infrastructures

Article

Deep-Learning-Based Temporal Prediction for Mitigating Dynamic Inconsistency in Vehicular Live Loads on Roads and Bridges

Ananta Sinha, Mi G. Chorzepa *, Jidong J. Yang, Sung-Hee Sonny Kim and Stephan Durham

College of Engineering, The University of Georgia, Athens, GA 30605, USA
* Correspondence: chorzepa@uga.edu

Abstract: Weigh-In-Motion (WIM) data have been collected by state departments of transportation (DOT) in the U.S. and are anticipated to grow as state DOTs expand the number of WIM sites in order to better manage transportation infrastructure and enhance mobility. Traditional approaches for monitoring the vehicle weight measured in WIM systems include conducting statistical tests between two datasets obtained from two calibration visits. Depending on the frequency of visits, these traditional approaches are ineffective or resource-demanding for identifying calibration needs. Excessive vehicle-weight drifts exceeding 10% are usually indicative of poor performance by WIM systems. However, it has been difficult to consistently monitor such performance due to the sheer amount of data. In Georgia, the number of WIM sites have expanded from 12 to 29 in the past 3 years. This paper proposes a deep-learning-based temporal prediction approach for modeling sequential data and monitoring the time-history of the live loads imposed on roads and bridges. In total, 29 WIM sites in Georgia are analyzed to examine the effectiveness of a proposed temporal prediction approach for evaluating observed live loads. This study finds that the Jensen–Shannon divergence method is more effective than statistical difference tests, particularly when screening for live load anomalies. It is concluded that a LSTM neural network is able to capture temporal dynamics underlying the sequential load patterns observed in the WIM data and serves as an effective model for consistently monitoring the performance of WIM systems over time.

Keywords: weigh-in-motion; WIM; RNN; LSTM; JS divergence; deep learning; dynamic inconsistency; live load; temporal

check for updates

Citation: Sinha, A.; Chorzepa, M.G.; Yang, J.J.; Kim, S.-H.S.; Durham, S. Deep-Learning-Based Temporal Prediction for Mitigating Dynamic Inconsistency in Vehicular Live Loads on Roads and Bridges. *Infrastructures* **2022**, *7*, 150. https://doi.org/10.3390/infrastructures7110150

Academic Editor: Fabrizio Scozzese

Received: 31 August 2022
Accepted: 21 October 2022
Published: 31 October 2022

1. Introduction

1.1. Background

Weigh-In-Motion (WIM) systems have been widely used by state agencies for quantifying infrastructure usage for weight enforcement, maintenance, traffic forecasting, infrastructure-investment decision-making, and transportation planning. They are installed on roadways and include quartz load cells for measuring axle weights. Quartz load cells are referred to as strain-gauge-type load sensors and are a piezoelectric device that involves electric polarization resulting from the application of mechanical force, such as a vehicle's weight. However, despite the advances made in sensor technology [1], WIM systems face challenges in obtaining accurate and reliable live load data because they are sensitive to disruptive events, driving patterns, weather conditions, speeding, and changes in surrounding pavement conditions [2]. In this study, evidence-based vehicle-weight-data quality-control (QC) measures are investigated to increase the reliability of WIM systems. One such approach emphasizes any measures that consistently reflect changes in the quality of live loads and trends of traffic demands over time.

In the state of Georgia (in the southeastern region of the United States), WIM data are anticipated to be used for managing an inventory of approximately 15,000 bridges and

25,000 centerline miles of roads. The Georgia Department of Transportation uses computer simulations and/or asset management systems to predict future bridge and pavement conditions and identify the most cost-effective methodology for its long-term asset management given the available funding. As a result, the reliability of asset usage and growth in usage depends on the accuracy of the vehicle weights measured by WIM systems. Bridges and pavements, which are subjected to repeated overstress and dynamic magnification of stresses and deflections, may experience sudden failure and/or require costly repairs or rehabilitation. Additionally, the performance and live-load-carrying capacity of a bridge or of a pavement at its design stage does not remain constant but reduces as time progresses. That is, accurately characterizing the live load spectra imposed on pavements and bridges and quantifying the associated uncertainties are important processes for evaluating the asset deterioration and traffic inputs required for the mechanistic–empirical pavement design guide (MEPDG) as well as assessing the live loads required for bridge design.

Sujon and Dai [3] have evaluated the need for dynamic vehicle-weight monitoring in highway infrastructure maintenance and emphasized the need for advanced machine learning and deep learning tools to monitor the WIM data quality. A recent study [4] also emphasized the importance of WIM data monitoring and proposed a methodology to implement a computer-vision-based bridge-model updating technique. Another study [5] proposes a WIM data-driven bridge-load-rating methodology. Therefore, there is a need to generate quality WIM data, such that they can be implemented for the improvement of transportation infrastructure.

This study investigates current vehicle-weight-monitoring practices used for WIM systems [6] and proposes a time-series weight-forecasting approach for consistently monitoring live loads over time. The current approach includes conducting statistical difference tests on two datasets produced between two calibration visits. Researchers have proposed another approach [7] to auto-calibrate WIM sensors by using Automatic Vehicle Identification (AVI) techniques and improve the quality of the WIM data collected. This study examines an improved method for measuring a divergence between two weight datasets. Subsequently, this study examines the benefits of reviewing a time history of vehicle weights and monitoring them in reference to time-series predictions of vehicular live loads. Such an approach is referred to as a temporal prediction method and uses a Recurrent Neural Network, which is a deep learning approach for modeling data with a temporal sequence. It specifically employs the Long Short-Term Memory (LSTM) method [8] for sequential learning, which has advanced considerably with promising machine learning approaches. A traditional time-history-forecasting method is also investigated for comparison. Lastly, a methodology employing a LSTM-based network architecture is successfully employed to examine vehicular live loads observed in 29 WIM sites in Georgia, including those in the Atlanta and Savannah regions, as shown in Figure 1.

Figure 2 shows photos taken during a WIM system calibration visit. Each WIM site is equipped with a data acquisition system and magnetic loop and load sensors embedded in the road, as shown in Figure 2c.

1.2. Need to Monitor Divergence in Vehicular Live Loads

Monitoring live loads imposed on public roads and bridges is important [9] for transportation asset management and risk/safety assessment [10]. Additionally, gaining a deeper understanding of evolving live-load patterns is critically important for future planning and design. For example, more e-commerce distribution centers such as Amazon warehouses have been established in the U.S., so the vehicle weight distribution in the vicinity of such locations has significantly changed. Closely spaced autonomous trucks are increasingly heading out on highways and are expected to operate in groups. Researchers have proposed methodologies to identify vehicles' loads by time–space-coupled distributed dynamic loads [11]. Thus, monitoring vehicular live loads observed in WIM systems is critically important for safety and understanding respective risk because the axle weight data are used for bridge load posting and the design of roads and bridges. Additionally, a weight drift or anomaly can indicate a potential quality issue in WIM systems. State DOTs often identify such weight drifts observed by WIM sensors, by comparing the weight difference between group means that are statistically significant. A control or reference dataset is generally established from the most recent calibration of weight sensors and is compared with a newly acquired dataset.

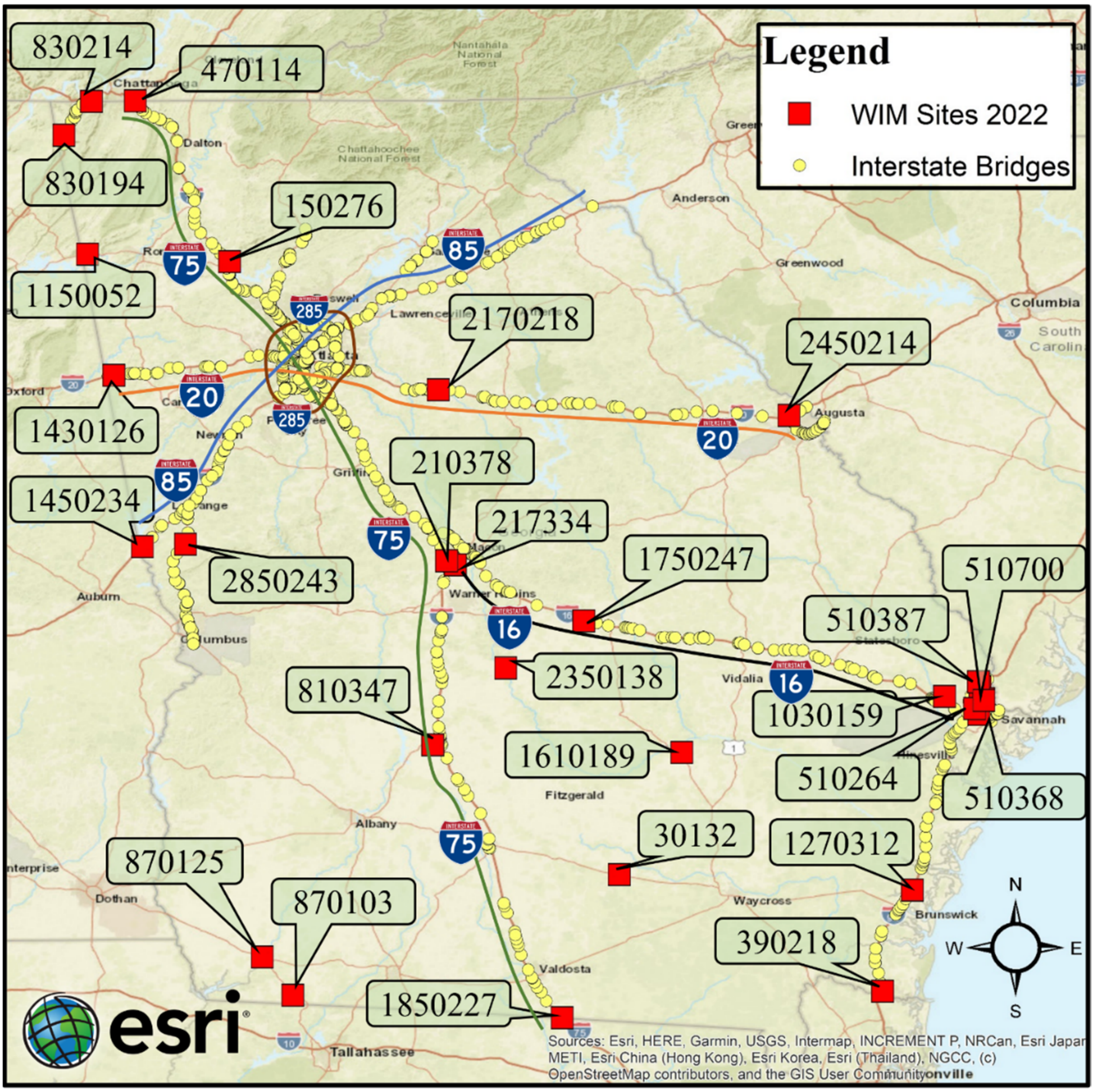


Figure 1. Map showing the locations of 29 WIM sites in Georgia.

1.3. Need to Mitigate Dynamic Inconsistency in Monitoring WIM Performance

Enforcing a 95% confidence level for the acceptable range of the vehicle-weight drift (<5%) measured in WIM systems appears to be a reasonable policy, due to the tendency to leave policies as they are. However, a problem arises when a decision-maker observes that the threshold level yields the need for a calibration adjustment of the weight sensors at a significant number of WIM sites. They may dynamically consider changing the policy or apply a different criterion (or method), either due to the benefit of the doubt or a lack of knowledge. A control dataset can contain monthly or yearly data, depending on calibration interval, and is collected immediately after the latest calibration. A lack of confidence in the control data, due to the dynamic nature of vehicle weights and other maintenance problems, contributes to an increased inconsistency of the decision-maker in measuring the difference between two datasets collected at two different time periods.

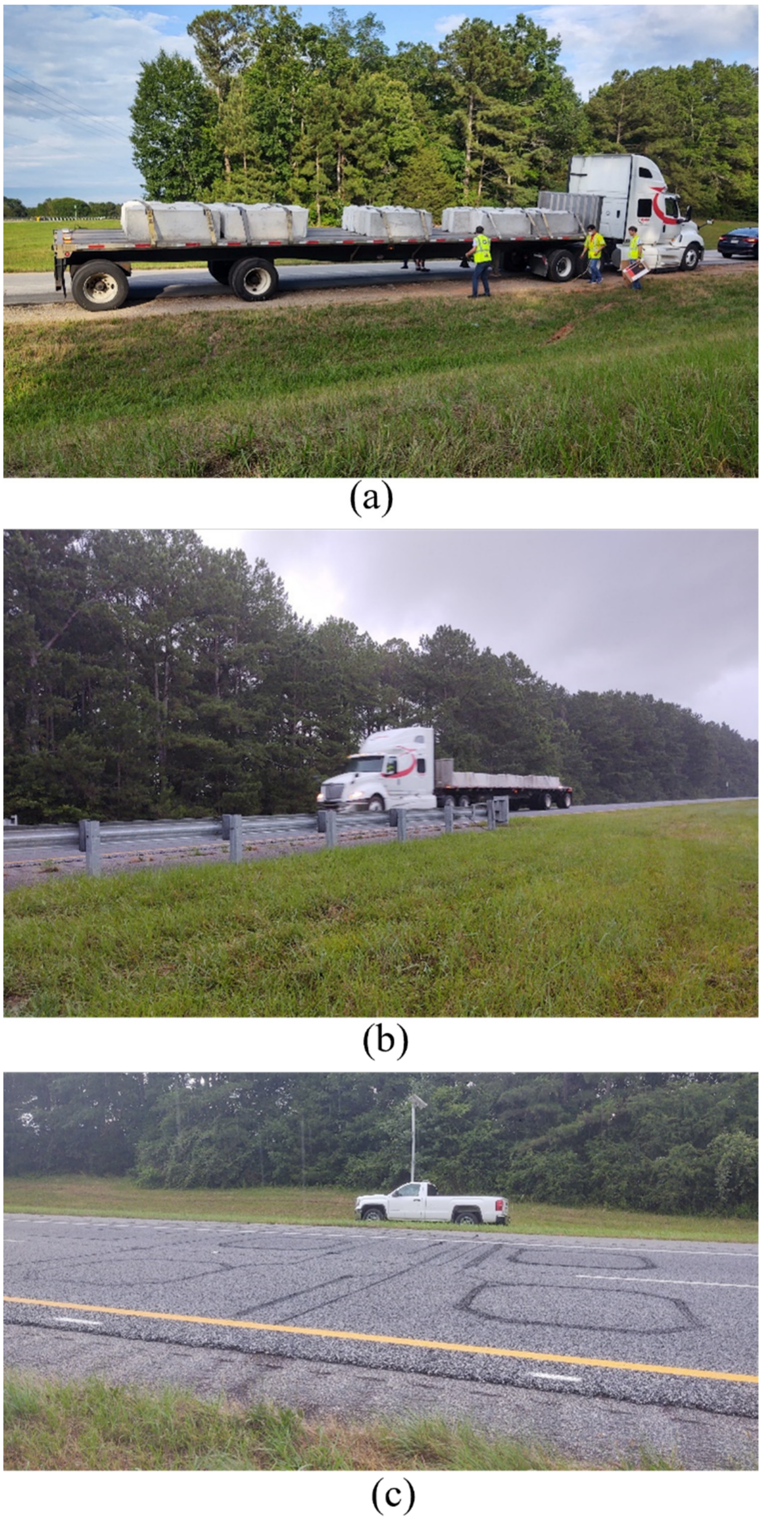


Figure 2. Photo illustration showing a WIM site: (**a**) a vehicle used for calibration of WIM systems, (**b**) calibration vehicle running on an instrumented lane, and (**c**) weight sensor installed on a roadway.

Such dynamic inconsistency is one of the most profound problems in behavioral economics and social science. It refers to a situation in which a decision-maker's preference changes over time. In this paper, it is hypothesized that, currently, the effectiveness of a weight-monitoring policy relies on the credibility of a commitment strategy to implement or sustain such a policy in the future. To sustain a time-consistent weight monitoring policy and detect system failures, a more precise method of measuring the drift in vehicle-weight distribution and capturing changes in weight patterns must be proposed to increase the credibility of the results. Contrary to a retrospective approach of comparing a vehicle-weight dataset to a control dataset, this study investigates a proactive approach of predicting time-series weight patterns in order to monitor evolving live loads. Once proven effective, state agencies can consistently review live loads in reference to time-series predictions and more holistically assess the associated errors.

1.4. Significance and Motivation

Figure 1 presents a map showing the locations of 29 WIM sites in the state of Georgia. The GDOT currently retrieves vehicle axle weight data annually and evaluates the need for calibration. It was speculated that some of the main highway bridges in Atlanta experienced damage to their concrete deck joints due to increased and repeated heavy truck traffic, based on visual observation and traffic volume. This study investigated the live loads measured at multiple WIM sites located on the routes between the Port of Savannah and Atlanta as well as those along the routes running between Atlanta and (1) Gainesville, FL, (2) Columbia, SC, (3) Chattanooga, TN, and (4) Birmingham, AL. With the existing weight drifts measured by WIM systems, there was no strong evidence to determine whether the bridges were experiencing heavier-than-expected live loads [12] in Atlanta, as the major Port of Savannah deepening project commenced in 2015. Savannah has become the third busiest U.S. port. As of December 2021, the live loads measured by the WIM systems near the port are very high, which appear to be affecting local roads near the port. Heavier container-truck weights appear to be redistributed in distribution centers near the Port of Savannah and, thus, do not appear to significantly affect the major interstate routes (I-16 and I-75) toward Atlanta. Additionally, Brunswick, a city south of Savannah, is the number one terminal in the nation for new automobile imports, processing 900,000 vehicles per year. Recently retrieved WIM data have indicated an increase in vehicle weights in the vicinity of Atlanta; however, such changes alone are not an absolute indicator of an increase in vehicle weight, because weight sensors may be significantly affected by other factors such as surface and weather conditions and dynamic amplification.

1.5. Research Questions and Scope

This study specifically aims to answer the following research questions:

- So far, an annual examination of WIM data was determined reasonable in Georgia due to the cost and logistics associated with the process. It has been observed that WIM data have significant month-to-month variations in weight distributions, which are a barrier for examining yearly data and consistently detecting weight anomalies. The possible cause may be due to a calibration error, a sensor malfunction, or other natural forces. How does one quantify a monthly (or yearly) weight drift at a particular WIM site using statistical approaches and develop a strategy for WIM sensor calibration?
- The current practice of monitoring a vehicle-weight drift involves performing a statistical significance test on two consecutive months (or years) of datasets. Are the results acceptable? If not, is there an improved approach? Is time-series prediction a better method for consistently monitoring future weight drifts than the current practice (performing a statistical test)?
- Seasonal Autoregressive Integrated Moving Average (SARIMA) is a very popular time-series forecasting method and, thus, is initially considered to predict a vehicle-weight trend or establish a reference dataset for a comparison with future weight data. With the latest advances in artificial intelligence, deep learning (DL) methods

are expected to improve time-series predictions. One of the promising DL models with growing popularity is LSTM. Does LSTM perform better than the traditional time-series prediction methods such as SARIMA?

- How does monitoring time-series weight data help mitigate the dynamic inconsistency problem described above?

Although statistical significance tests are widely accepted by state DOTs to evaluate a vehicle-weight drift observed in two WIM datasets between calibration visits, they consider a probability distribution and, thus, are not able to isolate a weight deviation alone. Hence, this study proposes using the JS divergence method, which normalizes the probability distribution of monthly (or yearly) vehicle weights and quantifies the divergence between two weight distributions. Nevertheless, such statistical tests on probability distributions do not explain the time component of weight data. A time-series forecast takes past observations and makes predictions based on what the expected live loads will be in the future, if the same load patterns and trends continue to hold true. Consequently, this study investigates a popular deep learning module (LSTM) to predict near-term gross vehicle weights using recent WIM data. This model is trained and validated with the most recent 6-month data and compared with the prediction outcomes obtained from a traditional model, SARIMA.

2. Current Practices and State of the Art

2.1. Statistical Tests for Comparing Vehicle-Weight Datasets in Different Time Periods

Statistical difference tests are normally performed to observe a significant vehicle-weight drift in WIM systems. Such tests involve two datasets between two calibration visits and, thus, include either two consecutive months or years of data. In this approach, the vehicle-weight data obtained from the latest calibration visit is considered as a control dataset. For example, Figure 2 shows a vehicle used for calibration at Site Number 1430126 on 5 June 2021. Therefore, the weight data observed between 5 June and 5 July are used as the control dataset for a statistical test. A control dataset is defined as the data that is used to evaluate another set of data. For instance, weight data observed from a subsequent year may be tested against a control dataset obtained from the prior year.

As WIM sensors can lose accuracy over several months due to temperature changes, rain, and other factors, it is important to determine how much vehicle-weight drift is present in the data. Mann–Whitney U and Kruskal–Wallis tests are the two nonparametric statistical tests initially performed in this study [13], with the weights of FHWA vehicle class 4 or greater vehicles (or truck traffic) from 29 WIM sites. The significance level was set to 0.05 (5%). In these statistical tests, we reject the null hypothesis, that the two samples come from populations with the same distribution, if the p-value is less than or equal to 0.05. If the p-value is greater than the significance level, we fail to reject the null hypothesis. There are limitations of these statistical tests (see the Results section), and, thus, an improved approach (JS Divergence) is investigated.

2.2. Jensen–Shannon (JS) Divergence

This approach considers two probability distributions of vehicle weights, P and Q, similar to other statistical tests. For discrete probability distributions, they are defined on the same probability space χ, and the relative entropy (or KL divergence) from Q to P is defined by Equation (1), where x denotes a vehicle weight.

$$D_{KL}(P \parallel Q) = \sum_{x \in \chi} P(x) \log\left(\frac{P(x)}{Q(x)}\right) \tag{1}$$

The Kullback–Leibler (KL) divergence [14] measures the nonmetric distance between two distributions. This divergence is not symmetrical, which means that the KL distance from *P(x)* to *Q(x)* is not the same as the distance between *Q(x)* and *P(x)*. In addition, it does not follow the triangular inequality [15], meaning that severe distortions are observed when measuring the degree of similarity. Furthermore, determining the KL distance can be challenging when $P \neq 0$, but $Q = 0$, because the divergence, $D_{KL}(P \parallel Q)$, must be defined as an infinity. This means that if one event is possible, the other event is not. As a result, the two distributions must be different. Meanwhile, the JS divergence [16] is a symmetrical and smoothed/normalized version of the KL divergence, defined by Equation (2), where $M = (P + Q)/2$. The JS divergence (JSD) is bounded by 0 and 1.

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

In this paper, monthly vehicle-weight distributions in vehicle classes greater than 4 are first plotted. Then, the weight data are multimodal and thus, are not expected to show a normal distribution in most cases. Therefore, nonparametric statistical significance tests may be performed. The significance level of 0.05 is used. A weight deviation between two months is determined. This deviation is measured using the JS divergence method. Monthly weight data are divided into equal-sized bins with a bin size of 907 kg (2000 pounds). The probability of occurrence of each bin is calculated by dividing the number of vehicles in each bin by the total number of vehicles. A JS distance is then calculated based on two (normalized) probability distributions representing the weight data from two consecutive months. The two probability distributions are considered identical if the JS distance is 0. Conversely, if the JS distance is 1, the two distributions must be uniquely different. The distance is converted into a percentage by multiplying by 100. Lastly, Site Number 510368 was not operational between July and December, possibly due to a maintenance issue. As a result, a figure will illustrate how the JS divergence method was able to differentiate two similar load spectra despite the missing data.

3. Proposed Methodology: Time-Series Weight Predictions

Two time-series forecasting models are mainly investigated to consistently oversee live loads.

3.1. Seasonal Autoregressive Integrated Moving Average Model

ARIMA models are generally fitted to the time-series data to better understand the data and make near-term future predictions. ARIMA is comprised of the autoregressive term (AR), the level of integration (I), and the moving average term (MA). Several researchers have used the ARIMA model to predict traffic patterns using WIM data [17,18]. The advantage of this model is the easy interpretation of outcomes. ARIMA models are mostly used where the mean is stationary. Highly nonstationary data are converted to stationary data after one or more differentiation steps. Kumar and Vanajakshi [19] have used the Seasonal ARIMA (SARIMA) model to predict short-term traffic flow. They found that the SARIMA model performed well with nonstationary traffic data. In this study, a multistep forecast is generated.

3.2. Deep Learning Approach

While autoregressive moving average methods are widely used in time-series forecasting due to the simplicity and interpretability of the forecasting process, these traditional methods lack capacity to capture complex temporal dynamics and are not able to accurately predict complex load patterns observed in the WIM data. These methods are also not effective in time-series predictions when there are missing data, and such an incident is very likely to occur in WIM data collection due to occasional road closures for maintenance and weather conditions. The Long Short-Term Memory (LSTM) [20] deep learning model can capture richer temporal dynamics and generally provides greater accuracy in time-series predictions when a large number of data are provided [21]. By using multi-layered nonlinear structures, LSTM models are proven to improve accuracy in time-series forecasting. LSTM is also known for sharing its parameters over time steps, which, hence, reduces the tendency of overfitting. In recent past, many researchers [22–24] have adopted deep learning frameworks in analyzing traffic data. A LSTM network is a special type of Recurrent Neural Network (RNN), which is capable of learning long-term dependencies through purposely designed gates. Figure 3 shows a LSTM network structure employed in this study.

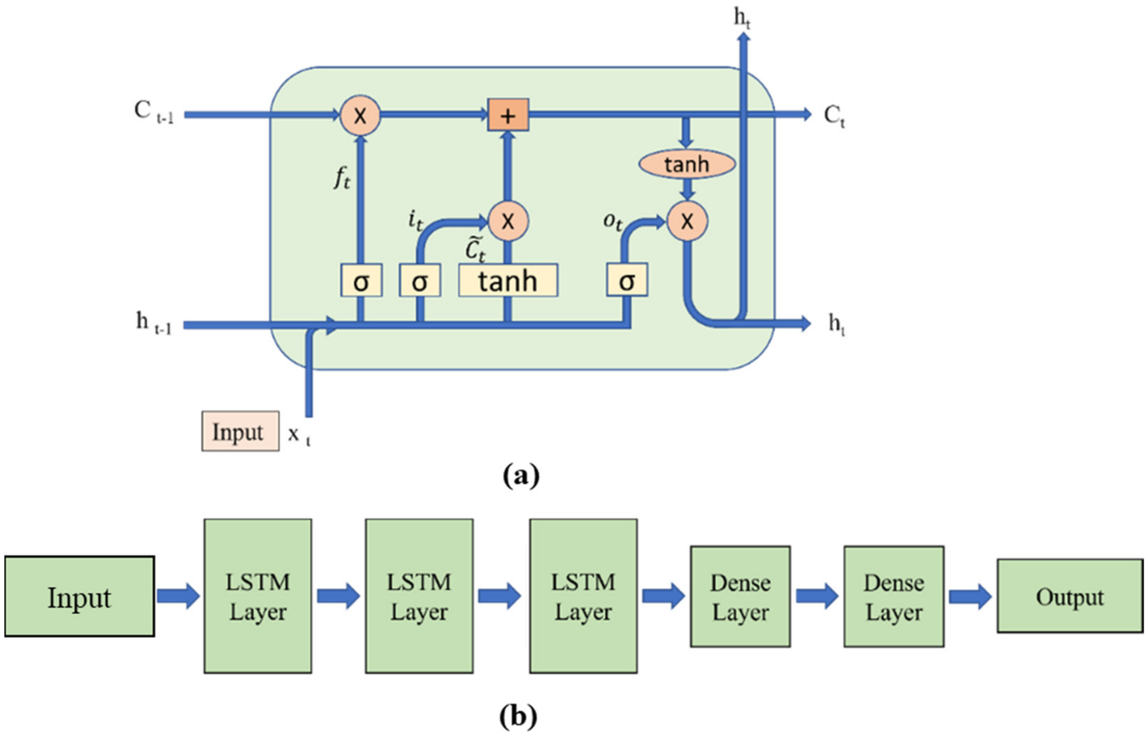


Figure 3. Recurrent neural network: (**a**) LSTM cell and (**b**) model architecture.

A time-series of length T is expressed as $X = (x_1,\ x_2,\ \ldots\ldots\ldots,\ x_T)$, where x_t represents the observation at time t. In this study, x_t represents a vehicle weight at time t. C_t is the memory cell or cell state, which contains the information at time step t. The cell is mainly operated by three types of gates: the Input gate, Forget gate, and Output gate. In Figure 3a, σ indicates the sigmoid function. f_t is the output of the Forget gate at time t. The first step in this LSTM structure is to decide the amount of information to be kept or thrown away from the cell state. The decision is made through the sigmoid (σ) function, as shown in Equation (3), where W_f is the weight matrix, h_{t-1} is the hidden output from the previous LSTM block at time $t - 1$, and b_f is the bias vector for the Forget gate.

$$f_t = \sigma\left(W_f.[h_{t-1},\ x_t] + b_f\right) \tag{3}$$

The subsequent step determines new information to be added in the cell state. This step combines two operations : sigmoid (Equation (4)) and tanh (Equation (5)). The former decides the amount of new information to be added in the memory, while the latter proposes the candidate cell state ($\widetilde{C}_t$).

$$i_t = \sigma\left(W_i.[h_{t-1},\ x_t] + b_i\right) \tag{4}$$

$$\widetilde{C}_t = \tan\mathrm{h}\left(W_c.[h_{t-1},\ x_t] + b_c\right) \tag{5}$$

In the next step, the old cell state (C_{t-1}) and the candidate cell state ($\widetilde{C}_t$) are updated by multiplying the respective gate outputs and combined to produce the new cell state C_t as shown by Equation (6). Note: $\odot$ indicates Hadamard product.

$$C_t = \left(f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t\right) \tag{6}$$

Finally, the hidden output at time t (h_t) is determined based on the Output gate (o_t) and the updated cell state C_t, as shown by Equations (7) and (8).

$$o_t = \sigma\left(W_o.[h_{t-1},\ x_t] + b_o\right) \tag{7}$$

$$h_t = o_t \odot \tanh(C_t) \tag{8}$$

The primary architecture of the LSTM model consists of three fully connected LSTM layers and two fully connected dense layers, as shown in Figure 3b. The observation variable is one-dimensional, and a sequence length of 50 is used in this study. The batch size is 16. All of the modeling parameters are scaled before feeding into the LSTM architecture. The 'Adam' optimizer is used for training, and the loss is defined by the mean squared error. A dropout rate p = 0.1 is adopted for the fully connected layers. Subsequently, the SARIMA and LSTM models are compared based on their test prediction accuracies in terms of root-mean-square error (RMSE) and mean absolute percentage error (MAPE).

4. Case Study Application: Comparing Two Live-Load Distributions

4.1. Statistical Tests

Weight data from 29 WIM sites are analyzed in this study using Mann–Whitney U and Kruskal–Wallis tests. The significance level, also denoted as α, is 0.05 for both. Two datasets (from two weight distributions representing two consecutive months of data) are compared. For all WIM weight data analyzed herein, we failed to reject the null hypothesis and concluded that the two distributions are statistically different at the 0.05 significance level. Contrary to the statistical test results, Figure 4a shows an example of monthly weight distributions where no major deviation is observed, specifically no shift to the left or right of the weights is shown in the figure. Statistical tests may observe changes in the frequency of the weights (or vertical shifts), but they are not able to exclusively capture lateral shifts such as the weight decrease/increase shown in Figure 4b, which is most relevant to weight calibration. The main objective of statistical tests is to detect monthly weight shifts, specifically horizontal shifts representing a noticeable weight decrease or increase. Figure 4c show that an increase in the frequency of weights should not be construed as a weight deviation (increase or decrease). Therefore, even though the statistical tests indicate a significant deviation in the weight distribution in June for Site Number 0217334, this outcome does not necessarily indicate that the weight sensor needs to be calibrated. This simply means that June did not record as many vehicles relative to other months. Conventional statistical tests alone are not sufficient to detect weight shifts or anomalies, similar to the ones shown in Figure 4c.

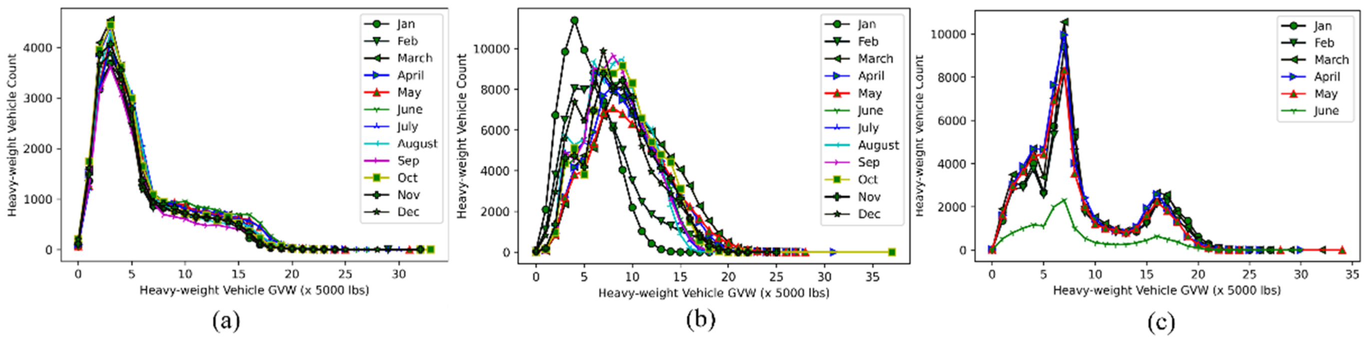


Figure 4. Monthly weight distributions at WIM Site Numbers (**a**) 0510700, (**b**) 0217334, and (**c**) 510368.

4.2. JS Divergence Method

The JS divergence method is a better method, as it compares two weight distributions and determines the distance between the two. This method accounts for the increase in the total number of vehicles—that is, the vertical shifts described with the statistical tests above—and normalizes weight density to screen for pertinent weight divergence.

Although the weight differences are adequately recognized with the divergence test, the 5% weight deviation suggested in the literature [25,26] appears to be a stringent threshold criterion for identifying a weight drift with 95% confidence, thus requiring calibration. Based on a review of the results and monthly weight distributions, requiring a threshold of 10–15% weight divergence appears to be more reasonable. Most of the month-to-month weight deviations observed in the 2021 WIM data are within 10–20%, with the exception of a few sites. The weight deviation observed at Site Number 217334 is within 10%, as visibly observed in Figure 4, which means the site's WIM systems may not require calibration.

5. Case Study Application: Time-Series Weight Forecasting

Table 1 provides a summary of performance metrics for both the SARIMA and LSTM models.

Table 1. Summary table of prediction using LSTM neural network and SARIMA model.

Site Number	Lane	LSTM		SARIMA	
		RMSE (Kips)	MAPE (%)	RMSE (Kips)	MAPE (%)
30132	EB	6.43	7.41	21.25	30.89
30132	WB	2.74	4.70	3.66	8.90
210378	NB	2.86	3.56	11.75	16.45
210378	SB	6.54	5.45	25.55	28.62
217334	NB	1.69	3.34	3.52	8.50
217334	SB	1.58	2.17	2.74	3.76
390218	NB	1.25	1.58	2.49	5.44
390218	SB	0.84	1.07	1.01	1.20
510368	EB	1.68	2.21	4.21	6.68
510368	WB	1.45	2.02	5.14	8.21
510387	NB	1.61	2.63	1.55	3.09
510387	SB	14.79	6.92	300.16	574.22
511113	EB	4.50	6.51	9.63	21.82
511113	WB	4.62	6.57	6.10	11.02
810347	EB	2.52	3.36	9.20	14.85
810347	WB	1.35	2.94	1.54	3.26
830214	EB	3.85	4.36	32.96	46.35
830214	WB	2.64	3.31	2.07	3.83
870103	NB	3.59	5.52	6.58	12.26
870103	SB	3.18	4.03	4.86	6.81
870125	NB	3.85	3.78	14.73	23.19
870125	SB	2.76	4.02	5.19	10.67
1030159	EB	5.60	7.70	6.31	9.87
1030159	WB	2.27	2.61	3.04	3.71
1150052	EB	3.43	4.10	4.81	6.00
1150052	WB	3.16	4.88	4.20	5.39
1270312	NB	1.20	1.84	4.93	8.44
1270312	SB	11.71	9.96	83.13	125.90
1450234	NB	4.38	5.64	7.37	9.97
1450234	SB	4.27	6.00	5.22	7.38
1610189	NB	2.30	3.98	2.82	6.23
1610189	SB	3.97	4.77	13.51	18.71
1750247	EB	12.73	5.19	16.18	24.62
1750247	WB	1.53	2.76	6.06	10.41
1850227	NB	0.93	1.58	2.94	5.03
1850227	SB	0.57	0.76	0.62	0.92
2170218	EB	1.50	2.05	1.72	2.62

Table 1. *Cont.*

Site Number	Lane	LSTM		SARIMA	
		RMSE (Kips)	MAPE (%)	RMSE (Kips)	MAPE (%)
2170218	WB	1.02	1.39	2.19	3.59
2350138	NB	6.21	6.87	15.09	15.49
2350138	SB	6.99	7.35	17.29	21.24
2850243	NB	2.14	3.66	12.17	21.64
2850243	SB	1.69	2.63	2.34	4.07

5.1. SARIMA Daily Average Gross Vehicle Weight Predictions

Figure 5 shows a time-series analysis (or a one-step-ahead forecast) of the gross vehicle weight (GVW) data sampled daily from the WIM data observed from three selected sites in Georgia. The solid gray line on the left shows the training data. The solid black line inside the box with an arrow shown on the right presents the test data, and the dashed black line within the same box shows the predicted gross vehicle weight. Specifically, 80% of the GVW data were in the training set, and 20% were in the test set. A sensitivity analysis was conducted to evaluate the different proportions of the training and test datasets and select the 80/20 train/test datasets. Table 2 shows the results. The test versus predicted vehicle weight plot is enlarged on the right-hand side to show the prediction accuracy for a period of 6 months, which is hypothetically assumed to be a future period (1 June–31 December 2021) in this study. The top and bottom lines of the shaded region indicate the maximum and minimum vehicle weight observed on a daily basis, respectively.

Table 2. Sensitivity of the training dataset size on the prediction accuracy.

Training Data (%)	LSTM		SARIMA	
	RMSE (Kips)	MAPE (%)	RMSE (Kips)	MAPE (%)
65	2.61	2.71	0.65	2.61
70	1.93	1.50	0.70	1.93
75	1.95	1.55	0.75	1.95
80	1.75	1.39	0.80	1.75
85	1.76	1.39	0.85	1.76

Figure 5a indicates that the time-series forecast increasingly deviates from the test data as the prediction interval increases. Figure 5b shows monthly trends in the vehicle-weight data. However, the SARIMA model does not predict the time history of GVW with great accuracy, as shown in Figure 5c. The month of December is expected to show the lowest moving average of gross vehicle weights due to the holiday season. Yet, the time-series forecast does not show any kind of reduction in weight in December. Overall, the SARIMA model fails to capture the effects of seasonality and weight trends in its predictions. The variance of prediction error increases with time. The residual plots are shown in Figure 6a,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. Nonetheless, it is observed that the SARIMA model does not yield good results for the sequential vehicle-weight predictions. This outcome is consistent with findings in the existing literature [27].

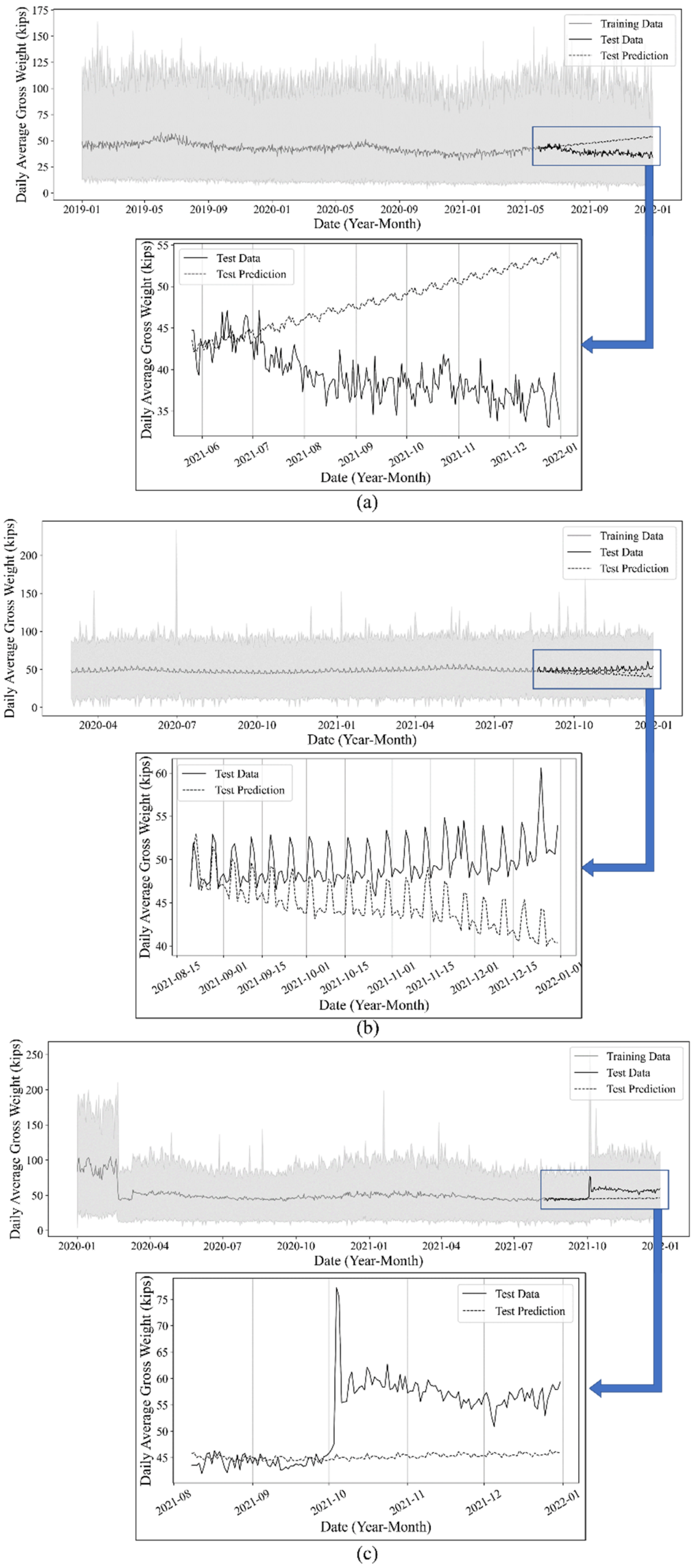


Figure 5. Daily average gross vehicle weight forecasting using SARIMA method for Site Numbers (**a**) 217334 (NB), (**b**) 390218 (NB), and (**c**) 870125 (SB).

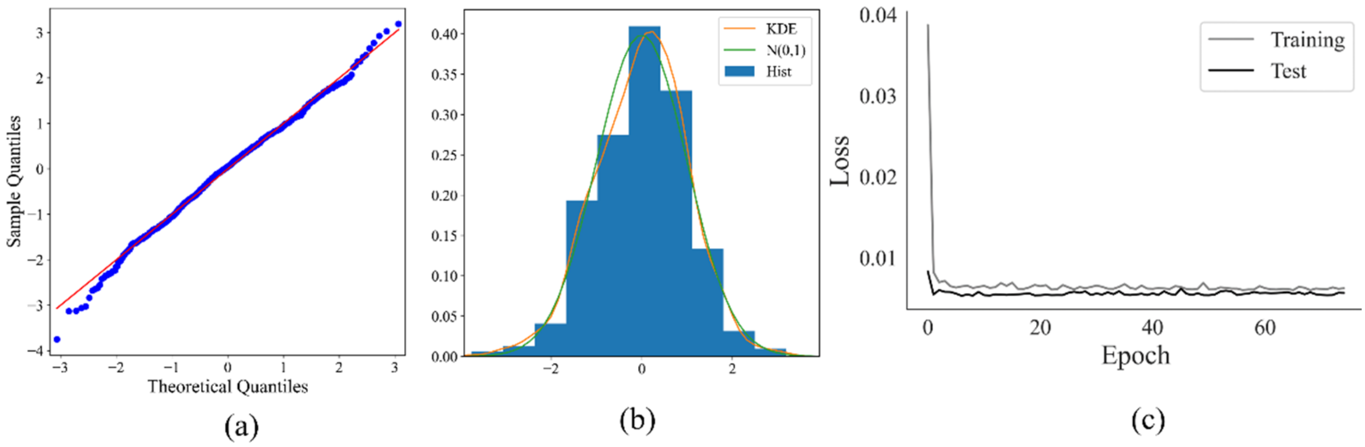


Figure 6. Evaluation metrics: (**a**) Q–Q plot of residuals in SARIMA model, (**b**) distribution of residuals in SARIMA model, and (**c**) loss versus epoch plot in the LSTM network for Site Number 217334 (NB).

5.2. Recurrent Neural Networks for Forecasting Gross Vehicle Weight

The Long Short-Term Memory (LSTM) model is known to overcome such deficiencies in the SARIMA model and is able to predict complex weight patterns in the time-series data.

Figures 7–9 present the time-series gross vehicle and tandem-axle weight data sampled daily from selected WIM sites. Their time-series predictions using a LSTM neural network are also presented on the right. NB and SB indicate the northbound and southbound lanes, respectively. Similar to the SARIMA prediction model, the gray solid line on the left shows the training data. The black solid line on the right shows the test data, and the dashed black line shows the LSTM model's predictions. The shaded region indicates a live-load range between the observed daily 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 figures.

In Figures 7–9, the weight plot includes only class 9 vehicles, noting that the gross permissible weight is 36,287 kg. In Figure 7, it is observed that the gross vehicle weight predictions agree with the test data for Site Number 217334 (NB) and Site Number 390218 (NB), and the prediction error is 567 and 767 kg, respectively. This error is considered acceptable because it is within 1% of the expected gross vehicle weight. The LSTM model performance is superior to the SARIMA model because the prediction error is lower. When the training data has anomalies such as a weight drift, they affect the LSTM predictions, as shown in Figure 7c. In the case of the time-series data observed at Site Number 870125 (SB), 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 Number 390218 (NB), as shown in Figure 7b, the LSTM model's predictions agree well with the test data, and no significant weight drift is observed.

Figure 8 shows the time history of tandem-axle weights using a LSTM neural network, noting that the permissible tandem-axle weight is 15,422 kg. The tandem weights are overall consistent with the gross vehicle-weight patterns shown in Figure 7 but are evaluated within a narrow range of the permissible tandem load. As shown in Figure 8a, a weekly weight pattern is observed in the WIM data obtained from Site Number 390218 (NB). A small downward weight shift (680 kg) is observed in the test data in the last week of the forecasting period, which may indicate a sudden change in the vehicle-weight pattern. In Figure 8b, the live loads appear to be within an expected tandem-axle weight range. Yet, the test data display no clear weight pattern. The prediction results show a similar trend reflecting no weekly weight pattern but rather show the noise in the data. Figure 8c shows the vehicle loads after calibration in 2021, but the data remain noisy without an identifiable pattern. Similar to Site Number 870125 (SB)'s gross vehicle-weight data shown in Figure 7c, Site Number 1270312 (SB)'s time-history plot shows a sudden increase in tandem-axle weight in year 2020, as shown in Figure 8d.

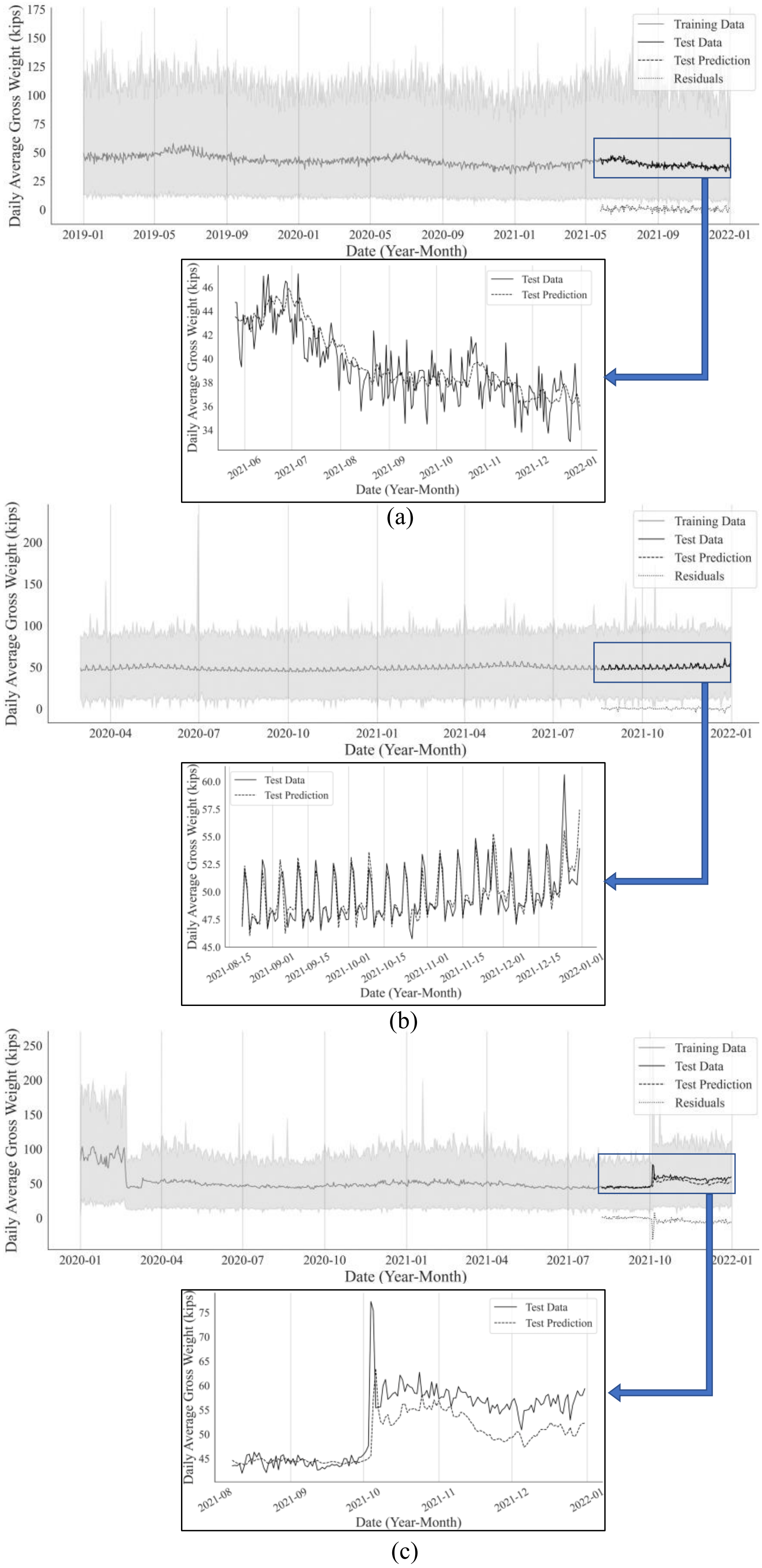


Figure 7. Daily average gross vehicle weight forecasting, utilizing LSTM architecture for Site Numbers (**a**) 217334 (NB), (**b**) 390218 (NB), and (**c**) 870125 (SB).

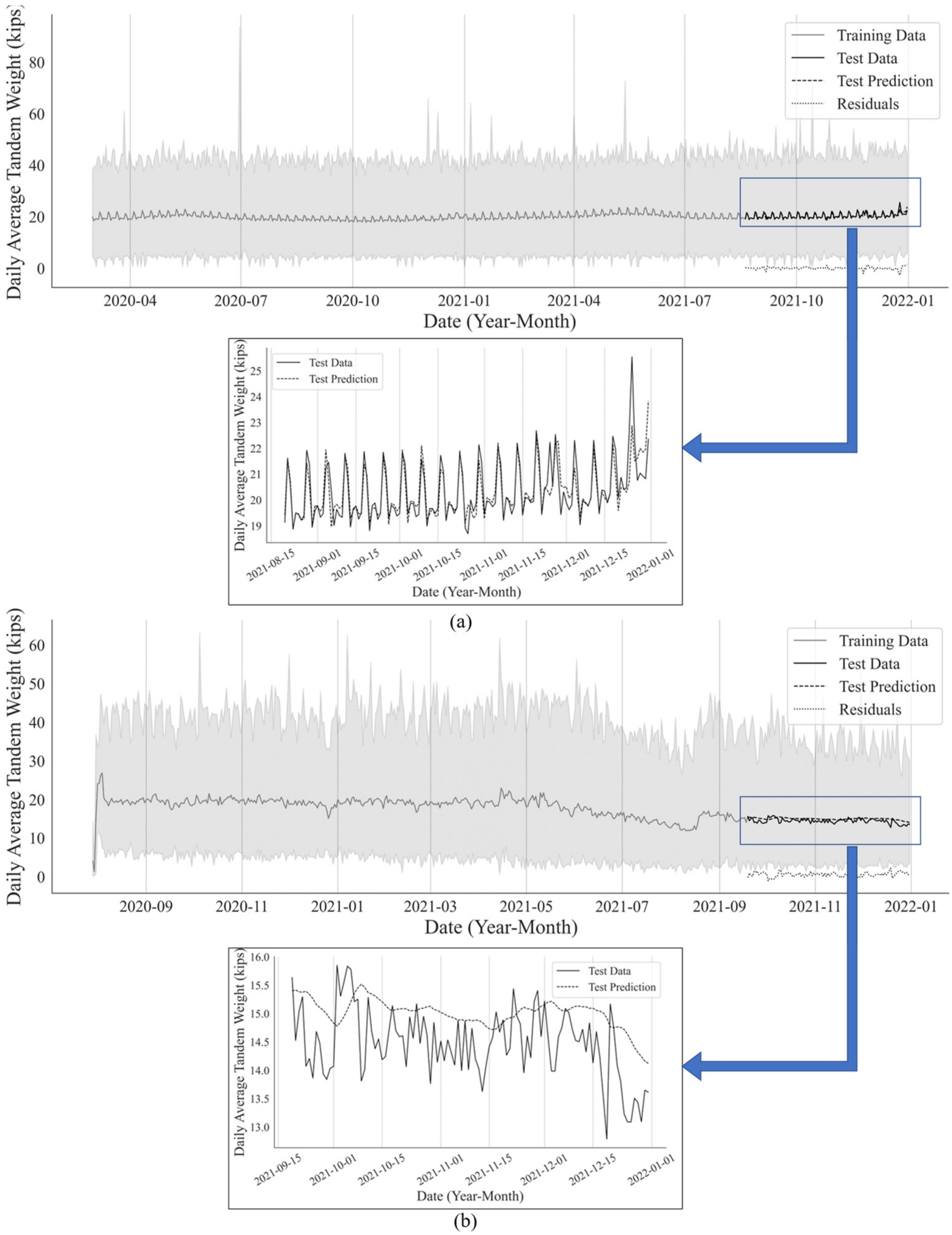
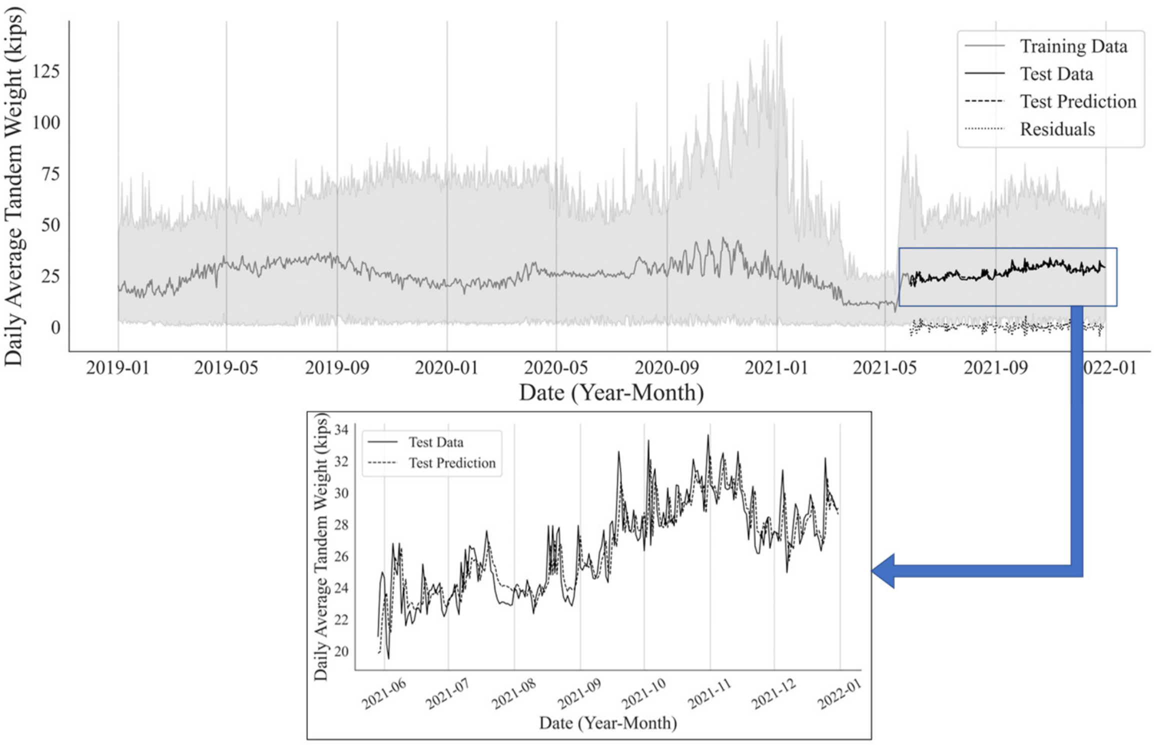
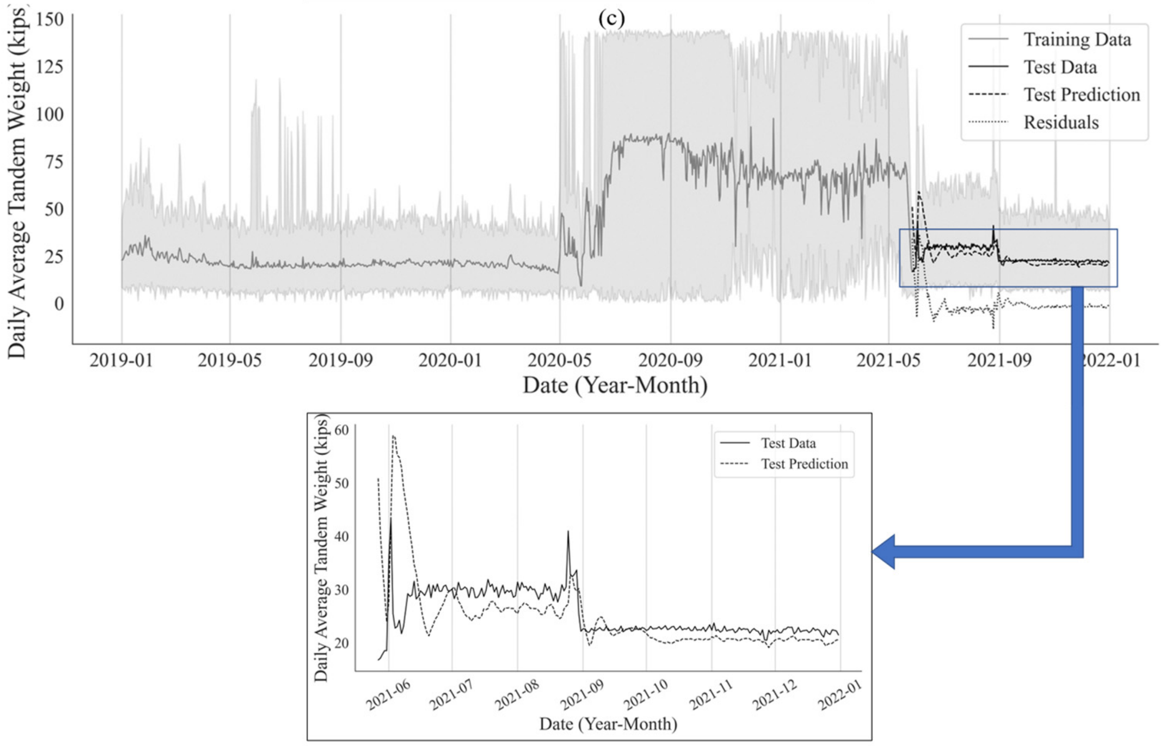


Figure 8. *Cont.*



(c)



(d)

Figure 8. Daily average tandem-axle vehicle weight forecasting utilizing LSTM architecture for Site Numbers (**a**) 390218 (NB), (**b**) 810347 (WB), (**c**) 830214 (EB), and (**d**) 1270312 (SB).

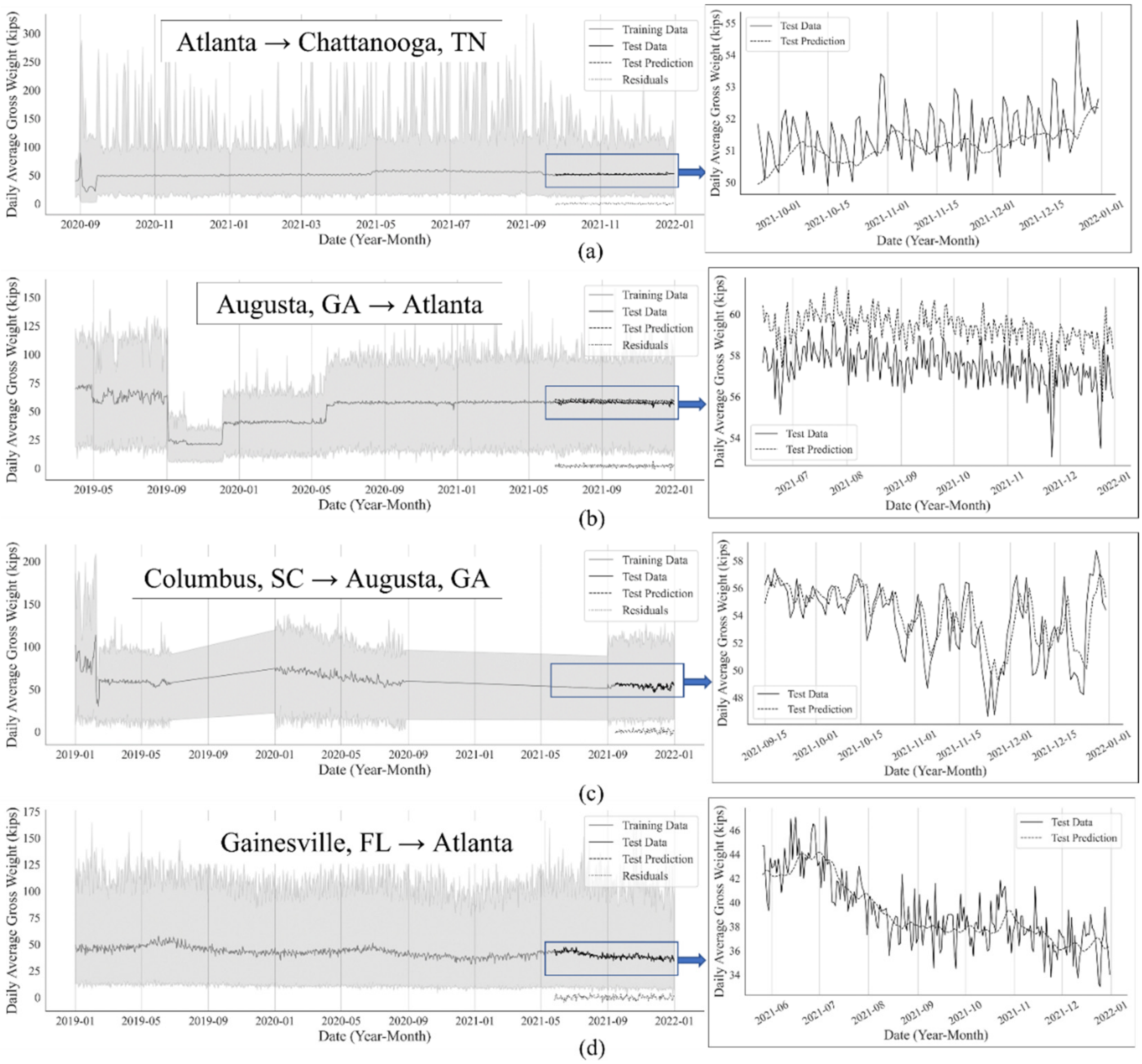


Figure 9. Average daily GVW forecasting for WIM sites located in the vicinity of Atlanta: Site Numbers (**a**) 0150276 (NB), (**b**) 2170218 (WB), (**c**) 2450214 (WB), and (**d**) 0217334 (NB).

5.3. Analysis of the Results

This section answers the research questions presented earlier, in conjunction with the results above. A statistical difference test is widely used for comparing two datasets (in this case, two vehicle-weight-occurrence distributions) and detecting weight anomalies. The JS divergence method enhances the screening process because it quantifies a vehicle-weight drift on a scale between 0 and 1. Using this method allows us to measure a deviation between two probability distributions more accurately compared to nonparametric statistical difference tests. For example, the weight deviation at Site Number 0510700 (SB) is less than 5%, according to the JS divergence, although statistical tests fail to demonstrate this. Therefore, the method measuring a JS divergence is more efficient for detecting a weight drift observed in WIM systems. However, these evaluation methods do not capture temporal elements such as seasonality or weight trends observed in the previous period (e.g., year or month).

The SARIMA model is a linear-regression-based time-series forecasting approach that is widely used due to its simplicity. However, it is not capable of understanding the complexity in the time-series vehicle-weight data. On the other hand, the LSTM forecasting model employing an RNN is very good at identifying the long-term and short-term temporal dependencies by learning the structure of training data and seasonality/trend in the data. Overall, the LSTM network model outperforms the traditional SARIMA model and significantly reduces the prediction errors (RMSE and MAPE), regardless of the WIM data quality observed in a dataset. That is, trends learned during disruptive events are in its memory and are reflected in the predictions. It is noted that the memory required was 1121 and 921 MB to run the SARIMA and LSTM models, respectively, for Site Number 217334, and that the time required to predict the weight was 58 and 10 s, as a result.

The proposed method of forecasting vehicle weight closes a gap created by missing values in the weight data (due to maintenance issues, for example), and its predictions yield a reference dataset that can be used to detect anomalies in future datasets. Dynamic inconsistency occurs when a decision-maker has incomplete data [28] and, thus, has low confidence for the results of a weight drift observed between two calibration measurements. This study not only proposes a new technique for quantifying a weight drift but also proposes a temporal monitoring strategy that increases the confidence in the results by providing a stable, yet time-dependent, reference dataset, whether it is for monitoring weight trends or detecting anomalies. A LSTM neural network can capture temporal dynamics and provides time-series predictions of vehicle weights as an benchmark, and, thus, is able to capture the instances that vehicle weights (or live loads) significantly deviate from the predicted weights.

6. Discussion on the Use of WIM Data

Figure 9 presents time-series predictions of gross vehicle weight at four selected WIM locations (see Figure 1).

The results show increased daily maximum live loads in the vicinity of Atlanta. Trucks are not allowed inside the perimeter of Atlanta, and, thus, all trucks traveling on I-75, I-85, or I-20 must take I-285 around the city. The average daily truck traffic (ADTT) has increased by 6% over the past 5 years. Consistently, the daily average weights had a slight but unnoticeable increase in the past 3 years. Overall, the westbound and northbound lanes toward Atlanta and Chattanooga indicate significantly high daily maximum vehicle weights measured in WIM systems. The dynamic load amplification (DLA) due to moving vehicles on highway may affect the maximum weights observed. However, the DLA factor ranges between 125% and 150% at other locations, whereas these four sites, shown in Figure 9, show significantly higher gross vehicle weights, which are equivalent to a DLA factor of 180%. The average daily weights remain fairly consistent across WIM sites. Weight data measured at Site Number 175247 (see Figure 1) on the I-16 route did not exhibit increased live loads, whereas most WIM sites in the vicinity of the Port of Savannah showed significantly higher live loads relative to the WIM systems located in other areas of Georgia. Therefore, the increased maximum daily live loads in the I-285 northbound route appear to be mainly attributed to truck traffic from Florida (or Brunswick via I-75) and South Carolina, traveling north or northwest. Therefore, based on the 2019–2021 WIM evaluation, increased daily maximum live loads observed at the four sites are expected to negatively impact bridges and roads on I-285 and may better explain the recurring joint damage on I-285's bridges than the live loads from the Port of Savannah. As of now, the latter appears to be redistributed before reaching the I-16 route toward Atlanta. To confirm such observations, it is necessary to review trends in time-series weight predictions and increase the reliability of WIM systems' performance.

7. Conclusions

Forecasting vehicle-weight patterns is expected to aid in transportation planning and developing strategies for future highway-infrastructure maintenance and asset management. Thus, understanding live loads imposed on transportation assets is critically important for infrastructure projects, public safety, and mobility. This study analyzes the

WIM data obtained from 29 WIM locations in Georgia and investigates effective methods for monitoring performance of the WIM systems over time and mitigating temporal biases, such as dynamic inconsistency when live loads are measured by WIM systems. Based on the findings of this study, the following conclusions are made:

- It is concluded that the JS divergence method is more suitable for comparing two vehicle-weight datasets and capturing a weight drift at WIM sites than conducting a statistical significance test of two independent data sets consisting of different sample sizes. The JS divergence method compares normalized probability distributions of vehicle weights and yields a more effective evaluation measure for quantifying the difference between two weight datasets.
- With two datasets from two different time periods, the JS divergence approach determines if a new dataset contains an acceptable amount of weight drift. The allowable divergence limit ranges between 5% and 10% in the literature but does not provide absolute assurance for detecting weight anomalies in WIM systems. Additionally, there is room for errors and temporal inconsistencies in decision-making, particularly when the time interval between two calibration visits varies.
- A deep-learning-based time-series prediction provides an easier, as well as more accurate and intuitive, measure for monitoring live loads over time and detecting anomalies in evolving weight data, for identifying WIM systems needing calibration. Compared with a SARIMA model, a Long Short-Term Memory (LSTM) model has a higher capacity and learns to retain and forget information to capture the temporal dynamics underlying time-series data. Predicting seasonality and changes in average weights are attainable when a LSTM model is used to monitor evolving vehicle-weight data.
- A deep-learning architecture enhances time-series predictions and provides a more complete picture of WIM systems' health and the spectrum of live loads that are expected to be imposed on public roads and bridges.

8. Future Work and Recommendations

As a result of this study, it is recommended that the JS divergence method be used in order to monitor the performance of WIM systems. More importantly, it is recommended that the proposed temporal prediction method employing a LSTM network be implemented to detect more complex live-load patterns and provide highly accurate near-term weight predictions. Besides LSTM, other promising modern architectures of recurrent neural networks such as gated recurrent unit (GRU) networks [29], liquid time-constant (LTC) networks [30], and transformer architecture [31] should be evaluated. Finally, both spatial and temporal validations should be performed to show that the model is transferable. Second, the authors find that training individual models on each FHWA class (e.g., class 9) is better than training all vehicle classes together in the LSTM model. This paper focuses on presenting the proposed temporal approach rather than presenting all of the available results. The impact of the presence of seasonality, weather, and pavement conditions on WIM data quality will need to be further investigated.

Author Contributions: All authors have contributed to the conceptualization, methodology, analysis, validation, and writing. Formal analysis, A.S.; investigation, A.S. and M.G.C.; resources, M.G.C. and J.J.Y.; data curation, A.S. and M.G.C.; writing—original draft preparation, A.S.; supervision, M.G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Georgia Department of Transportation, grant number RP20-04.

Data Availability Statement: Some or all the data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The study presented in this paper was conducted by the University of Georgia under the auspices of the Georgia Department of Transportation [32] (RP 20-04). The opinions, findings, and conclusions may not reflect the views of the funding agency or other individuals.

Conflicts of Interest: The funders had no role in the design of the study; in the analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Burnos, P.; Gajda, J.; Sroka, R.; Wasilewska, M.; Dolega, C. High Accuracy Weigh-In-Motion Systems for Direct Enforcement. *Sensors* **2021**, *21*, 8046. [CrossRef] [PubMed]
2. Gajda, J.; Burnos, P.; Sroka, R. Accuracy assessment of weigh-in-motion systems for vehicle's direct enforcement. *IEEE Intell. Transp. Syst. Mag.* **2018**, *10*, 88–94. [CrossRef]
3. Sujon, M.; Dai, F. Application of weigh-in-motion technologies for pavement and bridge response monitoring: State-of-the-art review. *Autom. Constr.* **2021**, *130*, 103844. [CrossRef]
4. Martini, A.; Tronci, E.M.; Feng, M.Q.; Leung, R.Y. A computer vision-based method for bridge model updating using displacement influence lines. *Eng. Struct.* **2022**, *259*, 114129. [CrossRef]
5. Hou, R.; Jeong, S.; Lynch, J.P.; Ettouney, M.M.; Law, K.H. Data-driven analytical load rating method of bridges using integrated bridge structural response and weigh-in-motion truck data. *Mech. Syst. Signal Process.* **2022**, *163*, 108128. [CrossRef]
6. Chorzepa, M.G.; Kim, S.S.; Durham, S.A.; Sinha, A. *Development of Weigh-in-Motion Data Quality Control Algorithms and Procedures*; Department of Transportation, Office of Performance-Based Management and Research: Atlanta, GA, USA, 2020.
7. Hernandez, S.; Baker, J.; Durandal, F.Z. *Final Report TRC1801 Evaluation of WIM Auto-Calibration Practices and Parameters*; Arkansas Department of Transportation: Little Rock, AR, USA, 2019.
8. Graves, A. Long short-term memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Studies in Computational Intelligence; Springer: Berlin/Heidelberg, Germany, 2012; pp. 37–45.
9. Wei, Y.-T.; Yi, T.-H.; Yang, D.-H.; Li, H.-N. Bridge Damage Localization Using Axle Weight Time History Data Obtained through a Bridge Weigh-in-Motion System. *J. Perform. Constr. Facil.* **2021**, *35*, 04021065. [CrossRef]
10. Han, W.; Yuan, Y.; Xie, Q.; Gao, G.; Chen, X.; Xu, K. Load effect, safety assessment, and traffic strategy of multigirder bridges under lateral eccentric customized transport vehicle. *J. Perform. Constr. Facil.* **2019**, *33*, 04018110. [CrossRef]
11. Liu, J.; Li, K. Sparse identification of time-space coupled distributed dynamic load. *Mech. Syst. Signal Process.* **2021**, *148*, 107177. [CrossRef]
12. Sinha, A.; Chorzepa, M.G.; Yang, J.J.; Sonny Kim, S.; Durham, S.A. Cognitive Approaches to Hyperbolic Discounting of High-Impact Low-Probability Bridge Overload Events and Live-Load Factors. *J. Perform. Constr. Facil.* **2022**, *36*, 04022009. [CrossRef]
13. Selezneva, O.; Wolf, D.J. *Successful Practices in Weigh-in-Motion Data Quality with WIM Guidebook [Volumes 1 & 2]*; Arizona. Dept. of Transportation: Phoenix, AZ, USA, 2017.
14. Joyce, J.M. Kullback-leibler divergence. In *International Encyclopedia of Statistical Science*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 720–722.
15. Raiber, F.; Kurland, O. Kullback-leibler divergence revisited. In Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, Amsterdam, The Netherlands, 1–4 October 2017; pp. 117–124.
16. Fuglede, B.; Topsoe, F. Jensen-Shannon divergence and Hilbert space embedding. In Proceedings of the International Symposium onInformation Theory, 2004 (ISIT 2004), Chicago, IL, USA, 27 June–2 July 2004; p. 31.
17. He, Q.; Wang, Y.; Tu, H.; Ruan, X. Incorporating the WIM Data into the Analysis of Road Traffic Risk. In *CICTP 2017: Transportation Reform and Change—Equity, Inclusiveness, Sharing, and Innovation*; American Society of Civil Engineers: Reston, VA, USA, 2018; pp. 4730–4739.
18. Smolarek, L.; Ziemska, M. Using information collected by weigh in motion for modeling traffic structure of vehicles. *Arch. Transp. Syst. Telemat.* **2017**, *10*, 32–37.
19. Kumar, S.V.; Vanajakshi, L. Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *Eur. Transp. Res. Rev.* **2015**, *7*, 21. [CrossRef]
20. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
21. Wang, W.; Zhou, C.; He, H.; Wu, W.; Zhuang, W.; Shen, X. Cellular traffic load prediction with LSTM and Gaussian process regression. In Proceedings of the ICC 2020—2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6.
22. Li, J.; Gao, L.; Song, W.; Wei, L.; Shi, Y. Short term traffic flow prediction based on LSTM. In Proceedings of the 2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP), Wanzhou, China, 9–11 November 2018; pp. 251–255.
23. Morris, C.; Yang, J.J.; Chorzepa, M.G.; Kim, S.S.; Durham, S.A. Self-Supervised Deep Learning Framework for Anomaly Detection in Traffic Data. *J. Transp. Eng. Part A Syst.* **2022**, *148*, 04022020. [CrossRef]
24. Tian, Y.; Zhang, K.; Li, J.; Lin, X.; Yang, B. LSTM-based traffic flow prediction with missing data. *Neurocomputing* **2018**, *318*, 297–305. [CrossRef]

25. Selezneva, O.; Von Quintus, H.; Brink, W. Exploring the Importance of Traffic Data Input Levels for Mechanistic-Empirical Pavement Design. *TechBrief* **2021**, 8–9.
26. Selezneva, O.I.; Ayers, M.; Hallenbeck, M.; Ramachandran, A.; Shirazi, H.; Von Quintus, H. *MEPDG Traffic Loading Defaults Derived from Traffic Pooled Fund Study*; Federal Highway Administration: Washington, DC, USA, 2016.
27. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. A comparison of ARIMA and LSTM in forecasting time series. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 1394–1401.
28. Ou, X.; Wu, Y.; Wu, B.; Jiang, J.; Qiu, W. Dynamic Bayesian Network for Predicting Tunnel-Collapse Risk in the Case of Incomplete Data. *J. Perform. Constr. Facil.* **2022**, *36*, 04022034. [CrossRef]
29. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint* **2014**, arXiv:1412.3555 2014.
30. Hasani, R.; Lechner, M.; Amini, A.; Rus, D.; Grosu, R. Liquid time-constant networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: Palo Alto, CA, USA, 2021; pp. 7657–7666.
31. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need, in Neural Information Processing Systems. 2017. Available online: https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (accessed on 1 October 2022).
32. GDOT. *Georgia's Traffic Monitoring Program*; Atlanta, N.W., Ed.; 600 West Peachtree Street, Office of Transportation Data: Atlanta, GA, USA, 2018; p. 30308.