DATA-INFORMED COGNITIVE APPROACHES TO THE EVALUATION OF WEIGH-IN-MOTION BASED LIVE LOADS FOR BRIDGE RELIABILITY

by

ANANTA SINHA

(Under the Direction of Mi Geum Chorzepa)

ABSTRACT

Evaluation of live load effects has a profound influence on the design, maintenance, and rehabilitation of bridges in the United States. This dissertation investigates three main topics related to live loads. First, statistical methods are used to evaluate high-impact and low-probability bridge overloading events. This study proposes a cognitive approach to evaluating live load factors because stakeholders discount the probability of observing overweight vehicles based on the prospect theory. It quantifies the likelihood of observing extreme weights on bridges from various Weigh-In-Motion (WIM) sites in Georgia using the Extreme Value Theory and examines the dependency of live load factors on the choice of a threshold or an extreme percentile. Subsequently, the process of predicting maximum live load factors is validated using another state's data. It is concluded that a live load factor is affected by a shape parameter and numerically quantifiable for each site, and near-term live load factors are more salient for preparing bridges for high-risk low-probability overloading events. In a subsequent study, National Bridge Inventory (NBI) data is used to gain better understanding of the correlation between bridge capacity and bridge load ratings by means of a supervised machine learning model. Lastly, a data-driven reliability analysis is conducted by employing WIM and NBI data, and their reliability is quantified

using real-life traffic data, in order to mitigate an adverse selection problem when maintaining and replacing bridges. It is concluded that the capacity of bridges is correlated with bridge load ratings, and a WIM data-driven reliability analysis can assist in minimizing information asymmetry and adverse selection problems in bridge safety. The last study analyzes anomalies in vehicle weight data using the Jensen-Shannon divergence method. Additionally, a comparative time series analysis utilizing machine learning and deep learning algorithms is performed to search for a pattern for a time-history of Class 9 vehicle weights, establish a control dataset, and ultimately overcome dynamic inconsistency that may exist when making a data-driven decision for transportation assets.

INDEX WORDS: Weigh-In-Motion Data, Live load Factors, Rating Factors, Bridge Reliability and Risk Assessment, National Bridge Inventory, Bridge Condition Ratings, Extreme Value Theory, Machine Learning for Bridges, Deep learning Networks, Long Short-Term Memory Networks, Weight Forecasting, Weight drift, Jensen-Shannon Divergence.

DATA-INFORMED COGNITIVE APPROACHES TO THE EVALUATION OF WEIGH-IN-MOTION BASED LIVE LOADS FOR BRIDGE RELIABILITY

By

ANANTA SINHA

B.Tech., Meghnad Saha Institute of Technology, Kolkata, India, 2014

M.Tech., Indian Institute of Technology, Guwahati, India, 2017

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2022

DATA-INFORMED COGNITIVE APPROACHES TO THE EVALUATION OF WEIGH-IN-MOTION BASED LIVE LOADS FOR BRIDGE RELIABILITY

by

ANANTA SINHA

Major Professor:	MI GEUM CHORZEPA
Committee:	SUNG-HEE "SONNY" KIM
	JIDONG YANG
	STEPHAN DURHAM

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2022

DEDICATION

This dissertation is dedicated to the loving memory of my mother, Chandra Singha, for her unconditional love and my father, Dibakar Sinha, for his unwavering support and encouragement throughout this academic journey.

ACKNOWLEDGEMENTS

To the extent possible, I would like to acknowledge those who have contributed immensely to the successful completion of my doctorate. I would like to extend a warm thanks to my advisor, Dr. Mi Geum Chorzepa, who has enthusiastically supported my research. Because of your strategic guidance and encouragement, I was able to demonstrate my ability to provide alternative approaches to further sustain bridge asset management. With your intelligence and vitality, I am confident that my research will lead to decades of ongoing analysis. I am not only grateful to have you as my mentor, but also for being there when I needed the most. You are truly the best. Moreover, I am grateful for the guidance and inspiration provided by the advisory committee, which consisted of Dr. Stephen Durham, Dr. Jidong Yang, and Dr. Sonny Kim. I am incredibly grateful for their commitment and their inputs on my research. Moreover, I am grateful to the Georgia Department of Transportation (GDOT) for providing funding for this study. I would like to thank all GDOT staff and personnel who helped make this research possible. A special thanks to Mr. Eric Conklin for his support. I also wish to acknowledge the technical contributions of my colleagues, Alexander Trammell, Abuzar Turabi, Hiwa Hamid, and Brian Oyegbile. I wish to acknowledge the contribution and emotional support of my friends, Subham Das, Swapnanil Banerjee, Arunava Samaddar and Somdut Roy. Finally, with the utmost honor, I acknowledge the support and great love of my family - my mother, Chandra Singha (deceased but always remembered and loved), my father, Dibakar Sinha and my uncle, Dilip Kumar Singha, who supported with all their effort to make my dream come true. To my siblings, Sucharita Sinha and Sudarshana Sinha, who have watched me grow up, I would like to let you know that your support

and friendship in my life and ongoing belief in me, matter. I would also like to express my sincere thanks to my brother-in-law for supporting my journey to a new country and believing in me.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

Chapter 1

1. Introduction

1.1 Background

1.1.1 Vehicle Weight Monitoring Practices

U.S. Department of Transportations are expanding their Weigh-In-Motion (WIM) networks throughout the United States. WIM sensors are capable of recording vehicles characteristics such as axle spacing, speed, axle weight, length, number of axle and many more. The main goal is to implement the weight regulations strictly, to ensure the highway safety, and utilizing the data for long term highway infrastructure maintenance. However, the quality of data depends on several factors, such as environmental forces, sensor malfunctioning etc. Therefore, monitoring the deviation in WIM data is important. Selezneva et al. (2016) provided several guidelines to monitor the WIM weight data, which suggests a confidence level of 95% in weight deviation and the bias should be within 5%. Several independent studies (Chorzepa et al. 2020; Liao et al. 2015) suggested strategies in delivering high quality WIM data and continuously improving WIM performance, which ultimately would help to generate good quality data and could be used for highway asset safety assessment and planning.

1.1.2 Evaluation of Live Load Factors

GDOT maintains an inventory of approximately 15,000 bridges and 25,000 centerline miles of roads. Managing such a huge network of bridges is a challenging task. WIM data provides a

continuous real-time information about the vehicle passing through the bridge network in Georgia. Bridge design and reliability analysis is heavily dependent on the live load models which is used to replicate the traffic load on bridges. Several studies have been carried out to address the impact of growing heavy weight vehicle live loads for the bridges in Georgia. For Service Limit Strength, NCHRP 368 (Nowak 1999; Zhou et al. 2018; Sinha et al. 2022) emphasized on the future extreme expected load, that a bridge can expect for its 75 years of lifetime. Due to the availability of considerable WIM data, it is important to use it for the development of the statistical parameter of live load, which could accurately reflect the future maximum vehicle loads. (AASHTO 2011). These studies emphasized the importance of site-specific live load for the design of bridges rather than using the national average load data, which in most of the cases they might not experience during the service life.

1.1.3 Bridge Condition Monitoring and Performance Measures

Bridges in the United States are continuously being monitored by bridge owners. Bridge condition and safety assessments are done through a combination of different methods mentions in the AASHTO's manual for bridge evaluation (AASHTO 2011). There are three different methods available for bridge evaluation such as allowable stress, load factor, and load and resistance factored methods. There are a significant number of bridges in the United States which are 50 years or older. To assure the safety, bridge risk and reliability, several studies (Wang et al. 2011; Frangopol et al. 2008) were carried out to improve the current bridge rating process. However, an easy approach for reliability assessment of existing bridges using available real-time traffic data and bridge inventory data does not exist. Therefore, there is a strong need for the application of

WIM data and available NBI bridge capacity data in assessing the risk associated with each bridge in Georgia.

1.2 Problem Statement

United States Department of Transportations (USDOT) are working tirelessly towards expanding their Weigh-In-Motion (WIM) networks throughout the United States. WIM sensors are capable of recording vehicles characteristics such as axle spacing, speed, axle weight, length, number of axle and many more. Having such a high-quality state-of-the-art WIM data is undoubtedly beneficial for the decision makers. However, it is observed that the number of heavy weight vehicle is growing significantly over the years, which poses a great risk to the aging bridge infrastructure in the United States, where a significant number of bridges are 50 years or older. Therefore, to ensure the public safety and maintain a smooth traffic flow, data driven reliability assessment of bridges using advanced statistical and machine learning techniques are necessary. Furthermore, utilizing the real-time traffic data and advanced numerical tools, it is possible to explore, observe seasonal trends and uncover hidden information about the data, which could further help the decision maker in monitoring the WIM data and incorporating it into highway infrastructure improvements in the future.

1.3 Research Objectives

This study aims to develop a data-driven time-dependent bridge reliability and risk assessment. The specific aims of this dissertation are listed below and repeated in each chapter.

Chapter 1 aims to:

a) Review the past methodologies utilizing WIM database and National Bridge Inventory database for bridge maintenance and risk assessment.

b) Review the scope of utilizing WIM database and National Bridge Inventory database in development of sustainable bridge maintenance strategy.

Chapter 2 aims to:

a) Focus on the quality control and quality assessment of WIM data.
b) Developing python scripts to observe the inconsistency in the data and filtering sites for further detailed investigations.

Chapter 3 aims to review following objectives.

a) There exist several methods for estimating future maximum live load factors on highway bridges. Determining the method which is best suited for determining maximum live load factors for bridges.
b) Expected maximum extreme live load factors for Georgia's bridges.
c) Strategic steps for effectively communicating the expected maximum vehicle weights on highway bridges with stakeholders (e.g., transportation agencies and departments of public safety).
d) How do a future high-impact and low probability event is perceived by the decision makers and how hyperbolic discounting becomes a barrier to action now-a-days.

Chapter 4 aims to:

a) Develop a methodology to continuously monitor the safety of bridges in each state utilizing the real time vehicle information from state-of-the-art WIM data and the NBI database.
b) Develop a strategy effectively communicate reliability, resilience, risk, or probabilities of failure based on current traffic [or vehicle weights] on highway bridges with decision makers and stakeholders (e.g., transportation agencies and departments of public safety).
c) Mitigate the adverse selection problem in bridge maintenance.

Chapter 5 aims to:

a) Determining a methodology to quantify monthly vehicle weight deviation and identify the need for calibration of a sensor.
b) Utilizing time series based advanced machine learning models to explore the data structure, seasonality and trend and predict short-term vehicle weight.
c) Ultimately overcome dynamic inconsistency that may exist when making a decision on the quality of the live load data in uncertain times and recalibration needs.
d) Develop a control or reference dataset using WIM data, which could be used for future bridge analysis.

1.4 Research Significance and Scopes

This study is primarily concerned with data-driven bridge reliability and risk assessment. It examines the benefits of using vehicle weight data for long-term/short-term predictions, which could assist in assessing bridge reliability in the near future. A methodology utilizing probability theory and supervised machine learning algorithm is developed to quantify bridge reliability for Georgia bridges. Additionally, this research aims to determine a control dataset from available WIM data using statistical methods and advanced numerical methods, which can be used in bridge design and maintenance work.

Another aspect of this study is to implement the knowledge of cognitive thinking and incorporate concepts such as hyperbolic discounting in decision-making, prospect theory, adverse selection and overconfidence problems, and dynamic inconsistency.

1.5 Research Methodology

Throughout the study, the following tasks are performed in order to meet the above research goals. The methods used are described below.

1.5.1 Literature Review

Before developing the methodology, an in-depth literature review is conducted. The objective is to gather insight into the latest research studies conducted by several DOTs, get information on current practices, learn about evolving WIM technologies, and bridge maintenance practices.

1.5.2 Preliminary Data Analysis

In order to understand the WIM and National Bridge Inventory Data, a preliminary data analysis is conducted. A series of plots is then designed as a means to visualize the seasonal variations in the data and to identify anomalies.

1.5.3 Determining the Influence of Hyperbolic discounting on high Impact Bridge Overloading Event and Estimate Live Load Factors

As part of this chapter, several statistical and numerical methods are employed to estimate the live load factor for several WIM sites in Georgia. This chapter also discusses human hyperbolic discounting in relation to high impact low probability events, as well as strategies to overcome this situation.

1.5.4 Data-driven Bridge Reliability and Risk Assessment

This chapter discusses the utilization of WIM data and NBI data to determine the reliability of bridges and quantify the risks. It also implements machine learning models to estimate the bridge

capacity based on several NBI data attributes. Additionally, it discusses the risks associated with asymmetric information and overconfidence and how they affect decision-making.

1.5.5 Developing a Control Data set and Dynamic Inconsistency

WIM sensors are susceptible to environmental changes and wear and tear from daily traffic. In this chapter, statistical tests and Jensen-Shannon Divergence method are used to quantify the deviation between consecutive months and derive a control dataset. Furthermore, this section discusses the implementation of time series-based forecasting models using advanced machine learning and deep neural networks, such as Autoregressive Integrated Moving Averages (ARIMA) and Long Short-Term Memory (LSTM) networks.

1.6 Organization of the Dissertation

The dissertation is divided into five chapters that describes the investigation for the proposed data-driven bridge reliability analysis and present results from the application of WIM and NBI data, which include:

Chapter 1 presents a background on the previous studies, necessity of the current research. Additional this chapter describes research scope and significance of this study.

Chapter 2 describes a preliminary data analysis conducted to understand the data type, quality, and filter out the usable data.

Chapter 3 describes the necessity of live load factor and proposed a cognitive approach to overcome hyperbolic discounting in human behavior. 10 WIM sites data in Georgia are used in this study. Extreme value theory is employed to estimate the future live load factor for different bridge lengths in Georgia.

Chapter 4 presents a reliability analysis carried out for the bridges in Georgia utilizing WIM and NBI data. NBI database is used to define a relationship between several NBI bridge attributes and bridge capacity using advanced machine learning techniques. This chapter offers an insight into the condition of Georgia bridges and proposes a technique to monitor the risk associated due to the evolving heavy weight traffic.

Chapter 5 presents the monitoring of vehicle weight deviation of weight data using advance numerical tool. This chapter quantifies the monthly weight deviation using a probabilistic approaches. In addition, time-history analysis is performed using advance deep neural networks to better understand the data and provide short-term vehicle weight forecasting.

Chapter 6 presents the conclusion and recommendations for future.

Chapter 2

2. Preliminary Data Analysis

2.1 Introduction

Weigh-In-Motion (WIM) devices are installed in roadways to record a vehicle information about gross weights, axle weights, and axle spacings as vehicles drive over measurement sites without having to divert traffic to static weigh stations. The Georgia Department of Transportation (GDOT)'s Office of Transportation Data (OTD) currently manages WIM data collected from more than 29 permanent sites in Georgia. In Fig. 1, GDOT's Traffic Analysis and Data Application (TADA) shows active WIM locations on a map as of June 2022.

2.1.1 Description of WIM sites

The site description is given in Table 1. The active sites produce large WIM data sets, each of which provides the date and timestamp of every passing vehicle at each sensor. The WIM systems record the number of axles, single axle weight, GVW, axle spacing, lane, vehicle class, length of vehicle, gap between vehicles, speed, and/or temperature of the site.

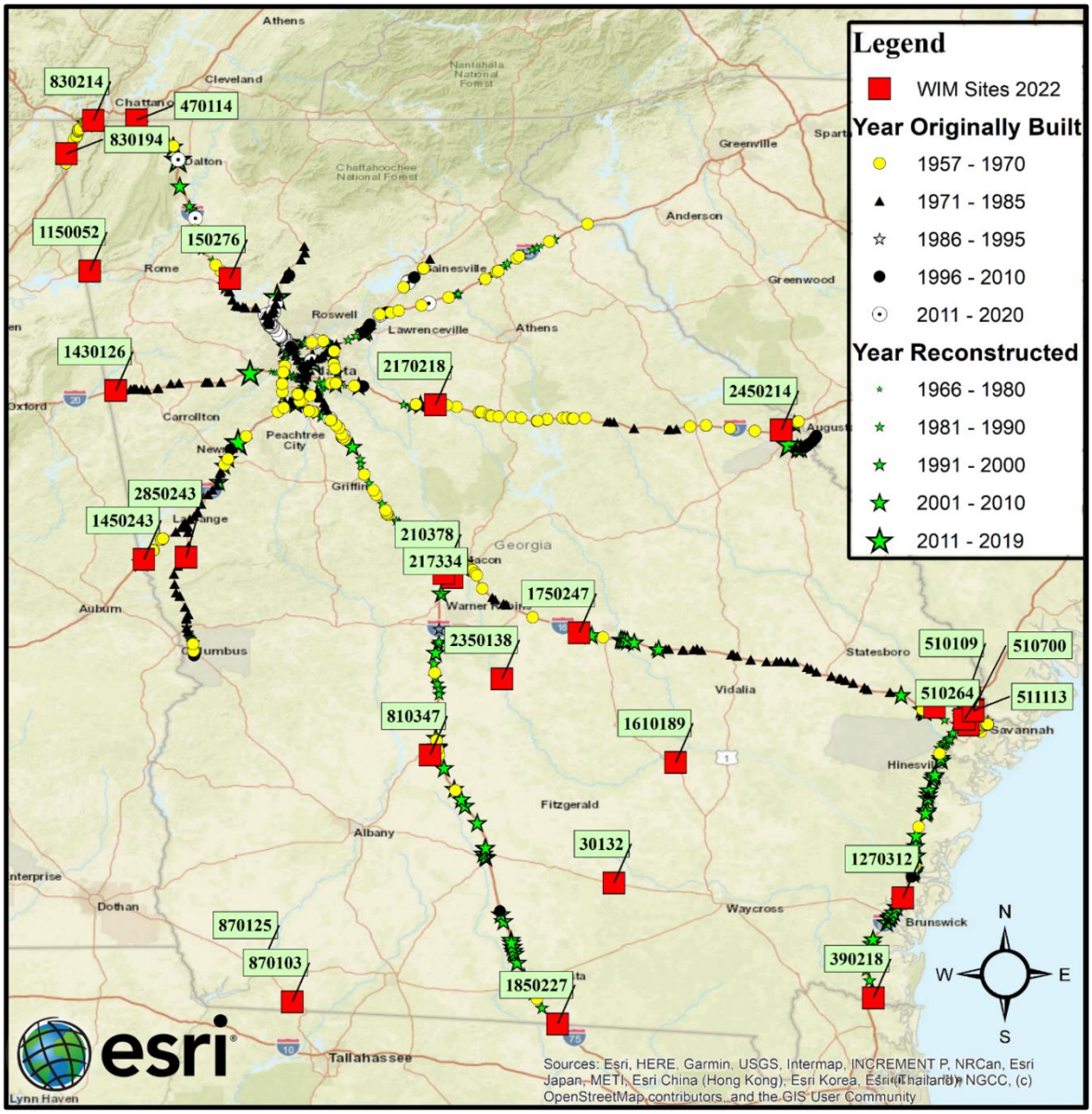


Fig. 1. Map showing WIM sites and Bridge locations in 2022

Table 1. WIM sites description and coordinates

Site ID	Site Description	Latitude	Longitude
510368	I-16 East of Dean Forest exit.	32.06899	-81.1928
1270312	I-95 btwn SR27 & Golden Isles Parkway SR 25 Spur M	31.23438	-81.5093
1430126	I-20 btwn Alabama State line & SR100 Veterans Mem Hwy	33.68077	-85.3022
1850227	I-75/SR401 @FLA SL, Lake Park, Lowndes Co	30.62671	-83.1731
217334	I-75 N of I-475 Split Dr, Macon	32.77959	-83.6806
2850243	I-185 N of SR18 @Dennis Smith Rd, Pine Mtn	32.87801	-84.9622
510387	I-95, 2 mi N. of SR-21 @ SC state line	32.2002	-81.1877
1750247	I-16, 1.4 miles East of SR-338 MP 43	32.51342	-83.0697
470114	I-75 btwn SR146 & Tennessee Line	34.98511	-85.2009
830214	I-24 bn TN State Line & SR299 W Side, Chattanooga	34.98289	-85.4098
830194	I-59/SR406 bn Pudding Rdg Rd & SR136 MP 8	34.82233	-85.5396
210378	I-475 (SR 408) btwn I-75 & SR 22 South of SR 22	32.79937	-83.7212
870103	US-27/SR1 S of US-27BU/SR1BU/E Griffin Ave	30.73387	-84.4519
510700	SR21AL/Jimmy Deloach Pkwy N of Grange Rd	32.14385	-81.1707
510264	US-80/SR26 E of I-95 @Pine Barren Rd, Savannah	32.09632	-81.2134
2450214	I-20 Columbia Co Line & SR415 Bobby Jones Expressway	33.49057	-82.0949
2450218	I-20 E of I-520 @SC State Line, Augusta	33.52746	-82.0191
2170218	I-20 WEST OF SR 11 BTWN SR 142	33.61157	-83.7616
150276	I-75 just above SR 20	34.22139	-84.7521
810347	SR300 W of I-75 & US-41/SR7 @Culpepper Rd, Cordele	31.92337	-83.787
1610189	US-23,341/SR19,27 W of US-221/SR135, Hazelhurst	31.8879	-82.6041
30132	US-82/SR-520 W of US-221,441 nr Sunnyside Ch	31.30711	-82.9015
390218	I-95; bn FL SL & St Marys Rd, Kingsland, GA	30.75257	-81.6508
510109	SR21 N of I-95 & Rice Mill Rd, Savannah MP 16.9	32.22384	-81.1949
511113	Grange Rd W of Jimmy Deloach Pkwy	32.13596	-81.1695
870125	US-27,84/SR1,38 N of SR253, Bainbridge, Decatur	30.91494	-84.5978
1030159	US-80/SR26,30 W of SR17, Bloomingdale, Effingham Co	32.1542	-81.3564
1150052	GA-20 W of SR100 at GA/AL St Line	34.25742	-85.4274
1450243	I-85/SR403 1 mi E of AL State Line, West Point	32.86648	-85.1663
2350138	US-129ALT/SR26,112 S/O SR257(2L)(PLP)	32.29008	-83.4405
1450234	I-85/SR403 1 mi E of AL State Line, West Point	32.86648	-85.1663

2.1.2 Type of WIM sensors

Two load sensor models are primarily used by the vendor: Kistler Lineas quartz sensors (hereafter referred to as quartz sensors) and Roadtrax BL Class 1 sensors (hereafter referred to as BL sensors). In Georgia, both sensors exist in the state's current WIM systems, and they are often compared for their performance. Lanes instrumented with quartz sensors and inductive loops are considered to provide information on both vehicle weight and class,

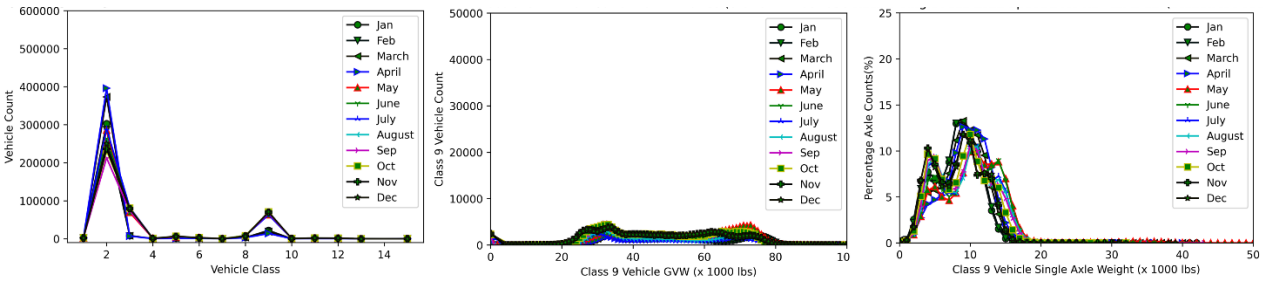
whereas lanes with BL sensors are only useful for obtaining vehicle class information. Although BL sensors record vehicle weight information, the vendor considers the weight data from quartz sensors more accurate and thus more reliable.

2.2 Data Characterization

This section reports an initial data analysis conducted to explore the raw data. The goals of this preliminary analysis is to understand the contents of a WIM dataset and describe the characteristics of WIM data records from all the WIM sites in Georgia. Raw data refers to data that has not been altered or processed for use. For this analysis, the raw data was acquired from Drakewell, the data-hosting services provider. The purpose of data characterization is to obtain information about the WIM records relevant to quality control. During this process, the study team considered the feasibility of reading comma-separated values (CSV) files using Python scripts. The data is then further explored for any missing value, zero weight entities, etc. “NaN” and “NA” rows are removed.

2.2.1 Data Quality Assessment and Quality Assurance

To observe the quality and consistency of the data, Quality control and Quality assurance tasks are performed. To observe the frequency of different vehicle class, monthly vehicle class distribution is plotted and shown in Fig. 2 (a). The monthly Gross Vehicle Weight (GVW) distributions are plotted for class 9 vehicle in Fig. 2 (b). Single axle load spectra for class 9 vehicle is shown in Fig. 2 (c).



(a) (b) (c)

Fig. 2. (a) Vehicle class distribution, (b) GVW distribution, and (c) Single axle load spectra

2.2.2 Gross Vehicle Weight and Vehicle Speed

WIM systems have been widely used by state agencies for quantifying infrastructure usage for maintenance, traffic forecasting, infrastructure investment decision-making, and transportation planning. However, these systems face challenges in obtaining accurate and reliable data due to sensor characteristics, which can be sensitive to driving patterns, weather conditions, speeding, and changes in surrounding pavement conditions (Chorzepa et al. 2020). This section investigates the influence of vehicle speed on GVW measurements.

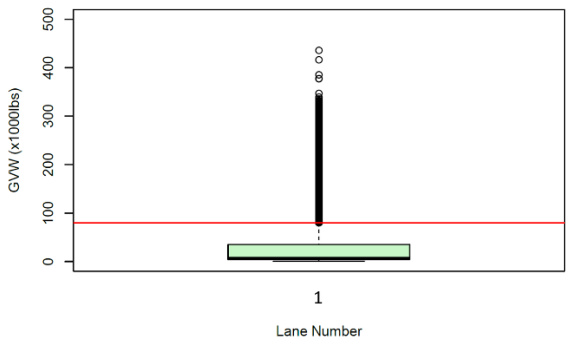
GVW is analyzed by creating boxplots of GVW by lane, scatterplots of class 9 GVW and speed, and boxplots of class 9 GVW by month. The monthly variation in GVW is reviewed to observe a temperature sensitivity in the absence of temperature data.

Fig. 3 (a) and (b) show a boxplot of GVW in each traffic direction, with the red lines indicating the 36,287 kg (80,000 lb) weight limit. Lane 3, the WB slow lane, recorded GVW reaching far beyond the upper whisker (Q3+1.5IQR). F

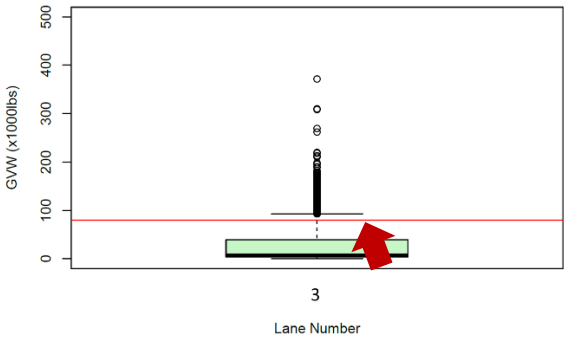
Fig. 3 (c) and (d) show scatter plots of class 9 GVW. The vertical blue lines indicate the 36,287 kg (80,000 lb) limit, and the dotted orange lines indicate the legal speed limit. The vehicles in the upper right quadrant exceed the legal weight and speed limits. According to these plots, as class 9

vehicles become heavier, the driving speed does not appear to decrease significantly. This indicates that increase in speed did not have a direct correlation with increase in vehicle weight although speed in general affects weight in the WIM data.

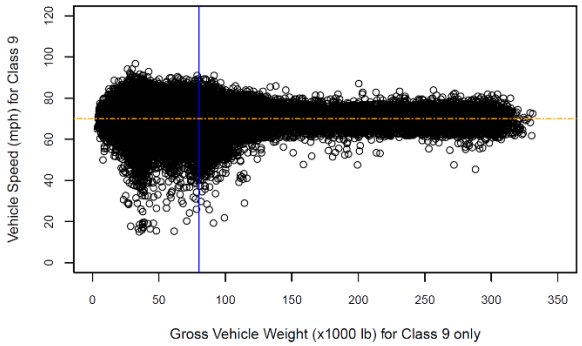
Fig. 3 (e) and (f) show boxplots of class 9 GVW by month. Although no significant month-to-month variation is observed, the median and the number of outliers increases between June and July.



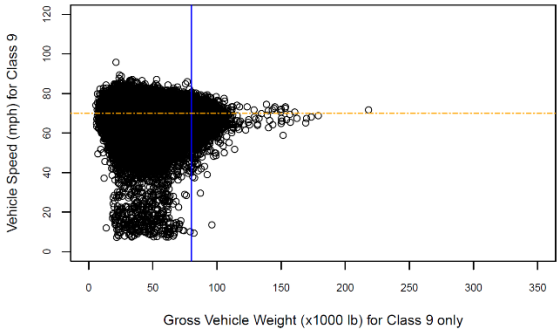
(a) EB lanes, GVW by lane



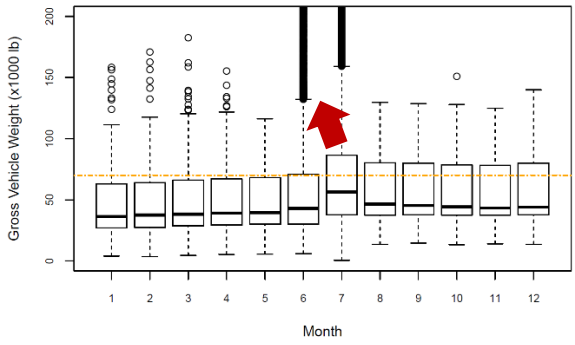
(b) WB lanes, GVW by lane



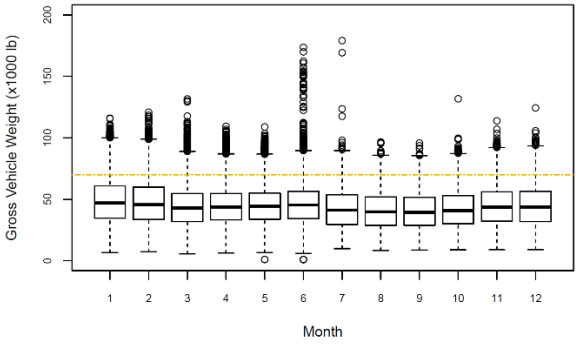
(c) EB lanes, GVW vs. speed



(d) WB lanes, GVW vs. speed



(e) EB lanes, GVW by month



(f) WB lanes, GVW by month

Fig. 3. Analysis of gross vehicle weight for site 175-0274.

Chapter 3

3. Cognitive Approaches to Hyperbolic Discounting of High Impact Low Probability Bridge Overload Events and Live Load Factors

3.1 Introduction and Definition

Many bridges in the United States are at risk due to aging, corrosion, increasingly heavy truck traffic, maintenance practices such as a worst-first approach, and more. All these factors impact the health of bridges and safety of citizens. Live loads (or vehicle weights) are of particular interest to state Departments of Transportation (DOTs) because bridges should be able to safely carry live loads and resist the impact from vehicles during their service life. To ensure the structural safety of bridges within each state's inventory, bridge load rating is required by the U.S. Department of Transportation (USDOT). The Manual for Bridge Evaluation (MBE) (AASHTO 2018) describes the procedure for bridge load rating and design, legal, and permit loads (e.g., HL-93, AASHTO, and state legal trucks) for safe design practices. When load rating bridges, state DOTs desire to limit restrictions for traffic mobility and thus minimize the number of posted bridges.

The design load rating procedure involves the ability of a bridge to withstand the load effect caused by a HL-93 truck defined in the AASHTO LRFD specifications (AASHTO 2017). Bridges are generally rated at two levels: inventory and operating rating. The former corresponds to the live load which safely utilizes an existing bridge for an indefinite period of time (AASHTO 2017), whereas the latter is the maximum allowed load to which a bridge is subjected. This means that

understanding live loads a bridge is expected to take is essential for bridge load rating. During load rating analyses, bridge elements are typically analyzed to be able to support HL-93 loadings magnified (or multiplied) by a live load factor – 1.3 and 1.75 for inventory and operating rating, respectively.

3.1.1 Live Load Effects and Factors

Live load effects refer to normalized shear and moment forces in this paper. Moment and shear force demands from passing a vehicle on a bridge are normalized with respect to moments and shear forces generated by HL-93 loading, respectively. In this study, normalized moment and shear forces represent live load multipliers in reference to the HL-93 loading and are referred to as live load effects on a bridge. Specifically, in this study, normalized moments are used (in place of vehicle weights or shear forces) to illustrate the analytical procedure.

Understanding site-specific bridge live load factors beyond current code requirements is important to decision makers for load rating, traffic mobility, and planning. As more Weigh-In-Motion (WIM) data become available, live load multipliers are numerically quantifiable – in terms of probabilities and expected extreme weights. Ghosn and Frangopol (1996) defined site specific bridge live load factors using WIM data. Several other studies (Enright and O'Brien 2013; Ghosn et al. 2013; Chi 2019) also concluded that current code specified load models underestimated live load factors. There is a need for in-depth qualitative and quantitative research on site-specific live load factors to ensure bridge safety.

3.1.2 Background and Motivation

Transportation agencies and departments of public safety are working to expand WIM networks in order to better understand live load demands on bridges and roads. WIM systems including Bridge WIM are capable of recording vehicle characteristics such as axle spacing, speed, axle weight, length, and the number of axles. It is indispensable to continue collecting high-quality WIM data to establish long-term continuous records, perform load rating for in-service bridges, and ensure highway safety. Researchers have emphasized the importance of recognizing the fact that the probability of observing trucks on a highway varies by WIM sites (Laman and Nowak 1997). The probability of observing weight exceedance (over a threshold) is often quantified for an observation period. Therefore, predicting a probability of such events occurring within a forecast period (of 10, 30, or 75 years) is a common strategy among state DOTs because decision makers (e.g., asset managers and a public finance team) are able to recognize a possible loss from the probability of occurrence of a hazardous event such as a bridge collapse from an overloaded truck(s) or rapid deterioration.

Decision analysis for monitoring and controlling overweight vehicles generally involves estimating the probability of such extreme events with a low probability of occurrence. New Jersey Department of transportation (NJDOT) reported an annual truck count of 45 million in 2011 in NJ, and approximately 6.4% trucks were overweight (Nassif et al. 2018). Evaluating risk of a bridge collapse from a small number of vehicles (e.g., 6.4%) requires decision making skills. Good decisions are typically regarded as those that lead to the identification of robust risk forecast methodologies. However, risk is complex and volatile, often leading to a gap between the calculated and perceived risk.

Probabilistic approaches are generally beneficial when they are incorporated in normative decision-making frameworks utilizing an interpretation of 'risk' as a knowledge practice for informing decision-making (Pidgeon and Butler 2009). Thus, these approaches are used as an instrument to provide appropriate justification for risk-bearing public policy decisions such as over-weight vehicle enforcement. In this context, "normative" decision making means that people maximize their own utility so as to best achieve their goals (Keeney and McDaniels 1992). Probabilistic approaches, in the normative framework, have three main limitations: (1) biases that affect analysts when defining probabilities, (2) difficulty of attaining an agreement on probability distributions, and (3) insufficiency in accounting for all the uncertainties (e.g., a weight drift since the last WIM sensor calibration) involved in a probabilistic analysis.

Furthermore, decision-making on the basis of probabilities requires an understanding of how probabilities change over time. Predictability across timescales and associated uncertainties are expected to vary. Thus, data collected for extreme events may be subject to significant uncertainties due to disruption in traffic patterns (e.g., a pandemic). Extreme events may also be localized. New technologies (e.g., state-of-the-art WIM systems) provide opportunities for real-time monitoring of more precise vehicle weights. Conversely, such non-stationarity data poses a challenge to the quantification of extreme event probabilities.

Nonetheless, as more WIM data become available, probabilistic approaches are expected to effectively communicate uncertainties to decision makers and make the decision analysis process more transparent by quantifying risks (Caprani and OBrien 2010). Therefore, statistical approaches undoubtedly are a useful tool for estimating extreme live loads with a low probability of occurrence. The probabilities characterizing the uncertainty in different states of traffic (i.e.,

overweight vehicles) are derived either from empirical data or by means of statistical modeling. Methods of particular interest involve a probability of exceedance for a threshold value (e.g., a specific vehicle weight) over a pre-defined period of time (Moses 2001).

3.1.3 Research Questions and Scope

The following research questions are developed:

e) There exist several methods for estimating future maximum live load factors on highway bridges. Most existing literature focuses on statistical methods for predicting future live loads and associated factors. How well do the methods predict live loads and how do they differ?

f) Which method is best suited for determining maximum live load factors for bridges?

g) What are the expected maximum extreme live load factors for Georgia's bridges? Live load factors should not be greater than 1.3 or 1.75 for operating and inventory load rating, respectively.

h) What are the strategic steps for effectively communicating the expected maximum vehicle weights on highway bridges with stakeholders (e.g., transportation agencies and departments of public safety)?

i) Based on the Prospect theory, people discount future probabilities and become risk-seeking (i.e., do nothing with a hope of never losing anything). Therefore, one would settle for a smaller present reward such as cost savings now. When presenting a future high-impact and low probability event, is such hyperbolic discounting a barrier to action today?

j) Would sampling a selected percentile or monitoring weekly maximum vehicle weights be more efficient than monitoring all data?

k) Is reporting near-term live load factors more salient and thus more cognitively noticeable when communicating expected maximum vehicle weights?

This study aims to answer and address the above questions. WIM data from 10 sites in Georgia (see Fig. 4) are evaluated to determine live load factors for bridges. Vehicle weight data from heavy truck traffic (FHWA Classes 8 through 13) are used to statistically estimate a live load factor for each site using forecasting periods of 1, 2, 10, and 75 years. A proposed cognitive behavioral approach is considered in association with presenting live load factors. Additionally, an effort is made to validate statistically projected live load factors in 7-year (bridge service) lifetime with vehicle weights obtained from another state's WIM data.

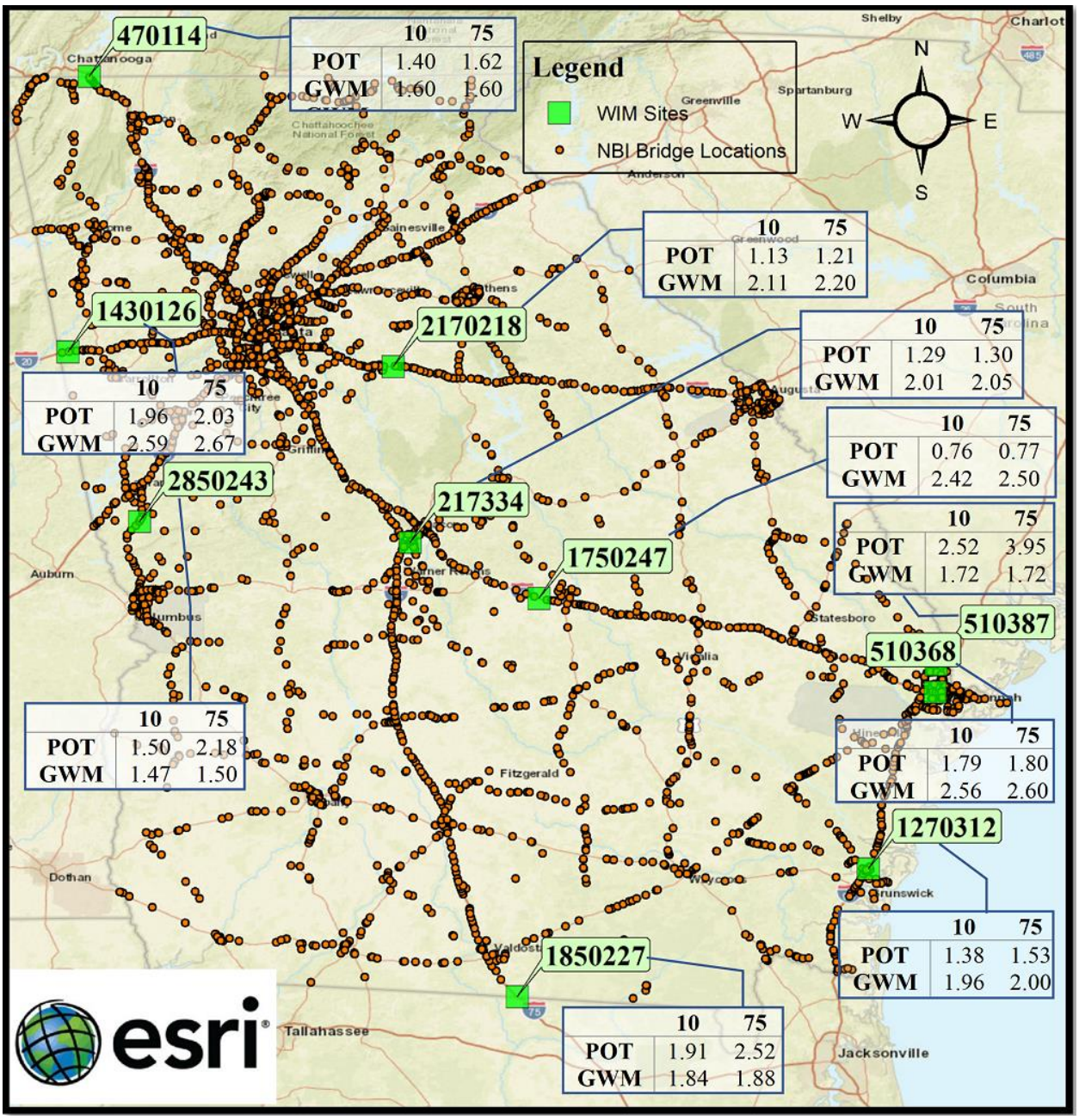


Fig. 4. Map showing locations of bridges and WIM sites.

3.2 Past Methodologies for Evaluating Extreme Live Loads Using WIM Data

This section describes four main approaches of which focus on determining expected extreme normalized live loads (i.e., live load factors).

3.2.1 Nowak's Theory

Nowak's theory (Nowak and Rakoczy 2013) is frequently used because it involves a simplified return-period approach applicable for a wide range of Average Daily Truck Traffic (ADTT) counts. This method predicts live load factors by a generalized approach (Nowak and Rakoczy 2013) using a normal probability plot (Haldar and Mahadevan 2000). That is, the probability of observing live loads (or vehicles) within a forecast period (Nowak and Rakoczy 2013) is determined by Eq. (1), where Φ^{-1} is the inverse normal distribution function, and n is the number of vehicles which is dependent on ADTT and a forecasting time period.

$$Z = -\Phi^{-1}\left(\frac{1}{n}\right) = -\Phi^{-1}\left(\frac{1}{ADTT \times Unique\ days \times year}\right) \quad (1)$$

In Nowak's return period approach, the expected maximum value of normalized live load (e.g., moment/shear) is predicted by Eq. (2) where recursive time periods of 1, 2, 10, and 75 years are selected for analysis (Sivakumar and Ghosn 2011).

$$L_{max} = Mean + Z \times Standard\ Deviation \quad (2)$$

3.2.2 Upper Quantile Selection and Extreme Value Theory

Although Nowak's return period approach is widely accepted in estimating the probability of occurrence of extreme live load effects (or normalized moment and shear forces), a normalized moment distribution is generally highly skewed to the right. Researchers have used a number of different approaches to analyze an α-upper quantile region data to underscore the amount of probability above the α-quantile to predict the future maxima (Zhou et al. 2012; Chen et al. 2015) and considered a Rice distribution, which is skewed to the right, to predict the future maximum

live load factors. Others fit a Gumbel distribution to the data. In the method of selecting upper quantile data, a return period is linearly extrapolated based on available traffic data, which is a shortcoming. Similar to Rice's method, the Extreme Value Theory (EVT) generally deals with a distribution of high quantiles of data. EVT is a popular method because of its simplicity in mathematical expression. EVT is known to be useful for analyzing live loads for small-to-medium span bridges (Enright and O'Brien 2013). However, the rate of utilization of the sample weight data is generally low (Caprani and OBrien 2010; Caprani et al. 2008; Fu and You 2011).

3.2.3 Peak-Over-Threshold (POT) Method

The exceedance-over-threshold, also known as Peak-Over-Threshold (POT) using Generalized Pareto Distribution (GPD), is an alternative approach to the extreme value distributions (Gumbel, Fréchet and Weibull distributions) and is widely used to predict the future possibility of an extreme event in ocean engineering, wind applications, and rainfall prediction (James 2003; Zhou et al. 2016). This approach fits a Generalized Pareto Distribution (GPD) to extreme live load data beyond a threshold value, u. The advantage of this method (over the block maxima method described below) is that it accounts for all live load data greater than a threshold value and does not lose important data (i.e., heavy vehicle weight). GDP is described by Eq. (3), where $\tilde{\sigma}$ is the positive scale parameter, ξ is the shape parameter, and h is the height over threshold value. The threshold is selected by $\sqrt{n}$ and verified using a graphical approach. This process is presented with the results. The shape parameter, ξ, has significant influence in describing qualitative behavior of GPD.

$$G(h) = \begin{cases} 1 - \left(1 + \frac{\xi h}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} & if\ \xi \neq 0, \\ 1 - exp\left(-\frac{h}{\tilde{\sigma}}\right), & if\ \xi = 0 \end{cases} \quad (3)$$

The parameters of the GPD that best describe vehicle weight data, are determined by the Maximum Likelihood Estimator (MLE), which is valid when $\xi > -1/2$. This parameter estimation is essential for determining the expected maximum live load effect as the parameter, ξ, explains the tail data distribution (with a fat or thin tail) and controls the outcome (or extreme live loads).

3.2.4 Gumbel Weekly Maxima (GWM) Method

Another approach is a block maxima sampling method. In this method, a forecast period is divided into n numbers of an equal unit of time. The unit time period is selected as one month (Fu and You 2011), 1 week, or 1 day, and a Gumbel distribution is fitted on maximum live loads per unit time. The block maxima method focuses on the largest value taken from each block and ignores all other high values in the block. A very short recursive period such as a 1-day maxima focuses on the lower end of a dataset rather than the extreme upper tail data, and such approach decreases the goodness of fit of a Gumbel distribution (Sivakumar and Ghosn 2011). On the other hand, the POT method selects live loads above a certain threshold value and thus does not lose much important information. In GWM, load effects (i.e., normalized moments) are represented as random variables ($X_1, X_2, X_2, X_4, \dots \dots X_n$), where n is the nth event (passing of one truck over a bridge). $Y_{n,max}$ is the maximum value. Eq. (4) describes the cumulative probability distribution of $Y_{n\,max}$.

$$F_{Y_{\,n\,max}}(y) = P(X_1 \leq y, X_2 \leq y, \dots \dots . X_n \leq y) = [F_X(y)]^n \quad (4)$$

In Eq. (4), n is the number of load events obtained from the WIM data. The accuracy of the GWM method is heavily dependent on the n value. According to previous studies (Fu and You 2011; Soriano et al. 2017), for repeated high frequency loading events, n, the future statistics of load effect are controlled by the (upper) tail end data in a normalized live-load moment distribution. According to Ang & Tang (Ang and Tang 2007), the extreme value from an exponentially decaying tail will converge asymptotically to a double exponential function if n is very large. The parameters of such functions such as u_n and α_n depend on the initial distribution (Normal/Gumbel), where u_n is the most probable value of Y_n and α_n is the inverse dispersion coefficient. Since the normalized live-load moment data agree with that of a normal distribution, the mean and standard deviation of $Y_{n\,max}$ is obtained by using the mean (μ_x) and standard distribution (σ_x) of the initial distribution as shown in Eqs. (5) and (6). The mean of the projected live load effect is given by Eq. (7), where $\gamma = 0.577216$ (Euler's number) and, the standard deviation is calculated by Eq. (8).

$$u_n = \mu_x + \sigma_x\left(\sqrt{2\ln(n)} - \frac{\ln(\ln(n)) + \ln(4\pi)}{2\sqrt{2\ln(n)}}\right) \tag{5}$$

$$\alpha_n = \frac{\sqrt{2\ln(n)}}{\sigma_x} \tag{6}$$

$$\mu_{max} = u_n + \frac{\gamma}{\alpha_n} \tag{7}$$

$$\sigma_{max} = \frac{\pi}{\sqrt{6\alpha_n}} \tag{8}$$

In the above equations, the forecast period is represented by means of n, where n is the expected number of load events during a forecast period. As an example, for a *R*-year forecast

period, n is calculated by multiplying R and the number of live load event present in a year. μ_{max} and σ_{max} represents the expected future mean and standard deviation, respectively.

3.3 Proposed Cognitive Behavioral Methods for Presenting Live Load Factors

3.3.1 Risk Perception, Prospect Theory, and Temporal Discounting

On the subject of probabilistic losses, the Prospect theory (Kahneman and Tversky 1979) suggests that those averse to risk tend to avoid the chance of losing, whereas those risk-seeking are likely to take chances, in order to possibly lose nothing, and thus feel that they can tolerate the possibility of losing more. For low probability events such as trucks exceeding federal weight limits, people tend to be risk-seeking (few actions on state weight regulations) resulting from a low likelihood of occurrence. The frequency of 'high-risk but low-probability' or extraordinary events has given way to a new 'normal' in the last decade. Recent pandemic, Japanese earthquake and associated nuclear accidents, and Hurricane Katrina are such events. They raise questions about the way in which people perceive risk with a low probability and identify and prepare for disruptive events. Additionally, risks associated with future infrastructure decisions are often discounted. People perceive future events with a low probability as nearly zero probability events – hyperbolic discounting which weighs the value of today's rewards or pleasure. This employs the standard hyperbolic discounting equation of rewards over time shown in Eq. (9) (Blanchard et al. 2013), where $LL_{Computed}$ is computed (or estimated) live load factor for a forecast period, $LL_{Perceived}$ is the discounted or perceived live load factor by the decision makers to the present period, k is the discount factor and D is the pre-reward delay.

$$LL_{Perceived} = \frac{LL_{Computed}}{1 + k * D} \tag{9}$$

Previous studies indicate that it is worthwhile to predict future traffic growth because it projects increased future probabilities of weight exceedance. However, due to a human cognitive bias, taking a near-term projection approach and presenting short-term probabilities of observing overweight vehicles should be more effective [than 75-year probabilities], particularly when state DOTs can react to rapidly growing information such as real-time and continuous flow of vehicle weight data. Ultimately, such an approach should have the advantage of having a positive impact on long-term maintenance.

3.3.2 Perceived vs. Computed Live Load Factors

People underestimate the probability that negatively impact our lives due to cognitive biases and associated optimism.

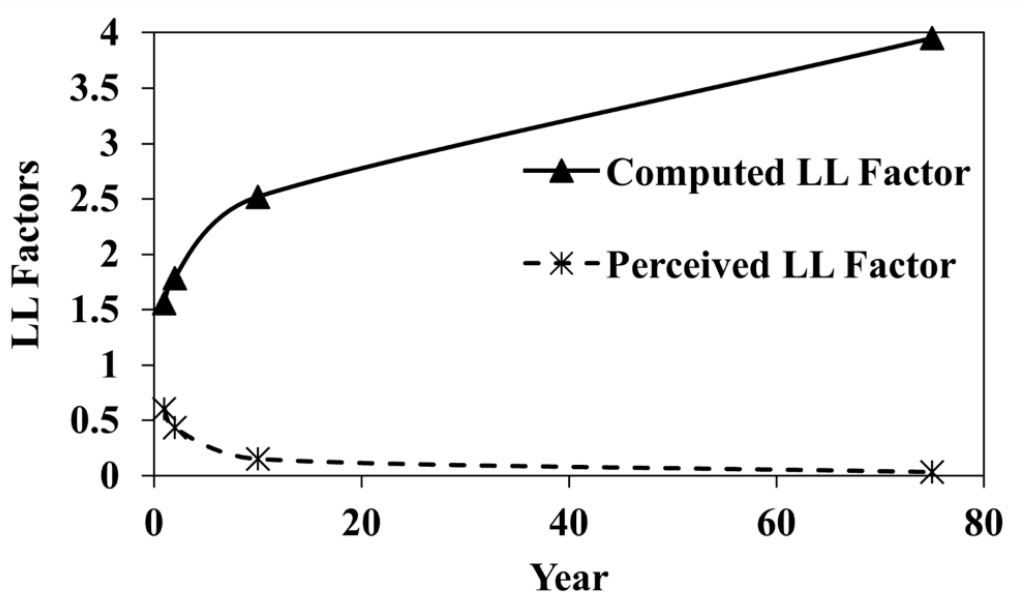


Fig. 5. Computed vs. perceived live load factors for a forecast period

Fig. 5 illustrates how people perceive live load factors in the present time and discount the probability of occurrence of high impact future live load events. For example, when a decision maker sees a live load factor of 3.95 in 75 years, it is perceived as 0.03 now, as a result of hyperbolic discounting (see Eq. (9)). This illustrates why people do not react to over-weight vehicles.

In the following section, live load statistics are determined for highway bridges (see Fig. 4) using available WIM data. Vehicle weight (or live load) distributions are generally right skewed. This means that accurately measuring the weight of heavy vehicles (e.g., an upper quintile data) is critical for predicting extreme live loads. Nowak's method focuses on predicting the extreme expected live load effects in 75-year lifetime. The probability of observing over-weight vehicles on a bridge is expected to increase due to increases in traffic and/or truck weights over time. To make matters worse for bridges, the load carrying capacity or resistance decreases due to deterioration in the members.

In this study, it is proposed to 1) determine a prior probability distribution – a distribution of vehicle weights that will be observed in the future – from past WIM data (with a duration of at least one year) in association with the five approaches listed below, 2) evaluate and compare live load factors, and 3) present expected maximum live load factors in the near term, in order to account for human discounting behavior.

1. Fit a Gumbel distribution to the full dataset.
2. Fit a Gumbel distribution to a selected (upper 5% tail) dataset.
3. Fit a normal distribution to the full dataset and use Nowak’s return period approach (Nowak 1999).

4. Fit a Generalized Pareto Distribution (GPD) to a selected dataset using the Peak Over Threshold (POT) approach.
5. Fit a Gumbel distribution to weekly maxima.

3.4 A Step-by-Step Approach for Determining Live Load Factors

The analysis is first performed using the full WIM data set (without sampling heavy trucks from the tail end) to predict the maximum expected normalized moments (i.e., live load effect). For each WIM site, moment demands for a simple span are calculated using an Influence Line Diagram (Caron et al.). Each vehicle was placed over a bridge to determine the moment and shear forces at critical locations. In Georgia, bridges span between 10m (30 ft) and 55m (180 ft). The maximum bending moment developed by the HL-93 loading (AASHTO 2017) is used as the reference load. In the first approach, a Gumbel distribution is used to characterize the probability of occurrence within an observation period (1 year) and determine the mean of a forecasted distribution using the Extreme Value Theory (see Eq. (4)). Fig. 10 and Fig. 11 show Quantile - Quantile (Q-Q) plots which explain how well the Gumbel distributions fit to the [normalized moment] data.

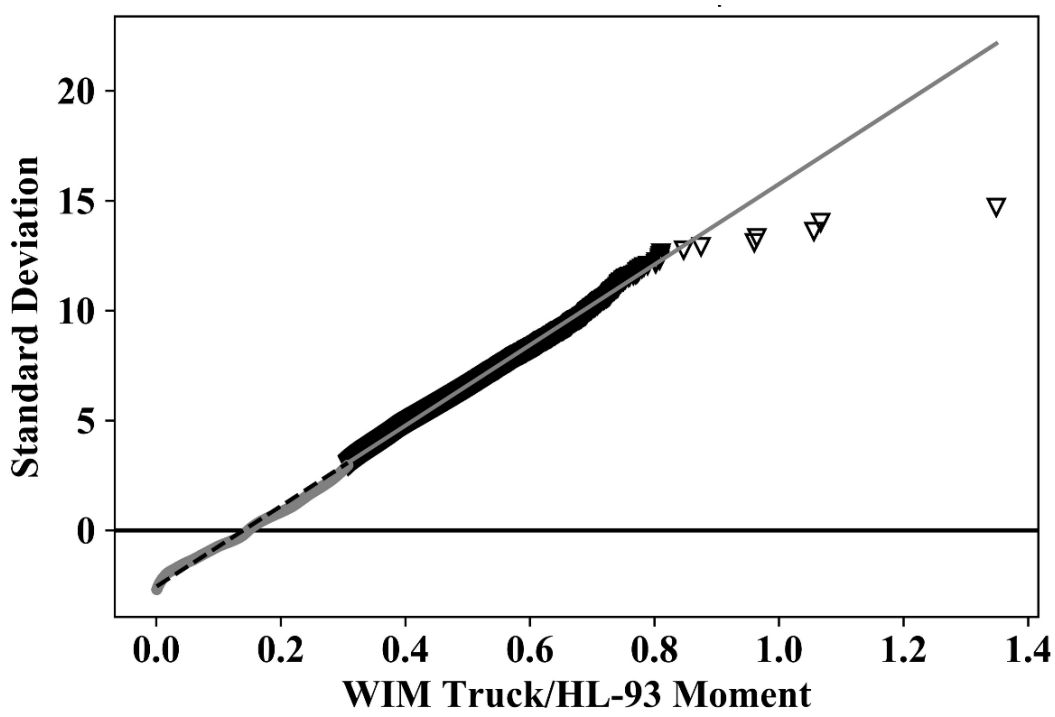


Fig. 6. Inverse CDF plot for site 1270312 using Gumbel distribution (in the methodology right after the Eq. (1) presented as an example).

To analyze data in the higher quantile region, the upper 5% tail end of normalized moment (Soriano et al. 2017) is fitted with a straight line in the second method (Gumbel fitted to the 5% tail data). A regression coefficient is derived with a slope of 1.0 for the Gumbel distribution for site 1270312. For example, Fig. 6 plots the inverse CDF versus normalized moment. μ_x is determined by a straight-line regression on the 5% right-tail data. σ_x is the corresponding standard deviation in the plot. The μ_x and σ_x are used to predict the expected future maximum live load effect using Eqs. (5) through (8). Nowak's return period approach is the third method used in this study. Nowak's approach determines the maximum live load effect (see Eqs.(1) and (2)) assuming a Normal distribution of normalized moments.

POT using a GPD is the fourth method used in this study to analyze extreme upper quantile data. In this approach, two parameters drive live load factors. Firstly, the shape parameter is used to estimate the R-year return level (see Eq. (3)). The determination of a threshold is another important step involved in this method and explained in the section, "Threshold Selected for POT". Once a threshold is selected, Eqs. (10) and (11) determine the mean and standard deviation, where σ and μ of the characteristic distribution are the location and scale parameters, respectively.

$$\sigma = \tilde{\sigma} m_u{}^{\xi} \tag{10}$$

$$\mu = u + \frac{(\sigma - \tilde{\sigma})}{\xi} \tag{11}$$

The block maxima approach is used as the fifth method with one week selected as the unit time. Maximum weekly values are plotted with a Gumbel distribution, and Eqs. (5) through (8) are used to predict the maximum expected live load effect.

Lastly, the five statistical approaches are validated in the next section by analyzing the WIM data available from the Indiana Department of Transportation (INDOT)'s website (INDOT 2020). The degree of variation in live load factors is evaluated with respect to varying forecasting methods and periods. The results are further discussed on the basis of underlying biases and statistical approaches.

3.5 Validation of the Proposed Approach Using Available Data

In addition to GDOT's data, 2013-2020 WIM data publicly available with the INDOT are used to validate the five statistical approaches and study the proposed cognitive approach. Live load

statistics for year 2020 are predicted using the 2013 site data and are compared with the live load statistics determined for Indiana's WIM site ID 951000.

Table 2. Mean and standard deviation of normalized moment (WIM/HL-93) for INDOT location 951000.

Location: 951000 I-64	**Original Data Year 2013**		**Estimated Statistics for Year 2020 using 2013 data**		**Year 2020 (% difference between estimated and actual)**	
Methods	**M**	**SD**	**M**	**SD**	**M**	**SD**
Gumbel (Full Data)	1.36	0.21	1.41	0.19	1.41	0.29
Gumbel (5% tail Data)	1.53	0.26	1.61	0.24	1.90	0.40
Normal dist. Using Nowak	1.34	None	1.40	None	1.37	None
POT	1.29	0.08	1.59	0.27	1.08	0.02
GWM	1.26	0.12	1.28	0.12	1.58	0.34

Table 2 shows the results and indicates that the proposed approach employing five models predict live load (WIM moment/ HL-93) statistics reasonably well. The results presented in the table are for bridges with a 30 ft span length. The first and the third methods (Gumbel and Nowak) predicts the maximum live load fairly well, whereas the second model with a Gumbel distribution fitted to the 5% tail data underestimates the maximum live load. The POT approach yields higher live load statistics when compared with the results generated by GWM. Fig. 7 shows the goodness of fit for the four models noting that the normal distribution used for Nowak's method is intentionally excluded here as the histograms are noticeably right skewed. The 5% tail end data fits reasonably well with a Gumbel distribution as shown in Fig. 7(b). Additionally, Fig. 7(d) indicates that a Gumbel distribution fits well with the weekly maximum data. However, the Indiana data do not fit well with the GPD model (fourth model). Neither do the first (Gumbel with full data) and fifth (GWM) models.

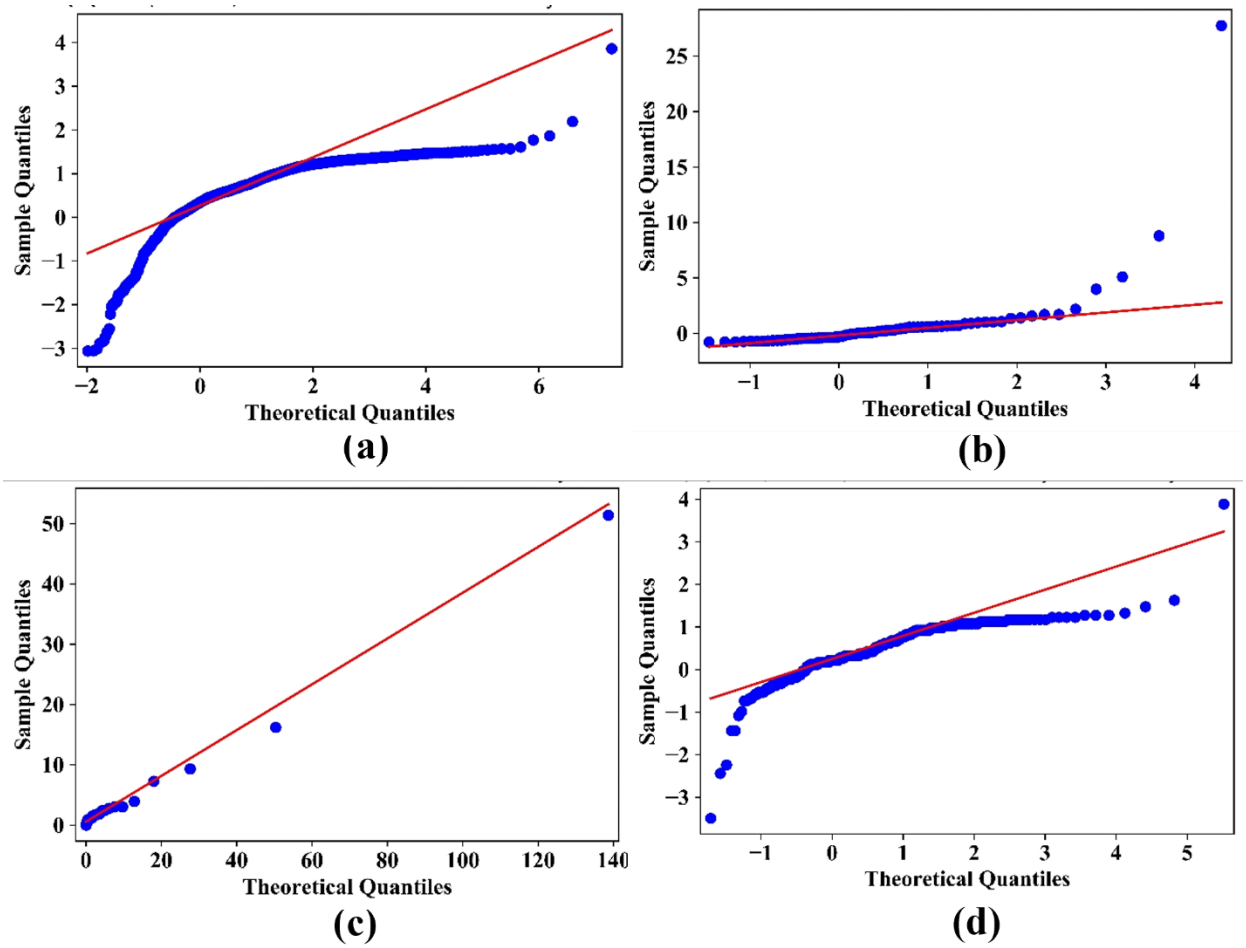


Fig. 7. Q-Q plots for (a) Gumbel distribution with full data, (b) Gumbel fitted on the 5% upper tail end, (c) POT using GPD, and (d) GWM for site 951000 (INDOT).

Table 3. Live load factors of normalized moment (WIM/HL-93) for 30ft span for INDOT location 951000.

Location: 951000 I-64	**Original Data Year 2013**	**Estimated Statistics for Year 2020 using 2013 data**	**Year 2020 (% difference between estimated and actual)**
Methods	M	M	M
Gumbel (Full Data)	2.49	2.49	1.09
Gumbel (5% tail Data)	1.70	1.70	
Normal dist. Using Nowak	1.34	1.39	
POT	1.40	1.97	
GWM	1.70	1.72	

Within the forecasting period of 7 years (i.e., from 2013 to 2020), the expected live load factor remains 1.09. Although this outcome may be attributed to no significant traffic growth at the selected site by 2020 (1.09), the statistical approaches (hypothetically employed in 2013) would have indicated that the expected live load factor reached 1.97 in 7 years according to the POT results in Table 3. The fact that a live load factor prediction exceeds 1.75 in 7 years is not likely to be salient in decision maker's cognitive environment, especially when there is little possibility of unfavorable events that could impact the public safety and traffic mobility. It is more cognitively efficient to present the current year's live load factor (e.g., 1.70 hypothetically back in 2013), according to the standard hyperbolic discounting equation (see Eq. (9)). The live load factor expected to see in 7 years is perceived as 0.16 [=1.97/(1+1.56*7)] now as shown in Fig. 2, which is very low compared to the recorded maximum live load factor (i.e., 1.09 in year 1). Thus, presenting a near-term live load factor of 1.09 is significantly more effective although it does not account for the potential growth in the number of overweight vehicles.

3.6 Results – Live Load Factors for Bridges in Georgia

Fig. 4 shows the expected maximum live load factors over the course of 10 and 75 years. Due to limited space on the map, the results employing POT and GWM are only presented. It is observed that the live load factors are higher at the WIM sites (e.g., 1850227, 0510387, and 1430126) on the interstate highways than those determined for sites on state routes. The following sections show how the live load factors are determined using vehicle weight data obtained from 10 WIM sites. Two sites are selected for illustrating the analysis process and live load statistics. In association with five methods presented earlier, Fig. 8 and Fig. 9 give a graphical display of the two sites' normalized live load data in a histogram and a probability distribution (see solid lines) fitted to the

data. The normal distribution used for Nowak's method is intentionally not presented here as the histograms are noticeably right skewed.

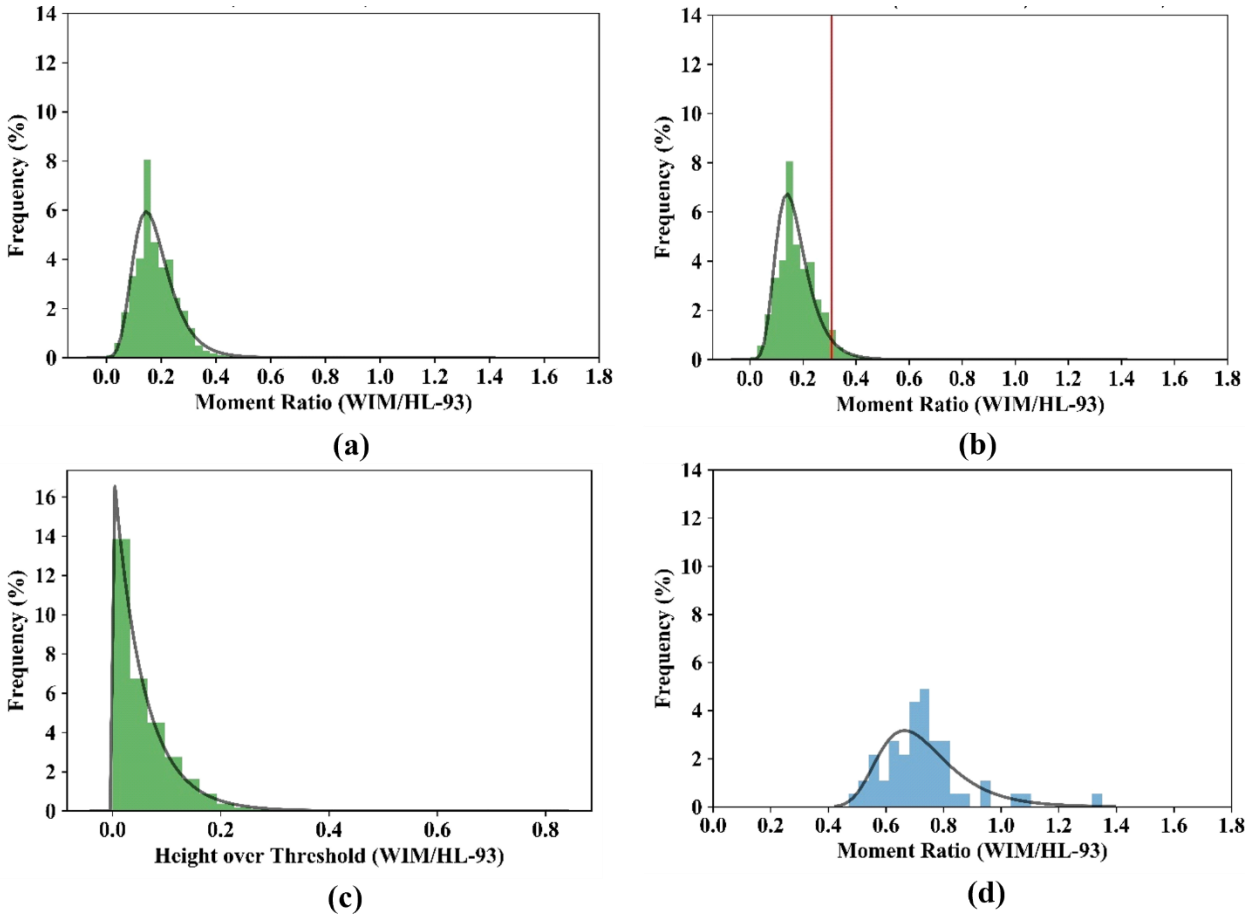


Fig. 8. Histograms of (a) Gumbel distribution with full data, (b) Gumbel fitted on the 5% upper tail end, (c) POT using GPD, and (d) GWM for site 1270312.

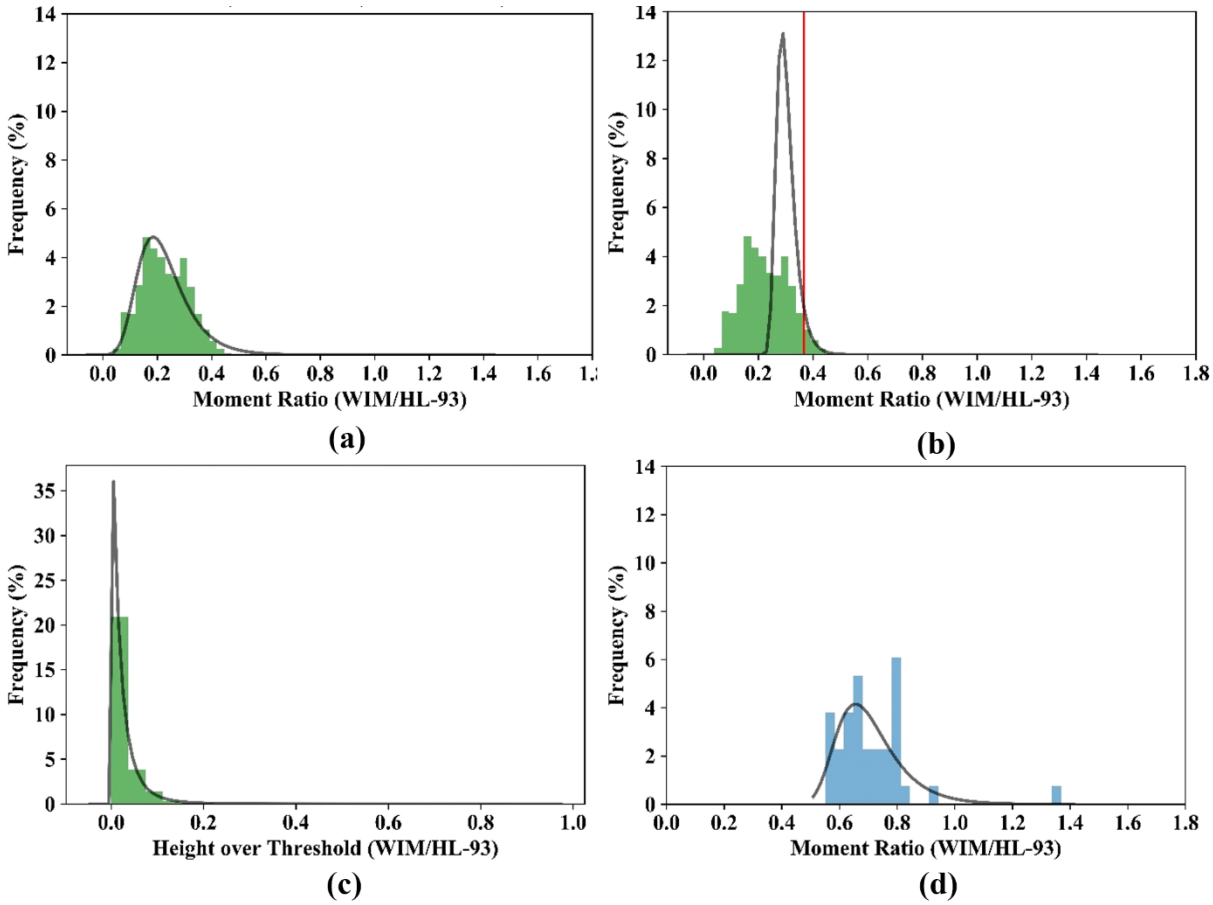


Fig. 9. Histograms of (a) Gumbel distribution with full data, (b) Gumbel fitted on the 5% upper tail end, (c) POT using GPD, and (d) GWM for site 0510387.

Table 4. Mean and standard deviation of normalized moment (WIM/HL-93) for 30ft span for selected sites.

Site ID (State) Approach		Expected maximum normalized moment							
		Year = 1		Year = 2		Year = 10		Year = 75	
		M	SD	M	SD	M	SD	M	SD
1270312 (GA)	Gumbel (Full Data)	0.58	0.16	0.59	0.15	0.61	0.15	0.64	0.15
	Gumbel (5% tail Data)	0.54	0.12	0.54	0.12	0.56	0.11	0.57	0.11
	Normal dist. Using Nowak	0.55	None	0.56	None	0.57	None	0.60	None
	POT using Gen Pareto Dist.	0.99	0.06	1.04	0.07	1.14	0.07	1.28	0.07
	GWM	1.25	0.19	1.27	0.19	1.30	0.18	1.34	0.18
0510387 (GA)	Gumbel dist. Using EVT (Full Data)	0.72	0.17	0.73	0.17	0.76	0.17	0.79	0.16
	Gumbel (5% tail Data)	0.46	0.11	0.47	0.11	0.48	0.11	0.49	0.10
	Normal dist. Using Nowak	0.64	None	0.65	None	0.68	None	0.71	None
	POT using Gen Pareto Dist.	1.11	0.19	1.25	0.23	1.70	0.34	2.60	0.57
	GWM	1.10	0.16	1.11	0.16	1.14	0.16	1.17	0.15

3.6.1 Goodness of Fit

For site 1270312, Fig. 8(b) fits a Gumbel distribution (RMSE = 0.065) to the 5% right-tail data. Fig. 8(d) shows that a Gumbel distribution (p-value = 0.40) fits well to weekly maximum live load data, although the number of data points are limited due to weekly sampling. Fig. 8(c) illustrates a GPD with a selected threshold value (=0.54), which is equivalent to the 0.06% (right tail) data. The following section presents how this threshold value is determined. The GPD fits well to the extreme live load data and shows a reasonable goodness of fit (p-value = 1.00). With a shape factor of 0.02, which is very close to zero, GPD exhibits an exponentially decaying tail as shown in Fig. 8(c).

On the other hand, Fig. 9 presents the results obtained by analyzing site 0510387 live load data. Fig. 9(b) indicates that fitting a Gumbel distribution to the 5% right-tail data (or method 2) does not appear reasonable due to the increased presence of extreme vehicle live loads. The fat tail data fits better with a GPD having a polynomial decaying tail (with a shape factor of 0.25) shown in Fig. 9(c). When compared to Fig. 8(c), the site 1270312 data show a relatively thinner tail.

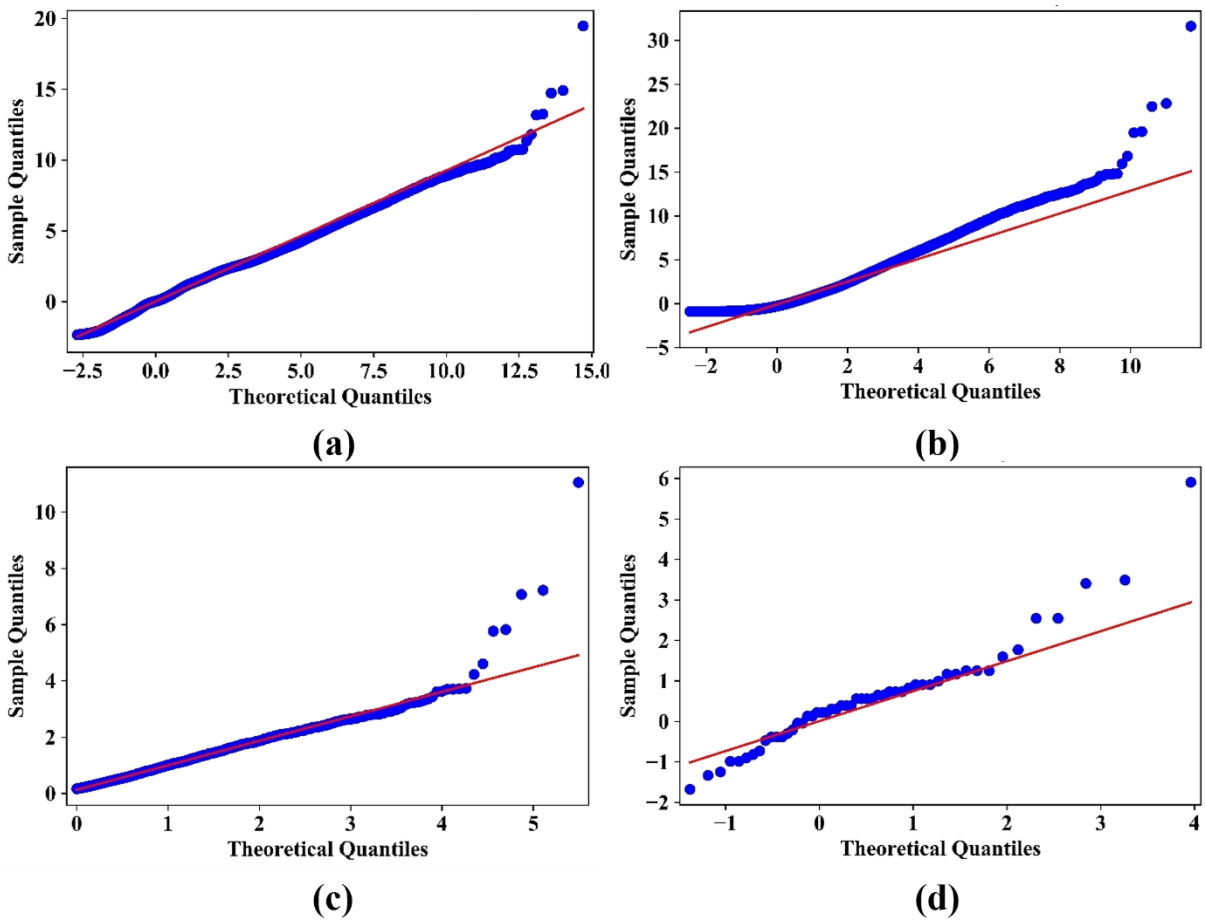


Fig. 10. Q-Q plots using (a) Gumbel distribution with full data, (b) Gumbel fitted on the 5% upper tail end, (c) POT using GPD, and (d) GWM for site 1270312.

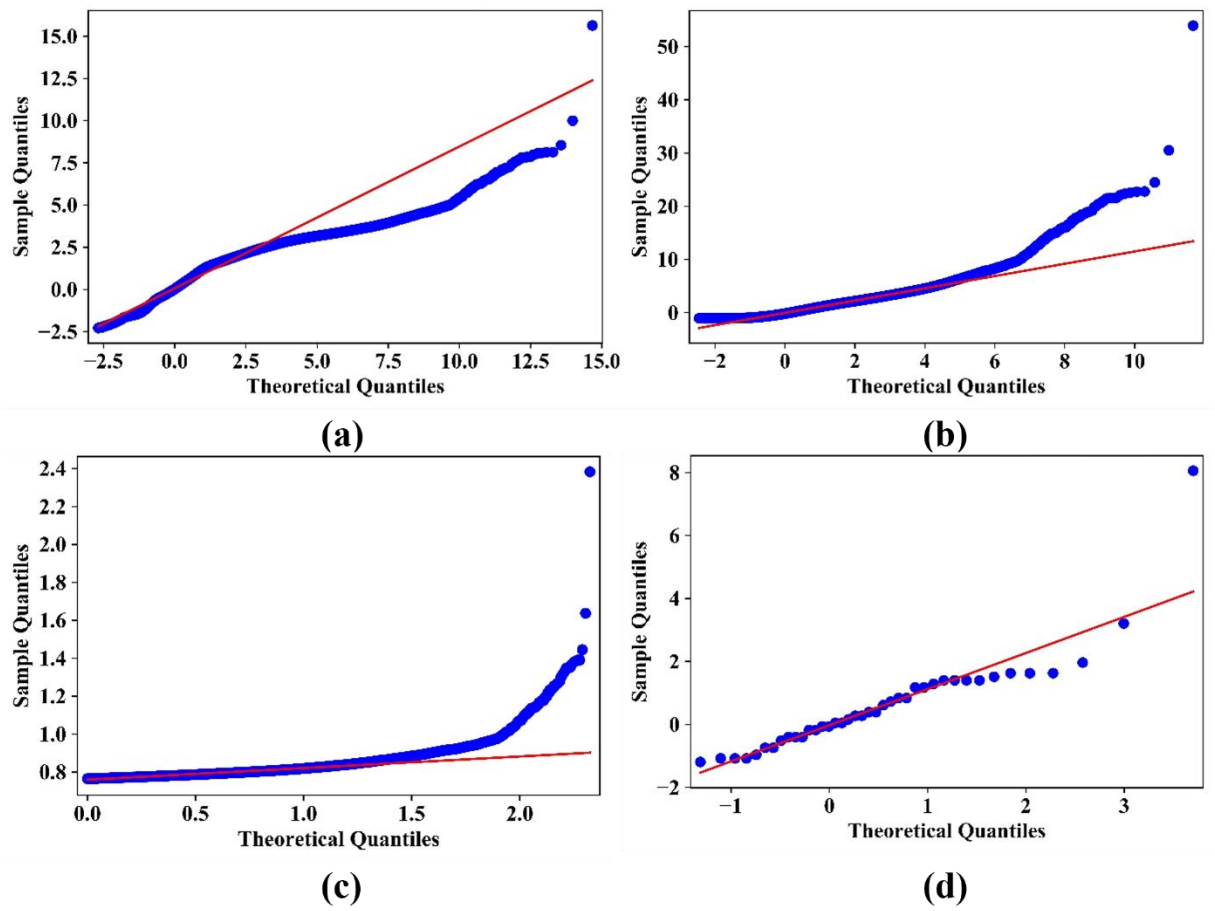


Fig. 11. Q-Q plots using (a) Gumbel distribution with full data, (b) Gumbel fitted on the 5% upper tail end, (c) POT using GPD, and (d) GWM for site 0510387.

Furthermore, it is observed from Q-Q plots that GPD and Gumbel distributions fit reasonably well to the data (see Fig. 10 and Fig. 11). However, they deviate from a couple of extreme live load data in the higher quantile regions. Overall, the degree of deviation is higher for the sites with the fatter tail or high kurtosis (see Fig. 8 and Fig. 9). Kurtosis defines the tail end data distribution. Higher kurtosis indicates that the tail end contains extreme values.

3.6.2 Threshold Selected for POT

The accuracy of the POT approach depends heavily on the selection of the threshold value. The threshold is selected by reviewing Fig. 12. The Mean Residual Life (MRL) plot (Davison & Smith,

1990) shown in Fig. 12(a) shows the relationship between the threshold and mean over the threshold. The parameter stability plot shown in Fig. 12(b) shows the relationship between the threshold and shape parameter. The parameter stability plot shows the change in linearity of shape parameter of GPD with increasing threshold values. After reviewing both plots, the threshold value is determined 1) in the region where the MRL plot plateaus, not in the erratic region, and 2) where the parameter stability plot has a constant slope. The threshold value selected for site 1270312 is 0.54.

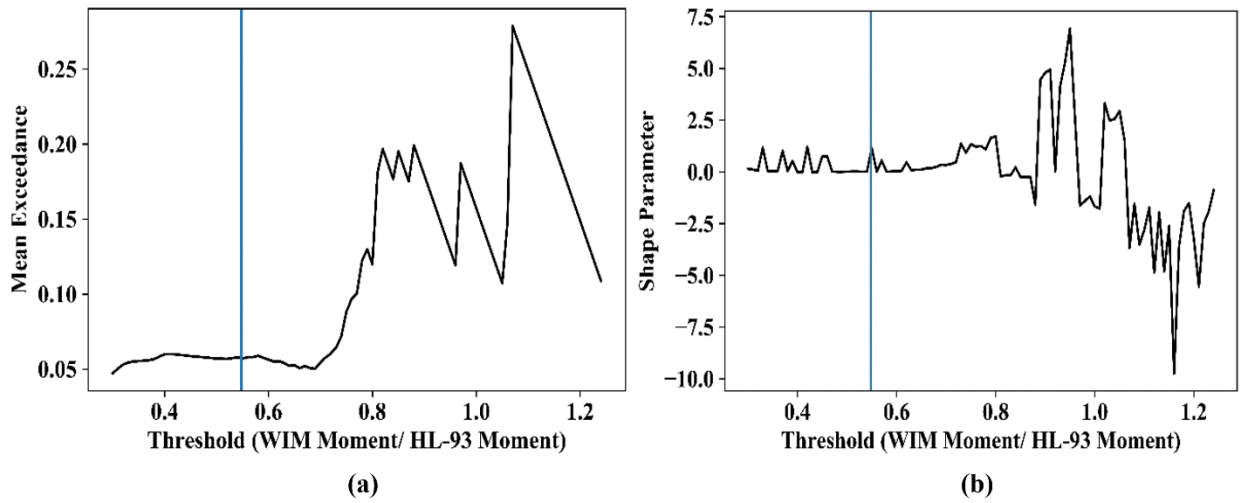


Fig. 12. (a) Mean exceedance of threshold, (b) variation of shape parameter with threshold for site 1270312.

3.6.3 Live Load Statistics

POT and GWM predict higher live load statistics as they focus on the upper tail end data (or extreme live loads). This is reasonable because the two methods use approximately 0.05% right tail data. In the POT approach, which uses a GPD, the outcome largely depends on the shape factor. A higher shape factor indicates the presence of a fat tail in the data. The shape factor for sites 1270312 and 0510387 is 0.02 and 0.25, respectively. In case of site 1270312, the shape factor is close to zero. Therefore, the GPD fits to the live load data with a relatively thin tail. When the

shape factor is significantly higher than zero, POT predicts the highest live load statistics. This is because the GPD fits to the data with a fat tail. For the sites (including 0510387) studied in this study, a GDP with a shape factor exceeding 0.25 results in higher live load statistics.

3.6.4 Sensitivity of Live Load Statistics on Percent Tail Selected

For cases where a Gumbel distribution is fitted to the data with a fat tail (site 0510387), the percent selection below 5% (in the upper tail) has no significant effect on live load statistics. On the other hand, the expected normalized live loads with a thin tail are affected by the percent extreme live load selection (see Table 5 and Table 6)

Table 5. Sensitivity of mean and standard deviation on the percentage in the right tail of Gumbel distribution for 30ft span for selected sites.

	Year=	**1**		**2**		**10**		**75**	
Site ID	Tail (%)	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
1270312	3.0	0.57	0.12	0.58	0.12	0.59	0.12	0.61	0.11
	4.0	0.55	0.12	0.56	0.12	0.57	0.12	0.59	0.11
	5.0	0.54	0.12	0.54	0.12	0.56	0.11	0.57	0.11
	6.0	0.52	0.12	0.53	0.11	0.54	0.11	0.56	0.11
	7.0	0.51	0.11	0.52	0.11	0.53	0.11	0.54	0.11
0510387	3.0	0.46	0.11	0.46	0.11	0.47	0.10	0.49	0.10
	4.0	0.46	0.11	0.46	0.11	0.47	0.10	0.49	0.10
	5.0	0.46	0.11	0.47	0.11	0.48	0.11	0.49	0.10
	6.0	0.46	0.11	0.47	0.11	0.48	0.11	0.49	0.10
	7.0	0.47	0.11	0.47	0.11	0.48	0.11	0.50	0.10

Table 6. Sensitivity of mean and standard deviation on the percentage in the right tail of Gumbel distribution for varying span length for site 1270312.

Site ID		Year =1	Year =2	Year =10	Year =75
Span (ft)	Tail (%)	Gumbel	Gumbel	Gumbel	Gumbel
30	3.0	0.72	0.73	0.74	0.77
	4.0	0.72	0.72	0.74	0.75
	5.0	0.71	0.72	0.75	0.75
	6.0	0.71	0.72	0.75	0.75
	7.0	0.71	0.72	0.73	0.75
60	3.0	0.90	0.92	0.93	0.93
	4.0	0.91	0.95	0.90	0.93
	5.0	0.91	0.93	0.91	0.93
	6.0	0.91	0.89	0.91	0.93
	7.0	0.87	0.88	0.91	0.93
120	3.0	0.80	0.80	0.78	0.80
	4.0	0.80	0.77	0.78	0.80
	5.0	0.76	0.77	0.78	0.76
	6.0	0.76	0.77	0.78	0.76
	7.0	0.76	0.77	0.78	0.76

3.6.5 Live Load Factors and Temporal Discounting

Table 6 shows the live load factors for selected sites in Georgia. The live load factors are determined using a 5% significance level (95% confidence interval) of the mean value for each site shown in Fig. 4. It is observed that sites (2850243 and 0510387) with data exhibiting a fat tail (or with larger shape factors) yield higher live load factors in extreme value predictions. The POT method fits a GPD to the peak data over the threshold value with a shape factor greater than 0.20 for Site No. 2850243. On the other hand, the GWM method predicts higher live load factors for sites having a shape factor of less than 0.20 (e.g., Sites 1270312, 217218, and 0470114). Overall, Nowak's return period approach using a normal probability distribution predicts the lowest live load factors.

Table 7. Estimated live load factors (WIM/HL-93 moment) for 30ft span for all sites.

Site ID (Shape Factor)	Span 30 ft Method	Year = 1	Year = 2	Year = 10	Discounted LL Factor to the Present Period	Year = 75	Discounted LL Factor to the Present Period
1270312 (0.02)	**Gumbel Full Data**	1.16	1.14	1.16	0.07	1.19	0.01
	Gumbel Fitted on 5% Tail End data	1.03	1.04	1.07	0.06	1.06	0.01
	Normal dist. Using Nowak	0.32	0.33	0.34	0.02	0.35	0.00
	POT (threshold = 0.06% upper tail)	1.22	1.27	1.38	0.08	1.53	0.01
	GWM	1.94	1.96	1.96	0.12	2.00	0.02
1430126 (-0.02)	**Gumbel Full Data**	1.54	1.56	1.56	0.09	1.60	0.01
	Gumbel Fitted on 5% Tail End data	2.35	2.39	2.42	0.15	2.47	0.02
	Normal dist. Using Nowak	1.20	1.23	1.28	0.08	1.35	0.01
	POT (threshold = 0.05% upper tail)	1.85	1.88	1.95	0.12	2.03	0.02
	GWM	2.55	2.58	2.59	0.16	2.67	0.02
1850227 (-0.07)	**Gumbel Full Data**	1.23	1.26	1.27	0.08	1.28	0.01
	Gumbel Fitted on 5% Tail End data	1.03	1.04	1.04	0.06	1.05	0.01
	Normal dist. Using Nowak	0.59	0.60	0.63	0.04	0.65	0.01
	POT (threshold = 0.06% upper tail)	1.14	1.16	1.21	0.07	1.26	0.01
	GWM	1.82	1.84	1.84	0.11	1.88	0.02
0217334 (-0.14)	**Gumbel Full Data**	1.25	1.27	1.28	0.08	1.29	0.01
	Gumbel Fitted on 5% Tail End data	1.37	1.38	1.40	0.08	1.45	0.01
	Normal dist. Using Nowak	0.63	0.64	0.67	0.04	0.71	0.01
	POT (threshold = 0.18% upper tail)	1.26	1.27	1.29	0.08	1.30	0.01
	GWM	1.99	2.00	2.01	0.12	2.05	0.02
0510368 (-0.15)	**Gumbel Full Data**	1.46	1.47	1.47	0.09	1.51	0.01
	Gumbel Fitted on 5% Tail End data	2.20	2.25	2.33	0.14	2.40	0.02
	Normal dist. Using Nowak	0.83	0.85	0.89	0.05	0.93	0.01
	POT (threshold = 0.10% upper tail)	1.78	1.78	1.79	0.11	1.80	0.02
	GWM	2.48	2.50	2.56	0.15	2.60	0.02
2850243 (0.23)	**Gumbel Full Data**	1.25	1.27	1.26	0.08	1.30	0.01
	Gumbel Fitted on 5% Tail End data	0.82	0.81	0.82	0.05	0.83	0.01
	Normal dist. Using Nowak	0.60	0.61	0.64	0.04	0.67	0.01
	POT (threshold = 0.15% upper tail)	1.03	1.15	1.50	0.09	2.18	0.02
	GWM	1.46	1.47	1.47	0.09	1.50	0.01
1750247 (-0.12)	**Gumbel Full Data**	1.44	1.42	1.46	0.09	1.47	0.01
	Gumbel Fitted on 5% Tail End data	1.03	1.04	1.03	0.06	1.05	0.01
	Normal dist. Using Nowak	0.68	0.70	0.73	0.04	0.76	0.01
	POT (threshold = 0.10% upper tail)	0.75	0.76	0.76	0.05	0.77	0.01
	GWM	2.36	2.38	2.42	0.15	2.50	0.02
2170218 (0.08)	**Gumbel Full Data**	1.35	1.37	1.37	0.08	1.40	0.01
	Gumbel Fitted on 5% Tail End data	0.86	0.86	0.84	0.05	0.86	0.01
	Normal dist. Using Nowak	0.67	0.69	0.72	0.04	0.76	0.01
	POT (threshold = 0.09% upper tail)	0.95	1.00	1.13	0.07	1.31	0.01
	GWM	2.05	2.10	2.11	0.13	2.20	0.02
0510387 (0.25)	**Gumbel Full Data**	1.34	1.35	1.38	0.08	1.38	0.01
	Gumbel Fitted on 5% Tail End data	0.86	0.87	0.88	0.05	0.86	0.01
	Normal dist. Using Nowak	0.37	0.38	0.39	0.02	0.41	0.00
	POT (threshold = 0.07% upper tail)	1.56	1.79	2.52	0.15	3.95	0.03
	GWM	1.68	1.69	1.72	0.10	1.72	0.01
0470114 (0.08)	**Gumbel Full Data**	1.21	1.22	1.22	0.07	1.24	0.01
	Gumbel Fitted on 5% Tail End data	1.26	1.27	1.28	0.08	1.32	0.01
	Normal dist. Using Nowak	0.57	0.58	0.60	0.04	0.63	0.01
	POT (threshold = 0.04% upper tail)	1.19	1.25	1.40	0.08	1.62	0.01
	GWM	1.56	1.57	1.60	0.10	1.60	0.01

Table 7 also shows the discounted live load factors for forecasted periods of 10 and 75 years and uses Eq. (9). For site 1270312, a forecasted maximum live load factor is 2.00 in 75 years and 1.96 in 10 years. However, it is 0.12 and 0.02 when they are perceived now when they are expected to occur in 10 or 75 years. The forecasted live load factor for the site 0510387 is 2.52 and 3.95 for a period of 10 and 75 years, respectively. In the present time, they are perceived as 0.15 and 0.03, respectively.

3.7 Analysis of Results and Discussion

Considering the results obtained in this study, the following findings are summarized in conjunction with research questions presented earlier in the paper:

a) Table 7 shows that both POT and GWM methods result in higher live load factors compared to the other methods used in this study. Although GWM gives the highest live load factors for most sites, it should be recognized that this method excludes important information about other significantly heavy loads due to sampling of weekly maxima. However, the GWM is expected to improve with additional WIM data. On the other hand, the POT method includes all loads greater than a threshold value, which is more data-efficient, but may not closely represent the data generation mechanism underlying the target extreme events. As such, it is critical to select a proper threshold that trades off between resembling true extreme load events and being data efficient. This issue may be alleviated as more WIM data are continuously collected.

b) The analysis results indicate that the expected maximum live load factor is anticipated to be significantly higher than 1.75 during the 75-year service life. This study reveals that there is an anticipated difference in live load factors given for the AASHTO design basis

(HL-93) and determined from site-specific bridge load data. Furthermore, the extreme value analysis based on the upper tail end data results in a higher live load factor. It is preferable to choose the maximum live load factor by comparing live load factors determined by all available methods used to consider human cognitive bias (temporal discounting). It has been observed that for two sites having a shape factor greater than 0.20 (sites 2850243 and 0510387), the POT method yields live load factor greater than that of GWM. In this case, the advantage of employing the POT method over GWM is the fact that it has a higher goodness of fit, which positively affect decision makers' confidence in live load factors.

c) Considering the 75-year bridge service-life, the maximum live load factor is 3.95 at Site 0510387 near Savannah. The live load factors are expected to be more than 1.75 at the other sites located except for Site 0470114. The fact – bridges are expected to see vehicle weights that exceed the live load limit within its service life – should be alarming.

d) However, in Table 7, when the maximum live load factor is discounted to the present period, it ranges between 0.01 and 0.02, which is extremely small and thus immaterial. This outcome indicates that it is more effective and thus salient to work with the maximum live load factor in the present period, which ranges between 1.0 and 2.0 (see year 1 results). Therefore, the live load factor of 1.68 at site 0510387 near Savannah is more alarming (than 3.95 in the next 75 years) because it is approaching 1.75, and thus bridge overloading events are perceived imminent. The higher live load factor poses a greater risk to the aging bridges in Georgia. Hence, stakeholders may consider selectively calibrating WIM sensors for heavier gross weight (> 356 KN) vehicles, in order to accurately characterize the upper tail data (0.05%~5%).

e) From Table 7, the results suggest that Georgia bridges will be at high risk if no action were to be taken, and strict regulations may be needed for maintenance and rehabilitation of bridges. Because the probability of observing overweight vehicles remains low, people seek risk in favor of avoiding costly maintenance, rehabilitation, or replacement of bridges. However, if a bridge failure were to occur due to overweight vehicles, it can cost human life, create a havoc in the area, and disrupt the traffic flow. Although this outcome appears highly anticipated based on the live load factors determined in this study, hyperbolic discounting of future probabilities imposes barriers to a timely implementation of [vehicle] weight regulations, resulting in avoidance of such a high-risk low probability event or negligence of care for bridges.

f) The tail end WIM (or extreme weight) data contains heavy weight vehicles, which is the main interest for bridge design and load rating. It is observed that the Gumbel weekly maxima method generally yields greater maximum expected live load factors than the results obtained using the full dataset. Therefore, sampling and storing weekly maximum vehicle weights or weights beyond a certain threshold (0.05%) may be considered to improve efficiency in determining live load factors.

g) State DOTs should review present (or near-term predictions of) live load factors determined from continuous WIM data because they are more salient to stakeholders. It should be recognized that the proposed approach is different from short-sighted bridge management because by cognitively processing high-impact and near-future events, transportation agencies can be better prepared for low probability events ahead of time. That is, the proposed cognitive approach should have a positive impact on long term bridge maintenance.

In closing, for a state transportation agency to benefit from the proposed cognitive behavioral approach, it should (1) understand live load distributions on highway bridges, (2) evaluate expected maximum live loads and probabilities, and (3) more realistically predict near-term live load factors with evolving information about vehicle weight rather than forecasting distant-future live load factors. The latter is not as efficient as the former due to human propensity to engage in temporal discounting.

3.8 Conclusions

In this study, the WIM data from ten sites on major highways in Georgia are investigated to statistically determine maximum live load factors for bridges. It is concluded that a highly right-skewed distribution exists in describing vehicle live load data, and the extent of kurtosis is numerically quantifiable. Within the service life of 75 years, it is concluded that live load factors are expected to exceed 1.75. In addition, it is concluded that the discounted live load factors (0.01-0.02) to the present period is perceived as extremely small vehicle weights on bridges due to hyperbolic discounting behavior. To overcome such cognitive bias, it is more perceptible for decision makers to work with bridge live load factors determined for the present (or near-term) period than obtaining 75-year service-life predictions. Based on the findings of this study, the following conclusions are made:

- Analyzing weekly maximum vehicle weight is efficient, although not sufficient, to predict the maximum live load factors because weekly maxima overall yield the highest live load factors, noting that this method is not able to capture all vehicle weights.
- While GWM predicts the highest live load factors, state agencies with a live load distribution exhibiting high kurtosis are more likely to benefit from the POT method in predicting expected

maximum live load factors. This is because a higher goodness of fit is observed when the POT approach is used to fit a GPD to heavy vehicle weight data.

- Near-term (or present) live load factors are more salient [than service-life predictions] to decision makers because perceived live load factors are significantly smaller than the forecasted (10-75 years) live load factors.

Chapter 4

4. Mitigating Adverse Selection and Overconfidence in Bridge Maintenance by A Data-Informed Reliability Assessment

4.1 Author keywords:

Reliability Index, Machine Learning, WIM Data, Bridge Maintenance, NBI Condition Rating, Adverse Selection

4.2 Introduction

We face a new challenge because a significant number of bridges in the United States are at risk due to deterioration of bridge components from aging and an increased volume of heavy truck traffic. Load rating and bridge inspection is required by the U.S Department of Transportation (USDOT) to ensure structural safety of bridges. In the U.S., the U.S Department of Transportation (USDOT) requires inspection and load rating of bridges. Due to aging and/or accelerating deterioration, many bridges are in need of being rehabilitated and replaced.

Risk is associated with every structures. However, risk is complex and difficult to predict. A useful indicator of quantifying the risk is to determine probability of failure, which in turn provides the reliability. Failure does not necessarily always mean the collapse of bridge or catastrophic failure, rather it tells us that bridge performance is no longer satisfying for which it was designed. Estimating the probability of failure requires a good level of decision-making skill. Good decision is considered as those decision, which leads to proper identification of robust risk prediction

methodologies. Decision making based on probabilities are generally beneficial if combined with the knowledge of how probabilities changes over time. Therefore, it is salient to utilize the state-of-the-art WIM system to accurately assess the associated risk. As more WIM data become available, probabilistic approaches are becoming more effective to quantifying and communicate risk with the decision makers (e.g., asset managers and a public finance team) and transforming the decision-making process more transparent (Caprani et al. 2008). Therefore, probabilistic approaches are undoubtedly a useful tool for estimating and quantifying the risk and reliability of a bridge. Akgül and Frangopol (2004) and Frangopol et al. (2008) emphasized the importance of reliability based assessment of deteriorating bridges because the live load increases over time whereas the capacity decreases with deterioration. Hence, there is an urgent need to perform a data-driven reliability analysis of bridges utilizing the most recent WIM data.

4.2.1 Background and Motivation

Weigh-In-Motion (WIM) data is useful for understanding the site-specific live load demand as it records information about vehicles such as axle spacing, axle weight, length, and the number of axles. Departments of public safety and transportation are working tirelessly towards expanding the WIM network throughout the United States to obtain the real time vehicle weight information on important highways. National Bridge Inventory (NBI) on the other hand is a database, compiled by Federal Highway Administration (FHWA), where all necessary bridge information is accessible with bridge component ratings.

Bridges are designed for an expected service life of 75 to 100 years including all the necessary load and safety factors and a reliability index of 3.5 (Inventory Level) (Moses 2001) and considered safe if the rating is above 2.5, which is the operating level reliability index (Frangopol

et al. 2008; Estes and Frangopol 2005). Bridges are expected to perform satisfactorily within their service life. Rating factor is an important index to assess the performance of bridge elements (such as deck, superstructure, and substructure). However, rating factors are time dependent and reduces over time due to the deterioration of bridge elements caused by natural forces. Despite the influence of environmental stressors, current heavy vehicular movement (WIM database) is also has a significant impact in rapid bridge deterioration and should be taken into considerations while making bridge maintenance decisions.

However, it can be perceived from decision makers policy regarding bridge maintenance decisions that, they are unaware about the real time traffic growth and information asymmetry is clearly visible. This insufficient information might lead to the overconfidence and adverse selection problem of bridge owners about the bridge capacity and undermine the hidden risk of bridge failure due to rapid vehicle growth. Therefore, there is a need for decision makers and stakeholders to consider the growing vehicle weight demand (WIM Data) and capacity of bridge infrastructure statistically along with the NBI data in the real-time. This will ensure the public safety and optimize long term bridge maintenance cost (Hu et al. 2015; Xie et al. 2018).

4.2.2 Asymmetric Information and Adverse Selection Problem in Bridge Maintenance

Asymmetric information, also known as information failure, occurs when two parties involved in a transaction do not have an equal level of knowledge about the product. Asymmetric information leads to an adverse selection situation, a prevalent problem in behavioral economics, especially in the insurance industry (Sandroni and Squintani 2013). A problem of adverse selection exists in making bridge maintenance and rehabilitation decisions. Bridge owners and decision-makers are not always aware of the current traffic flow and GVW carried by a bridge due to not having access

to the real-time WIM data, which leads to an asymmetric information problem. Asymmetric information, in turn, leads to overconfidence about the strength of bridges.

Furthermore, it is observed that the number of heavyweight vehicles has been growing rapidly over the past few years, which is in many cases more than the anticipated load level. Bridge owners are still not considering the WIM data while making bridge maintenance, which ultimately leads to adverse selection, which could undermine the hidden risk of bridge failure. The issue of adverse selection in bridge maintenance needs to be addressed as it is linked to the public safety and needs to be avoided by conducting data-informed reliability analysis, where decision makers are provided with symmetric information regarding more accurate bridge capacity and true bridge traffic load demand by utilizing the WIM data.

4.2.3 Research Questions and Scope

Although, there are several studies (Frangopol et al. 2008; Estes and Frangopol 2005) related to the reliability of bridges in the United States, proper utilization of real time WIM data and NBI database in developing the strategy of reliability-based bridge maintenance has not been found. This study aims to perform an accurate data driven assessment of demand and capacity of bridges, which eventually yield more accurate reliability assessment for prioritizing bridge maintenance and rehabilitation. Specifically, this study aims to address the following research questions:

1. In recent years, heavy weight traffic is growing significantly on the interstate and major state highways. Is there any way to continuously monitor the safety of bridges in each state utilizing the real time vehicle information from state-of-the-art WIM data and the NBI database?
2. It is hypothesized that bridge element (Deck, Superstructure, and Substructure) condition rating has significant influence on the load carrying capacity of a bridge. Is operating rating is

a true indicator of bridge capacity? What are the attributes in the NBI database that are affecting the bridge capacity?

3. How does one effectively communicate reliability, resilience, risk, or probabilities of failure based on current traffic [or vehicle weights] on highway bridges with decision makers and stakeholders (e.g., transportation agencies and departments of public safety)?
4. Finally, how could we mitigate the adverse selection problem in bridge maintenance?

4.3 Procedures Involved in Mitigating the Adverse Selection Problem - Methodology

4.3.1 Establishing a fair representation of bridge capacity from the NBI data

A reliability analysis of bridges is performed by characterizing live load demands (i.e., vehicle weights) from 15 WIM sites and capacities from the NBI's operating load rating data. Any occurrences where a reliability index is less than the threshold reliability index warrant further investigation on either vehicle loads or bridge capacities as one of the two components is likely to be erroneous. This reliability analysis includes truck traffic (class 4 or greater vehicles).

This study hypothesized that the NBI operating rating, as per the definition given in the NBI guidelines (NBI 2020), provides a fair representation of bridge capacity. To test this hypothesis and establish a relationship between bridge capacity and NBI attributes (e.g., age) associated with condition ratings, a multivariable linear regression analysis is performed between NBI attributes and operating rating. The significance of the regression analysis is to observe the effect of the change in operating load ratings on the change in bridge condition ratings. The process is summarized in Fig. 13 as a flowchart. This hypothesis test supports the fact that the change in bridge load carrying capacity due to deterioration is characterized by the change in NBI operating load ratings.

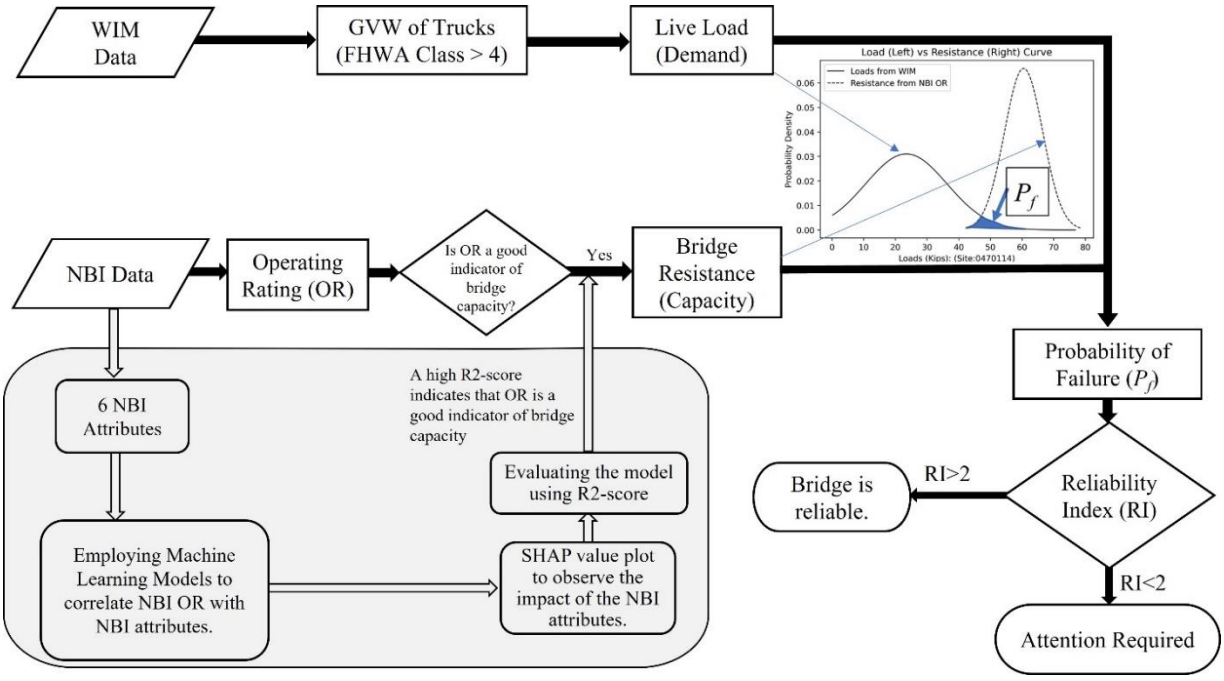


Fig. 13. Flow Chart of the data driven reliability assessment.

As described in the methodology, the strength of a linear relationship is investigated first. If found to be weak, a machine learning model such as eXtreme Gradient Boosting (XGBoost) is incorporated. The predictors are condition ratings for deck, superstructure, and substructure, and the target variable is the bridge operating load rating, which is hypothesized to represent bridge capacity. A SHAP plot is used to determine the impact of predictors on the target variable. Machine learning model parameters such as tree depth and learning rate are tuned using a test data set, and the performance is evaluated by multiple error estimation methods including the Root Mean Squared Error (RMSE), R-squared (R2) value, and K-fold cross validation (Anguita et al. 2012; Jootoo and Lattanzi 2017). Bridges are analyzed with weight data available at the nearest WIM site based on satellite coordinates as shown in Fig. 14.

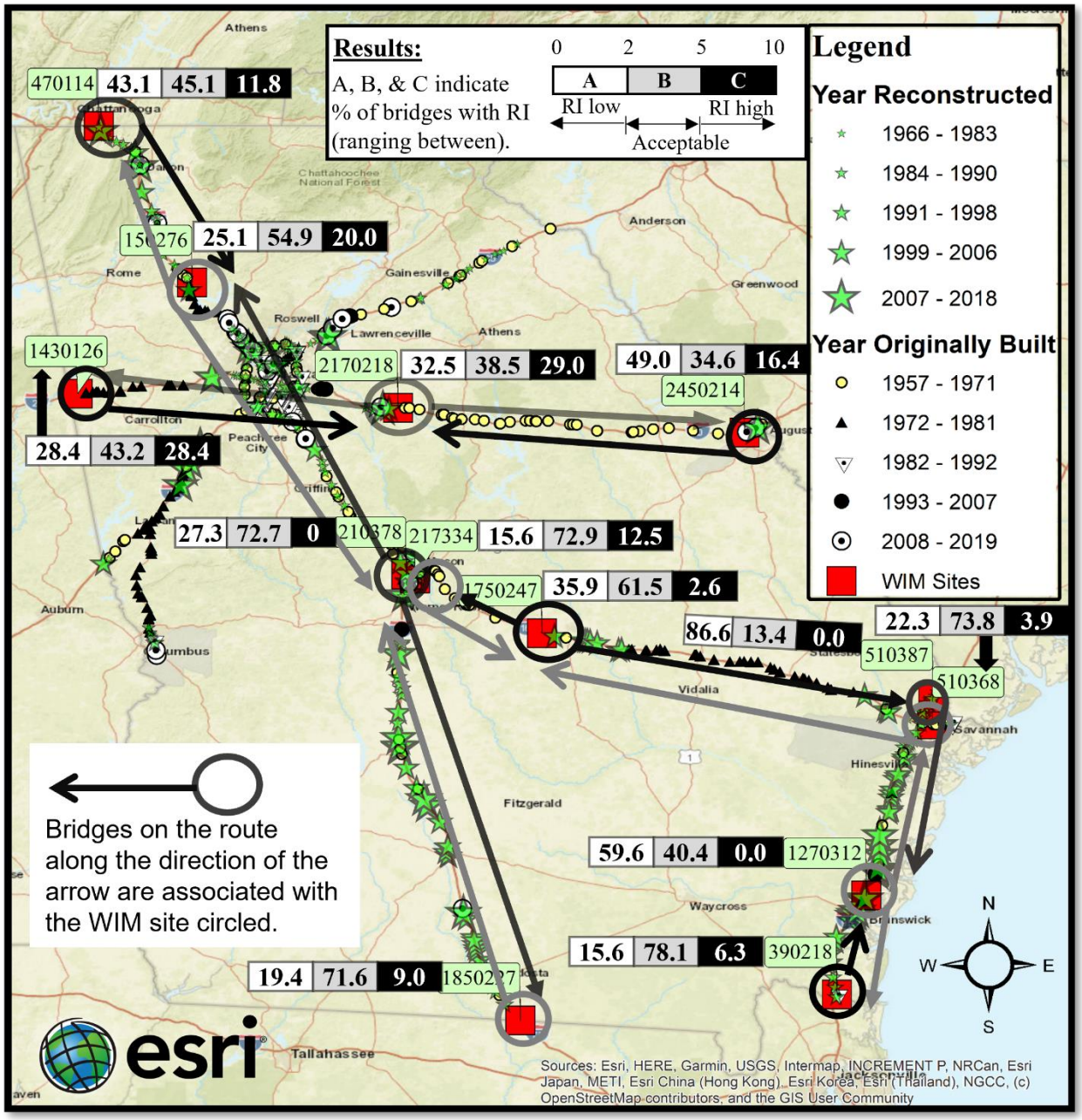


Fig. 14. Map showing locations of old and reconstructed bridges in Georgia.

4.3.2 Procedure for Determining a Reliability Index to Close the Information Gap

A data-driven reliability analysis uses WIM data and quantifies the risk of live loads exceeding structural capacity by calculating a reliability index. Therefore, the resulting outcome should provide a quantitative metrics to the decision-makers regarding the associated risk for each bridge in Georgia and enable decision makers mitigate overconfidence problems in measuring the bridge capacity. Furthermore, it should reduce the miscalculated risk which is hinged on incomplete information resulting from uncertainties in predicting the rate of vehicle growth and bridge deterioration, ultimately reducing the likelihood of adverse selection problems.

The limit state function, g, that forms the basis for a reliability analysis is defined by Eq. (12), where the strength of a bridge and load effect is represented by R and Q, respectively. It is noted that both R and Q are random variables, and thus it is assumed that their actual value cannot be determined with absolute certainty.

$$g = R - Q \tag{12}$$

Structural behavior is considered as satisfactory if the limit state, g, is greater than 0 or $(R - Q) > 0$, and unacceptable when $(R - Q) < 0$. Since the probabilistic distribution of R and Q is not precisely known, a method is derived which is known as the First Order Second Moment (FOSM) probabilistic analysis (Galambos and Ravindra 1981; Nowak 2004), employing the mean and standard deviation of R and Q. In the FOSM analysis, the probability of failure (p_f) is calculated by Eq. (13), where μ and σ are the mean and standard deviation. The reliability Index (β) is determined by Eq. (3), and ϕ^{-1} is the inverse normal distribution function.

$$p_f = \phi\left(-\frac{\mu_R - \mu_Q}{\sqrt{{\sigma_R}^2 + {\sigma_Q}^2}}\right) \tag{13}$$

$$\beta = \phi^{-1}(1 - p_f) \quad (14)$$

While bridges are designed for a reliability index of 3.5, an acceptable reliability index ranges between 2.0 and 5.0. Bridges with a reliability index below 2.0 need special attention as they no longer meet the load carrying capacity required for current traffic. In this study, the following assumptions are made to determine reliability indices of bridges in Georgia:

WIM truck weight and NBI capacity are random quantities, and a normal distribution is used to represent weight data.

Operating load rating is taken as the mean of a load capacity distribution, and the standard deviation is assumed to be 10 percent of the mean value (Estes and Frangopol 2005; Nowak 2004). This is a fairly conservative approach but recognizes that the load rating provided in the NBI is not an absolute measure but is expected to vary due to uncertainties associated with parameters used for bridge load rating.

4.4 Results – Reliability Indices Using WIM Data and Regressions to Characterize Resistance

4.4.1 Correlating Operating Load Rating with Attributes Available in the NBI Data

A supervised ML algorithm implementing the XGBoost model (Chen and Guestrin 2016) is used to define a relationship between the two variables—changes in condition ratings and changes in operating ratings. This is achieved by predicting the operating ratings using the six variables shown in Fig. 15. XGBoost is a decision-tree-based ensemble ML algorithm that uses a gradient boosting framework. This algorithm minimizes loss by employing the mean squared error between two iterative steps. Twenty-nine years (1992–2020) of NBI data are analyzed. Fig. 15 presents a plot showing SHAP values. The plot consists of many dots, where each dot has three main

characteristics. First, the vertical location shows the feature importance, ranked in downward descending order. Second, a different shade of gray exhibits whether the variable has high or low impact on the prediction outcome. Black and white dots indicate a high and low impact variable, respectively, and its position shows the strength of impact on the prediction of the target variable (i.e., the operating rating). Finally, the horizontal axis shows the positive and negative impact of the predictors (e.g., variables such as condition rating) on the target variable. Ultimately, a SHAP plot illustrates the impact of predictors on the target variable prediction.

The results shown in Fig. 15(a) are generated using the maximum depth of 13, the learning rate of 0.3, the estimator parameter of 150, and the lambda of 1 for the XGBoost model. Three methods are used to evaluate the model accuracy: RMSE, R2-score, and K-fold cross-validation. The RMSE value is 6.17, and R2-score is 0.87, both of which are considered good scores for a regression model. In this study, ten folds, with k-1 being the training data (Anguita et al. 2012), are used, and the average accuracy is 87.25 percent with a standard deviation of 0.26 percent.

The XGBoost model results are compared with the results from two other ML models: gradient boosting ML algorithm CatBoost (Prokhorenkova et al. 2018) and simple Random Forest Regression (Breiman 2001) technique. CatBoost is a robust gradient boosting algorithm that yields excellent results in forecasting using fewer data points. On the other hand, the Random Forest Regression is a popular supervised decision-tree-based machine learning algorithm that uses the ensemble technique to determine the final outcome. The CatBoost yields an R2-score of 0.75 and an RMSE of 8.40. The Random Forest model, however, produces an R2-score of 0.86 and reaches an RMSE of 6.20, similar to the results from the XGBoost model.

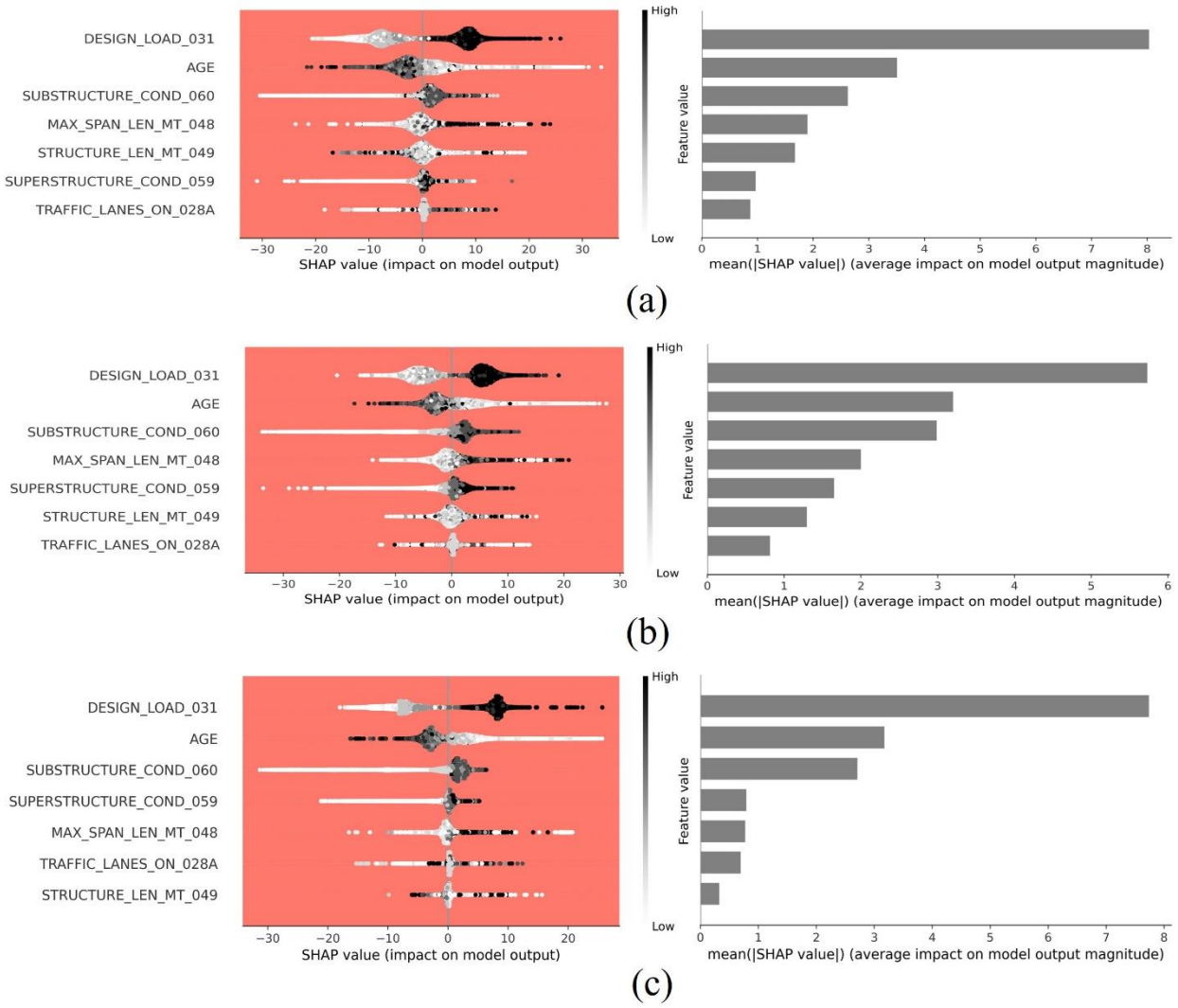


Fig. 15. SHAP value plot of attributes from NBI Database using (a) XGBoost, (b) CatBoost, and (c) Random Forest algorithm.

4.4.2 Reliability Analysis of Bridges in Georgia

As shown in Fig. 14, bridges are associated with a WIM site based on the geographic information system (GIS) coordinates obtained from the NBI database (NBI 2020). A histogram representing GVW distribution is plotted for site 0210378 as shown in Fig. 16. A curve representing the distribution of bridge capacities is also shown in Fig. 17 for each bridge that is located between

two WIM sites. In Fig. 17, the thin dashed gray line shows the capacity of a bridge in 2021, and the thin solid black line shows the capacity when the bridge was constructed. The purpose of this plot is to show the reduction (or increase) of load carrying capacities over time. The blue solid (thick) line shows the live load on a bridge observed in the WIM data, and the red dashed (thick) line shows the bridge capacity when a reliability index of 2.5 is selected for the live load distribution observed in the WIM data.

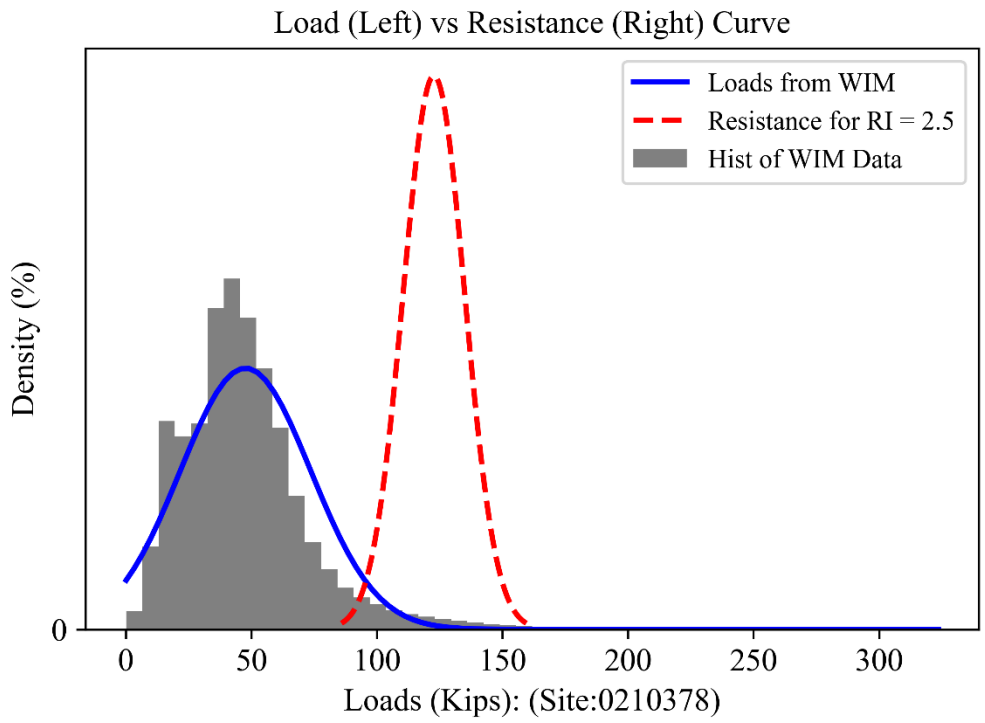


Fig. 16. WIM live load histogram fitted with normal distribution.

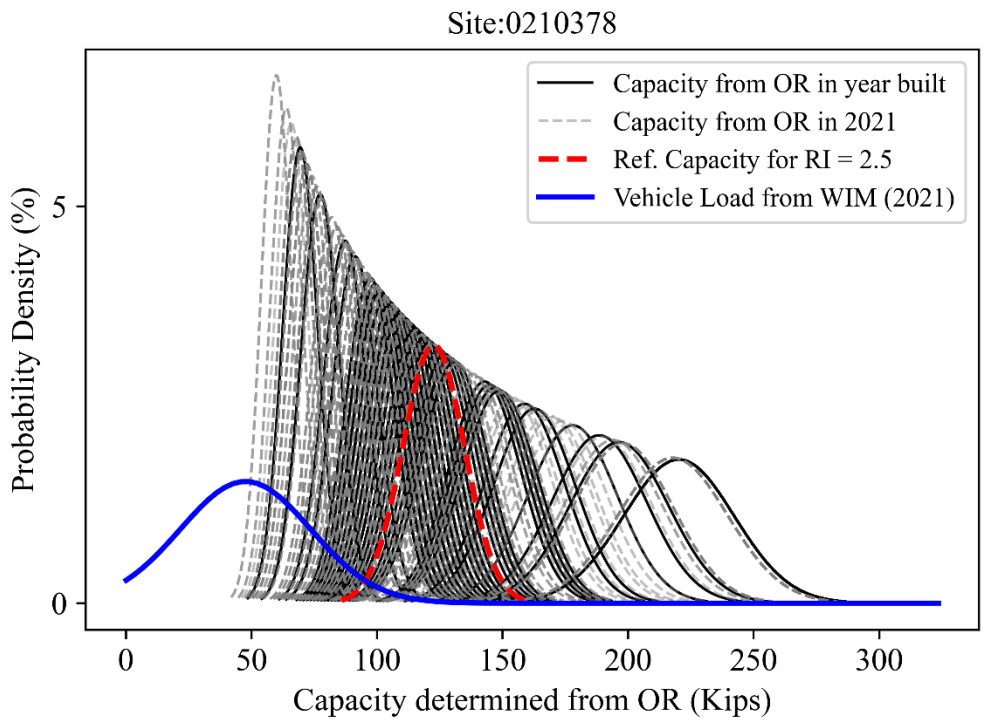


Fig. 17. Load and capacity curve of bridges (Site ID: 0210378).

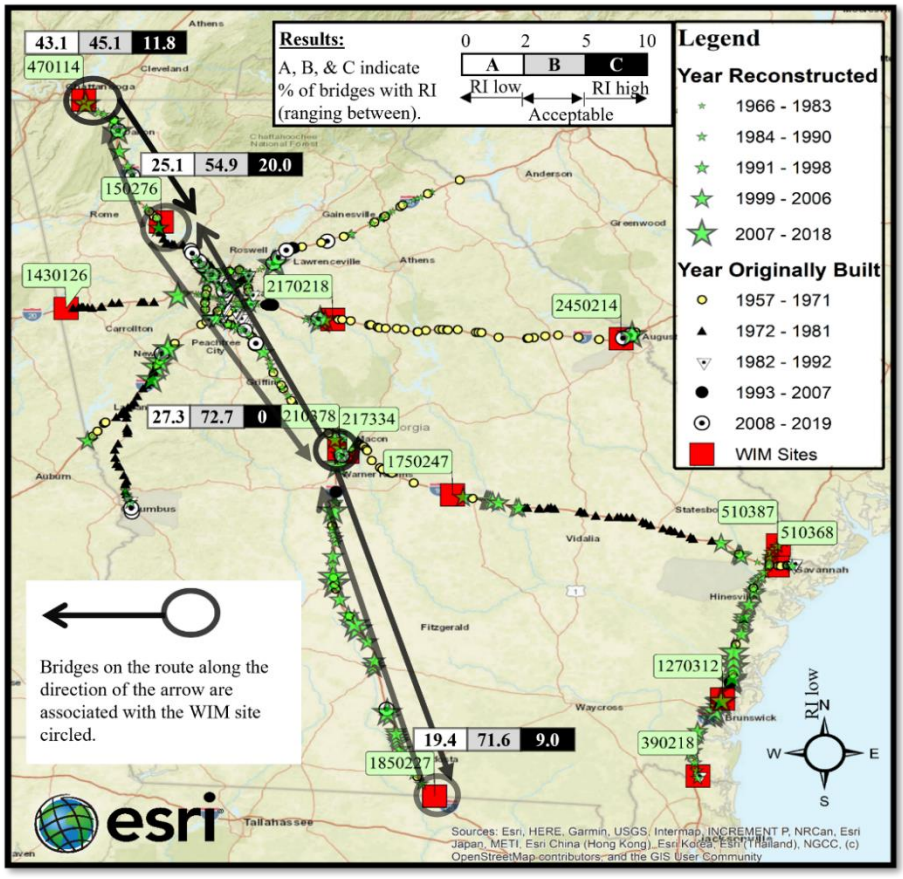
The Reliability Index (RI) is calculated along with the probability of failure for each bridge using Eqs. (13) and (14). As shown in Table 8, a significant number of bridges have an RI less than 2.0, which requires a review of either the operating load ratings or the weight data available in the WIM data sets.

Table 8. Reliability Index of bridges in Georgia

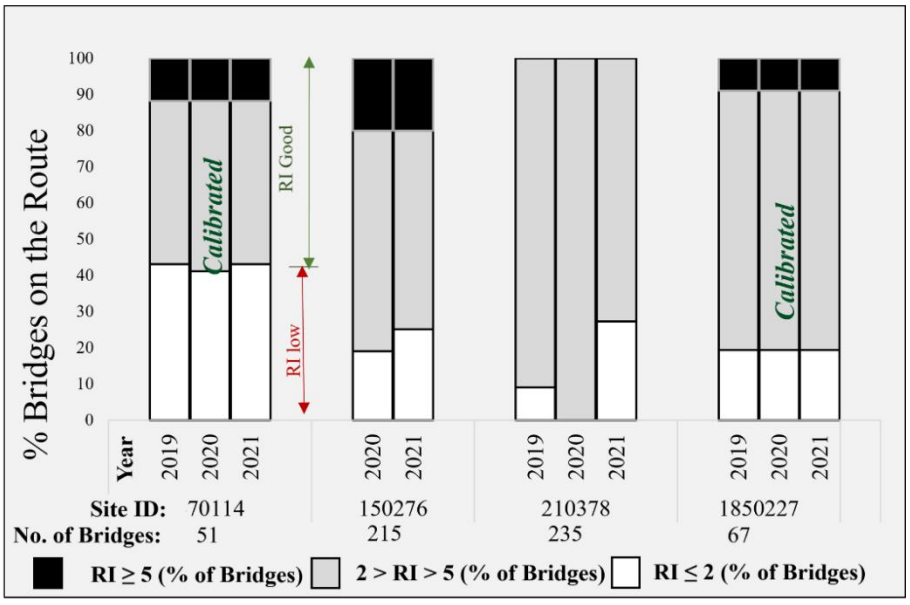
Route	Site ID	Site Name	Site Description	% Bridges with RI		
				<2.0	2.0-5.0	>5.0
I-16	217334	021-w334	I-75 N of I-475 Split Dr, Macon	14.6	72.9	12.5
	1750247	175-0247	I-16, 1.4 miles East of SR 338 MP 43	35.9	61.6	2.6
	510368	051-0368	I-16 East of Dean Forest exit	22.3	73.8	3.9
I-20	1430126	143-0126	I-20 between Alabama SL & SR 100 Veterans Mem Hwy	28.4	43.2	28.4
	2170218	217-0218	I-20 West of SR 11 between SR 142 and SR 11	32.5	38.5	29.1
	2450214	245-0214	I-20 Columbia Co Line & SR 415 Bobby J. Expressway	49.1	34.6	16.4
I-75	470114	047-0114	I-75 between SR 146 & Tennessee SL	43.1	45.1	11.8
	150276	015-0276	I-75 just above SR 20	25.1	54.9	20.0
	210378	021-0378	I-475 (SR 408) between I-75 & SR 22 South of SR 22	27.3	72.7	0.0
	1850227	185-0227	I-75/SR401 @Florida SL, Lake Park, Lowndes Co	19.4	71.6	8.9
I-95	510387	051-0387	I-95, 2 miles North of SR 21 @ South Carolina SL	86.6	13.4	0.0
	1270312	127-0312	I-95 SR 27 & Golden Isles Parkway SR 25 Spur M	59.6	40.4	0.0
	390218	039-0218	I-95 between Florida SL & St Mary's Rd, Kingsland, GA	15.6	78.1	6.3

The Reliability Index is calculated along with the probability of failure for each bridge using Eq. (13) and (14) shown in Table 8. It is observed that a significant number of the bridges have a reliability index less than 2.0 (Table 8), which requires a review of either the operating load rating or the weight data available in the WIM datasets.

To observe the change in reliability index over the year, coupled stacked bar charts are created for different routes shown in Fig. 18 to Fig. 21. The x-axis in Fig. 18 represents the year in which the WIM data were observed for multiple WIM sites. The y-axis represents the percentage of bridges with a RI in 3 different categories, when analyzed with the associated WIM site data. The white stack represents the percentage of bridges with a reliability index less than 2.0. Similarly, the gray stack represents the percentage of bridge with a reliability index greater than 2.0 and less than 5.0. Finally, the black stack represents the percentage of bridges with a reliability index greater than 5.0.

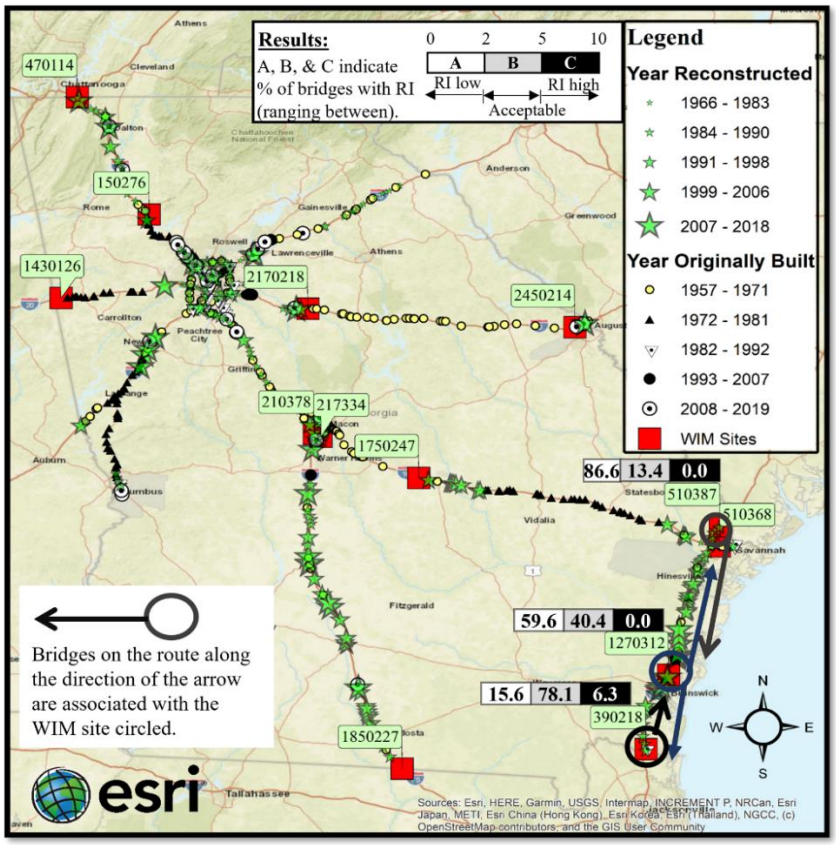


(a) Map showing bridges on the I-75 route

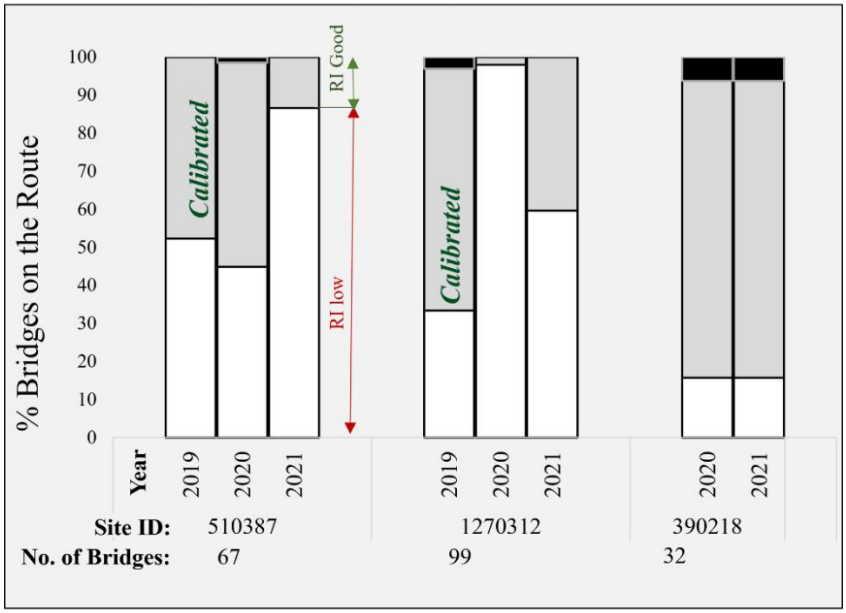


(b) Barcharts showing the % of bridges

Fig. 18. Reliability analysis for I-75 route.

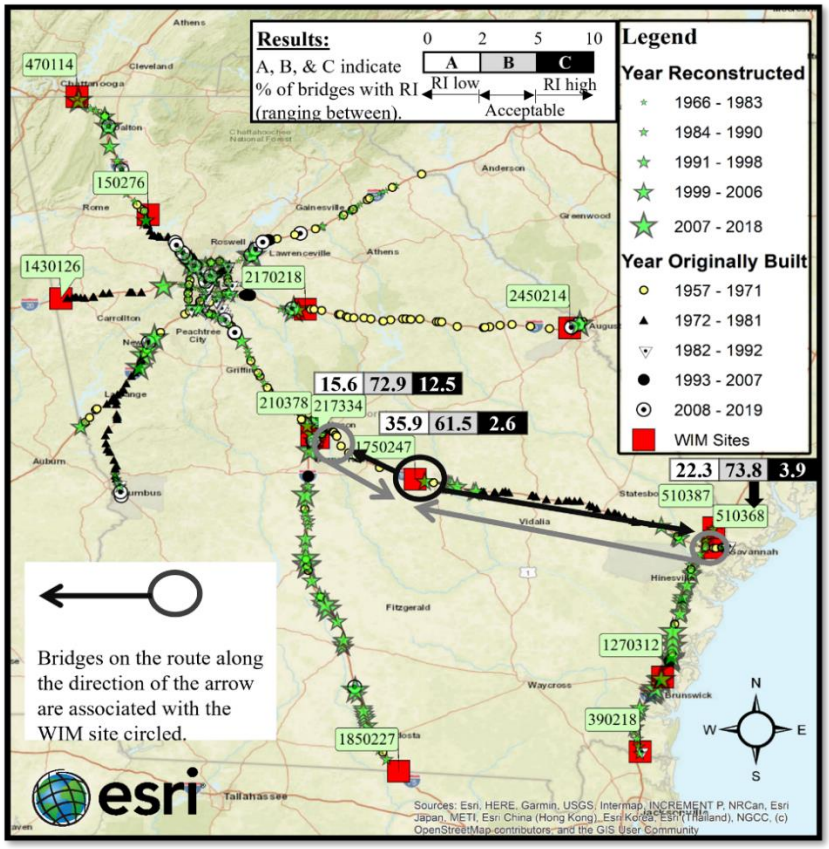


(a) Map showing bridges on the I-95 route

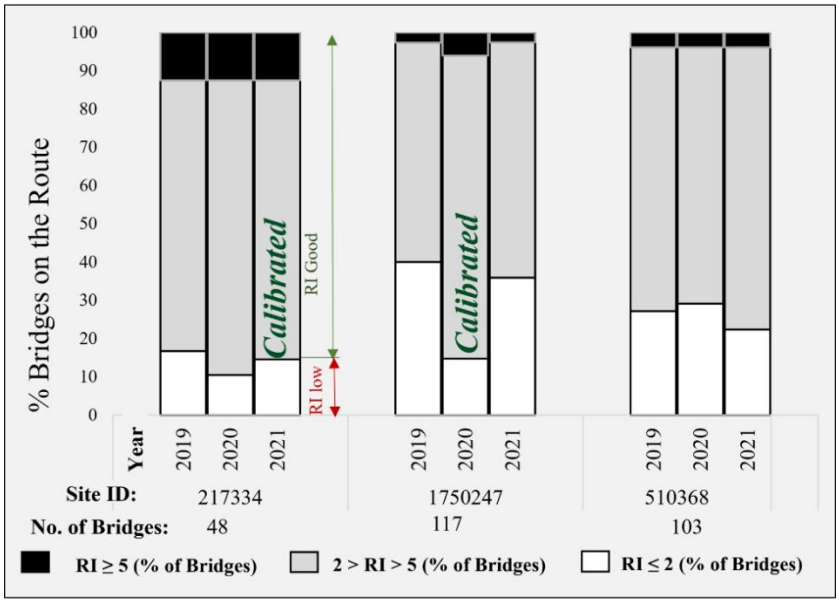


(b) Barcharts showing the % of bridges

Fig. 19. Reliability analysis for I-95 route.

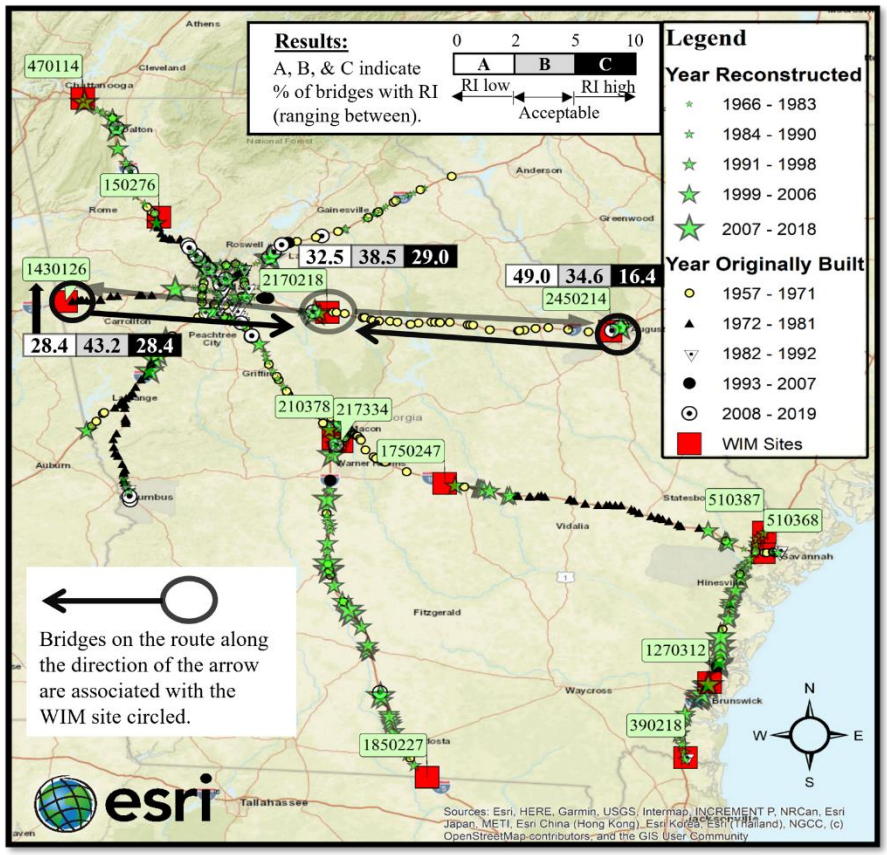


(a) Map showing bridges on the I-16 route

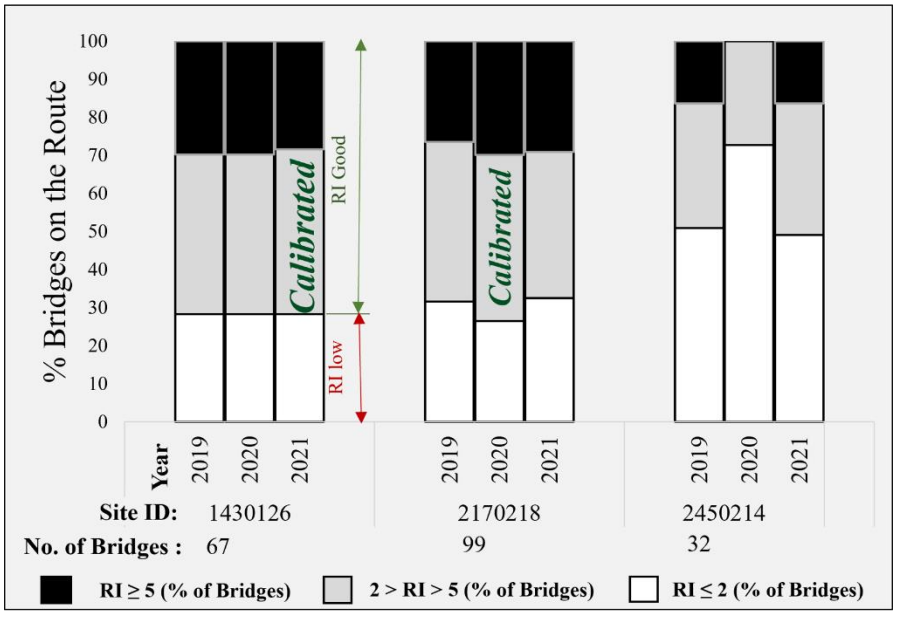


(b) Barcharts showing the % of bridges

Fig. 20. Reliability analysis for I-16 route.



(a) Map showing bridges on the I-20 route



(b) Barcharts showing the % of bridges

Fig. 21. Reliability analysis for I-20 route.

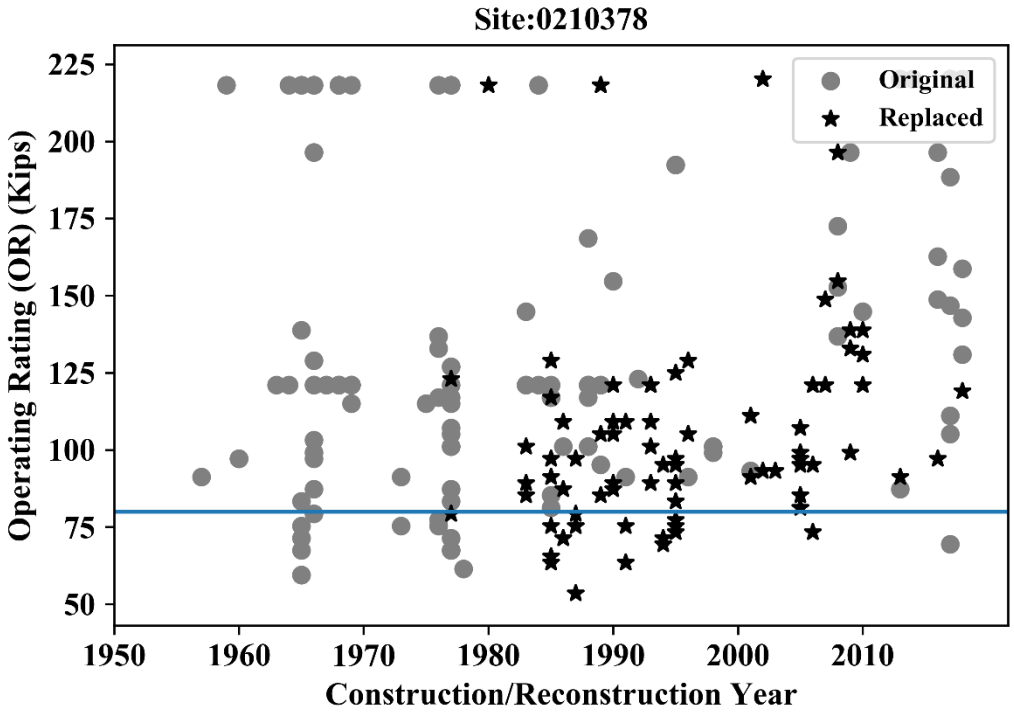


Fig. 22. Operating rating vs year of construction/reconstruction.

Bridges are constructed for a life span of 75-100 years. As, bridge gets old, its capacity gets reduced due to several factors. To observe the relationship between the time of construction and capacity of bridges, Fig. 22 is plotted. The design vehicle used in this study is HL-93, which has a gross vehicle weight of 80 kips. It is observed that some bridges at Site No. 0210378 on I-75 route have a capacity less than 80 kips and are constructed before 1980. Some bridges have a capacity as low as 50 kips even after reconstruction.

Furthermore, bridges with a low reliability index (RI < 2.0) are presented by age groups in Fig. 23. It is observed that most of the bridges with a low reliability index are 40 years old or older. Fig. 23 shows the correlation between a reliability index and bridge age by the routes studied.

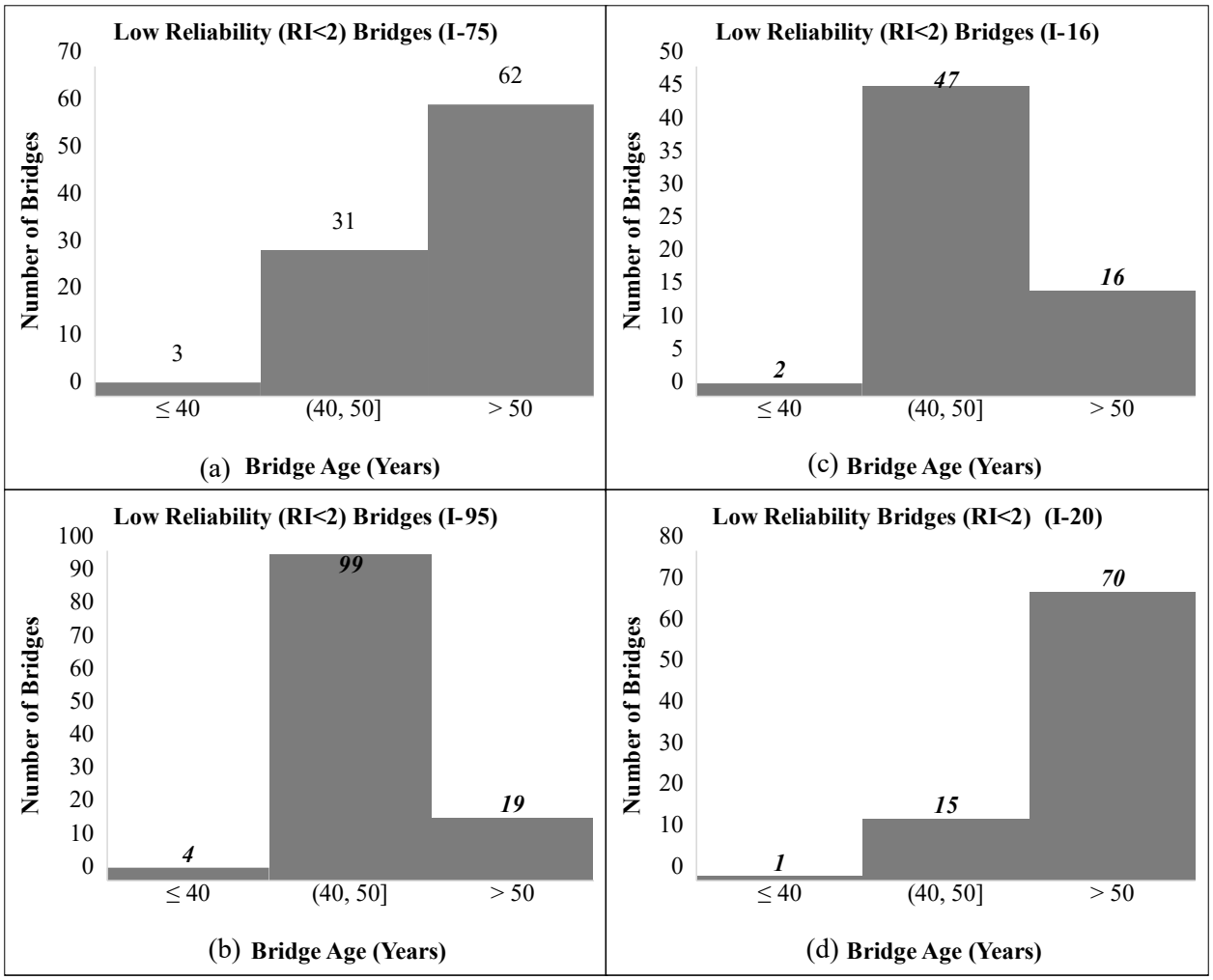


Fig. 23. Plot of low reliability bridge count and age.

Fig. 24 is also created using the ArcGIS application to show the locations of bridges with low reliability index on main interstate routes in Georgia. Bridges with reliability index less than 2 is shown in red circles and green stars shows the location of bridges having reliability index greater than 2.

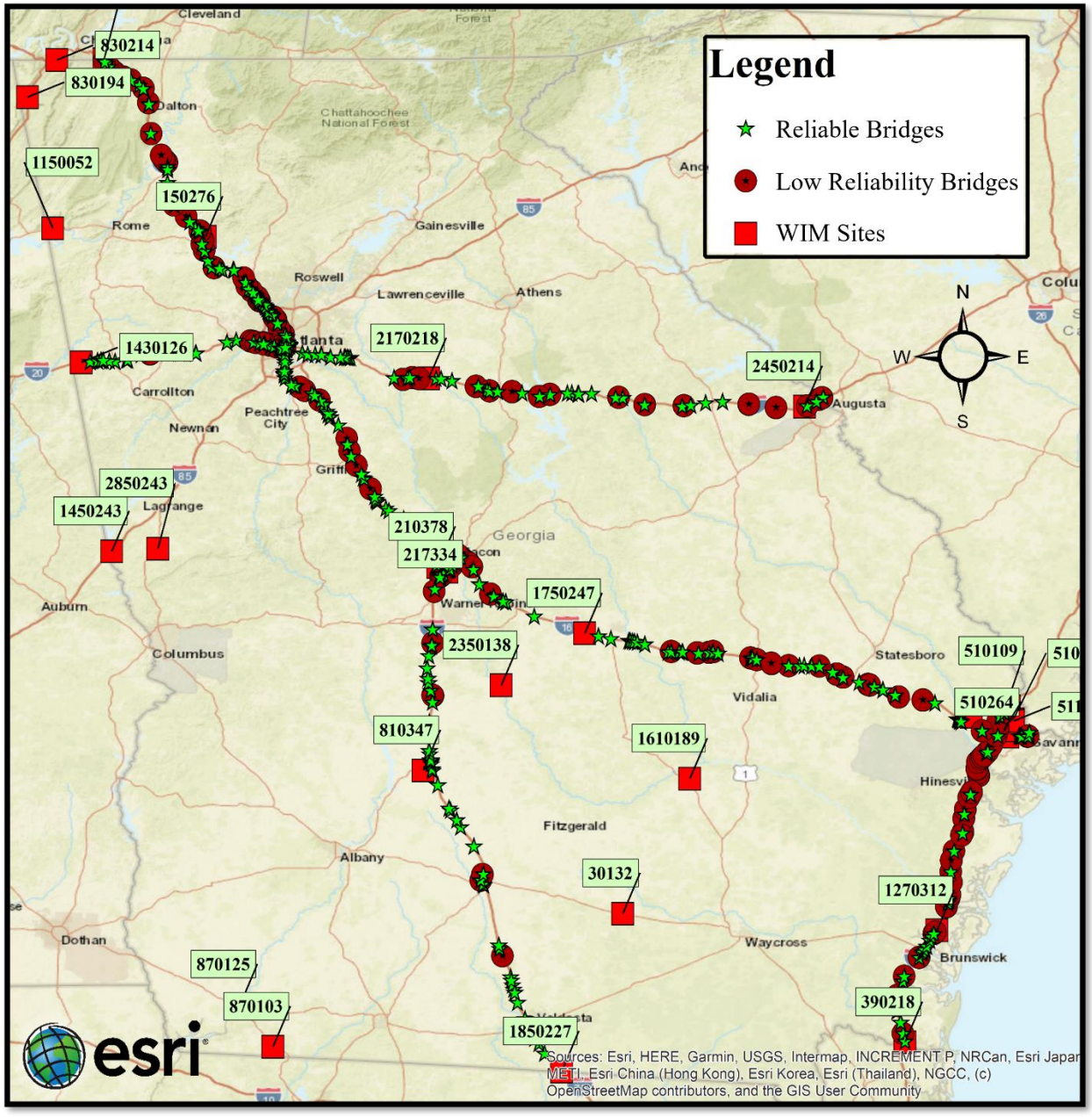
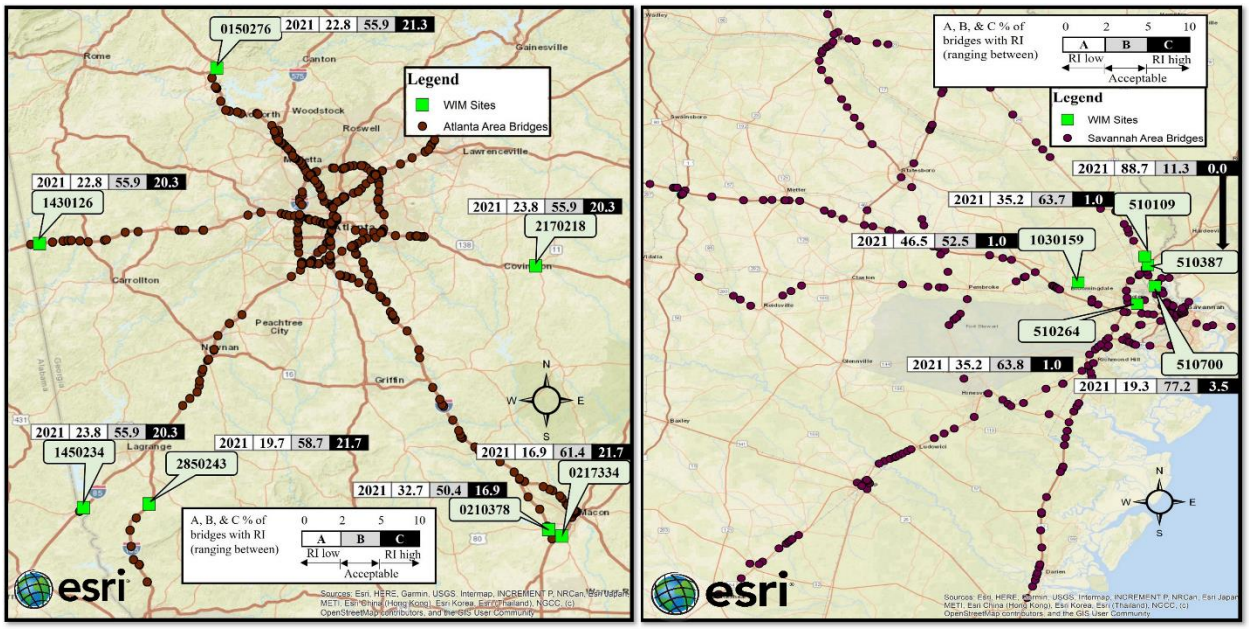


Fig. 24. Locations of low and high reliability bridges in Georgia interstate routes

Reliability indices for bridges in the vicinity of Atlanta and Savannah areas are also determined. Fig. 25 presents a map showing Atlanta and Savannah area bridges as well as the adjacent WIM sites. The bridges shown in Fig. 25 (a) and (b) are paired with the WIM sites in the vicinity of Atlanta and Savannah areas, respectively. It is observed that some of the WIM sites in

the vicinity of the Savannah region shows extremely high live loads and hence are associated with bridges with a low reliability. These two areas need further investigation of traffic logistics such as commercial activities of transporting goods and other factors including the locations of distribution centers.



(a) **(b)**

Fig. 25. Reliability Index for bridges in the vicinity of Atlanta area and Savanah area.

4.5 Summary of Results and Analysis

Following finding are summarized considering the results obtained in this study in conjunction with the research questions raised in the paper.

The easy and quick way to determine the safety of bridges is to evaluate the reliability index using available WIM data as load and NBI bridge operating rating as resistance. Fig. 14 shows the WIM site and bridge locations. Table 8 illustrates the results of reliability analysis. It is observed that Georgia has a significant number of bridges under reliability index 2.0, which raises concern to the public safety. As traffic load is increasing rapidly over the recent years, it is more effective

and salient to evaluate bridges more frequently on important routes using the available WIM data as it provides the real time vehicle information at any site. Furthermore, bridges older than 50 years should be on the top priority for evaluation as it can be seen from Fig. 23 that most of the low reliability bridge are 50 years or older. Furthermore, sites with bridges replaced or reconstructed recently (shown as stars in Fig. 14 and Fig. 24) has less number of bridges under reliability index 2 (Bridges on I-75 routes). Although, site 1270312 on route I-95 (Fig. 24)shows a very high number of bridges with low reliability despite having newer bridges, it is observed that the load sensors were out of calibrations at that site and were not capturing the load accurately.

In defining the relationship between the changes in Substructure condition rating (CR) and changes in Operating load rating (AASHTO), a linear regression model is not adequate because multiple factors including the substructure CR affects the operating LR. When several other factors are included as predictors such as the year built, CRs, types of bridge, and design loads, a nonlinear relationship identified by a XGBoost machine learning model yields a high R-squared value of 0.87, a low RMSE value of 6.20, and a prediction accuracy of 87.25%. Other models also produced similar results. The regression analysis indicates that a nonlinear regression involving machine learning algorithms provides a stronger relationship between the operating LRs and bridge attributes such as CRs and age.

Table 8 shows reliability indices (RI) and the percentage of bridges (with a RI less than 2.0) associated with a WIM site. Many bridges in Georgia have a reliability index greater than 5.0. A reliability index between 2 and 5 is considered within an acceptable safety range. However, bridges with a reliability index less than 2.0 are considered under risk and possess a greater risk to the transportation network. It is observed that bridges with a low-reliability index are relatively older

(see Route I-20 in Fig. 23). Therefore, WIM sites associated with bridges with a low reliability (RI < 2) should be calibrated to manage risk and improve the reliability of the WIM data. Furthermore, the weight of the flowing traffic should be monitored carefully, and the number of overweight vehicles needs to be validated against the number of permits issued on these routes with high importance, as these bridges are a matter of concern to the public safety.

Nonetheless, the data driven reliability assessment of bridges should be the new normal in bridge maintenance. The advantage of data-driven assessment is that it increases the transparency about bridge safety between owners and users by providing symmetric information, and it considers the influence of growing heavy weight traffic in highways. Furthermore, symmetric information also eliminates the possibility of overconfidence of owners and helps to reduce the risk of adverse selection.

4.6 Conclusion

In this study, WIM data from 4 major interstate routes in Georgia are investigated with 21 years of NBI database bridge capacity to determine a relationship between operating load rating and bridge attributes and determine the reliability index for bridges in Georgia. The following conclusions are made:

- It is concluded that a robust nonlinear relationship exists between NBI bridge attributes and bridge capacity.
- Reliability analysis results show that a significant number of bridges have a reliability index of less than 2, which poses a great risk to the transportation infrastructure. Coincidentally, low-reliability bridges are also relatively older. Therefore, the performance associated with bridges

with a low-reliability index needs to be evaluated more frequently to increase the reliability and avert the risk associated with each bridge in Georgia.

- The reliability analysis results are a great way to communicate the associate risk with the bridge owner, which would eventually be helpful resolve the information asymmetry and mitigate the adverse selection problem in decision making.

Chapter 5

5. Mitigating Dynamic Inconsistency in Vehicular Live Loads on Bridges over Time and Vehicle Weight Forecasting Employing a Recurrent Neural Network

5.1 Introduction

Weigh-In-Motion (WIM) is a system which measures the per-axle weight, overall weight of vehicles, axle-spacing, vehicle speed, and more. With the aid of such systems, transportation agencies monitor and enforce vehicle weight regulations on the highways. The Georgia Department of Transportation (GDOT) has expanded the number of WIM stations on a wide range of highway routes. WIM technology is in place for more than 20 years. Although there are many advanced WIM sensors are in place, there are limitations in the load cell accuracy due to various factors, which calls for regular calibration of sensor (Chorzepa et al. 2020). However, due to lack of confidence/knowledge, many WIM sensors are still out of calibration and generating misleading data, which are to be used for various transportation development projects and maintenance of existing bridge infrastructure. Therefore, there is a need to develop a control dataset for each site utilizing advanced statistical and computational tools available using existing dataset and remove misleading data from the database.

There are several ways to determine weight deviations when analyzing the WIM data. Statistical tests determine monthly and annual variations in the weight data and thus may be used

to monitor a weight deviation. For example, two sample distributions (January and December data) may be compared in a hypothesis test with null hypothesis that they are of the same distribution. However, statistical tests such as a chi-square test and a t-test assume that they follow a particular type of probability distribution. Thus, nonparametric statistical tests are used to analyze these types of data. A nonparametric test (such as Mann-Whitney U and Kruskal-Wallis tests) is applied to datasets that do not follow a specific distribution. Although statistical significance tests are used to determine the difference in central tendency (e.g., mean or median), they do not provide a quantitative measure of the deviation between two datasets.

Relative entropy or Kullback-Leibler (KL) Divergence based probabilistic approaches are also frequently used to quantify randomness in continuous time series, characterizing relative entropy in information system and information gain in comparing two statistical models. A Jensen-Shannon (JS) Divergence test is a normalized version of the KL divergence approach and is performed along with the statistical significance test. JS divergence is a special type of tool to measure the statistical distance and rank the strength of relation [between two probability distributions] 0 to 1 (0 being the best).

Despite weight deviation tests being useful in detecting weight drift, they do not provide us with the complete picture. There is a seasonal trend and a complex pattern to the weight distribution of vehicles. By analyzing time series data, we can identify anomalies in vehicle weights, understand the reasons for complex seasonality and trend, and understand the structure of the data. Sinha et al. (2022) shows in a different study using Georgia WIM data that heavy weight vehicles are encountered more often on bridges than expected and predicted live load factors are significantly higher than a bridge is designed to carry in their lifetime. Time-series analysis could help us to determine how consistently overloaded vehicles appear on a bridge.

This study compares two consecutive months' WIM data and identifies a month when a significant weight deviation is first observed. Subsequently, it discusses challenges with weight variations observed in the WIM data. It should be recognized that a stringent requirement for calibration will not practice as most of the existing WIM instruments are required to be calibrated or re-calibrated. This study employs a statistical analysis model which uses time series vehicle weight data to better understand the weight (or live loads) pattern as well as predict future weight trends. Additionally, a deep learning model is used in this study the complex vehicle weight pattern and unravel the hidden information about the anomalies in the data.

5.2 Procedures for Monitoring Weight Deviations

5.2.1 Statistical tests

Statistical tests are performed to observe a significant vehicle weight deviation between two consecutive months. The main objective is to define a control dataset. A control dataset is defined as the data which is used to select, execute, or modify another set of data. As WIM sensors can be out of calibration for several months due to temperature changes, rain, and other factors, it is important to determine how much deviation in vehicle weights of a WIM data exists. Mann-Whitney U test and Kruskal -Wallis (KW) tests are the two statistical tests performed in this study (Selezneva and Wolf 2017). Vehicles with class 4 or greater from the 29 WIM sites are used in this study. The significance level is set to 0.05 (5%). The hypothesis is as follows:

If the p-value is less than 0.05, reject the null hypothesis. If p-value is greater than the significance level, we fail to reject null hypothesis.

5.2.2 Jensen-Shannon (JS) Divergence

Consider two probability distributions, P and Q. For discrete probability distributions, P and Q defined on the same probability space χ, the relative entropy (or KL Divergence) from Q to P is defined as:

$$D_{KL}(P \parallel Q) = \sum_{x \in \chi} P(x) \log\left(\frac{P(x)}{Q(x)}\right) \tag{15}$$

The KL divergence (Joyce 2011) measures the distance between two distributions noting that it is not the metric distance. The divergence is also not symmetrical, which means that the KL distance between $P(x)$ to $Q(x)$ is not same as the distance between $Q(x)$ to $P(x)$. Futhurmore, it does not follow the triangular inequality (Raiber and Kurland 2017). Another problem occurs in determining the KL distance. If $P \neq 0$, but $Q = 0$, $D_{KL}(P \parallel Q)$ is defined as ∞. This means that if one event is possible, and the other event is absolutely impossible. As a result, the two distributions must be different.

The Jensen-Shannon Divergence (Cai and Lim 2022) is a symmetric and smoothed/normalized version of the KL divergence ($D_{KL}(P \parallel Q)$). It is defined by:

$$\text{JSD}(P \parallel Q) = \frac{1}{2} D_{KL}(P \parallel M) + \frac{1}{2} D_{KL}(Q \parallel M) \tag{16}$$

, where M = $\frac{1}{2}(P + Q)$.

The JS divergence is bounded by 0 and 1 for two probability distributions.

$$0 \leq \text{JSD}(P \parallel Q) \leq 1 \tag{17}$$

5.3 Dynamic Inconsistency

Understanding live loads on our public roads and bridges is important for transportation asset management. Thus, monitoring a vehicle weight drift indicating a potential quality issue in WIM

systems is critical for bridge safety. State DOTs often identify such a deviation by comparing the weight difference between group means (control vs. WIM data sets) that are statistically significant. A control or reference data set is generally established from a most recent calibration of weight sensors. Enforcing a 95% confidence level in identifying an acceptable weight drift range appears to be a reasonable policy due to our propensity to do nothing. However, a problem arises when a decision maker observes that the threshold level yields a recalibration of weight sensors at a significant number of WIM sites. He or she may dynamically consider changing the policy or applying a different criterion either for the benefit of doubt or lack of knowledge. A lack of confidence in the control data, due to dynamic nature of vehicle weights, contributes to an increased inconsistency of the decision maker.

Such dynamic inconsistency is one of the most profound problems in behavioral economics and social science. In this paper, it is hypothesized that the effectiveness of a weight-monitoring policy today relies on the credibility of a commitment strategy to implement or sustain that policy in the future. To sustain a time-consistent weight monitoring policy, a more precise method of measuring a weight drift must be proposed to increase the credibility of the results stating that vehicle weights are significantly misrepresented by WIM sensors, warranting a recalibration.

5.4 Procedures for Weight Forecasting using Advanced Time Series Forecasting Models

Vehicle weight forecasting is considered as an important and challenging topic in asset management. The vehicle weight forecasting will aid in planning and/or developing strategies for future highway infrastructure maintenance.

5.4.1 Seasonal Auto-Regressive Integrated Moving Average (SARIMA) Model in Time-series Analysis

ARIMA models are generally fitted to the time series data to better understand the data and make short-term future predictions based on that. ARIMA models are first introduced by Box George et al. (1976), the model comprised of the autoregressive term (AR), the level of integration (I), and the moving average term (MA). Many past researchers utilized ARIMA model to predict traffic pattern using WIM data (He et al. 2018; Smolarek and Ziemska 2017). The advantage of this model is the easy interpretation of outcomes. ARIMA models are mostly used where data mean is stationary. Highly non-stationary data are converted to stationary data after one or more differentiation steps. Kumar and Vanajakshi (2015) have utilized the Seasonal ARIMA (SARIMA) model to predict short-term traffic flow. The SARIMA model performs well with the non-stationary traffic data. The success of accurate prediction lies in the proper selection of non-seasonal hyperparameters (p, d, q) and seasonal hyperparameter (P, D, Q, and m). The seasonal and non-seasonal hyperparameters are determined by analyzing the autocorrelation (ACF) and partial-autocorrelation plots (PACF).

5.4.2 Application of Deep Neural Network in Time-series Analysis

Various machine learning techniques (such as ARIMA, Markov Chain models, etc.) are in use in time-series prediction modeling. Auto regressive moving average methods is one of them and widely used in forecasting due to its simplicity and interpretability of the whole forecasting process. However, due to the multimodal and complex nature of the data, these traditional methods lack accuracy and not efficient to unveil the complex pattern in the data. Furthermore, these shallow networks need high dimensional space to model to increase the accuracy and hence

increase the chance of overfitting problem. These methods are also not helpful when there are missing data, which is very likely in the WIM data due to occasional road closers for maintenance. Additionally, they are not good for long time forecasting and perform very poorly at turning points. If the data is big, this method can be very expensive computationally.

To overcome this problem, deep learning, especially Long-Short-Term-Memory neural network (LSTM) (Hochreiter and Schmidhuber 1997) has been implemented with great accuracy in time-series forecasting problems. LSTM neural network is a special type of Recurrent Neural Networks (RNN), where 4 different interacting layers communicating in a different way, learn and hold information from long term sequence of observation with an advantage of memorizing the sequence of data. By using multi-layered non-linear structures, LSTM neural network proved to be a very useful technique with great accuracy in time-series forecasting. LSTM also reduce the number of dimensions and hence reduce the tendency to overfit.

In recent past, many researchers (Li et al. 2018; Morris et al. 2022; Tian et al. 2018) have emphasized in the application of LSTM neural network-based framework in traffic data. This subsection specially focused on the key concepts of a LSTM network and the learning process. Fig. 26 shows a sample LSTM network structure.

A time-series of length T is expressed as $X = (x_1, x_2, \dots \dots \dots, x_T)$, where x_t represent the t-th observation. C_t is the memory cell or cell state, which contains information at time step t. The memorization of past trend is feasible through some gates along with the memory line. The upper line in each cell [see Fig. 26(a)] connects one module to another one passing the information from past and collecting them for the present state. The gates are the forget gate (f_t), input gate (i_t) and output gate (o_t), and are used to dispose, filter, or addition for the next cell. In Fig. 26(a) σ is the sigmoid function. f_t is the output of forget gate at time t.

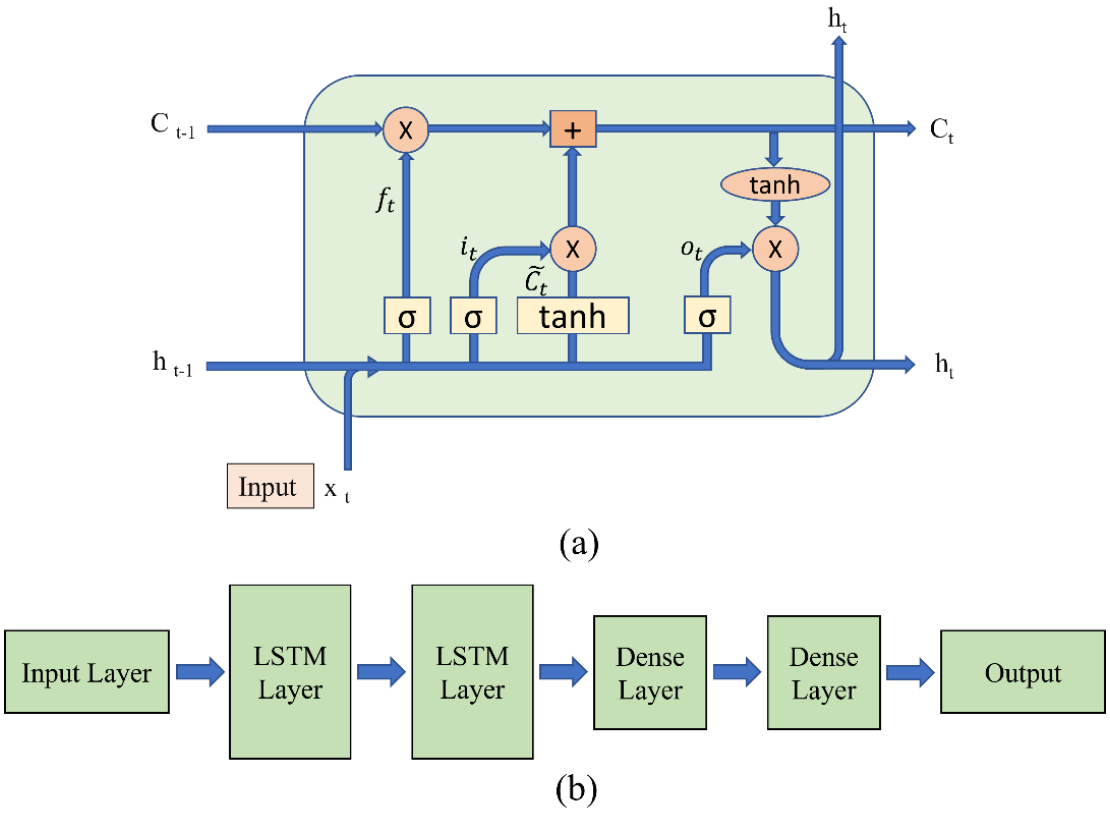


Fig. 26. A sample (a) LSTM network structure and (b) architecture of LSTM network for weight forecasting

The first step in this LSTM structure is to decide the information to be kept of thrown away from the cell state by assigning a value of 0 or 1. The decision is made by the sigmoid layer (σ).

$$f_t = \sigma\left(W_f.[h_{t-1}, x_t] + b_f\right) \tag{18}$$

The next step is to decide what new information is going to be stored in the current state. This step combines two layers, one sigmoid layer (σ) and one tanh layer. First sigmoid layer decides which values to let through (0 or 1) and second tanh layer adds weightage to the information which are passed through based on their level of importance (ranging -1 to 1) and creates an update on to the cell state ($\tilde{C}_t$).

$$i_t = \sigma\,(W_i.[h_{t-1}, x_t] + b_i) \tag{19}$$

$$\widetilde{C_t} = \tanh\,(W_c.[h_{t-1}, x_t] + b_c) \tag{20}$$

Now in the next step, the old cell state (C_{t-1}) is updated into a new cell state (C_t).

$$C_t = (f_t * C_{t-1} + i_t * \widetilde{C_t}) \tag{21}$$

Finally, these steps decide what will be our outputs. The first step is to pass the information through a sigmoid layer and help to decide what parts of the cell state contributes to the output. The information stored in the cell state are pushed through the tanh layer and multiplied by the output of sigmoid layer illustrated in Eqs. (22) and (23).

$$o_t = \sigma\,(W_o.[h_{t-1}, x_t] + b_o) \tag{22}$$

$$h_t = o_t * \tanh(C_t) \tag{23}$$

The performance in both the SARIMA and LSTM is compared using Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE).

5.5 Research Questions and Scope

This study specifically aims to answer the following questions.

a) It has been observed that WIM data has a drift in monthly weight distributions. The possible reason could be the calibration error or malfunctioning sensors. How to quantify the weight deviation between months at a particular site using statistical approaches and develop a strategy for WIM sensor calibration?

b) The current practice of weight deviation monitoring is to perform the statistical significance tests on two consecutive year data. Is time-series prediction a better method of monitoring weight deviation than the current practice?

c) With the latest advances in artificial intelligence deep learning methods are expected to improve temporal dependencies of time-series data. One of the promising DL models used for time-series prediction is LSTM. Does LSTM perform better than the traditional time-series prediction methods such as ARIMA?

d) How could monitoring weight deviation help to mitigate the dynamic inconsistency in decision making?

Although, statistical significance tests are widely accepted by state DOTs to determine weight deviation between calibration visits, it does not quantify the true deviation between two vehicle weight frequency distribution. Hence this study proposes a novel JS Divergence method to quantify the deviation between monthly weight distributions along with the statistical significance test. This study also proposes a deep learning (LSTM) technique which could accurately predict the short-term gross vehicle weight utilizing the real time WIM data and compared with the popular SARIMA model outcomes.

5.6 A step-by-step Approach for Weight Deviation and Weight Forecasting

The analysis is first performed with all the available WIM sites data since 2019. The WIM data are filtered for vehicle class 4 and above. The non-parametric statistical significance tests are performed first. Mann-Whitney U test and Kruskal -Wallis (KW) tests are the two statistical tests performed in this study (Selezneva and Wolf 2017). The significance level is set to be at 0.05. The hypothesis is explained as follows.

a) Null Hypothesis (H0): Two consecutive months vehicle weight data are same.

b) Alternative hypothesis (Ha): Two consecutive months data differs from one another.

If the p-value less then significance level, then reject the null hypothesis. If p-value is greater than significance level, then failed to reject null hypothesis.

The vehicle gross weight frequency distribution is created by considering 2000 kips weight bins. The frequency in each bin is plotted for each month. Two consecutive months frequency distribution is then compared using JS Divergence method by employing Eq. (16) and (17). The distance is then converted to percentage by multiplying 100.

The next step is the gross vehicle weight forecasting using advanced statistical and deep learning tools. In this step, the daily average gross vehicle weight data for class 9 vehicle is filtered for all the sites. The SARIMA model is used as statistical tool for daily average gross vehicle weight forecasting. Well established python packages are utilized to perform the forecasting. The train data is considered as 85% of the data and rest of the 15% data are used to test the model. The model hyperparameter are tuned and fitted on train data. The model accuracy is validated on test data. Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) are utilized the evaluation criteria to assess the forecasting accuracy.

Finally, a prediction model utilizing LSTM network is demonstrated for each WIM site. The primary architecture consists of two fully connected LSTM layers and two fully connected dense layers as shown in Fig. 26 (b). 80% of the data is selected as the train data and 20% of the data is considered as the test data. The hyperparameters are selected based on the data and trials. The observation variable is one dimensional and the sequence length is considered as 50 throughout the study. The batch length is selected as 16 and epochs are decided by observing the loss function for each site. All the parameters are scaled before feeding into the architecture. Adam algorithm is used for optimization and the mean squared error is utilized as the error function. Adam algorithm for optimization is employed as the data is very noisy. To avoid overfitting problem, a dropout

value is selected as 0.1. The LSTM model prediction accuracy is gauged by MAPE and RMSE. The SARIMA model and LSTM model are then compared based on their test prediction accuracy.

5.7 Results of Weight Deviation

5.7.1 Statistical Testing

In this study, statistical tests are performed to observe the deviation of vehicle weights between two consecutive months. The main objective is to define a control dataset. A control dataset is defined as the data which is used to select, execute, or modify another set of data. As many WIM sensors are found to be out of calibration for some months due to change of temperature, rain, and other factors, it is important to determine how much deviation in vehicle weights of a WIM data exists.

Weight data from 29 WIM sites are analyzed in this study. The statistical tests are performed using Mann-Whitney U and Kruskal-Wallis tests. Two consecutive months data are compared. For all WIM site weight data analyzed herein, we fail to reject the null hypothesis and conclude that the two distributions are statistically different at the 0.05 significance level. In the table, the first two months data which fails to reject the hypothesis are shown in parenthesis.

Contrary to the statistical test results, Fig. 27 shows an example of monthly weight distributions at a WIM site in year 2021, where no deviation, specifically a shift to left or right in weights, is observed. Statistical tests may observe a vertical weight deviation (or changes in the frequency of weights) but not the lateral shifts such as a weight decrease shown in Fig. 27 (c), which are relevant to weight calibration. The main objective of statistical tests is to detect monthly weight shifts, and Fig. 27 (a) and (b) show that an increase in the frequency of weights should not be construed as a weight deviation (increase or decrease). Therefore, despite of the fact that the

statistical test indicates a significant weight deviation for Site 0217334, the outcome does not necessarily indicate that the weight sensor needs to be calibrated. Therefore, statistical tests alone are not sufficient to detect pertinent weight shifts shown in Fig. 27 (c) and (d).

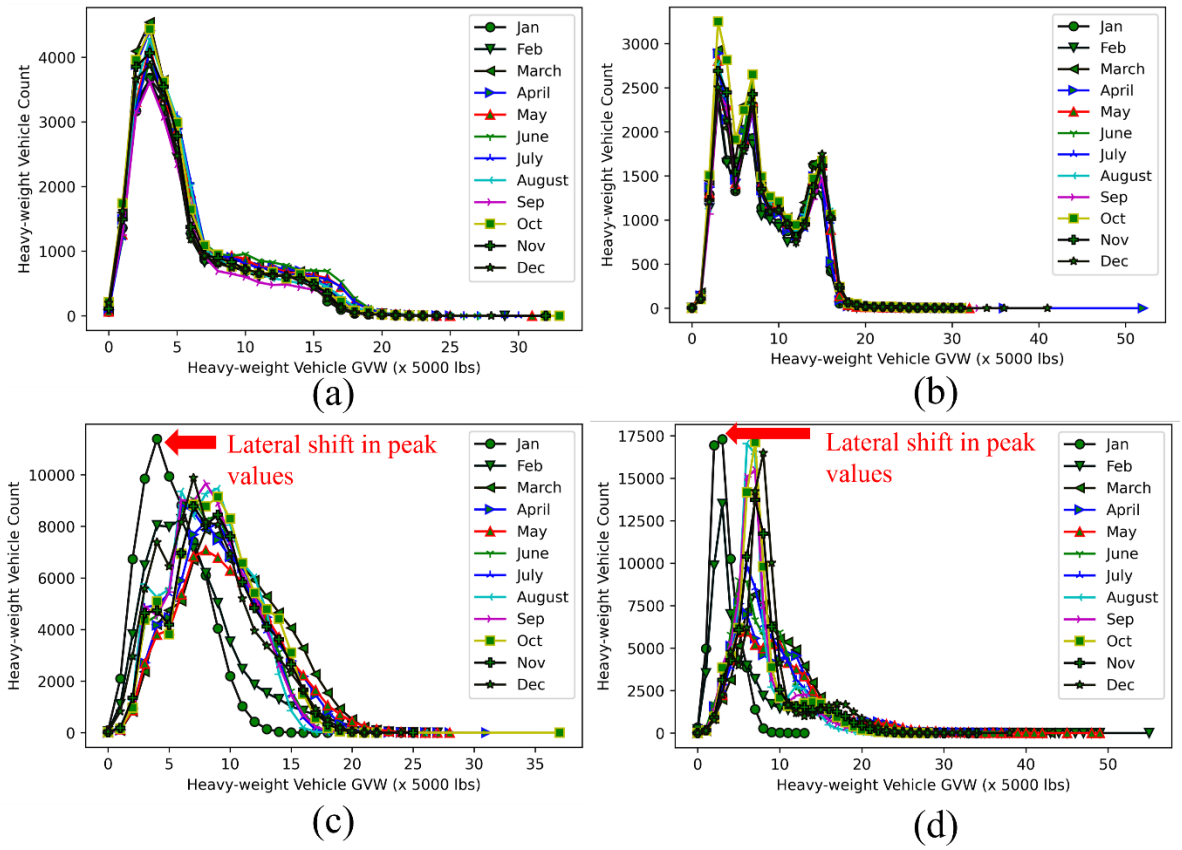


Fig. 27. Monthly Weight Distributions for Sites 0217334 and 0510700.

5.7.2 JS Divergence Test

A JS divergence test compares two weight distributions and provides the distance between the two. Each month weight data is divided into equal sized bins with an increasement of 2000 lbs. The probability of occurrence is determined by dividing a vehicle count at each bin with the total number of vehicles. This method accounts for the increase in the total number of vehicles (or vertical deviations described with statistical tests) and normalizes weight distributions.

Table 9 shows the analysis results. Although the weight shifts are adequately recognized with the divergence test, 5% weight deviation appears too stringent for applying a threshold criterion for a significant weight change with 95% confidence and requiring calibration. Based on a review of the analysis results and monthly weight distributions, 20-30% weight deviations deem more reasonable.

Table 9. Weight Deviation Results using Jenson-Shannon (JS) Divergence (Year 2021)

Site ID	Lane	Year	Wt. Dev. Within 5%	Wt. Dev. Within 10%	Wt. Dev. Within 20%	Wt. Dev. Within 30%
0030132	EB	21	No (Jan & Feb)	No (Jan & Feb)	No (Jan & Feb)	Yes
	WB	21	No (Jan & Feb)	No (Feb & March)	No (Feb & March)	Yes
0150276	NB	21	No (April & May)	No (April & May)	No (April & May)	Yes
	SB	21	No (March & April)	No (April & May)	No (April & May)	No (April & May)
0210378	NB	21	No (Jan & Feb)	No (Feb & March)	No (Feb & March)	No (Feb & March)
	SB	21	No (Jan & Feb)	No (Jan & Feb)	No (March & April)	Yes
0217334	NB	21	No (April & May)	Yes	Yes	Yes
	SB	21	No (April & May)	Yes	Yes	Yes
0390218	NB	21	No (April & May)	Yes	Yes	Yes
	SB	21	Yes	Yes	Yes	Yes
0470114	NB	21	No (Feb & March)	No (Feb & March)	Yes	Yes
	SB	21	No (Feb & March)	No (May & June)	Yes	Yes
0510109	NB	21	No (Jan & Feb)	No (Feb & March)	Yes	Yes
	SB	21	No (Jan & Feb)	No (Feb & March)	No (May & June)	No (May & June)
0510264	EB	21	No (Jan & Feb)	No (Feb & March)	No (Feb & March)	Yes
	WB	21	No (Jan & Feb)	No (Jan & Feb)	No (Sep & Oct)	No (Sep & Oct)
0510368	EB	21	No (March & April)	Yes	Yes	Yes
	WB	21	No (March & April)	No (March & April)	Yes	Yes
0510387	NB	21	No (Feb & March)	No (April & May)	Yes	Yes
	SB	21	No (Feb & March)	No (March & April)	No (March & April)	No (March & April)
0510700	NB	21	No (Jan & Feb)	No (Jan & Feb)	No (Jan & Feb)	No (Feb & March)
	SB	21	No (Jan & Feb)	No (Jan & Feb)	No (Jan & Feb)	No (Jan & Feb)
0511113	EB	21	No (Jan & Feb)	No (Jan & Feb)	No (Feb & March)	No (April & May)
	WB	21	No (Jan & Feb)	No (Jan & Feb)	No (Feb & March)	Yes
0810347	EB	21	No (Jan & Feb)	No (Feb & March)	No (March & April)	Yes
	WB	21	No (March & April)	No (May & June)	Yes	Yes
0830194	NB	21	No (Jan & Feb)	No (Jan & Feb)	No (Jan & Feb)	No (Feb & March)
	SB	21	No (Feb & March)	No (Feb & March)	No (Feb & March)	No (Nov & Dec)
0830214	EB	21	No (Jan & Feb)	No (Jan & Feb)	No (Feb & March)	No (March & April)
	WB	21	No (Jan & Feb)	No (Feb & March)	Yes	Yes
0870103	NB	21	No (Jan & Feb)	No (Jan & Feb)	Yes	Yes
	SB	21	No (Jan & Feb)	No (Feb & March)	No (Feb & March)	No (Feb & March)
0870125	NB	21	No (Feb & March)	No (April & May)	No (Sep & Oct)	No (Sep & Oct)
	SB	21	No (Feb & March)	No (Feb & March)	Yes	Yes
1030159	EB	21	No (June & July)	No (June & July)	No (June & July)	No (June & July)
	WB	21	No (June & July)	No (June & July)	No (June & July)	Yes
1150052	EB	21	No (July & August)	No (July & August)	Yes	Yes
	WB	21	No (July & August)	No (July & August)	Yes	Yes
1270312	NB	21	No (June & July)	Yes	Yes	Yes
	SB	21	No (Jan & Feb)	No (March & April)	No (April & May)	No (April & May)
1430126	EB	21	No (March & April)	No (March & April)	No (March & April)	No (April & May)
	WB	21	No (March & April)	No (April & May)	No (April & May)	Yes
1450234	NB	21	No (June & July)	No (June & July)	Yes	Yes
	SB	21	No (June & July)	No (June & July)	Yes	Yes
1610189	NB	21	No (Jan & Feb)	No (Jan & Feb)	No (Feb & March)	Yes
	SB	21	No (Jan & Feb)	No (Jan & Feb)	No (Feb & March)	No (Feb & March)
1750247	EB	21	No (Feb & March)	No (May & June)	No (June & July)	No (June & July)
	WB	21	No (Feb & March)	No (June & July)	No (June & July)	Yes
1850227	NB	21	No (Feb & March)	No (March & April)	Yes	Yes
	SB	21	Yes	Yes	Yes	Yes
2170218	EB	21	No (Oct & Nov)	Yes	Yes	Yes
	WB	21	Yes	Yes	Yes	Yes
2350138	NB	21	No (June & July)	No (June & July)	No (Sep & Oct)	No (Sep & Oct)
	SB	21	No (June & July)	No (June & July)	No (Sep & Oct)	No (Sep & Oct)
2450214	EB	21	No (August & Sep)	No (August & Sep)	Yes	Yes
	WB	21	No (August & Sep)	No (August & Sep)	Yes	Yes
2850243	NB	21	No (Feb & March)	No (Feb & March)	No (April & May)	No (May & June)
	SB	21	No (Feb & March)	Yes	Yes	Yes

It is observed that most of the weight deviations observed in the 2021 WIM data is within 30% except for a few sites. The weight deviation observed at Site No. 217334 is within 10% as visibly observed in Fig. 28. On the other hand, Fig. 28 (a) and Fig. 28 (b) show that the weight distribution peaks are continuously drifting over the months for both the NB and SB directions. Table 9 results support that argument by showing evidence of weight deviation is more than 30% using JS divergence for site 510700. Therefore, the WIM sensors at site 0510700 need to be calibrated to generate reliable weight data.

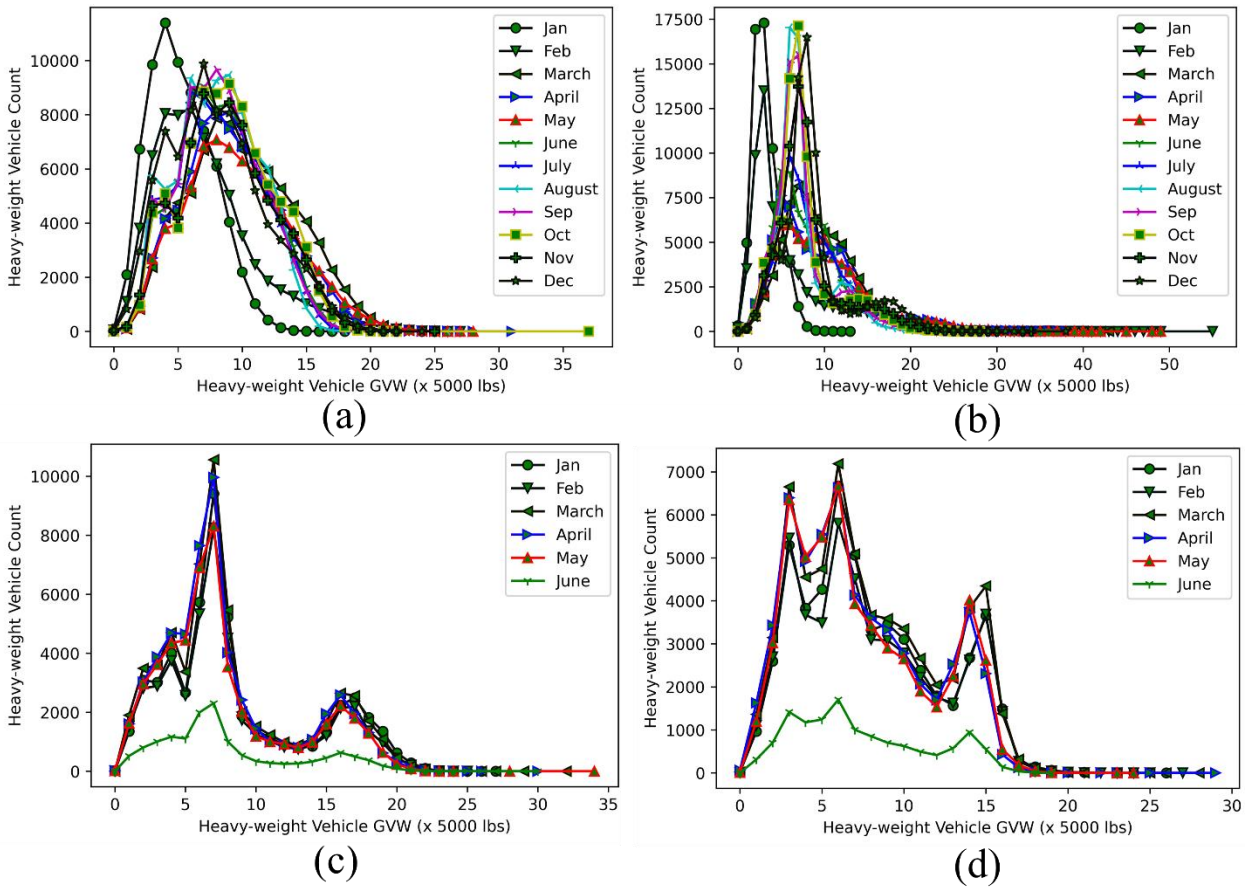


Fig. 28. Weight distribution plot for site 510700 and 510368

An example of a WIM site that does not warrant a call for calibration is shown in Fig. 28 (c) and Fig. 28 (d) for Site No. 510368. The site monthly gross vehicle weight distribution shows a

reduction in the total number of vehicles in June. However, the weight drift (to left or right) is not significant. In the WB bound lane, there is a small drift in weight from March to April, which is within 20% as shown in Table 9. EB lanes have a weight deviation greater than 5% and less than 10%.

5.8 Results of Weight Forecasting

5.8.1 SARIMA Daily Average Gross Vehicle Weight prediction models

Fig. 29 shows a time series analysis of the gross weight data sampled daily from the WIM data observed at site 390218. The solid gray line shows the training data on which the model is trained. The solid black line shows the test data, and the dashed black line shows the predicted gross weight. The test versus prediction plot is enlarged on the right-hand side to show the prediction accuracy. The top and bottom lines of the shaded region indicate the observed maximum and minimum vehicle weight observed on a particular day, respectively. The residual plots are shown in Fig. 30 (a) and Fig. 30 (b). It is observed that the residuals are uncorrelated and have a zero mean, which implies that the forecasts are not biased. Additionally, the residuals are normally distributed and have a constant variance.

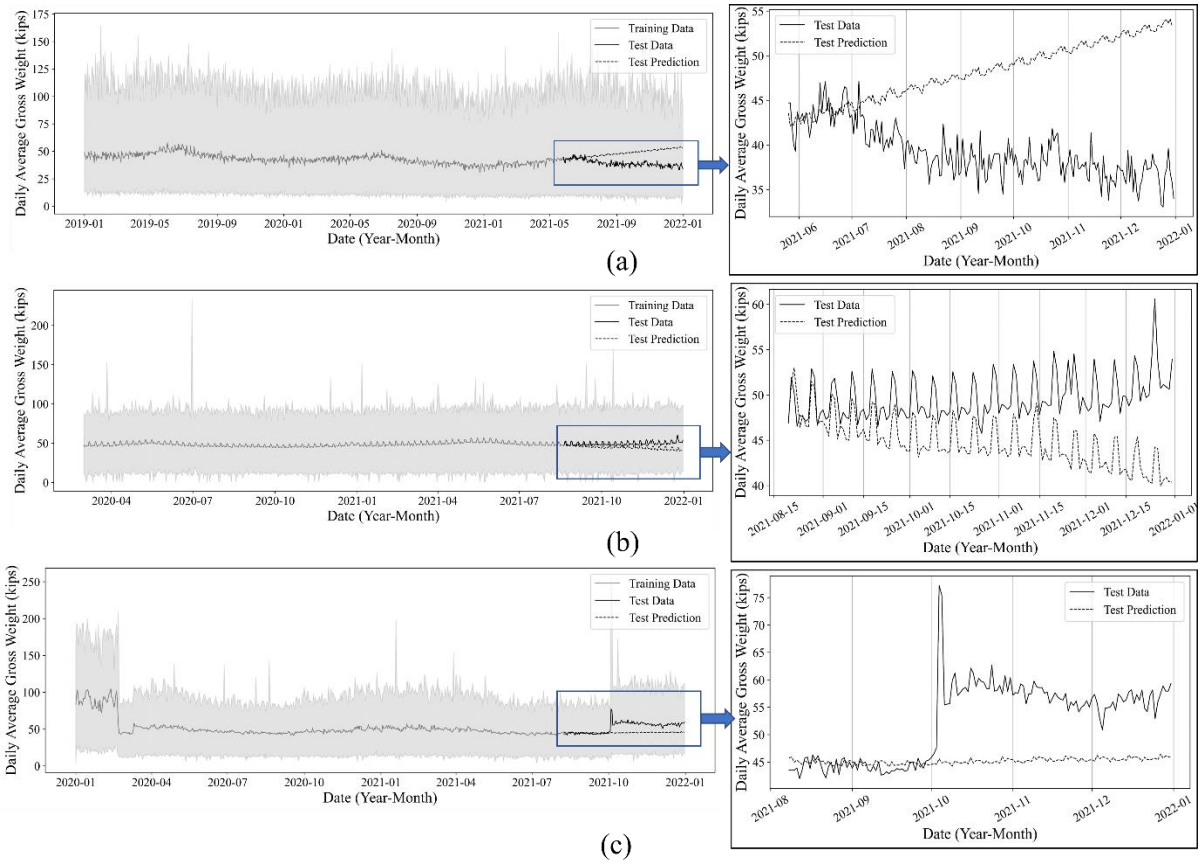


Fig. 29. Daily Average Gross Vehicle weight forecasting utilizing SARIMA method for site (a) 217334 (NB), (b) 390218 (NB), and (c) 870125 (SB)

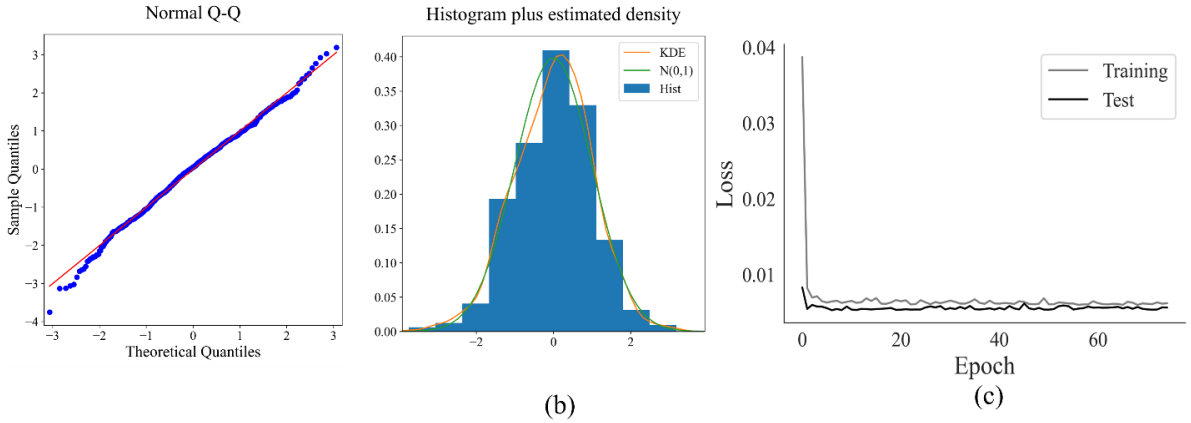


Fig. 30. (a) Q-Q Plot of Residuals in SARIMA model, (b) Distribution of residuals in SARIMA model, and (c) Loss vs Epoch plot in LSTM network for site 217334 (NB) lane

Fig. 29 (a) shows that the time series forecast deviates from the test data as the prediction length increases. Fig. 29 (b) shows that there are clear monthly trends in the vehicle weight data. However, the SARIMA model does not predict the time history of vehicle gross weight with great accuracy. The month of December shows the lowest moving average of gross vehicle weight. However, the time series forecast does not show any kind of reduction in weight in December. Thus, the SARIMA model fails to capture the effects of seasonality and trend in the weight data. The variance of prediction error grows with time, and Table 10 shows prediction errors calculated for the 29 WIM sites in Georgia. It is concluded that the ARIMA model does not give good results on the sequential vehicle weight data.

5.8.2 Application of Neural Network in Prediction of Daily Average Gross Vehicle Weight

The Long Short-term Memory (LSTM) model overcomes the deficiencies in the SARIMA model and predicts the complex weight pattern in the time series data. A loss function is a measure of how well the LSTM model predicts the expected weight outcome. The loss function is monitored with varying epoch (25-150) number to overcome an overfitting problem. The epoch number is selected when the loss function plateaus as shown in Fig. 30(c). This is a point where the gradient stops descending, and the analysis can no longer reduce the loss function. The "earlystop" function available in Keras is used to select the epoch number shown in Table 10. Once the epoch is selected, the model is fitted on the training data, and the prediction accuracy is evaluated using a test dataset.

Fig. 31 Fig. 32 presents the time-series gross and tandem axle weight data sampled daily as well as their time-series predictions by LSTM. Similar to the SARIMA prediction model, the gray solid line shows the training data. The black solid line shows the test data, and the dashed black

line shows the LSTM model prediction. The shaded region indicates a range between the observed maximum and minimum vehicle weights. An enlarged plot showing a comparison between the test data and prediction results is shown on the right of the figure.

In this study, the weight data only includes class 9 vehicles, and the gross permissible weight is 80 kips. Fig. 32 shows another important weight measure, the permissible tandem axle weight of 34 kips. The LSTM model predictions for 3 WIM sites are shown in . It is observed that the prediction agrees with the test data for site 217334-NB and site 390218-NB, and the prediction error is 1.69 kips and 1.25 kips, respectively. This error is considered acceptable because it is within 1% of the expected gross vehicle Fig. 31 weight (+/-80 kips). The model performance is satisfactory as compared to the SARIMA model because the prediction error is relatively low. Table 10 provides a summary of weight forecasting errors for the WIM sites analyzed in this study. It is concluded that the prediction accuracy improves when a LSTM model is used.

When the training data has anomalies such as a weight drift, they affect the LSTM prediction as shown in Fig. 31 (c). In case of time series data observed site 870125 in SB lane, the LSTM prediction captures the overall weight pattern as well as a downward shift trend resulting from the past (March 2020) weight drift observed in the training data. On the other hand, for site 390218 (NB), the LSTM model prediction agrees well with the test data, and no significant weight drift is observed.

Fig. 32 illustrates the time-series analysis of tandem axle weight using a LSTM neural network. A weekly weight pattern is observed in the WIM data for site 390218 in NB lane [see Fig. 32 (a)]. A small downward (1.5 kips) weight shift is observed in the test data during the last week of the forecasting period, which may indicate a sudden change in the vehicle weight pattern. The overall prediction was very accurate for that site. Fig. 32 (b) shows an example of the tandem

axle weight that exhibit a reasonable range of vehicle weight. Yet, the test data display no clear weight pattern and appear to show noise in the data. The prediction results show a similar trend reflecting the noise and no weekly weight pattern. Similar to site 870125 (SB)'s weight data shown in Fig. 31 (c), site 1270312 (SB) shows a sudden increase of tandem axle weight in Fig. 32 (d).

The influence of the calibration error or sensor malfunction is also reflected in the prediction as there is a constant deviation in between prediction and test data. Hence, by observing the deviation in test and prediction, it is possible to comment about the quality of the data, which eventually will be helpful in establishing the control dataset.

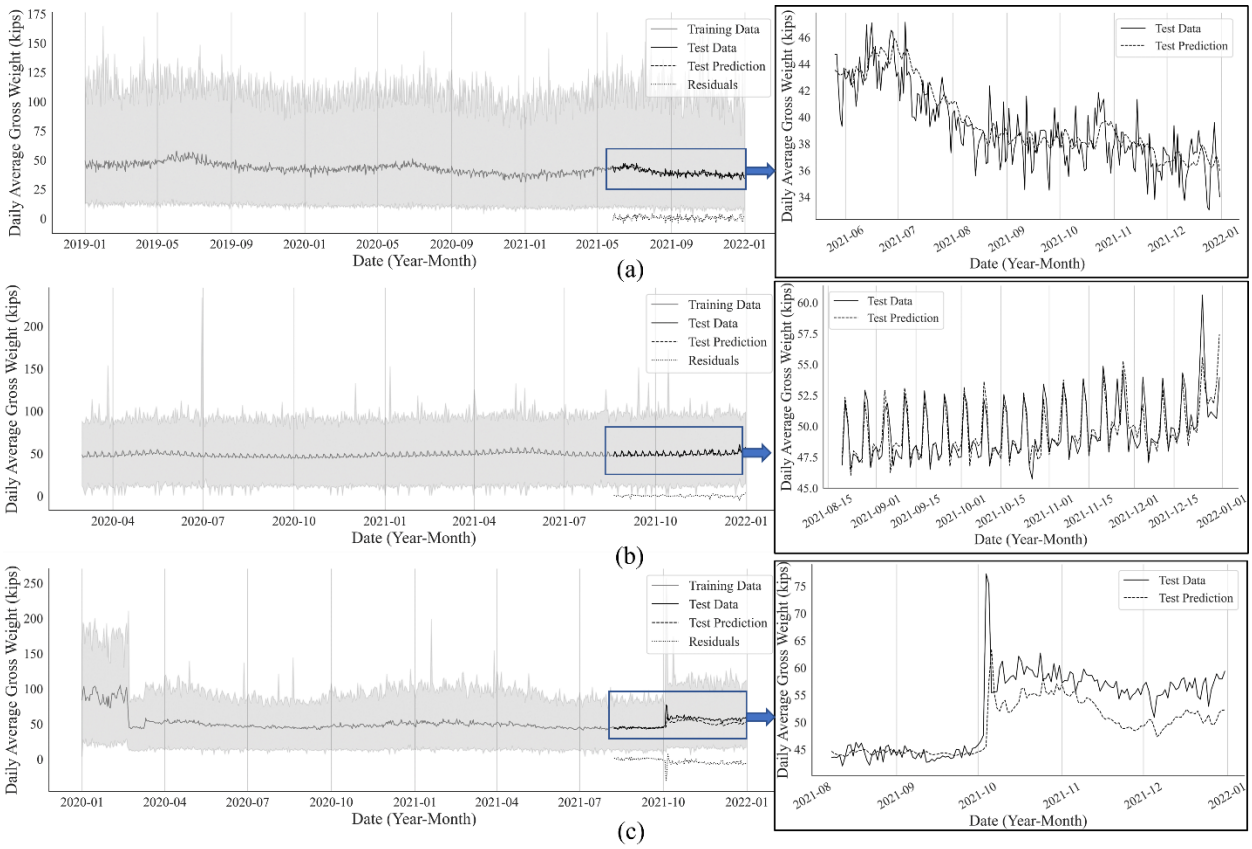


Fig. 31. Daily Average Gross Vehicle weight forecasting utilizing LSTM architecture for site (a) 217334 (NB), (b) 390218 (NB), and (c) 870125 (SB)

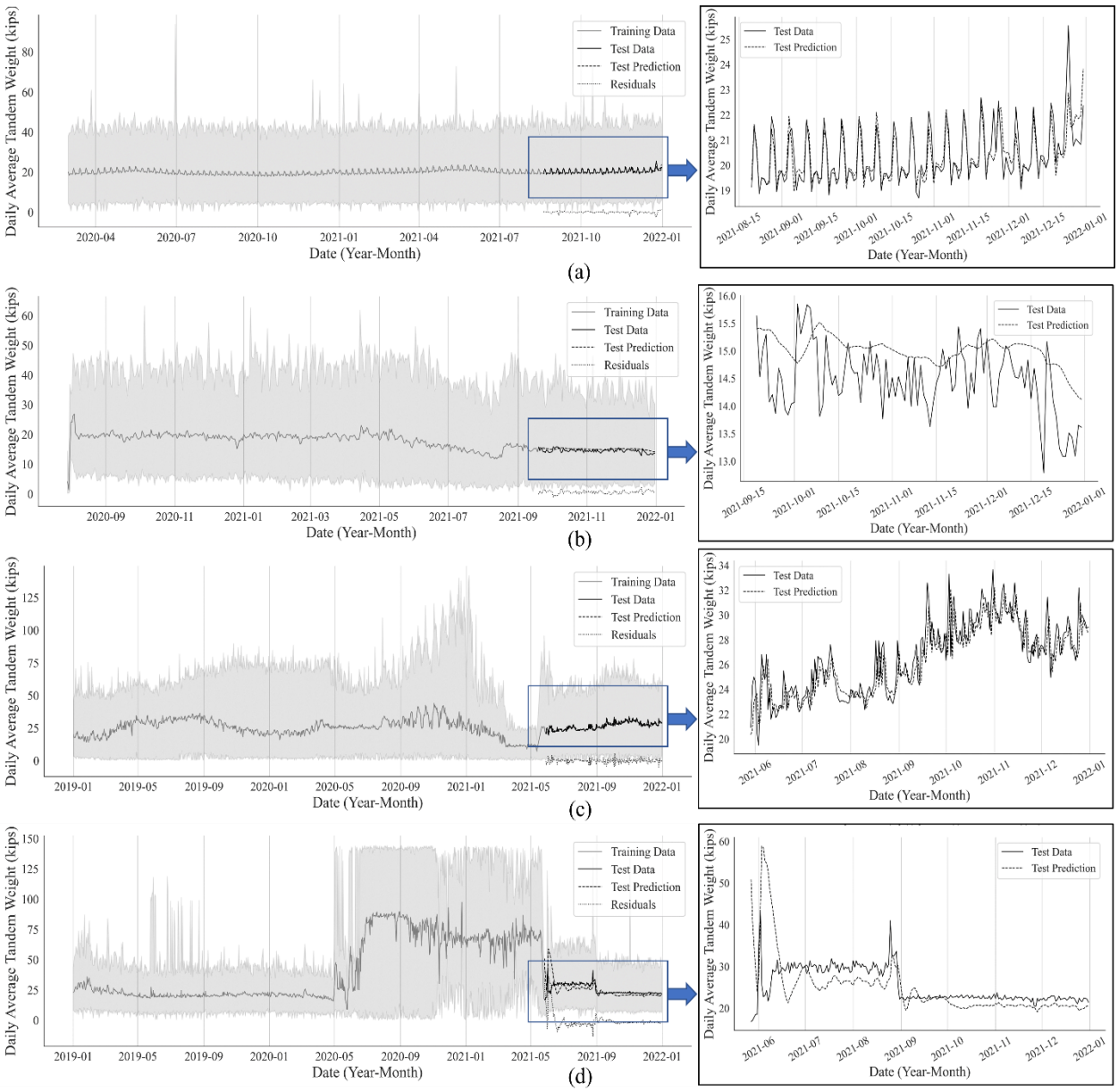


Fig. 32. Daily Average Gross Vehicle weight forecasting utilizing LSTM architecture for site (a) 390218 (NB), (b) 810347 (WB), (c) 830214 (NB), and (d) 1270312 (SB)

Table 10. Summary table of prediction using LSTM neural network and SARIMA model

Site ID	Lane	LSTM			SARIMA		
		Epoch (range)	RMSE (kips)	MAPE (%)	Model (SARIMA(p,d,q)(P,D,Q)[m])	RMSE (kips)	MAPE (%)
30132	EB	50-150	6.43	7.41	SARIMA(0,1,0)(7,1,0)[12]	21.25	30.89
30132	WB	50-150	2.74	4.70	SARIMA(2,1,0)(8,1,0)[12]	3.66	8.90
210378	NB	50-150	2.86	3.56	SARIMA(4,1,0)(8,1,0)[12]	11.75	16.45
210378	SB	50-150	6.54	5.45	SARIMA(3,1,3)(8,1,0)[12]	25.55	28.62
217334	NB	50-150	1.69	3.34	SARIMA(6,1,0)(8,1,0)[12]	3.52	8.50
217334	SB	50-150	1.58	2.17	SARIMA(3,1,0)(8,1,0)[12]	2.74	3.76
390218	NB	50-150	1.25	1.58	SARIMA(2,1,1)(6,1,3)[12]	2.49	5.44
390218	SB	50-150	0.84	1.07	SARIMA(8,1,3)(8,1,0)[12]	1.01	1.20
510368	EB	50-150	1.68	2.21	SARIMA(2,1,0)(4,1,1)[12]	4.21	6.68
510368	WB	50-150	1.45	2.02	SARIMA(2,1,0)(4,1,1)[12]	5.14	8.21
510387	NB	50-150	1.61	2.63	SARIMA(0,1,1)(6,1,1)[12]	1.55	3.09
510387	SB	50-150	14.79	6.92	SARIMA(0,1,1)(6,1,2)[12]	300.16	574.22
511113	EB	50-150	4.50	6.51	SARIMA(7,1,0)(8,1,0)[12]	9.63	21.82
511113	WB	50-150	4.62	6.57	SARIMA(6,1,0)(6,1,1)[12]	6.10	11.02
810347	EB	50-150	2.52	3.36	SARIMA(1,1,1)(8,1,0)[12]	9.20	14.85
810347	WB	50-150	1.35	2.94	SARIMA(0,1,0)(6,1,1)[12]	1.54	3.26
830214	EB	50-150	3.85	4.36	SARIMA(3,1,1)(8,1,0)[12]	32.96	46.35
830214	WB	50-150	2.64	3.31	SARIMA(5,1,1)(6,1,1)[12]	2.07	3.83
870103	NB	50-150	3.59	5.52	SARIMA(7,1,0)(8,1,0)[12]	6.58	12.26
870103	SB	50-150	3.18	4.03	SARIMA(6,1,0)(8,1,0)[12]	4.86	6.81
870125	NB	50-150	3.85	3.78	SARIMA(5,1,0)(7,1,0)[12]	14.73	23.19
870125	SB	50-150	2.76	4.02	SARIMA(2,1,2)(7,1,0)[12]	5.19	10.67
1030159	EB	50-150	5.60	7.70	SARIMA(2,1,2)(4,1,0)[12]	6.31	9.87
1030159	WB	50-150	2.27	2.61	SARIMA(1,1,0)(0,1,1)[12]	3.04	3.71
1150052	EB	50-150	3.43	4.10	SARIMA(0,1,0)(0,1,1)[12]	4.81	6.00
1150052	WB	50-150	3.16	4.88	SARIMA(0,1,2)(0,1,1)[12]	4.20	5.39
1270312	NB	50-150	1.20	1.84	SARIMA(5,1,2)(6,1,3)[12]	4.93	8.44
1270312	SB	50-150	11.71	9.96	SARIMA(8,1,0)(8,1,0)[12]	83.13	125.90
1450234	NB	50-150	4.38	5.64	SARIMA(2,1,0)(4,1,0)[12]	7.37	9.97
1450234	SB	50-150	4.27	6.00	SARIMA(2,1,1)(4,1,1)[12]	5.22	7.38
1610189	NB	50-150	2.30	3.98	SARIMA(3,1,1)(6,1,1)[12]	2.82	6.23
1610189	SB	50-150	3.97	4.77	SARIMA(5,1,3)(7,1,1)[12]	13.51	18.71
1750247	EB	50-150	12.73	5.19	SARIMA(0,1,1)(8,1,1)[12]	16.18	24.62
1750247	WB	50-150	1.53	2.76	SARIMA(2,1,0)(8,1,0)[12]	6.06	10.41
1850227	NB	50-150	0.93	1.58	SARIMA(6,1,1)(6,1,1)[12]	2.94	5.03
1850227	SB	50-150	0.57	0.76	SARIMA(0,1,2)(6,1,1)[12]	0.62	0.92
2170218	EB	50-150	1.50	2.05	SARIMA(0,1,1)(6,1,3)[12]	1.72	2.62
2170218	WB	50-150	1.02	1.39	SARIMA(0,1,0)(8,1,0)[12]	2.19	3.59
2350138	NB	50-150	6.21	6.87	SARIMA(2,1,0)(5,1,0)[12]	15.09	15.49
2350138	SB	50-150	6.99	7.35	SARIMA(1,1,0)(4,1,0)[12]	17.29	21.24
2850243	NB	50-150	2.14	3.66	SARIMA(7,1,0)(7,1,0)[12]	12.17	21.64
2850243	SB	50-150	1.69	2.63	SARIMA(2,1,0)(8,1,0)[12]	2.34	4.07

5.9 Analysis of the Results

Although statistical significance tests are widely used in comparing vehicle weight distributions between months, it does not accurately tell the weight shift. According to the non-parametric test, none of the sites in Georgia has consistent weight measurements. However, it is clearly visible in Fig. 27 (a) and (b) that the site 217334 has very low deviation in weights. Therefore, statistical significance test is not suitable to determine weight deviation.

The JS Divergence method, on the other hand, not only tells us about the quality of the data but also quantifies the weight deviation on a scale of 0 to 1. Using this method allows us to determine deviations in probability distributions more accurately. The weight deviation at site 217334 is less than 30% according to JS divergence, although statistical tests fail to demonstrate this. Therefore, JS divergence methods need to be employed to continuously monitor the weight deviation at every WIM location and sites with a weight deviation of 30% or higher should be considered for sensor recalibration to continuously generate good quality data.

SARIMA model is very useful tool in analyzing the trend and seasonal pattern in time series data and forecasting. One of the main advantages of this method is that it is easy to understand, and results are easily interpretable. However, SARIMA model is not capable of understanding the complexity in the time-series data. Whereas, in this study, time-series forecasting employing modern sophisticated machine learning models, more precisely LSTM network is proved to be more efficient in identifying the long term and short-term temporal dependencies and understanding the structure of data and complexity in time series forecasting. From Table 10, it is clear that LSTM network model outperforms the traditional SARIMA in terms of the reduction in errors (RMSE and MAPE) and improves the accuracy. The accuracy of forecasting using LSTM network improves even further when more quality data is available.

Time series base weight forecasting methods help us to better understand the everchanging complex traffic pattern over the year. This study illustrated the accurate forecasting capability of LSTM network model. LSTM is efficient in exploiting the data and finds out many complex hidden features, which yields better prediction. With the development of accurate forecasting techniques, vehicle weight forecasting method could help to create a control dataset by filling out missing values in the data and could be used as a reference dataset to identify or modify a different set of WIM data in future. Furthermore, traffic forecasting could also help the decision makers to schedule maintenance of highway infrastructure and allocate funds based on the forecasted traffic demand.

Dynamic inconsistency occurs when the decision maker has low confidence on the results of a weight deviation measurements. This study not only proposes a new technique in monitoring weight deviation but also increases the confidence in the results by quantifying the month weight drift. A novel JS divergence method is proposed to analyze monthly weight deviation and captures the first month when weight deviates significantly (say 30%) from the previous month. Instead of comparing the mean of a dataset with the mean of a control dataset, JS divergence method compares between two of the vehicle weight frequency distribution. It allows the decision maker to identify when the sensor began to malfunction and increases the credibility of the results. By providing more information regarding the weight shift, decision makers will feel more confident in implementing a strict weight monitoring policy.

5.10 Conclusion

This study analyzes WIM data from several sites in Georgia to statistically determine weight deviation. A JS divergence method (or similar statical tests) may be used to monitor a weight drift

in the data; however, a time-series model using a Long Short-Term Memory neural network is better for the following reasons.

- A time-series-based vehicle weight forecasting method employing an LSTM network efficient to unravel more information about the complex vehicle weight data pattern and provide highly accurate short-term weight prediction.
- Comparing the mean values of a WIM dataset with a reference dataset, which is the current practice in weight deviation monitoring provides a reduced amount of information about the data's quality, reduces the credibility of a result and ultimately increasing the potential for dynamic inconsistency. LSTM neural network-based time-series analysis enhances the confidence of a decision-maker in the accuracy of the results by providing more detailed information about weight deviations and the data quality. It is therefore crucial to conduct LSTM neural network-based time-series analysis to provide more insight about a range of live loads on our public roads and bridges.

Chapter 6

6. Conclusion and Recommendations

6.1 Conclusion

This dissertation investigates the WIM data from 29 sites on major highways in Georgia to statistically determine maximum live load factors and reliability indices for bridges. The National Bridge Inventory (NBI) data is used to establish operating load rating as the bridge capacity by means of advanced machine learning algorithms. Additionally, strategies for monitoring weight drifts are developed using deep neural networks. Based on the findings presented in this dissertation, the following conclusions are made:

- Highly right-skewed distributions exist in describing vehicle live load data, and the extent of kurtosis is numerically quantifiable. Within the service life of 75 years, it is concluded that live load factors are expected to exceed 1.75.
- Analyzing weekly maximum vehicle weight is efficient, although not sufficient, to predict the maximum live load factors because weekly maxima overall yield the highest live load factors, noting that this method is not able to capture all vehicle weights.
- While GWM predicts the highest live load factors, state agencies with a live load distribution exhibiting high kurtosis are more likely to benefit from the POT method in predicting expected maximum live load factors. This is because a higher goodness of fit is observed when the POT approach is used to fit a GPD to heavy vehicle weight data.

- Near-term (or present) live load factors are more salient [than service-life predictions] to decision makers because perceived live load factors are significantly smaller than the forecasted (10-75 years) live load factors.

- In defining the relationship between the changes in substructure condition rating and changes in operating load rating, a linear regression model is not adequate because multiple factors including the substructure condition rating affect the operating load rating. When several other factors are included as predictors—year built, condition ratings, bridge type, and design load—a nonlinear relationship identified by an XGBoost machine learning model yields a high accuracy. Other algorithmic models such as CatBoost and Random Forest produce similar results. The regression analysis indicates that a nonlinear regression involving machine learning algorithms provides a stronger relationship between the operating load ratings and bridge attributes such as condition ratings and age.

- Reliability analysis results show that a significant number of bridges have a low reliability index (less than 2), which poses a great risk to the transportation infrastructure. Coincidentally, low-reliability bridges are also relatively older. Therefore, the performance associated with bridges with a low-reliability index needs to be evaluated more frequently to increase the reliability and avert the risk associated with bridges.

- Statistical tests are appropriate to compare the medians of two distributions but are not sufficient to monitor weight drifts, specifically a weight reduction or increase at peak traffic. In addition, a statistical significance test is not adequate for quantifying the difference between two monthly weight distributions; it captures both vertical and horizontal weight shifts, but only horizontal drifts are relevant to weight calibration.

- The JS divergence method is an efficient approach for monitoring monthly weight deviations and erroneous weight data. It quantifies the distance between two probability distributions. Based on a visual inspection of monthly weight deviations observed at the 29 WIM sites, a reasonably acceptable and practical weight deviation threshold appears to be 30 percent although the ultimate goal may be 10 %. For the WIM sites exhibiting a weight deviation greater than 30 percent, their vehicle weight distributions present a noticeable drift. Such sites warrant a calibration or re-calibration of their weight sensors.
- A time-series-based vehicle weight forecasting method employing a LSTM network is more efficient to unravel hidden information about the complex vehicle weight data pattern and provide highly accurate short-term weight predictions which reflect the past weight trend and seasonality.
- Comparing the mean values of a WIM dataset with a reference dataset, which is the current practice in monitoring a weight drift provides a reduced amount of information about the data's quality, reduces the credibility due to uncertainty, and ultimately increases the potential for dynamic inconsistency. The LSTM neural network-based time-series forecasting method is expected to enhance the confidence providing more detailed information about weight drifts over time and the data quality.

6.2 Recommendations

Hyperbolic discounting plays an important role in human behavior in decision making. The discounted live load factors (0.01-0.02) to the present period are perceived as extremely small vehicle weights on bridges due to hyperbolic discounting behavior. To overcome such cognitive

bias, it is more perceptible for decision makers to work with bridge live load factors determined for the present (or near-term) period than obtaining 75-year service-life predictions.

Bridge infrastructure in the United States is threatened by aging, wear and tear, and other natural forces. Hence, there need to be significant changes in the highway infrastructure and asset maintenance policy. Furthermore, bridges relatively older with poor condition rating should be on the top priority for evaluation as it can be seen from the reliability analysis. Moreover, the vehicle weight should be monitored carefully, and the number of overweight vehicles needs to be validated against the number of permits issued on these routes with high importance, as these bridges are a matter of concern to the public safety. Additionally, data driven reliability assessment of bridges should be the new normal in bridge maintenance. The advantage of data-driven assessment is that it increases the transparency about bridge safety between owners and users by providing symmetric information. Such asset management practice considers the influence of growing heavy weight traffic in highways and thus able to account for demand risks. Symmetric information also eliminates the possibility of overconfidence of owners and helps to reduce the risk of adverse selection.

Lastly, to gain a deeper understanding of the complex vehicle weight patterns changing over time and overcome temporal inconsistencies, it is recommended that state agencies monitor weight data with reference to highly accurate near-term weight predictions using the latest deep neural networks such as LSTM and liquid time-constant networks.

7. References

1, E. (1994). "Basis of design and actions on structures, Part 3: Traffic loads on bridges." European Prestandard ENV 1991–3: European Committee for Standardisation, TC.

AASHTO (2017). "LRFD bridge design specifications." AASHTO, Washington, DC.

AASHTO (2018). "The Manual for Bridge Evaluation." AASHTO, Washington, DC.

Ang, A. H.-S., and Tang, W. H. (2007). *Probability concepts in engineering planning and design: Emphasis on application to civil and environmental engineering*, Wiley.

Blanchard, T. C., Pearson, J. M., and Hayden, B. Y. (2013). "Postreward delays and systematic biases in measures of animal temporal discounting." *Proceedings of the National Academy of Sciences*, 110(38), 15491-15496.

Caprani, C. C., and OBrien, E. J. (2010). "The use of predictive likelihood to estimate the distribution of extreme bridge traffic load effect." *Structural safety*, 32(2), 138-144.

Caprani, C. C., OBrien, E. J., and McLachlan, G. J. (2008). "Characteristic traffic load effects from a mixture of loading events on short to medium span bridges." *Structural safety*, 30(5), 394-404.

Chen, W., Ma, C., Xie, Z., Yan, B., and Xu, J. (2015). "Improvement of extrapolation of traffic load effect on highway bridges based on Rice's theory." *International Journal of Steel Structures*, 15(3), 527-539.

Chi, J. (2019). "WIM Based Live Load Factors for Consistent Illinois Bridge Reliability." Illinois Institute of Technology.

Enright, B., and O'Brien, E. J. (2013). "Monte Carlo simulation of extreme traffic loading on short and medium span bridges." *Structure and Infrastructure Engineering*, 9(12), 1267-1282.

Fu, G., and You, J. (2011). "Extrapolation for future maximum load statistics." *Journal of Bridge Engineering*, 16(4), 527-535.

GDOT "Georgia's Traffic Monitoring Program." Office of Trasportation Data, 600 West Peachtree Street, N.W., Atlanta, Georgia 30308.

Ghosn, M., and Frangopol, D. M. (1996). "Site-specific live load models for bridge evaluation." *Probabilistic Mechanics & Structural Reliability*, 30-33.

Ghosn, M., Sivakumar, B., and Miao, F. (2013). "Development of state-specific load and resistance factor rating method." *Journal of Bridge Engineering*, 18(5), 351-361.

Haldar, A., and Mahadevan, S. (2000). *Probability, Reliability, and Statistical Methods in Engineering Design*, J. Wiley & Sons, Incorporated.

INDOT (2020). "Indiana Department of Transportation." <https://www.in.gov/indot/2349.htm>. Accessed June 7, 2020.

James, G. (2003). "Analysis of traffic load effects an railway bridges." Byggvetenskap.

Kahneman, D., and Tversky, A. (1979). "On the interpretation of intuitive probability: A reply to Jonathan Cohen."

Keeney, R. L., and McDaniels, T. L. (1992). "Value-focused thinking about strategic decisions at BC Hydro." *Interfaces*, 22(6), 94-109.

Laman, J. A., and Nowak, A. S. (1997). "SITE-SPECIFIC TRUCK LOADS ON BRIDGES AND ROADS." *Proceedings of the Institution of Civil Engineers-Transport*, 123(2), 119-133.

Moses, F. (2001). *Calibration of load factors for LRFR bridge evaluation*.

Nassif, H., Ozbay, K., Na, C., Lou, P., Fiorillo, G., and Park, C. (2018). "Monitoring and Control of Overweight Trucks for Smart Mobility and Safety of Freight Operations."

Nowak, A. S. (1999). "Calibration of LRFD bridge design code, NCHRP Report 368." *Transportation Research Board, Washington, DC*, 208.

Nowak, A. S., and Rakoczy, P. (2013). "WIM-based live load for bridges." *KSCE Journal of Civil Engineering*, 17(3), 568-574.

Pidgeon, N., and Butler, C. (2009). "Risk analysis and climate change." *Environmental Politics*, 18(5), 670-688.

Sivakumar, B., and Ghosn, M. (2011). "Recalibration of LRFR live load factors in the AASHTO manual for bridge evaluation." *NCHRP Project*(20-07).

Soriano, M., Casas, J. R., and Ghosn, M. (2017). "Simplified probabilistic model for maximum traffic load from weigh-in-motion data." *Structure and Infrastructure Engineering*, 13(4), 454-467.

Zhou, X.-Y., Schmidt, F., Toutlemonde, F., and Jacob, B. (2016). "A mixture peaks over threshold approach for predicting extreme bridge traffic load effects." *Probabilistic Engineering Mechanics*, 43, 121-131.

Zhou, X. Y., Schmidt, F., and Jacob, B. (2012). "Extrapolation of traffic data for development of traffic load models: assessment of methods used during background works of Eurocode."

NBI (2020). "National Bridge Inventory." Federal Highway Administration, Washington DC.

GDOT "Georgia's Traffic Monitoring Program." Office of Transportation Data, 600 West Peachtree Street, N.W., Atlanta, Georgia 30308.

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

Chorzepa, M. G., Kim, S. S., Durham, S. A., and Sinha, A. (2020). "Development of Weigh-in-motion Data Quality Control Algorithms and Procedures." Georgia. Department of Transportation. Office of Performance-Based ….

Liao, C.-F., Chatterjee, I., and Davis, G. A. (2015). "Implementation of traffic data quality verification for WIM sites."

Nowak, A. S. (1999). "Calibration of LRFD bridge design code, NCHRP Report 368." *Transportation Research Board, Washington, DC*, 208.

Zhou, J., Shi, X., Caprani, C. C., and Ruan, X. (2018). "Multi-lane factor for bridge traffic load from extreme events of coincident lane load effects." *Structural Safety*, 72, 17-29.

Sinha, A., Chorzepa, M. G., Yang, J. J., Sonny Kim, S., and Durham, S. A. (2022). "Cognitive Approaches to Hyperbolic Discounting of High-Impact Low-Probability Bridge Overload Events and Live-Load Factors." *Journal of Performance of Constructed Facilities*, 36(2), 04022009.

AASHTO, L. (2011). "Manual for bridge evaluation." *Association of State Highway and Transportation Officials, Washington DC*.

AASHTO (2011). *The manual for bridge evaluation*, AASHTO.

Wang, N., O'Malley, C., Ellingwood, B. R., and Zureick, A.-H. (2011). "Bridge rating using system reliability assessment. I: Assessment and verification by load testing." *Journal of Bridge Engineering*, 16(6), 854-862.

Frangopol, D. M., Strauss, A., and Kim, S. (2008). "Bridge reliability assessment based on monitoring." *Journal of Bridge Engineering*, 13(3), 258-270.

AASHTO (2018). "The Manual for Bridge Evaluation." AASHTO, Washington, DC.

AASHTO (2017). "LRFD bridge design specifications." AASHTO, Washington, DC.

Ghosn, M., and Frangopol, D. M. (1996). "Site-specific live load models for bridge evaluation." *Probabilistic Mechanics & Structural Reliability*, 30-33.

Enright, B., and O'Brien, E. J. (2013). "Monte Carlo simulation of extreme traffic loading on short and medium span bridges." *Structure and Infrastructure Engineering*, 9(12), 1267-1282.

Ghosn, M., Sivakumar, B., and Miao, F. (2013). "Development of state-specific load and resistance factor rating method." *Journal of Bridge Engineering*, 18(5), 351-361.

Chi, J. (2019). "WIM Based Live Load Factors for Consistent Illinois Bridge Reliability." Illinois Institute of Technology.

Laman, J. A., and Nowak, A. S. (1997). "SITE-SPECIFIC TRUCK LOADS ON BRIDGES AND ROADS." *Proceedings of the Institution of Civil Engineers-Transport*, 123(2), 119-133.

Nassif, H., Ozbay, K., Na, C., Lou, P., Fiorillo, G., and Park, C. (2018). "Monitoring and Control of Overweight Trucks for Smart Mobility and Safety of Freight Operations."

Pidgeon, N., and Butler, C. (2009). "Risk analysis and climate change." *Environmental Politics*, 18(5), 670-688.

Keeney, R. L., and McDaniels, T. L. (1992). "Value-focused thinking about strategic decisions at BC Hydro." *Interfaces*, 22(6), 94-109.

1, E. (1994). "Basis of design and actions on structures, Part 3: Traffic loads on bridges." European Pre-standard ENV 1991–3: European Committee for Standardisation, TC.

Caprani, C. C., and OBrien, E. J. (2010). "The use of predictive likelihood to estimate the distribution of extreme bridge traffic load effect." *Structural safety*, 32(2), 138-144.

Moses, F. (2001). *Calibration of load factors for LRFR bridge evaluation*.

Nowak, A. S., and Rakoczy, P. (2013). "WIM-based live load for bridges." *KSCE Journal of Civil Engineering*, 17(3), 568-574.

Haldar, A., and Mahadevan, S. (2000). *Probability, Reliability, and Statistical Methods in Engineering Design*, J. Wiley & Sons, Incorporated.

Sivakumar, B., and Ghosn, M. (2011). "Recalibration of LRFR live load factors in the AASHTO manual for bridge evaluation." *NCHRP Project*(20-07).

Zhou, X. Y., Schmidt, F., and Jacob, B. (2012). "Extrapolation of traffic data for development of traffic load models: assessment of methods used during background works of Eurocode."

Chen, W., Ma, C., Xie, Z., Yan, B., and Xu, J. (2015). "Improvement of extrapolation of traffic load effect on highway bridges based on Rice's theory." *Int. J of Steel Structures*, 15(3), 527-539.

Caprani, C. C., OBrien, E. J., and McLachlan, G. J. (2008). "Characteristic traffic load effects from a mixture of loading events on short to medium span bridges." *Structural safety*, 30(5), 394-404.

Fu, G., and You, J. (2011). "Extrapolation for future maximum load statistics." *Journal of Bridge Engineering*, 16(4), 527-535.

James, G. (2003). "Analysis of traffic load effects an railway bridges." Byggvetenskap.

Zhou, X.-Y., Schmidt, F., Toutlemonde, F., and Jacob, B. (2016). "A mixture peaks over threshold approach for predicting extreme bridge traffic load effects." *Probabilistic Engineering Mechanics*, 43, 121-131.

Soriano, M., Casas, J. R., and Ghosn, M. (2017). "Simplified probabilistic model for maximum traffic load from weigh-in-motion data." *Structure and Infrastructure Engineering*, 13(4), 454-467.

Ang, A. H.-S., and Tang, W. H. (2007). *Probability concepts in engineering planning and design: Emphasis on application to civil and environmental engineering*, Wiley.

Kahneman, D., and Tversky, A. (1979). "On the interpretation of intuitive probability: A reply to Jonathan Cohen."

Blanchard, T. C., Pearson, J. M., and Hayden, B. Y. (2013). "Postreward delays and systematic biases in measures of animal temporal discounting." *Proceedings of the National Academy of Sciences*, 110(38), 15491-15496.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., and Joulin, A. "Emerging properties in self-supervised vision transformers." *Proc., Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9650-9660.

INDOT (2020). "Indiana Department of Transportation." <https://www.in.gov/indot/2349.htm>. (June 7, 2020).

Akgül, F., and Frangopol, D. M. (2004). "Time-dependent interaction between load rating and reliability of deteriorating bridges." *Engineering Structures*, 26(12), 1751-1765.

FHWA (2020). "Bridge & Structures."Federal Highway Administration, Washington DC.

Estes, A. C., and Frangopol, D. M. (2005). "Load rating versus reliability analysis." *Journal of Structural Engineering*, 131(5), 843-847.

Hu, X., Daganzo, C., and Madanat, S. (2015). "A reliability-based optimization scheme for maintenance management in large-scale bridge networks." *Transportation Research Part C: Emerging Technologies*, 55, 166-178.

Xie, H.-B., Wu, W.-J., and Wang, Y.-F. (2018). "Life-time reliability based optimization of bridge maintenance strategy considering LCA and LCC." *Journal of cleaner production*, 176, 36-45.

Sandroni, A., and Squintani, F. (2013). "Overconfidence and asymmetric information: The case of insurance." *Journal of Economic Behavior & Organization*, 93, 149-165.

Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., and Ridella, S. "The 'K'in K-fold cross validation." *Proc., 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 441-446.

Jootoo, A., and Lattanzi, D. (2017). "Bridge type classification: Supervised learning on a modified NBI data set." *Journal of Computing in Civil Engineering*, 31(6), 04017063.

Galambos, T. V., and Ravindra, M. (1981). "Load and resistance factor design." *Engineering Journal, AISC*, 18(3), 78-84.

Nowak, A. S. (2004). "System reliability models for bridge structures." *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 321-328-321-328.

Chen, T., and Guestrin, C. "Xgboost: A scalable tree boosting system." *Proc., Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. (2018). "CatBoost: unbiased boosting with categorical features." *Advances in Neural Information Processing Systems*, 31.

Breiman, L. (2001). "Random forests." *Machine Learning*, 45(1), 5-32.

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

Joyce, J. M. (2011). "Kullback-leibler divergence." *International encyclopedia of statistical science*, Springer, 720-722.

Raiber, F., and Kurland, O. "Kullback-leibler divergence revisited." *Proc., Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, 117-124.

Cai, Y., and Lim, L.-H. (2022). "Distances between probability distributions of different dimensions." *IEEE Transactions on Information Theory*.

Box George, E., Jenkins Gwilym, M., Reinsel Gregory, C., and Ljung Greta, M. (1976). "Time series analysis: forecasting and control." *San Francisco: Holden Bay*.

He, Q., Wang, Y., Tu, H., and Ruan, X. (2018). "Incorporating the WIM Data into the Analysis of Road Traffic Risk." *CICTP 2017: Transportation Reform and Change—Equity, Inclusiveness, Sharing, and Innovation*, American Society of Civil Engineers Reston, VA, 4730-4739.

Smolarek, L., and Ziemska, M. (2017). "Using information collected by weigh in motion for modeling traffic structure of vehicles." *Archives of Transport System Telematics*, 10.

Kumar, S. V., and Vanajakshi, L. (2015). "Short-term traffic flow prediction using seasonal ARIMA model with limited input data." *European Transport Research Review*, 7(3), 1-9.

Hochreiter, S., and Schmidhuber, J. (1997). "Long short-term memory." *Neural computation*, 9(8), 1735-1780.

Li, J., Gao, L., Song, W., Wei, L., and Shi, Y. "Short term traffic flow prediction based on LSTM." *Proc., 2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP)*, IEEE, 251-255.

Morris, C., Yang, J. J., Chorzepa, M. G., Kim, S. S., and Durham, S. A. (2022). "Self-Supervised Deep Learning Framework for Anomaly Detection in Traffic Data." *Journal of Transportation Engineering, Part A: Systems*, 148(5), 04022020.

Tian, Y., Zhang, K., Li, J., Lin, X., and Yang, B. (2018). "LSTM-based traffic flow prediction with missing data." *Neurocomputing*, 318, 297-305.