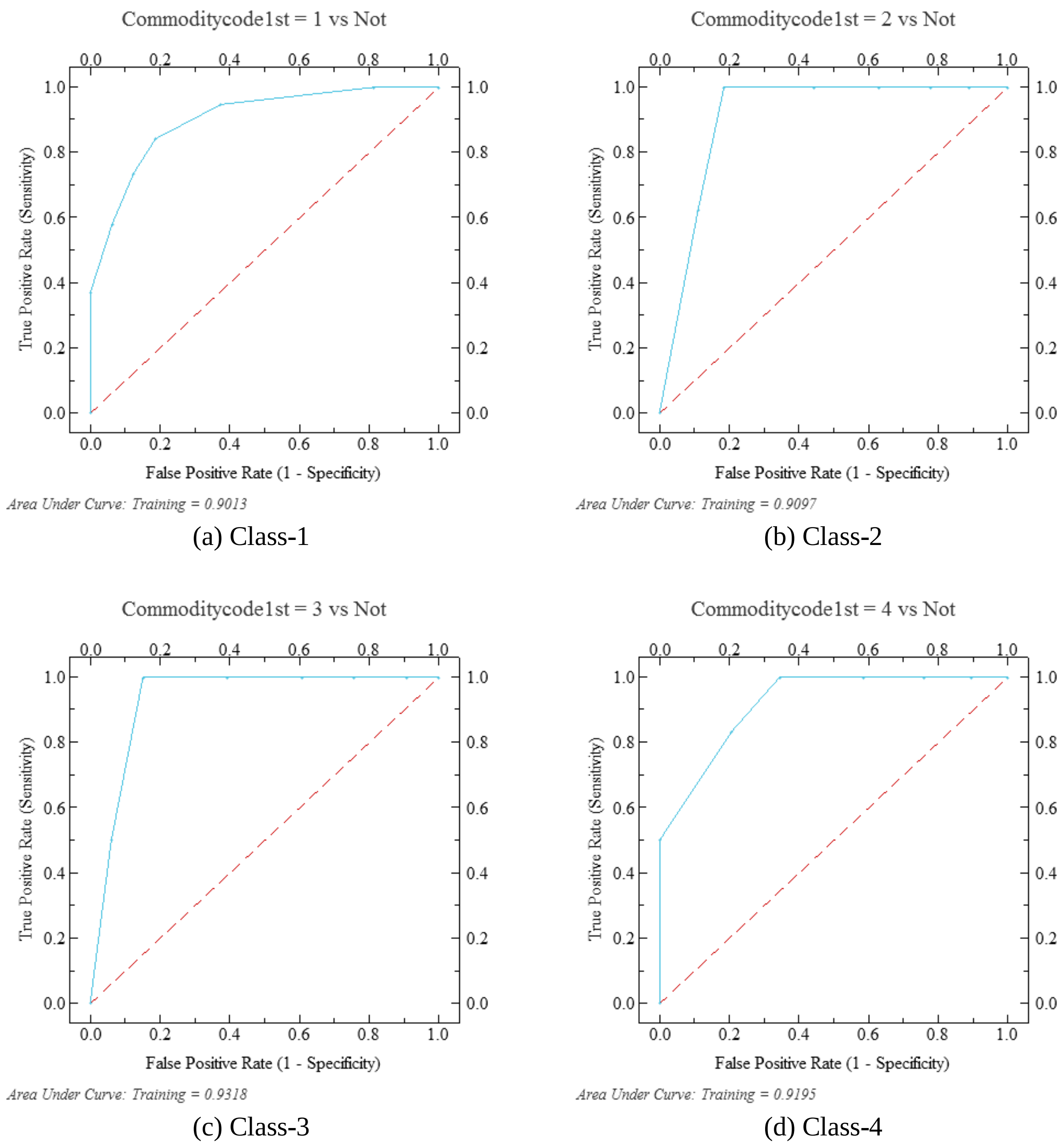


Figure 7-11 CART Classification model Decision Tree.



(a) Class-1 (b) Class-2

(c) Class-3 (d) Class-4

Figure 7-12 ROC Curves for different commodity classes.

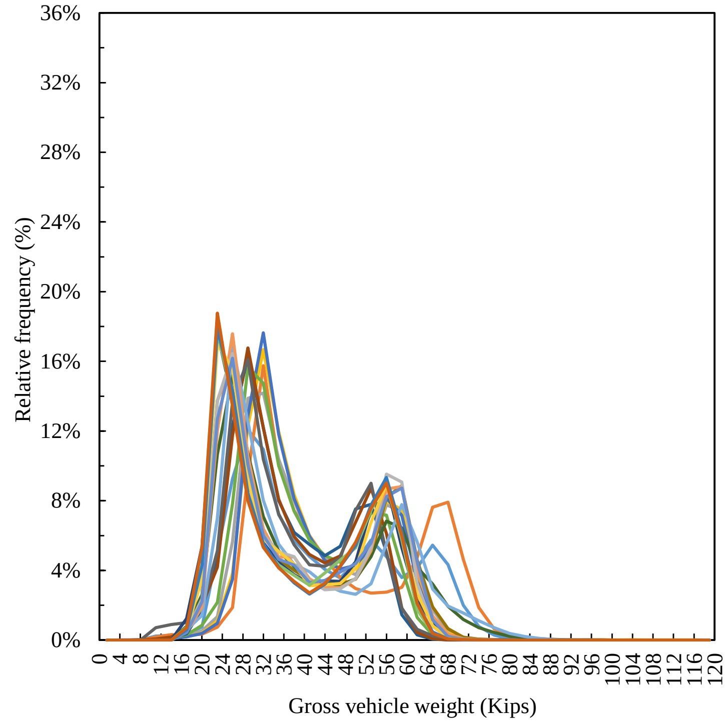
7.7.1 GVW Distributions and Freight Estimates for LTPP WIM sites

Figures 7-13 to 7-15 present the yearly GVW distributions for three LTPP WIM sites. The Class 9 GVW distributions show different shapes at the LTPP WIM sites. The Michigan and Washington sites show a higher percentage of empty trucks, whereas; the Ohio WIM site shows a somewhat similar frequency for empty and fully loaded trucks. The GVW distribution variability over time is also less at the LTPP Ohio WIM site. Figure 7-16 presents the freight for the LTPP WIM sites predicted using Equation 7.7. Also, it shows the count of days in a year. The freight values were consistent over time except for the LTPP-OH WIM site. The primary reason for fluctuations in freight over time was fewer days in a calendar year. Figure 7-17 presents the estimated freight comparisons for the LTPP WIM sites. Figure 7-17(a) compares the Transearch freight tonnage and the LTPP WIM site in Michigan. At this LTPP site, the freight carried by Class 9 trucks was 5.01 MT for the most recent year, i.e., 2016. At the same site, the Transearch average, maximum, and MDOT WIM-based (Class 9) freight for 2018 were 4.22, 5.32, and 4.94 MT, respectively. The results show that the freight estimates from the three data sources are comparable. Figures 7-17(b) to (d) show the freight tonnage predicted for Class 9 trucks only, freight predicted for all trucks, and all trucks excluding Class 9, respectively. The results show that the Class 9 trucks carry the maximum freight at the LTPP WIM sites in Michigan and Ohio [see Figure 7-17(b)]. In contrast, the other truck classes have the maximum contribution at the LTPP WIM site in Washington. The percentage of Class 9 trucks was obtained at all the sites to investigate this pattern (see Figure 7-18). On average, 26, 45, 48, and 71% of the traffic comprises Class 9 trucks for the LTPP-WA, MDOT, LTPP-MI, and LTPP-OH WIM sites, respectively. The findings imply that most freight at the Washington WIM site is carried by other trucks.

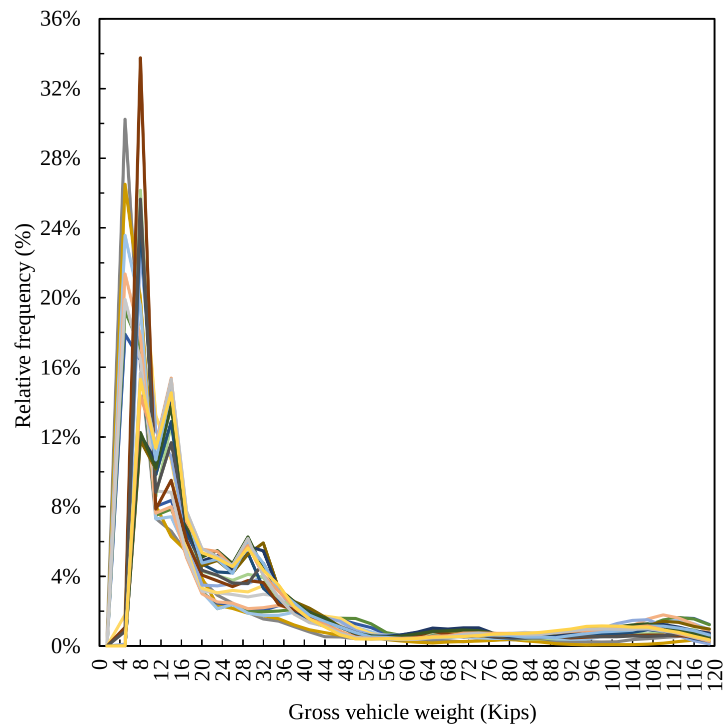
The predictions at the LTPP-WA site also show the Class 9 freight estimation model limitations. This model was developed for the MDOT WIM and Transearch freight data and may not fully capture Washington Class 9 traffic trends. Therefore, a careful evaluation of estimates is recommended, and these should be compared independently against actual freight tonnage where possible. The percentage of Class 9 trucks should be within the range of data for the model development.

7.7.2 Class 9 GVW Shape Factors for the LTPP WIM sites

This final section briefly discusses the GVW shape factors for empty, partially loaded, and fully loaded Class 9 trucks. The comparison data set includes three LTPP and 35 MDOT WIM sites. Figure 7-19 to 7-21 shows the 95% confidence interval plots for individual distributions for GVW mean, SD, and coefficient of variation (COV). Statistically significant differences were observed in GVW mean and SD values at these sites [see Figures 7-19(a) to (c) and 7-20(a) to c)]. The differences are more pronounced for partially loaded and fully loaded trucks. The LTPP-OH and LTPP-WA WIM sites generally show the lowest and highest variability in GVW data. The average loaded peak values in ascending order are 67, 72, 71, and 77 Kips for LTPP-WA, LTPP-MI, MDOT-MI, and LTPP-OH WIM sites, respectively.

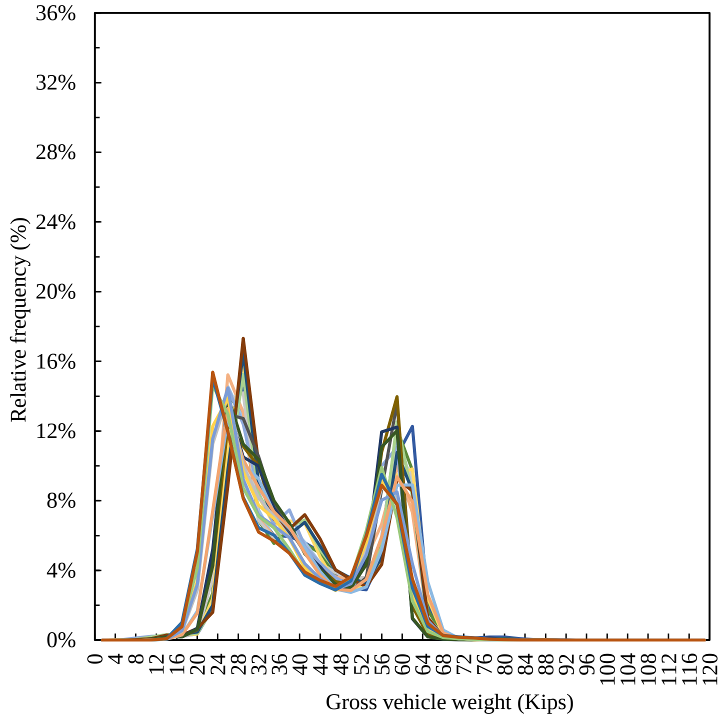


(a) LTPP-MI (Class-9)

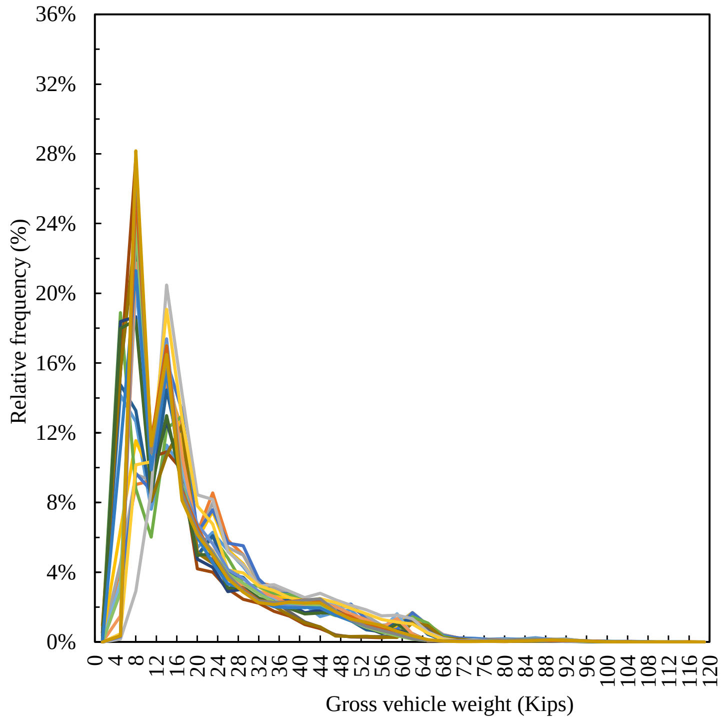


(b) LTPP-MI (Others)

Figure 7-13 GVW weights for Class 9 and other trucks-Michigan.

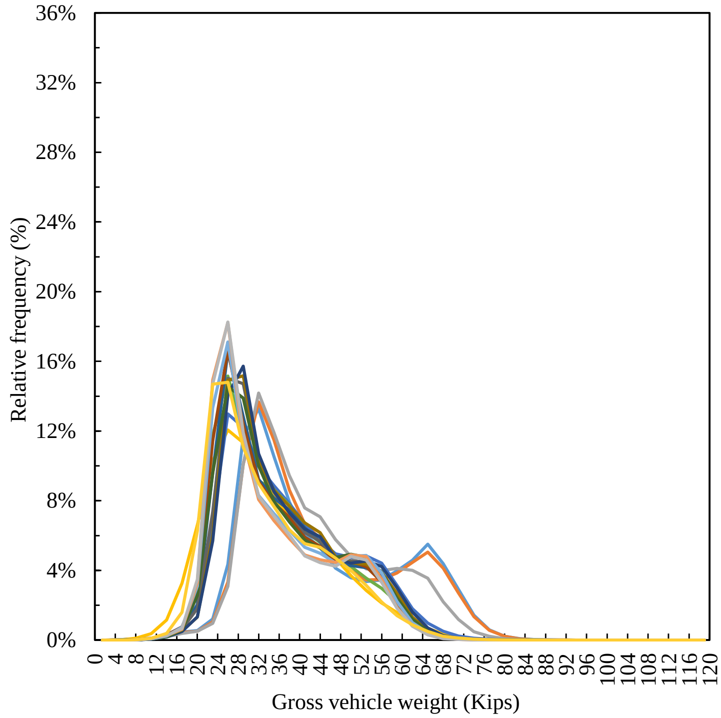


(a) LTPP-OH (Class-9)

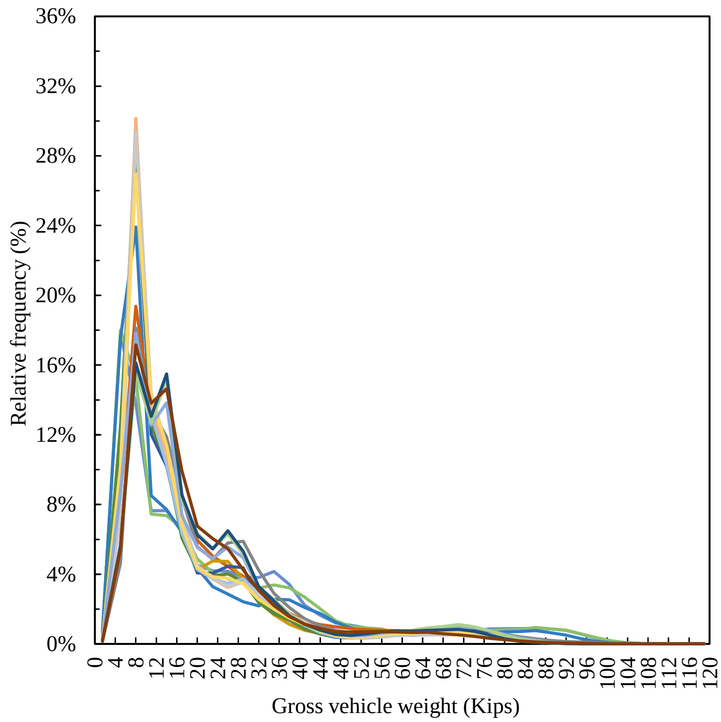


(b) LTPP-OH (Others)

Figure 7-14 GVW weights for Class 9 and other trucks-Ohio.

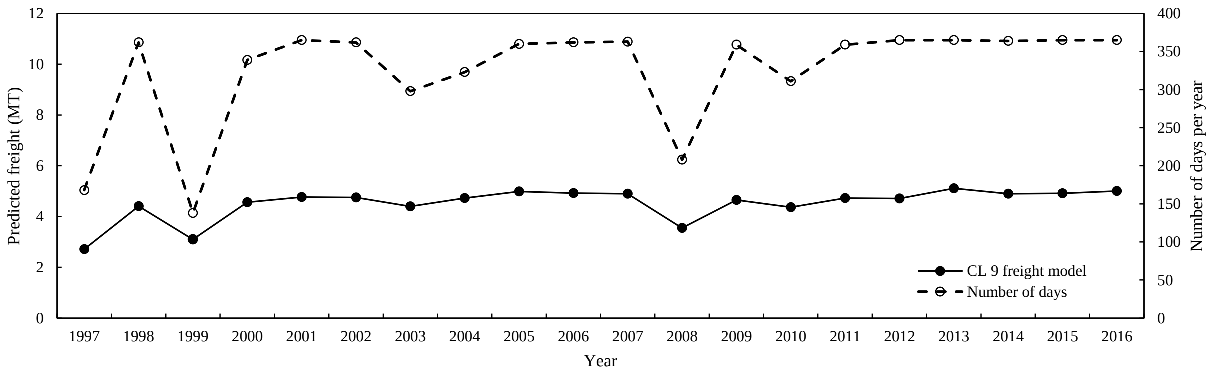


(a) LTPP-WA (Class-9)

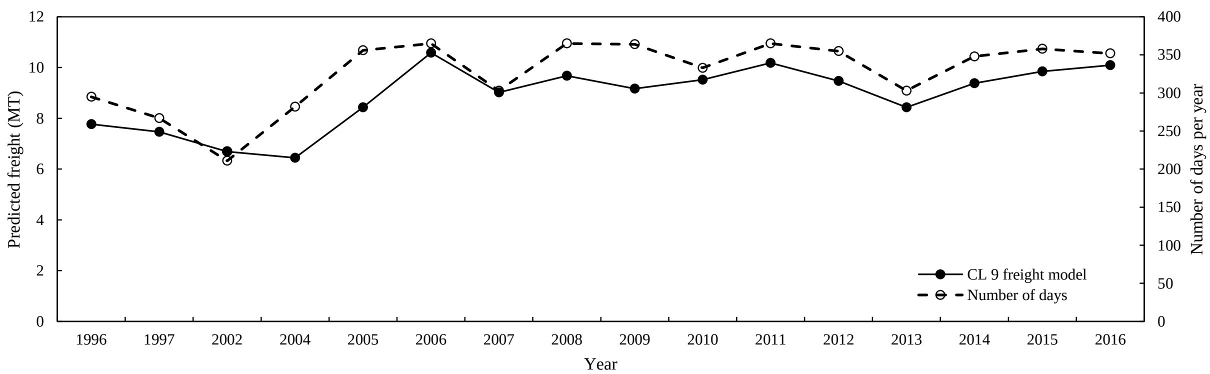


(b) LTPP-WA (Others)

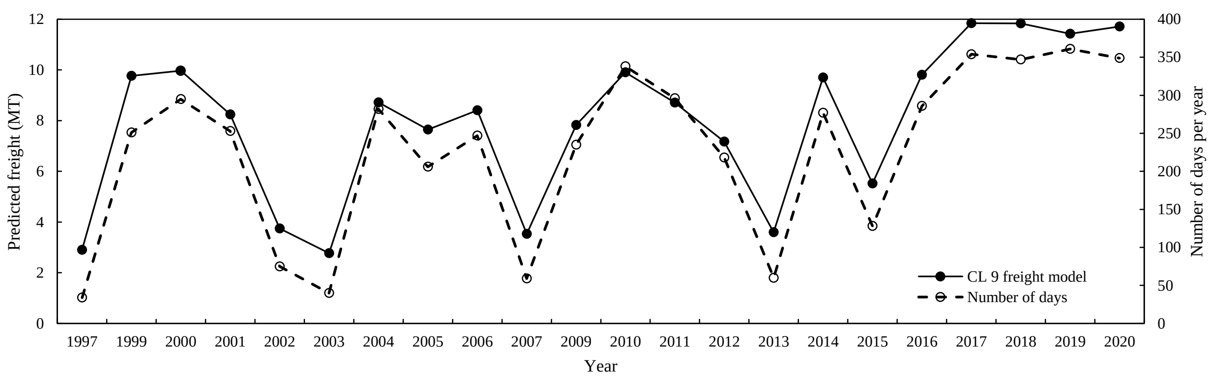
Figure 7-15 GVW weights for Class 9 and other trucks-Washington.



(a) LTPP- MI (26-0113) US-27 (South) Rural Principal Arterial – All lanes

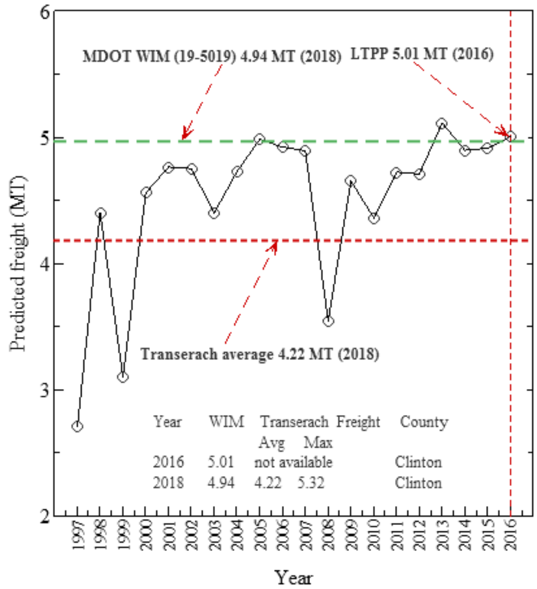


(b) LTPP-WA (53-6049) US-27 (South) State-167 – Urban Other Principal Arterial – All lanes)

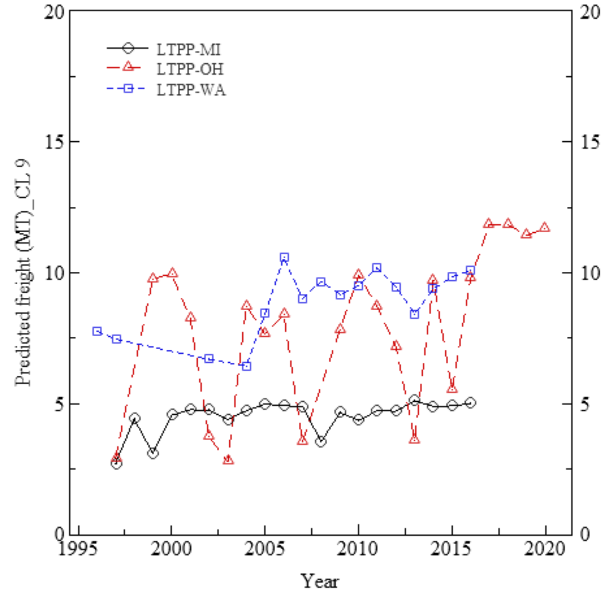


(c) LTPP-OH (39-0101) US-23 (South) Rural Principal Arterial – Lane 1

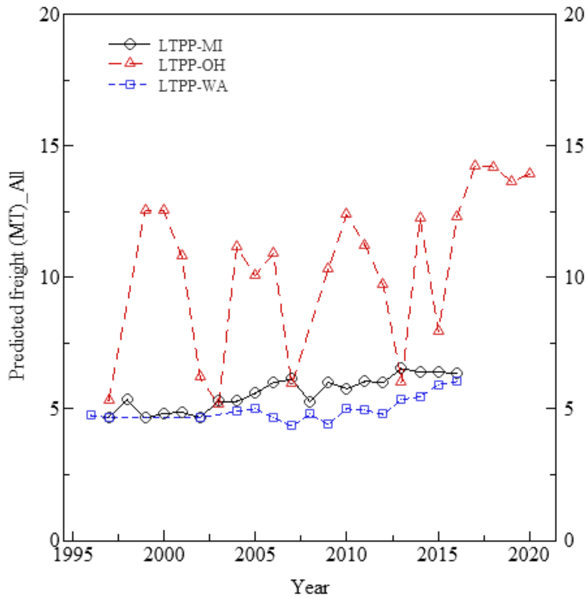
Figure 7-16 Predicted freight for Class 9 trucks and number of days.



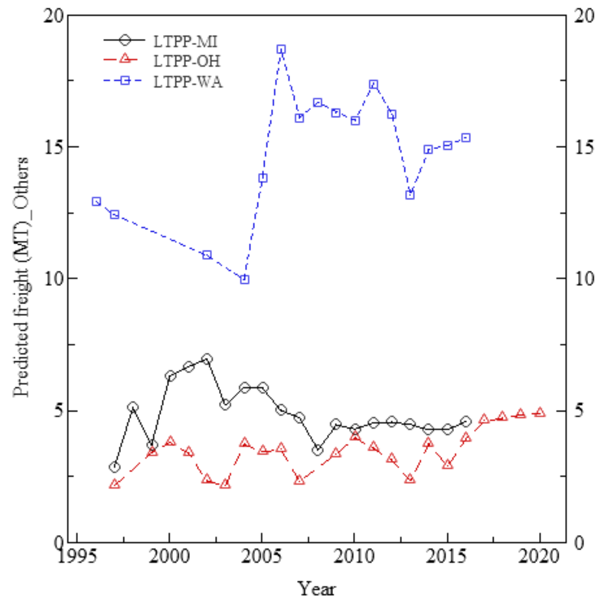
(a) LTPP-MI (Class-9)



(b) Predicted freight based on Class 9



(c) Predicted freight based on all trucks



(d) Predicted freight based on other trucks

Figure 7-17 Comparisons of predicted freight – LTPP WIM sites.

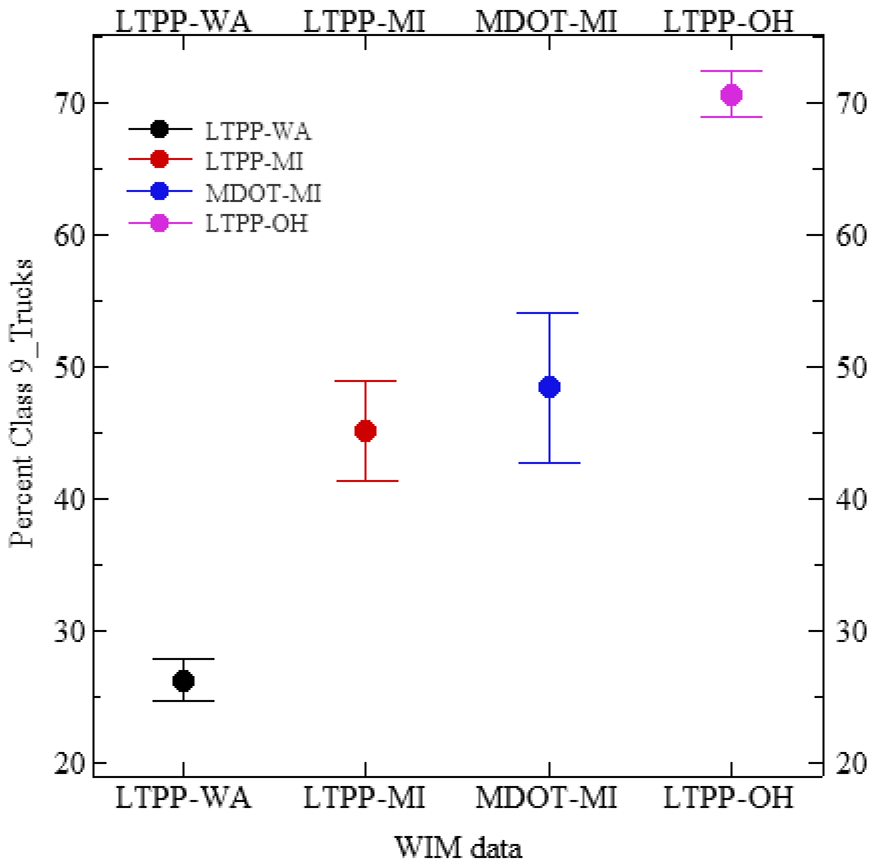
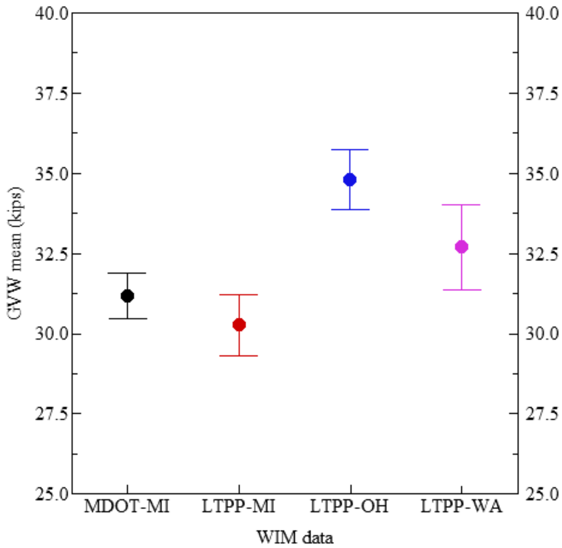
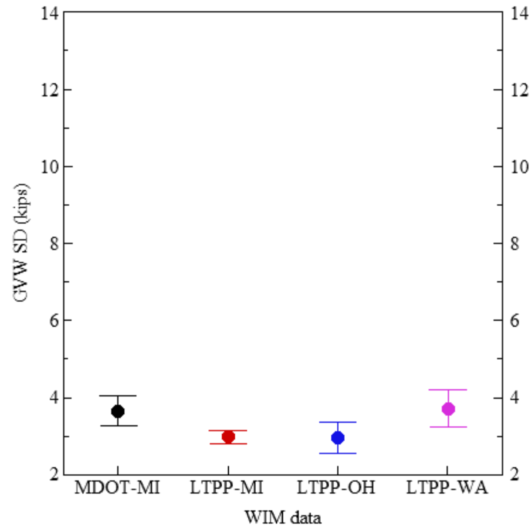


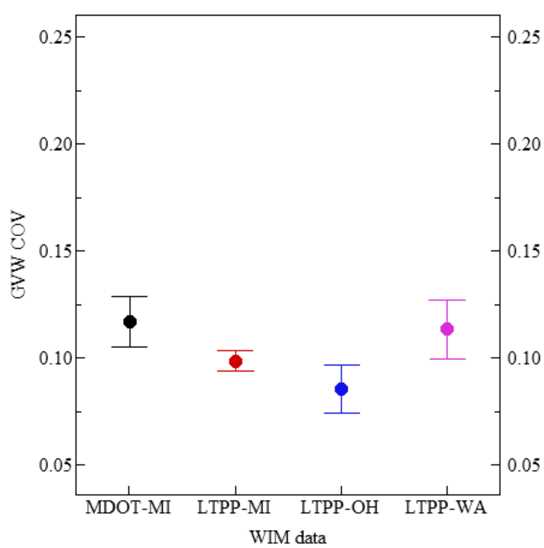
Figure 7-18 Percent of Class 9 trucks at LTPP and MDOT WIM sites.



(a) Empty trucks GVW mean

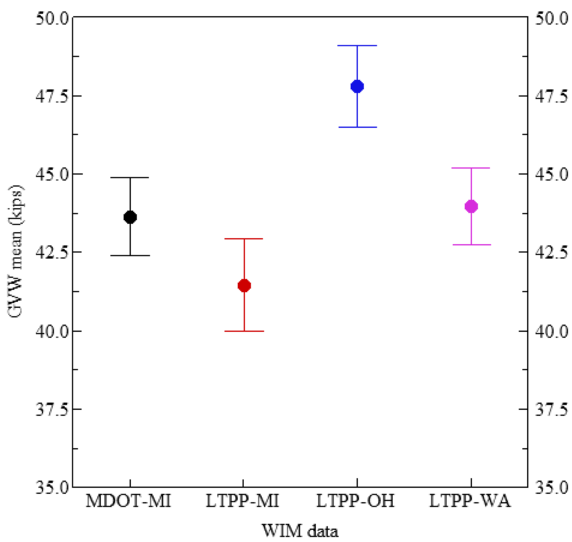


(b) Empty trucks GVW SD

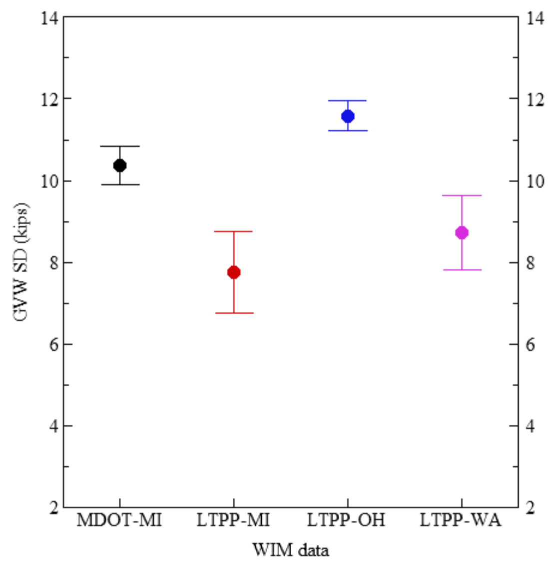


(c) Empty trucks GVW COV

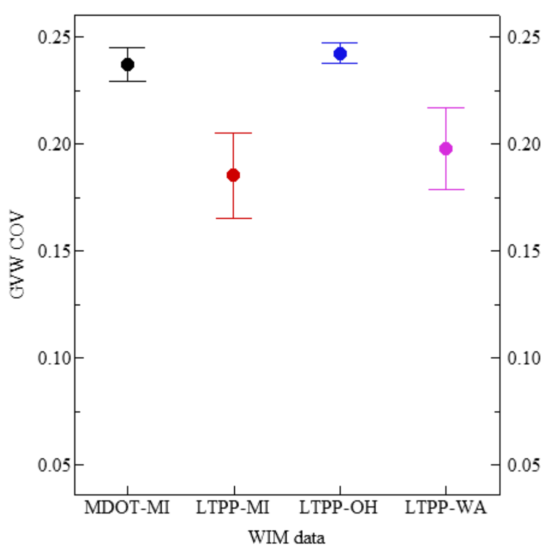
Figure 7-19 Comparisons of GVW shape factors – Empty trucks distributions.



(a) Partially loaded trucks GVW mean

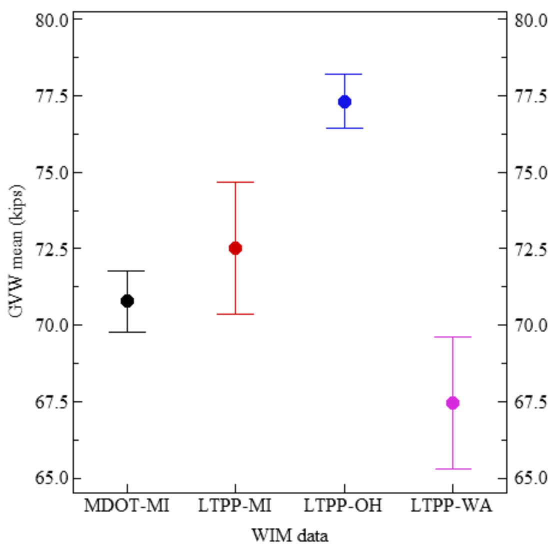


(b) Partially loaded trucks SD

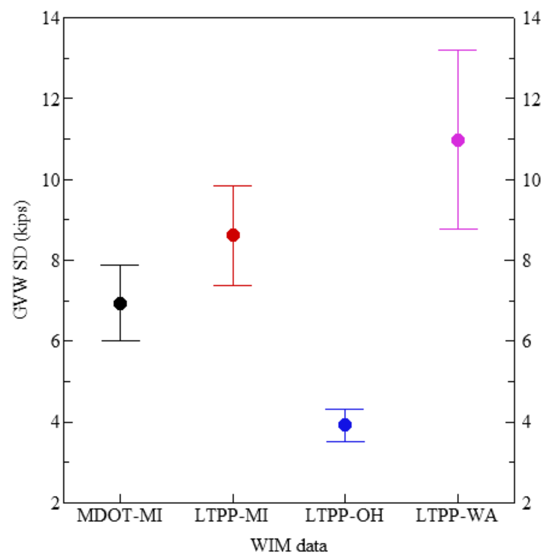


(c) Partially loaded trucks COV

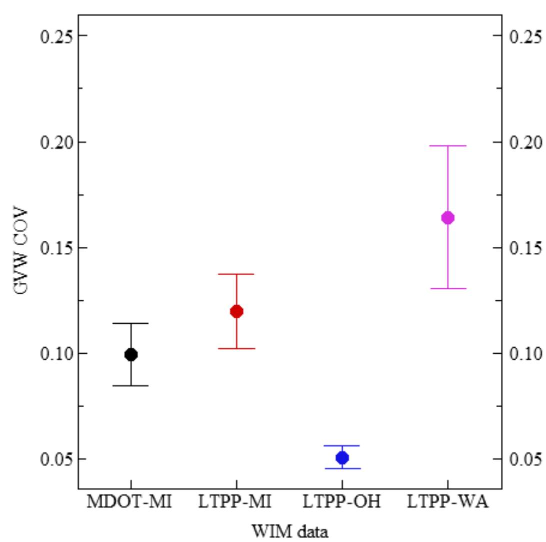
Figure 7-20 Comparisons of GVW shape factors – partially loaded trucks distributions.



(a) Fully loaded trucks GVW mean



(b) Fully loaded trucks SD



(c) Fully loaded trucks COV

Figure 7-21 Comparisons of GVW shape factors – fully loaded trucks distributions.

7.8 KEY FINDINGS

The following are the key findings based on the analyses of freight and GVW WIM data:

- This study presents a practical application of WIM data as an additional approach to estimating freight tonnage.
- The investigation used one year of WIM data collected at 35 WIM sites within

Michigan to estimate freight tonnage (payload) carried by Class 9 and other trucks.

- The freight (payload) computed for Class 9 trucks from GVW data strongly correlated with actual average freight tonnage.
- The regression model presented in the study can be used with reasonable accuracy (R^2 =0.84) to estimate freight tonnage using GVW data for Class 9 trucks.
- The research also presents a procedure to classify freight commodities based on normalized GVW shape factors for Class 9 trucks.
- The decision tree model correctly classified 24 out of 35 events (69%).
- The case studies from the LTPP WIM data show the potential of model application to estimate freight.
- The Michigan and Washington sites show a higher percentage of empty trucks, whereas; the Ohio WIM site shows an almost similar frequency for empty and fully loaded trucks.
- The results show that the freight estimates from three data sources (Transearch, MDOT, and LTPP) are comparable.

7.9 CHAPTER SUMMARY

This chapter further extends applications of WIM data to address an important issue related to freight data, i.e., how to estimate freight tonnage and classify commodities based on GVW WIM data. The methodology uses GVW loading data to estimate vehicle payload and commodity type. The investigation used one year of WIM data collected at 35 WIM sites within Michigan to estimate freight tonnage (payload) carried by Class 9 and other trucks. The freight (payload) computed for Class 9 trucks from GVW data strongly correlated with actual average freight tonnage. The regression model presented in the study can be used with reasonable accuracy (R2

=0.84) to estimate freight tonnage using GVW data for Class 9 trucks. The results show that the freight estimates from three data sources (Transearch, MDOT, and LTPP) are comparable. The presented methodology has good potential for application at WIM sites collecting GVW data. The use of WIM data is a different approach to traditional freight data collection methods like truck surveys, consumer reports, vehicle inventory and use surveys (VIUS), commodity flow survey (CFS), freight analyses framework (FAF), and other commercial data sources. The user can independently verify the freight estimates from surveys at locations close to WIM sites. The developed method can estimate freight at any route provided WIM data are available.

CHAPTER 8 CONCLUSIONS AND RECOMMENDATIONS

8.1 PROBLEM STATEMENT

Highway agencies use WIM technology to collect vehicle and axle weights on highways. WIM sensors measure the transient dynamic tire forces transmitted by vehicles moving at highway speeds. The WIM controller uses the signals from the WIM sensors to estimate the vehicle's static weight and axle loads at rest.

Because WIM technology estimates static weight for a moving vehicle, there are many potential sources of measurement error. Some errors are due to the variation in the forces transferred by the moving truck to the sensor that is caused by truck movement and pavement or bridge characteristics. Other factors affecting the accuracy of WIM measurement are related to WIM equipment operating characteristics, site design, installation, maintenance, and calibration. State and other highway agencies collect WIM data for highway planning, pavement and bridge design, freight movement studies, motor vehicle enforcement screening, and vehicle size and weight regulatory studies. The data collected must be accurate and consistent with so many potential uses.

8.2 OBJECTIVES

The objective of this research was to conduct an analysis of different factors affecting WIM measurement accuracy and develop practical tools and procedures to improve accuracy and increase the reliability of WIM data through more appropriate:

- WIM site selection.
- WIM system selection.
- WIM installation quality assurance.
- WIM calibration and maintenance.

- WIM data analysis methods and QC/QA processes.

8.3 ADVANCING STATE OF KNOWLEDGE IN MANAGING WIM DATA ACCURACY

There is a need to understand the relative importance of various sources of error on WIM data accuracy and for methods that could help minimize the effect of external factors on WIM data quality. Several factors affecting WIM data quality were identified through the literature review. A comprehensive and robust data analysis study was conducted to quantify the effect of multiple factors on WIM data accuracy and evaluate the relative significance of different factors on WIM performance. WIM calibration is an essential activity for maintaining WIM data accuracy. Statistical analysis and machine learning techniques were used to develop a data-driven method for identifying WIM calibration needs based on analysis of statistical attributes computed based on WIM data reported by the WIM system for FHWA Class 9 trucks. The models developed in this research investigation use axle load spectra attributes to assess the systematic changes (bias) in WIM measurements for gross vehicle weight (GVW), single axle (SA) load, and tandem (TA) load. This methodology can save significant time and resources required for field validation of WIM performance using test trucks when applied in practice. Additionally, depending on the extent of information related to the site, sensor, and calibration-related factors, the decision tree models developed in this study can help highway agencies to optimize WIM sensor type and array selection. This information can be integrated with WIM equipment installation costs and life cycle costs to determine the most reliable and economical equipment while also considering WIM data accuracy requirements received from WIM data users.

The scope of the chapter includes:

- Summary of conclusions from the data analysis task
- Description of benefits of estimating WIM errors using influential factors, and potential application
- Description of benefits of using NALS shape factors to estimate changes in WIM measurement errors, and potential application
- Key findings related to WIM calibration guidelines and freight data analyses
- Data limitations and their effect on data analysis results
- Recommendations for future data collection and research

8.4 DATA SETS USED IN ANALYSES

The WIM sites used in the analysis were categorized as follows:

- LTPP WIM sites providing Research Quality Data (RQD) from TPF5(004) study and SPS 10 sites (LTPP RQD): The WIM sites consistently meet the ASTM type 1 performance requirements (i.e., GVW total error ≤ ± 10 % for ≥ 75% of the calibration events were included in this data set). This data set consisted of 170 calibration records from 36 WIM sites that are part of the SPS TPF 5(004) and SPS-10 studies. These sites represent the highest quality WIM data due to the stringent LTPP WIM calibration protocol and daily WIM data review implemented by the LTPP program. This subset contains WIM data for BP, QP, and LC sensors.
- State-owned WIM sites providing high-quality WIM data (RQD Equivalent): This category included the state-owned WIM sites with the available calibration data meeting or exceeding the LTPP RQD data accuracy standards. The data set included 164

calibration records from 94 WIM sites. Four sensor types, i.e., BP, QP, LC, and PC, were included in this data set.

- State-owned WIM sites providing data of lesser quality than LTPP RQD sites (Less than RQD): The state-owned WIM sites with calibration data not meeting the LTPP RQD accuracy standards were considered in this category. The subset includes 80 calibration records from 40 WIM sites. This subset contains WIM data for BP (two sites with one calibration record each) and PC sensors (predominantly PC data with 38 sites and 78 calibration records).

8.5 REPRESENTATIVE RANGES OF WIM MEASUREMENT ACCURACY AND CONSISTENCY AFTER CALIBRATION

The representative ranges of WIM measurement accuracy and consistency achievable after calibration were developed based on the available data sets. Tables 8-1 to 8-3 show the key results for different sensors. The following conclusions were derived based on the observations of the representative ranges of WIM measurement accuracy and consistency.

- The results also show that, immediately after successful calibration, the GVW total error for all available sensors in LTPP RQD and RQD equivalent data set were within ± 5.8%, which is well within ASTM type 1 thresholds (± 10.0% for GVW total error). This included all BP, LC, and QP WIM sites and some PC WIM sensors.
- The results show that, when the WIM system was calibrated, the mean errors (i.e., measurement bias) in GVW were significantly reduced (all values within ± 1.60%) for all sensors available in the LTPP RQD set and RQD-equivalent WIM data set for state-owned WIM sites. However, even after calibration, a relatively higher bias was observed for all available sensors in less than the RQD WIM category. The highest average bias values were observed for PC sensors.

- Overall, bending plates (BP) sensors showed the best data accuracy and consistency results, followed by the load cell LC and QP sensors.
- Based on calibration results, the PC WIM sites included in the RQD equivalent data set showed low errors for all GVW data attributes, compared to the PC sites in less than the RQD category. However, the data showed that these errors tend to increase after calibration with the seasonal changes. Practitioners describe this phenomenon as a calibration drift.
- Only a limited number of LTPP WIM sites had measurement error data collected during pre-calibration test truck runs. These data were collected and reported before each routine field equipment validation or calibration. These data show that WIM measurement accuracy and consistency degrade over time for all WIM sensors in this investigation.
- These findings have an immediate practical application by providing highway agencies with the benchmark values demonstrating the practically achievable accuracy and variability of WIM measurements for different WIM sensor types after successful calibration.

Table 8-1 A representative range for GVW mean measurement errors (bias) observed for available WIM sites after calibration.

Data type	Sensor type			
	BP	LC	QP	PC
LTPP RQD	± 0.82%	± 1.60%	± 0.92%	-
RQD Equivalent	± 0.81%	± 1.00%	± 1.12%	± 1.50%
Less than RQD	-	-	-	± 4.51%
All except LTPP RQD	± 0.81%	± 1.00%	± 1.12%	± 3.01%
All combined	± 0.82%	± 1.30%	± 1.02%	± 3.01%

Table 8-2 A representative range for GVW random errors observed for available WIM sites after calibration.

Data type	Sensor type			
	BP	LC	QP	PC
LTPP RQD	± 3.65%	± 3.80%	4.86%	-
RQD Equivalent	± 3.20%	± 4.80%	4.22%	± 4.20%
Less than RQD	-	-	-	± 8.64%
All except LTPP RQD	± 3.20%	± 4.80%	4.22%	± 6.42%
All combined	± 3.43%	± 4.30%	4.54%	± 6.42%

Table 8-3 A representative range for GVW total observed for available WIM sites after calibration.

Data type	Sensor type			
	BP	LC	QP	PC
LTPP RQD	± 4.47%	± 5.40%	± 5.78%	-
RQD Equivalent	± 4.01%	± 5.80%	± 5.34%	± 5.70%
Less than RQD	-	-	-	± 13.15%
All except LTPP RQD	± 4.01%	± 5.80%	± 5.34%	± 9.43%
All combined	± 4.25%	± 5.60%	± 5.56%	± 9.43%

8.6 WIM PERFORMANCE OVER TIME-BASED ON WIM VALIDATION DATA AND AXLE LOAD SPECTRA ANALYSIS

- Findings from the WIM performance data analysis show that to objectively evaluate WIM measurement accuracy and consistency, it is critical to consider data collected before and after calibration.
- The analysis results show that data accuracy deteriorates between calibration events for all sensor types included in this investigation (see Table 8-4).

- Calibration scheduling should be data-driven to prevent significant calibration drift. This could be accomplished by monitoring changes in axle load spectra and other GVW and axle loading summary statistics over time.

Table 8-4 Pre and post-calibration GVW WIM data (average values).

Data type	Data set	Sensor type	
		BP	QP
GVW bias	Pre- Calibration	± 2.98%	± 4. 98%
	Post-Calibration	± 0.84%	± 1.10%
GVW SD	Pre- Calibration	2.5	3.10
	Post- Calibration	2.0	2.71
GVW total error	Pre- Calibration	± 8.01%	± 11.13%
	Post- Calibration	± 4.87%	± 6.58%

Pre = pre-calibration, Post = post-calibration.

8.6.1 WIM Sensor Performance and Calibration Frequency

Based on the analysis of NALS results (see Table 8-5), the following conclusions about the recommended frequency of field calibration were made:

- Calibration frequency longer than 1 year may be acceptable for the sites with BP sensors, provided the equipment maintenance schedule follows the manufacturer's specification (typically every 6 months).
- Annual calibration frequency is recommended for sites with QP sensors.
- Due to significantly higher NALS inconsistencies, the sites with PC sensors may need multiple calibrations in a year, especially in climates with high differences in seasonal temperatures.

Table 8-5 Percentage Change in SA and TA NALS over time after calibration.

Sensor type	Number of sites	Calibration records	Time after calibration (months)	Average SA bias using NALS (%)	Average TA bias using NALS (%)
BP	12	36	4	± 1.75	± 1.37
			8	± 2.39	± 1.60
			12	± 1.86	± 1.46
LC	3	6	4	± 1.88	± 0.35
			8	± 2.33	± 0.81
			12	± 3.20	± 1.18
QP	23	60	4	± 3.00	± 2.00
			8	± 3.69	± 2.41
			12	± 4.12	± 2.51
PC	12	18	4	± 3.50	± 3.48
			8	± 4.40	± 4.41
			12	± 4.92	± 4.52

8.7 INFLUENTIAL FACTORS AFFECTING WIM SYSTEM PERFORMANCE

8.7.1 Climatic Factors Affecting WIM Measurement Accuracy and Consistency

The effect of climatic factors was investigated, and the conclusions were summarized in Table 8-6.

Table 8-6 Effect of climate-related factors on WIM errors.

Factor	Sensor type	Statistical significance (Yes/No)	Comments
Climate	BP and LC	No	BP and LC errors are not affected by climate.
	QP, PC	Yes	Both sensors showed better precision in wet climates.
Calibration season	All sensors	No	Generally, calibrations performed in Fall (i.e. a season with moderate temperatures vs. seasons with extremely high or low temperatures) yield low WIM errors during a calibration event.
Calibration temperature (pavement)	BP, LC, QP	No	Generally, low WIM errors were observed with average pavement temperatures during calibration ranging between 75 to 100oF and with a differential of 30 to 40oF.

8.7.2 Road and Pavement Factors Affecting WIM Measurement Accuracy and Consistency

The effect of pavement-related factors on WIM measurement errors was investigated, and conclusions were summarized in Table 8-7.

8.7.3 Traffic Speed and WIM System Features Affecting WIM Measurement Accuracy and Consistency

The effects of traffic speed and WIM system features on WIM measurement errors were investigated, and conclusions were summarized in Table 8-8.

Table 8-7 Effect of pavement-related factors on WIM errors.

Factor	Sensor type	Statistical significance (Yes/No)	Comments
Pavement type	BP, LC	-	All BP and LC sensors were installed in PCC pavements.
	QP	No	Lower errors were observed in PCC pavements.
	PC	No	Lower errors were observed in AC pavements.
Pavement thickness	BP, LC, QP, PC	No	No significant impacts of surface thickness on WIM precision were observed based on the available data. However, based on the data analyses, the BP sensors can be installed in 10 inches or thicker PCC slabs to yield ASTM Type I accuracy. Irrespective of pavement type, 8 inches or above (PCC or HMA thickness) is recommended for QP sensors to obtain accurate WIM data.
Longitudinal grade	BP, LC, QP	Yes	Generally, flatter pavement (low grades, i.e., 1% or less) showed better precision.
Transverse slope	BP, LC, QP	No	No significant impacts of the transverse slope on WIM precision were observed based on the available data.
IRI, WRI	BP, LC, QP	No	No consistent trends were observed between IRI or WRI and consistency in WIM measurements based on the available data. Roughness data and WIM data were not collected at the same time for most sites.
FWD	QP	No	No consistent trends were observed between measured deflection and consistency in WIM measurements based on the available data for 8 WIM locations in Indiana.

Table 8-8 Effect of traffic speed and WIM system features on WIM errors.

Factor	Sensor type	Statistical significance (Yes/No)	Comments
Sensor type	BP, LC, QP, PC	Yes	PC sensor accuracy and consistency were significantly different compared to other sensors.
Sensor array	BP, LC, QP, PC	Yes	Significant differences amongst sensor arrays were observed during the analysis. Sensor array design is a critical factor in achieving the desired WIM data accuracy.
Calibration Speed points	BP, LC, QP, PC	Yes	WIM controllers with multiple speed points could significantly improve WIM precision and reduce measurement bias. However, some inconsistencies were observed for the PC sensor.
Calibration speed	BP, LC, QP	No	A speed range between 5 to 10 mph at the time of calibration showed less variability in calibration data. However, the use of a narrow speed range may lead to incorrect computation of WIM measurement error for the sites with a wide range of operating speeds.

8.7.4 Benefits of Estimating WIM Errors Using Influential Factors, and Potential Application

This analysis aimed to evaluate if effective statistical or logical models could be developed and used to quantify the effects of essential site, sensor, and calibration-related factors on the variability of WIM measurement error. Due to limited data availability, the analysis was focused on the following independent variables:

- Climate
- Pavement type
- Sensor array
- Sensor type
- Calibration speed points

The above list does not include several essential site factors (road geometry, pavement smoothness, and pavement strength information).

The following dependent variables were used in the analyses:

- GVW mean measurement errors (bias)
- GVW standard deviation of measurement errors
- GVW total measurement errors

The decision tree models show a vital application of site and sensor-related factors based on comprehensive data. The presented methodology utilizing the decision tree models shows good potential for estimating the WIM measurement error range using information about the WIM site and sensor-related factors. The decision tree models can help highway agencies to make an optimal WIM equipment selection giving due consideration to achievable WIM errors, climatic conditions, pavement type, and equipment life cycle costs.

8.7.5 Benefits of Estimating WIM Errors Using NALS Attributes, and Potential Application

A data analysis study was conducted to develop statistical attributes (NALS shape factors) and procedures to aid in identifying and quantifying changes in WIM measurement bias (calibration drift) based on analysis of changes in axle load spectra attributes between WIM equipment calibration events. The pre-and post-calibration data and axle load spectra were used in these analyses. The models developed using axle load spectra shape factors can be used to estimate measurement bias with reasonable accuracy (R^2 is about 80%). The results show that single and tandem axle load spectra attributes (SA mean axle load and TA mean load for the loaded axles weighing over 26,000 lbs.) can be effectively used to assess the systematic changes (bias) in WIM measurements for GVW, SA, and TA.

The methodology of WIM accuracy estimation through axle load spectra analysis can facilitate identifying the WIM equipment calibration requirements. The NALS analysis based on changes in shape factors can be used to estimate the changes in the SA, TA, and GVW WIM measurement bias. This methodology can save significant time and resources required for field validation of WIM performance using test trucks. The statistical models developed in this study for the prediction of WIM weight measurement bias should be used in combination with the visual inspection of SA and TA peak loads, along with information about the expected seasonal changes in traffic loading due to land use (if any).

8.8 KEY FINDINGS RELATED TO WIM CALIBRATION GUIDELINES

Successful WIM equipment calibration can eliminate weight, speed, and axle spacing errors. Following are the conclusions and recommendations based on data analyses.

- The results show that the effect of sample size on WIM errors was negligible, especially when the sample size is sufficiently large (n>=10) for QP and BP sensors. faltered
- The WIM site calibration can be performed using one test truck to achieve a representative range of BP and QP sensor errors. A single test truck with 12 runs (4 at each speed point) can be used for equipment calibration.
- The current LTPP filed operation guide recommendations of calibrating a WIM site at different speed levels should continue, preferably at three-speed points 50, 60, and 70 mph or as per the recommendations of the posted speed limits.
- Pre and post-calibration data can be collected on the same day for BP sensors, as no apparent effect of temperature was seen on BP WIM sites. If possible, the pre and post-calibration data can be collected for an extended period for QP sensors to account for higher temperature fluctuations.

- The representative post-calibration data can be collected accurately using one test truck with 12 passes at 3-speed points for QP and BP sensors. If a site shows higher speed dependency, the number of test truck runs may be increased to 20.
- The ASTM and the LTPP accuracy estimation methods agree; however, the methods should be compared when the sample size is small and in the presence of potential outliers.
- The developed models showed that the GVW errors could be accurately predicted using SA and two TAs.
- The results show that if GVW errors are within ASTM Typ1 I tolerance, the SA and TA errors will also be within acceptable limits. Therefore, the practice of calibrating a WIM site using GVW errors should continue.

The suggested changes in current WIM procedures can significantly reduce time and resources for successful equipment calibration. The preliminary models developed in this study can be validated in the field and improved further by adding more data in the future.

8.9 KEY FINDINGS BASED ON FREIGHT DATA ANALYSES

The following are the key findings based on the analyses of freight and GVW WIM data:

- This study presents a practical application of WIM data as an additional approach to estimating freight tonnage.
- The investigation used one year of WIM data collected at 35 WIM sites within Michigan to estimate freight tonnage (payload) carried by Class 9 and other trucks.
- The freight (payload) computed for Class 9 trucks from GVW data strongly correlated with actual average freight tonnage.
- The regression model presented in the study can be used with reasonable accuracy (R^2

=0.84) to estimate freight tonnage using GVW data for Class 9 trucks.

- The research also presents a procedure to classify freight commodities based on normalized GVW shape factors for Class 9 trucks.
- The decision tree model correctly classified 24 out of 35 events (69%).
- The case studies from the LTPP WIM data show the potential of model application to estimate freight.
- The Michigan and Washington sites show a higher percentage of empty trucks, whereas; the Ohio WIM site shows an almost similar frequency for empty and fully loaded trucks.
- The results show that the freight estimates from three data sources (Transearch, MDOT, and LTPP) are comparable.

The presented methodology has good potential for application at WIM sites collecting GVW data. The use of WIM data is a different approach to traditional freight data collection methods like truck surveys, consumer reports, vehicle inventory and use surveys (VIUS), commodity flow surveys (CFS), freight analyses framework (FAF), and other commercial data sources. The user can independently verify the freight estimates from surveys at locations close to WIM sites. The developed method can estimate freight at any route provided WIM data are available.

8.10 DATA LIMITATIONS

The following data availability limitations were noted during the data analysis task. As these data become available, an extended analysis may be beneficial:

- Pavement stiffness or other structural data were not available for any of the LTPP WIM site locations since no FWD testing or pavement coring and testing was conducted at

WIM site locations. Therefore, this factor was not considered in the network-level analysis.

- The IRI data were available for a limited number of WIM sites, and this factor was not considered at network-level analysis. The IRI data collection schedule was not coordinated with WIM field calibrations. Ideally, IRI and WIM performance data should be available for the same climatic conditions.
- The limited availability of sensor removal/replacement information resulted in missing or inaccurate sensor age calculations. Therefore, the effect of sensor age could not be assessed on WIM system performance.
- Limited or non-availability of pavement thickness at the WIM site locations resulted in eliminating this factor at network-level analysis.
- Pavement distress and FWD data collection efforts at LTPP pavement experiments did not cover the exact WIM site locations, limiting the number of known pavement factors at WIM site locations (except for pavement roughness data collected at LTPP TPF and Indiana DOT WIM sites).
- Adding state-owned WIM data for BP and QP sensors resulted in an unbalanced design. Most of the additional data for BP and QP sensors WIM data were provided by the states of California and Michigan, located in dry or wet climates, respectively.
- The distribution of WIM data for PC sites was not uniform for different factors, as most of these data were only available in wet climates.
- The non-availability of continuous variables was another challenge because most of the variables available for the data analysis were categorical, i.e., climate, pavement, sensor,

sensor array, and speed points, providing challenges for using regression modeling techniques.

8.11 RECOMMENDATIONS FOR FUTURE DATA COLLECTION AND RESEARCH

Based on the findings of this study, the following recommendations are made in order of priority for future data collection efforts to support additional analyses to improve the models developed in this study:

Table 8-9 Recommendations for future data collection.

Order of priority	Comments
1	The sensor array and controller functionality are essential factors affecting WIM data accuracy and consistency and should be documented for all LTPP WIM sites. This expanded data set will allow a more comprehensive analysis of the effect of sensor arrays on WIM measurement error and will help identify optimum sensor arrays for different WIM applications.
1	Additional data should be collected for WIM sites installed in dry climates (DF/DNF) for all four sensor types to investigate the effect of climate on WIM performance (especially for QP and PC WIM sites that show some sensitivity to climatic effects).
2	The sensor installation, repair, and removal/replacement dates should be documented to help in determining the sensor age and sensor performance over time accurately. These data will provide means for WIM equipment life cycle cost-benefit analysis and development of guidelines for sensor selection based on cost and length of data collection.
2	The pre-calibration data provide valuable information to assess the effectiveness of a previous calibration event, as well as the quantification of changes in WIM measurement accuracy and consistency over time. These data should be routinely collected by state highway agencies during WIM validation and calibration events, especially for PC sensors that show high variability in measurements between calibrations.
3	To analyze the effect of pavement smoothness on WIM measurement error, detailed pavement profile data should be collected in conjunction with WIM calibration or validation visit (to assure similar climatic conditions for collecting both WIM and pavement profile data) for road segments 400 feet before and 100 feet after WIM sensors.
3	To analyze the effect of pavement strength on WIM measurement error, detailed FWD data is needed for road segments 50 feet before and 50 feet after WIM sensors.
3	The pavement structure, grade, and slope at the WIM sites should be recorded in WIM installation documentation. This information is important for assessing the effect of pavement-related factors on WIM performance.
4	The calibration speed and temperature data for WIM sites outside of the LTPP RQD set should be recorded for each calibration event to support the analysis of the effect of speed and temperature on WIM measurement accuracy and consistency.
5	More QP and PC WIM sites installed in PCC pavements should be added to the dataset to evaluate the effect of pavement type on WIM performance.

REFERENCES

[1] G. G. Otto *et al.*, "Weigh-in-motion (WIM) sensor response model using pavement stress and deflection," *Construction and Building Materials,* vol. 156, pp. 83-90, 2017.

[2] D. Hazlett, N. Jiang, and L. Loftus-Otway, "Use of Weigh-in-Motion Data for Pavement, Bridge, Weight Enforcement, and Freight Logistics Applications," 0309481252, 2020.

[3] B. Jacob and V. Feypell-de La Beaumelle, "Improving truck safety: Potential of weigh-in-motion technology," *IATSS research,* vol. 34, no. 1, pp. 9-15, 2010.

[4] P. Burnos and J. Gajda, "Thermal property analysis of axle load sensors for weighing vehicles in weigh-in-motion system," *Sensors,* vol. 16, no. 12, p. 2143, 2016.

[5] P. Burnos, J. Gajda, and R. Sroka, "Accuracy criteria for evaluation of Weigh-in-Motion Systems," *Metrology and Measurement Systems,* vol. 25, no. 4, 2018.

[6] J. Gajda, R. Sroka, and P. Burnos, "Designing the Calibration Process of Weigh-In-Motion Systems," *Electronics,* vol. 10, no. 20, p. 2537, 2021.

[7] J. Gajda, R. Sroka, and T. Żegleń, "Accuracy analysis of WIM Systems Calibrated Using Pre-Weighed Vehicles Method," *Metrology and Measurement Systems,* vol. 14, no. 4, pp. 517-527, 2007.

[8] J. Gajda, R. Sroka, T. Zeglen, and P. Burnos, "The Influence of Temperature on Errors of Wim Systems Employing Piezoelectric Sensors Keywords: Piezoelectric Sensors, Temperature Influence, Temperature Error of Wim Systems, Error Correction," *Metrology and Measurement Systems,* vol. 20, no. 2, pp. 171-182, 2013.

[9] J. Gajda, R. Sroka, T. Zeglen, and P. Burnos, "The influence of temperature on errors of WIM systems employing piezoelectric sensor," *Metrology and Measurement Systems,* vol. 20, no. 2, pp. 171--182, 2013.

[10] AustRoads, "Weigh-In-Motion Technology," AP–R168, AP–R168, 2000.

[11] W. S. Calibration, "A Vital Activity for LTPP Sites," *Federal Highway Administration TechbriefFHWA-RD-98-104. US Department of Transportation, Washington, DC,* 1998.

[12] P. Davies and F. Sommerville, *Calibration and Accuracy Testing of Weigh-in-Motion Systems* (no. 1123). 1987.

[13] FHWA, "Traffic Monitoring Guide FHWA-PL-95-031 US Department of Transportation," *Federal Highway Administration, Washington, DC,* 1995.

[14] F. B. Roy Czinku, "Talking Traffic Webinar- , WIM Sensors, Arrays, and Applications," ed, 2020.

[15] G. E. Elkins, P. Schmalzer, T. Thompson, and A. Simpson, "Long-term pavement performance information management system pavement performance database user guide," *McLean: Federal Highway Administration,* 2003.

[16] O. I. Selezneva, M. Ayers, M. Hallenbeck, A. Ramachandran, H. Shirazi, and H. Von Quintus, "MEPDG Traffic Loading Defaults Derived from Traffic Pooled Fund Study," United States. Federal Highway Administration, 2016.

[17] S. W. Haider, R. S. Harichandran, and M. B. Dwaikat, "Impact of Systematic Axle Load Measurement Error on Pavement Design Using Mechanistic-Empirical Pavement Design Guide," *Journal of Transportation Engineering,* vol. 138, no. 3, pp. 381-386, 2011.

[18] ASTM, "Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Methods E 1318-09," *2007 Annual Book of ASTM Standards. Edited by ASTM Committee E17-52 on Traffic Monitoring. ASTM International, USA,* 2009.

[19] B. Jacob, "Assessment of the Accuracy and Classification of Weigh-in-Motion Systems Part 1: Statistical Background," *International Journal of Heavy Vehicle Systems,* vol. 7, no. 2-3, pp. 136-152, 2000.

[20] B. Jacob and E. J. O'Brien, "European Specification on Weigh-in-Motion Of Road Vehicles (COST323)," in *Second European Conference on Weigh-In-Motion of Road Vehicles, Held Lisbon, Portugal 14-16 September 1998*, 1998.

[21] S. W. Haider and M. M. Masud, "Accuracy Comparisons Between ASTM 1318-09 and COST-323 (European) WIM Standards Using LTPP WIM Data," in *Proceedings of the 9th International Conference on Maintenance and Rehabilitation of Pavements—Mairepav9*, 2020: Springer, pp. 155-165.

[22] S. W. Haider, M. M. Masud, O. Selezneva, and D. J. Wolf, "Assessment of Factors Affecting Measurement Accuracy for High-Quality Weigh-in-Motion Sites in the Long-Term Pavement Performance Database," *Transportation Research Record,* vol. 2674, no. 10, pp. 269-284, 2020, doi: 10.1177/0361198120937977.

[23] M. M. Masud and S. W. Haider, "Estimation of Weigh-in-Motion System Accuracy from Axle Load Spectra Data," in *Airfield and Highway Pavements 2021*, 2021, pp. 378-388.

[24] M. M. Masud and S. W. Haider, "Performance of Weigh-in-Motion (WIM) Sensors in Rigid and Flexible Pavements and Guidelines for Recommended Pavement Thickness," in *International Conference on Transportation and Development 2022*, 2022, pp. 224-232.

[25] M. M. Masud, S. W. Haider, O. Selezneva, and D. J. Wolf, "Impact of WIM Systematic Bias on Axle Load Spectra–A Case Study," in *Advances in Materials and Pavement Performance Prediction II: Contributions to the 2nd International Conference on*

Advances in Materials and Pavement Performance Prediction (AM3P 2020), 27-29 May, 2020, San Antonio, TX, USA, 2020: CRC Press, p. 64.

[26] FHWA-LTPP Technical Support Services Contractor, "LTPP Field Operations Guide for SPS WIM Sites Version 1.0 Draft Office of Infrastructure Research, Development, and Technology, Federal Highway Administration, McLean, Virginia," May 2009.

[27] T. G. Butcher, *Specifications, Tolerances, and Other Technical Requirements for Weighing and Measuring Devices*. US Department of Commerce, National Institute of Standards and Technology, 2001.

[28] NMi, "NMi International WIM Standard: Specification and Test Procedures for Weigh-in-Motion Systems," 2016.

[29] P. Burnos and D. Rys, "The Effect of Flexible Pavement Mechanics on the Accuracy of Axle Load Sensors in Vehicle Weigh-in-Motion Systems," *Sensors*, vol. 17, no. 9, p. 2053, 2017.

[30] FHWA. "WIM Pocket Guide Part-1 WIM Technology, Data Acquisition, and Procurement Guide." (accessed.

[31] B. Jacob, E. J. O'Brien, and W. Newton, "Assessment of the Accuracy and Classification of Weigh-in-Motion Systems. Part 2: European Specification," *International Journal of Heavy Vehicle Systems*, vol. 7, no. 2-3, pp. 153-168, 2000.

[32] M. Glover and W. Newton, "Evaluation of a Multiple-sensor Weigh-in-motion System," 0266-7045, 1991.

[33] T. Qin, M. Lin, M. Cao, K. Fu, and R. Ding, "Effects of Sensor Location on Dynamic Load Estimation in Weigh-in-Motion System," *Sensors*, vol. 18, no. 9, p. 3044, 2018.

[34] M. Y Darestani, D. P. Thambiratnam, A. Nataatmadja, and D. Baweja, "Experimental Study on Structural Response of Rigid Pavements Under Moving Truck Load," 2006.

[35] D. Rys, "Investigation of weigh-in-motion measurement accuracy on the basis of steering axle load spectra," *Sensors*, vol. 19, no. 15, p. 3272, 2019.

[36] A. Papagiannakis, E. Johnston, and S. Alavi, "Fatigue performance of piezoelectric Weigh-in-Motion sensors," *Transportation research record*, vol. 1769, no. 1, pp. 87-94, 2001.

[37] F. Scheuter, "Evaluation of Factors Affecting WIM System Accuracy," in *Proceedings of the Second European Conference on COST*, 1998, vol. 323, pp. 14-16.

[38] Olga I. Selezneva, Ramachandran A., Mustafa E., and Carvalho R., "Impact of Various Trucks on Pavement Design and Analysis: Mechanistic-Empirical Pavement Design

Guide Sensitivity Study with Truck Weight Data," *Transportation Research Record*, vol. 2339, no. 1, 2014.

[39] O. Selezneva and D. J. Wolf, "Successful Practices in Weigh-in-Motion Data Quality with WIM Guidebook [Volumes 1 & 2]," Arizona. Dept. of Transportation, 2017.

[40] O. Selezneva and A.-M. Mcdonnell, "Weigh-in-Motion, Advancing Highway Traffic Monitoring Through Strategic Research," 2017.

[41] O. Selezneva and D. Wolf, "Successful Practices in Weigh-in-Motion Data Quality with WIM Guidebook Vol 1," 2017.

[42] A. Papagiannakis, E. Johnston, S. Alavi, and J. Mactutis, "Laboratory and field evaluation of piezoelectric Weigh-in-Motion sensors," *Journal of testing and evaluation*, vol. 29, no. 6, pp. 535-543, 2001.

[43] A. Papagiannakis and N. Jackson, "Traffic Data Collection Requirements for Reliability in Pavement Design," *Journal of Transportation Engineering*, vol. 132, p. 237, 2006.

[44] N. H. Tran and K. D. Hall, "Development and influence of statewide axle load spectra on flexible pavement performance," *Transportation Research Record*, vol. 2037, no. 1, pp. 106-114, 2007.

[45] FHWA, "WIM Scale Calibration-A Vital Activity for LTPP Sites," *Federal Highway Administration TechbriefFHWA-RD-98-104. US Department of Transportation, Washington, DC*, 1998.

[46] R. Quinley, "WIM data analyst's manual," United States. Federal Highway Administration, 2010.

[47] FHWA, "A Vital Activity for LTPP Sites," *Federal Highway Administration TechbriefFHWA-RD-98-104. US Department of Transportation, Washington, DC*, 1998.

[48] FHWA, "LTPP Data Analysis: Optimization of Traffic Data Collection for Specific Pavement Design Applications," TechBrief FHWA-HRT-06-111, 2006.

[49] FHWA, "LTPP Field Operations Guide for SPS WIM Sites Version 1.0," FHWA-LTPP Technical Support Services Contractor, AMEC Earth and Environmental, 12000 Indian Creek Court, Suite F, Beltsville, Maryland 20705, , 2012.

[50] FHWA, "WIM Pocket Guide," Federal Highway Administration, Washington DC., Publication No. FHWA-PL-018-008, 2018.

[51] H. Refai, N. Bitar, J. Schettler, and O. A. Kalaa, "The study of vehicle classification equipment with solutions to improve accuracy in Oklahoma," Oklahoma. Dept. of Transportation. Materials and Research Division, 2014.

[52] D. Gupta, X. Tang, and L. Yuan, "Weigh-in-Motion Sensor and Controller Operation and Performance Comparison," Minnesota. Dept. of Transportation. Research Services & Library, 2018.

[53] S. W. Haider and M. M. Masud, "Effect of moisture infiltration on flexible pavement performance using the AASHTOWare Pavement-ME," in *Advances in Materials and Pavement Prediction: Papers from the International Conference on Advances in Materials and Pavement Performance Prediction (AM3P 2018), April 16-18, 2018, Doha, Qatar*, 2018: CRC Press, p. 31.

[54] S. W. Haider and M. M. Masud, "Use of LTPP SMP Data to Quantify Moisture Impacts on Fatigue Cracking in Flexible Pavements [summary report]," United States. Federal Highway Administration. Office of Research …, 2020.

[55] S. W. Haider, M. M. Masud, and K. Chatti, "Influence of moisture infiltration on flexible pavement cracking and optimum timing for surface seals," *Canadian Journal of Civil Engineering*, vol. 47, no. 5, pp. 487-497, 2020.

[56] S. W. Haider, M. M. Masud, and G. Musunuru, "Effect of Water Infiltration Through Surface Cracks on Flexible Pavement Performance," 2018.

[57] M. M. Masud, *Quantification of Moisture Related Damage in Flexible and Rigid Pavements and Incorporation of Pavement Preservation Treatments in AASHTOWare Pavement-ME Design and Analysis*. Michigan State University, 2018.

[58] M. M. MASUD, "IRF GLOBAL R2T Conference," 2019.

[59] M. M. Masud and S. W. Haider, "Long-Term Pavement Performance: International Data Analysis Contest, 2017–2018 Graduate Category: Use of LTPP SMP Data to Quantify Moisture Impacts on Fatigue Cracking in Flexible Pavements," 2020.

[60] M. M. Masud, S. W. Haider, and K. Chatti, "Incorporation of Pavement Preservation Treatments in AASHTOWare Pavement-ME Analysis and Design," 2018.

[61] E. Masad, A. Bhasin, T. Scarpas, I. Menapace, and A. Kumar, *Advances in Materials and Pavement Prediction: Papers from the International Conference on Advances in Materials and Pavement Performance Prediction (AM3P 2018), April 16-18, 2018, Doha, Qatar*. CRC Press, 2018.

[62] E. ASTM, "Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Methods E 1318-09," *2007 Annual Book of ASTM Standards. Edited by ASTM Committee E17-52 on Traffic Monitoring. ASTM International, USA*, 2009.

[63] P. Davies and F. Sommerville, "Calibration and Accuracy Testing of Weigh-In-Motion Systems," *Transportation Research Record*, vol. 1123, pp. 122-126, 1987.

[64] A. Bergan, C. Berthelot, and B. Taylor, "Effect of weigh in motion accuracy on weight enforcement efficiency," 1995.

[65] J. Prozzi, F. Hong, and A. Leung, "Effect of Traffic Load Measurement Bias on Pavement Life Prediction: A Mechanistic-Empirical Perspective," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2087, pp. 91-98, 2008.

[66] A. Bergan, C. Berthelot, and B. Taylor, "Effect of Weigh In Motion Accuracy on Weight Enforcement Efficiency," 1995.

[67] S. W. Haider and M. M. Masud, "Accuracy Comparisons Between ASTM 1318-09 and COST-323 (European) WIM Standards Using LTPP WIM Data," in *Proceedings of the 9th International Conference on Maintenance and Rehabilitation of Pavements--Mairepav9*, 2020, vol. 76: Springer Nature, p. 155.

[68] H. Gong, Y. Sun, X. Shu, and B. Huang, "Use of random forests regression for predicting IRI of asphalt pavements," *Construction and Building Materials,* vol. 189, pp. 890-897, 2018.

[69] S. W. Haider and R. S. Harichandran, "Relating Axle Load Spectra to Truck Gross Vehicle Weights and Volumes," *ASCE Journal of Transportation Engineering*, vol. 133, no. 12, pp. 696-705, 2007.

[70] S. W. Haider and R. S. Harichandran, "Quantifying the Effects of Truck Weights on Axle Load Spectra of Single and Tandem Axle Configurations," *the Fifth International Conference on Maintenance and Rehabilitation of Pavements and Technological Control*, pp. 73-78, 2007.

[71] S. W. Haider and R. S. Harichandran, "Characterizing Axle Load Spectra by Using Gross Vehicle Weights and Truck Traffic Volumes," *CD ROM, 86th Annual Meeting of Transportation Research Record,* 2007.

[72] S. W. Haider, R. S. Harichandran, and M. B. Dwaikat, "Closed-form solutions for bimodal axle load spectra and relative pavement damage estimation," *Journal of transportation engineering,* vol. 135, no. 12, pp. 974-983, 2009.

[73] FHWA, "WIM Data Analyst's Manual Publication No. FHWA-IF-10-018," United States. Federal Highway Administration, 2010.

[74] S. Gey and E. Nedelec, "Model selection for CART regression trees," *IEEE Transactions on Information Theory,* vol. 51, no. 2, pp. 658-670, 2005.

[75] USDOT, "Freight-demand Modeling to Support Public-sector Decision Making, Research Innovative Technology Administration National Cooperative Freight Research

Program (NCHFRP) Cambridge Systematics GeoStats, LLP. United States. Department of Transportation," *Transportation Research Board,* 2010.

[76] S. McLeod, J. H. Schapper, C. Curtis, and G. Graham, "Conceptualizing freight generation for transport and land use planning: A review and synthesis of the literature," *Transport Policy,* vol. 74, pp. 24-34, 2019.

[77] M. P. Boile, L. N. Spasovic, and K. Ozbay, "Estimation of Truck Volumes and Flows," New Jersey. Dept. of Transportation, 2004.

[78] H. B. Rai, T. Van Lier, D. Meers, and C. Macharis, "Improving urban freight transport sustainability: Policy assessment framework and case study," *Research in Transportation Economics,* vol. 64, pp. 26-35, 2017.

[79] S. Hernandez, "Estimation of average payloads from weigh-in-motion data," *Transportation research record,* vol. 2644, no. 1, pp. 39-47, 2017.

[80] J. D. Regehr, K. Maranchuk, J. Vanderwees, and S. Hernandez, "Gaussian mixture model to characterize payload distributions for predominant truck configurations and body types," *Journal of Transportation Engineering, Part B: Pavements,* vol. 146, no. 2, p. 04020017, 2020.

[81] L. F. Macea, L. Márquez, and H. LLinás, "Improvement of axle load spectra Characterization by a mixture of three distributions," *Journal of Transportation Engineering,* vol. 141, no. 12, p. 04015030, 2015.

[82] E. I. Kaisar, D. Liu, and T. Ardalan, "Evaluation of truck tonnage estimation methodologies," 2019.

[83] N. Eluru *et al.*, "Freight data fusion from multiple data sources for freight planning applications in Florida," 2018.

[84] M. Olfert, "Feasibility of a portable weigh-in-motion system for axle load data collection on secondary highways," 2021.

[85] D. Rivera-Royero, M. Jaller, and C.-M. Kim, "Spatio-Temporal Analysis of Freight Flows in Southern California," *Transportation Research Record,* vol. 2675, no. 9, pp. 740-755, 2021.

[86] FDOT, "Truck Empty Backhaul A Florida Freight Story-Transportation Data and Analytics Office Florida Depart of Transportation," 2018.

[87] J. C. Anderson, A. Unnikrishnan, M. A. Figliozzi, and S. Hernandez, "Develop New Methods to Use ODOT Weigh-in-Motion Data for Predicting Freight Flow and/or Commodity Patterns," Oregon. Dept. of Transportation. Research Section, 2020.

[88] H.-L. Hwang, H. Lim, S.-M. Chin, C. R. Wang, and B. Wilson, "Exploring the Use of FHWA Truck Traffic Volume and Weight Data to Support National Truck Freight Mobility Study," Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2019.

[89] S. Hernandez and K. Hyun, "Fusion of weigh-in-motion and global positioning system data to estimate truck weight distributions at traffic count sites," *Journal of Intelligent Transportation Systems,* vol. 24, no. 2, pp. 201-215, 2020.

[90] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 233-240.