Original Research Paper

Estimating the frequency of traffic overloading on road bridges

Roberto Ventura[*], Benedetto Barabino, Giulio Maternini

Department of Civil, Environmental, Architectural Engineering and Mathematics, University of Brescia, Brescia 25123, Italy

HIGHLIGHTS

- Traffic overloading events on road bridges are predicted to support failure prevention.
- A frequency metric of overloading is proposed as a driver of failure probability.
- Econometric and machine learning frequency models are set-up and compared by 2 million + weigh-in-motion raw records.
- The machine learning model overperformed the econometric on the considered dataset.

ARTICLE INFO

ABSTRACT

Load limits, which appear to be routinely exceeded by trucks, occasionally result in road bridge failures. Therefore, predicting failures is crucial for safeguarding road safety. Past studies have largely focused on forecasting bridge failure event probability using the reliability analysis method, whilst occasionally accounting for vehicular overloading effects. Only recently, a study has investigated design traffic overloading event frequency using generalised linear regression models (GLRMs), including a power component and negative binomial regressions (NBRs). However, as far as the authors know, artificial neural network models (ANNMs) have never been applied to this field. This paper is an attempt to fill in these gaps. First a frequency-based metric of traffic overloading was adopted as a driver of failure probability. Second, two alternative 'frequency' models were specified, calibrated, and validated. The former was based on a GLRM, the latter on ANNMs. Then, these models were compared using regression plots (RPs), measures of errors (MoEs) and the ratio between the number of observed vs predicted design load overcoming events to evaluate their performance. The models analysed more than 2 million weigh-in-motion (WIM) data records from a pilot station on a bridge on a heavily used ring road in Brescia (Italy). Results showed that ANNMs outperformed GLRMs. ANNMs have a higher correlation coefficient (between predicted and target frequencies), lower MoEs, and a closer-to-unity ratio (between predicted and target frequencies). These findings may increase prediction accuracy of design traffic overloading events and give road authorities more effective traffic management to protect bridges from load hazards.

---

* Corresponding author.
  E-mail addresses: roberto.ventura@unibs.it (R. Ventura), benedetto.barabino@unibs.it (B. Barabino), giulio.maternini@unibs.it (G. Maternini).

Please cite this article as: Ventura, R et al., Estimating the frequency of traffic overloading on road bridges, Journal of Traffic and Transportation Engineering (English Edition), https://doi.org/10.1016/j.jtte.2023.11.005

ARTICLE IN PRESS

2   J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

Nomenclature

ANNMs	Artificial neural network models
CoV	Coefficient of variation
EMs	Econometric models
FP	Filtering procedure
GLRMs	Generalized linear regression models
GVM	Gross vehicle mass
MAE	Mean absolute error
MLMs	Machine learning models
MoEs	Measures of errors
MSE	Mean squared error
NBR	Negative binomial regression
OIML	International organization of legal metrology
PDF	Probability density function
PFI	Permutation feature importance
QCA	Quality control algorithm
RAM	Reliability analysis method
RA	Road authority
RMSE	Root mean squared error
RPs	Regression plots
SVM	Support vector machine
TC	Traffic code
WIM	Weigh-in-motion

1. Introduction

Apart from hydraulic action, vehicular traffic is the primary hazard that threatens road bridge safety. Overload and vehicle collisions are two of the top five most frequent causes of bridge failures, according to a recent survey accounting for 4500+ bridge collapses worldwide between 2010 and 2016 (Zhang et al., 2022). Focusing on overload hazard, extremely heavy vehicles represent a crucial threat to bridge integrity (Ventura et al., 2020). As traffic volumes increase, truck loads often exceed legal limits and sometimes result in road bridge failures when design traffic loads are surpassed (Zhang et al., 2022).

Therefore, adopting promising technologies is relevant to gathering data about extremely high traffic loads on bridges. Whilst on the one hand, these data could be stored and analysed for traffic load monitoring to aid in bridge design and management, on the other, these data are a valuable resource for the development of failure prediction models used to support bridge safety management.

Weigh-in-motion (WIM) technologies have been shown to be useful for collecting data for traffic load monitoring (Sujon and Dai, 2021). According to ASTM (2017), a WIM system refers to "a collection of sensors and supporting devices that measure the presence of a moving vehicle and the associated dynamic tyre forces at fixed locations concerning time". Besides the total vehicle mass, a WIM system can measure other parameters such as class, passing speed, axle number, axle type, type loads, and interaxle. Several WIM technologies are currently available, such as piezoelectric, capacitive mats, bending plates, load cells and fibre optic

sensors, among others (Yannis and Antoniou, 2005). The primary benefit of WIM systems is their capacity to measure these parameters continuously without needing a human operator to infer them by choosing among random samples of vehicles and conducting a manual weighing procedure (Ventura et al., 2023a). WIM devices are widespread in some American and Asian countries (e.g., United States, Canada, and China), where they are used for a variety of applications, including bridge design and management, road pavement design and management, weight enforcement, freight movement analysis and traffic flow simulation (Fiorillo and Ghosn, 2014; Guo et al., 2011; Hernandez et al., 2016; Huang et al., 2022; Liao, 2014; Ren et al., 2019; Roh et al., 2016). In contrast, few studies have been conducted in Europe, where WIM systems are still regarded as pilot installations (Schmidt et al., 2016; Ventura et al., 2023a, 2023b).

Focusing on bridge design and management applications, several studies have processed WIM data to compare site-specific traffic load effects with design traffic load models, as well as to update national traffic load models (or to build new ones) by processing site-specific traffic data. These analyses started from findings that existing bridge design codes (e.g., the US AASHTO and the European Eurocodes) were based on narrow samples of old traffic data (Iatsko and Nowak, 2021). Hence, it has become essential to propose site-specific traffic load models based on WIM data that consider the spatial and temporal uncertainties posed by the specific traffic conditions at the bridge locations and by the progressive changes in the surrounding environment, respectively, as well as the increasing number of multi-lane bridges (Kim and Song, 2019, 2021; Zhou et al., 2020a). Moreover, because current bridge design codes could fail to represent the effects of present-day traffic fatigue action, damage phenomena induced by cyclic traffic loading have often been incorporated into new traffic load models based on WIM data (Maljaars, 2020). Additionally, some authors have focused on long-span cable-stayed bridges, pointing out that current design traffic load models are primarily conceived for short to medium-span structures (Hwang and Kim, 2019; Lu et al., 2018, 2019; Micu et al., 2019). Processing WIM data, they pointed out that whilst for short bridges it is likely that a very heavy truck or a few heavy trucks could generate the maximum load effects, for long-span bridges it is necessary to consider more complex scenarios (such as traffic congestion) during the design phase. Furthermore, almost all previous studies analysed the traffic loads acting on bridges under conventional serviceability conditions. Conversely, Zhou et al. (2021) set up traffic microsimulations based on WIM data to investigate the negative load effects induced by bridge maintenance works, such as the hostile traffic load actions generated by lane closures and by the presence of construction machinery on bridge decks.

As for failure prediction models, previous studies largely focused on the estimation of the probability of bridge failure events by applying the reliability analysis method (RAM), whilst vehicular overloading effects were accounted for only sometimes (Fiorillo and Ghosn, 2019, 2022). This method involves the calculation of the reliability index, which is a parameter related to the likelihood of failure and which is a conventional tool for evaluating the safety of existing

ARTICLE IN PRESS

J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx   **3**

structures according to current standards, e.g., the Eurocodes (European Committee for Standardization, 2002). This index is generally computed from the structural performance function, which is defined as the difference between structural capacity and structural demand functions: a negative value of the performance function would denote the condition of structural failure (Fiorillo and Ghosn, 2018). Whilst the structural capacity function considers geometrical and material properties, the structural demand function considers the maximum traffic load acting on the bridge for a specific return period. This latter is recurrently inferred through the extreme value theory by elaborating experimental WIM data (Zhou et al., 2022; Mandić Ivanković et al., 2019). Some authors have modelled structural demand in a refined way by accounting for the differences in traffic loads over multiple lanes, proving that the difference in the reliability indexes associated with each girder in the same bridge increases when the number of lanes increases as well (Zhou et al., 2020b). Nevertheless, estimating the probability of bridge failure events using RAM requires a structural model accounting for the geometrical and mechanical properties of the bridge's components. Since budget constraints often prevent road authorities (RAs) from calibration of such a refined structural model, implementing rapid traffic load management strategies based on RAM might be tricky.

Predicting the frequency of design traffic overload events induced by traffic load hazards could provide an addition to RAM. Indeed, once the design traffic load thresholds have been identified according to the traffic load schemes prescribed by current structural design codes, a frequency metric can be defined as the number of times these thresholds are exceeded over a specific time frame. This metric could well be a driver towards the indirect measurement of the probability of potential bridge failure events prompted by traffic load hazards. In this study, a "failure event" is not necessarily intended as a collapse but rather as any circumstance that could prevent an existing bridge or one of its components from performing as required by the design and construction standards. Moreover, this metric could be evaluated using appropriate mathematical models that link it to several predictors grouped into exposure, bridge side, temporal context, and traffic load hazards. Mining WIM datasets, these models could then enhance the theoretical knowledge of the bridge overload phenomenon by understanding the effect of predictors on overload frequency. Practically, these models could be essential in supporting RAs in improving the safety of their assets by suggesting traffic management strategies that would mitigate bridge overload events. Indeed, without the ability to predict these overload events before they occur, no traffic management actions can be taken to reduce them. Moreover, their frequency, together with their severity, is one of the two primary components of risk prediction models (Ventura et al., 2024).

Frequency models can be of two different types: econometric models (EMs) and machine learning models (MLMs). For EMs, only Ventura et al. (2024) specified, calibrated, and validated a generalised linear regression model (GLRM) to predict the frequency of events where the design load has been overcome. This model was one of the components that

make up a general framework used to evaluate the risk of traffic load hazards on bridges. According to ISO 39001 Road Traffic Safety Management Systems (2012), intermediate safety factors and risk exposure factors were identified as predictors of frequency. Results showed that the GLRM had good predictive ability and that factors related to compliance with mass limits prescribed by traffic code (TC) were those with the more significant effect on frequency predictions. Conversely, no study has yet employed MLMs to predict the frequency of design load overcoming events induced by traffic load hazards on bridges. Nonetheless, there are several applications of MLMs that can aid risk evaluations, such as elaborating vehicle accident data to forecast crash frequency or severity, processing textual data to detect key messages in accident investigation reports to identify combinations of attributes that contribute to injuries, or detecting anomalies in structural performance (Marucci-Wellman et al., 2017; Salazar et al., 2017; Wen et al., 2021; Zeng et al., 2016).

According to Hegde and Rokseth (2020), MLMs could be classified by the accuracy of their predictions using few or many data. On the one hand, boosting techniques and support vector machine (SVM) methods may show the best options if data are scarce, even though their performance worsens while dealing with large data sets due to long training times (Susmita, 2019; Wen et al., 2021). On the other hand, artificial neural network-based models (ANNMs) are those that have been adopted most often to forecast risk components due to their generally strong predictive performance. Indeed, if there is enough data, even if they are noisy, ANNMs generally prevail because they can automatically learn latent features, which are essential to ensure good prediction accuracy (Wen et al., 2021). This could easily be the case when applied to large datasets provided by WIM systems, which usually involve millions of records yearly (Ventura et al., 2023a).

ANNMs have strengths and drawbacks when compared to GLRMs. On the one hand, ANNMs have generally higher predictive performance and greater capacity for modelling non-linear phenomena (Alqatawna et al., 2021). Moreover, ANNMs do not require assumptions of underlying relationships between the predictors and the observed frequencies. Furthermore, they can more efficiently manage correlation issues among predictors than the GLRMs. Specifically, when a high correlation exists among predictors, the variability of estimated coefficients of a GLRM will be seriously inflated. This leads to a challenging interpretation of the relationship between the predictors and the response variables Conversely, ANNMs do not suffer from multicollinearity issues (Chang, 2005). On the other hand, ANNMs are viewed as a "black box" with respect to GLRMs. Black box models provide for functional relationships between system inputs and system outputs, whilst these functions' parameters do not have any physical significance (Zhang, 2010). Hence, it becomes more difficult to understand the effect that each predictor has on the response variable. Nonetheless, techniques based on the feature importance indicator can be effectively utilised to rank the predictors according to their significance (Strobl et al., 2007). Therefore, investigating the applicability of ANNMs for the prediction of the frequency of design load overcoming events and comparing their

ARTICLE IN PRESS

4   J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

performance with respect to GLRMs are turning out to be intriguing challenges.

Summing up, all previous studies have undoubtedly contributed to the analysis of the threat posed by extremely high traffic load hazards on bridges from different angles and have provided valuable findings. Measurement of site-specific traffic loads, comparison of site-specific and code-design traffic load effects on bridges, assessment of lifetime extreme traffic load effects, evaluation of bridge reliability under complex traffic load effects, and recommendation of specification-level traffic load models are key examples. However, some gaps persist. First, existing failure probability models might be complex and very expensive when used by RAs. Therefore, adopting frequency-based models as a driver of failure probability may be an addition to the RAM. Second, whilst the frequency design load overcoming events was estimated by GLRMs, there were no such estimations using ANNMs. Third, no literature has compared the fitting and predictive performance of GLRMs and ANNMs.

The objective of this study is to fill these gaps. Specifically, this study specifies, calibrates, and validates two frequency prediction models based on GLRMs and ANNMs, respectively. Next, it provides a detailed comparison of the related performance of each model, which was estimated using 2 million of WIM raw data collected on an important bridge along a heavily transited ring road in Brescia (Italy). The remainder of the paper is organised as follows. Section 2 presents the materials and methods used to process the WIM data, identifying frequency predictors, build GLRMs and ANNMs, and compare their performance. Section 3 shows and discusses the results, and then last, Section 4 draws some conclusions and provides a view of some future perspectives.

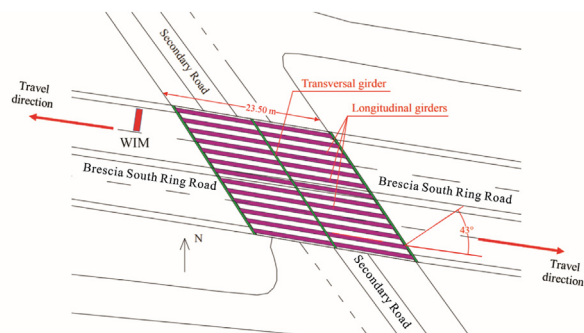**Fig. 1** − **The case study bridge on the south ring road in the city of Brescia (Lombardy Region, Italy).**

Fig. 2 − **Plan of the case study bridge (23.50 m is the span length of the bridge, whereas 43° represents the width of the skew angle).**

2. Materials and methods

2.1. Data type and collection

Data for this research were gathered on a currently functioning bridge along the South Ring Road in the city of Brescia (Lombardy Region, Italy). Brescia is one of Italy's most significant industrial and economic hubs and the second-most populous city in the Region (De Aloe et al., 2022; Martinelli et al., 2022). The South Ring Road is one of the city's arterial roads with the highest traffic volume and quota of commercial vehicles in the province (Faccin et al., 2011). The case-study bridge is a simply supported (23.50 m span length) overpass structure with a secondary road running beneath it (Fig. 1).

The road section on the case study bridge has two separate carriageways, with two lanes in each direction. A WIM system was placed on the north side westbound carriageway to collect experimental traffic load data on the bridge during the monitoring period (denoted by T). Due to budget restrictions, instrumentation was installed only in the right lane since heavy vehicle transit is not permitted in the left lane (Fig. 2). The WIM system adopted was composed of two stainless steel plates placed on the road surface, equipped with fibre optic sensors, and linked to a data logger. The system was rated as an "accuracy class 10" instrument according to the OIML standards for WIM devices (OIML, 2006).

WIM raw data were pre-processed to remove anomalies and outliers, according to the filtering procedure (FP) and the quality control algorithm (QCA) proposed by Ventura et al. (2024). Next, since the vehicular load was time dependent, the monitoring period (T) was partitioned into equal-width time frames (referred to as timeslots). This subdivision enabled the examination of how the features of the vehicles passing on the bridge and the frequency of design load overcoming events varied over time.

More formally:
- S is the set of timeslots and $T(s)$ is the subset of T in the timeslot $s \in S$.
- F is the set of frequency predictors, $f_i \in F$ is the ith predictor and $f_i(s)$ (predictor specific) is the value of $f_i \in F$ observed in $T(s)$. About the predictor specific, the unit of measurement varies according to the specific frequency predictor. For more details, please see Table A1 in the Appendix.
- $H(s)$ (events/timeslot) is the frequency of design load overcoming events observed during $T(s)$.

According to Ventura et al. (2024), the set of frequency predictors should include intermediate safety factors and one risk exposure factor. Intermediate safety factors describe the global geometrical properties of the bridge (bridge side factors), on the day and at the time when each

ARTICLE IN PRESS

J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx   **5**

timeslot was recorded (temporal context factors), and the traffic load hazard on the monitored bridge (traffic load hazard factors). Traffic load hazard factors were further organised into traffic flow characteristics, vehicular characteristics, interaction between vehicular and bridge characteristics, compliance with TC prescriptions and actions induced on the structure. The risk exposure factor measured the quantity of vehicles that could induce or could be involved in a design load overcoming event.

To sample the response variable of the frequency model, the frequency of design overloading events was measured. To perform this task, the traffic load acting on the monitored bridge lane in each instant $t \in T$ was first computed by processing the axle load data provided by the WIM device. It is noteworthy that disaggregated axle load data were preferred over aggregated vehicular load data, because non-integer fractions of vehicles might be present on the bridge lane at any given instant. Since the bridge deck runs along a certain length in space and the WIM system only provided records of axle loads at a single point in space, the motion law analysis of each axle was performed to determine the application point of each axle load at times other than those when it was passing over the WIM device.

Then, for each instant $t \in T$, the overall traffic load was computed by summing the loads acting on the axles that were on the bridge in the lane at $t \in T$. Second, different design traffic lane load thresholds were defined according to the load combinations prescribed by the Eurocode 1 standard (European Committee for Standardization, 2003) for different limit states (i.e., ultimate limit state, irreversible serviceability limit state and reversible serviceability limit state). Finally, for each $s \in S$, the frequency of design overloading events ($H(s)$) was measured by counting the number of times in which the traffic load on the monitored bridge lane exceeded one or more design load thresholds. More details about the procedure for building the frequency response variable were reported by Ventura et al. (2024).

2.2. Econometric approach

Once $H(s)$ was computed, the frequency prediction model was built in a traditional way by adopting an econometric approach. It was found that Poisson and negative binomial formulations could be applied to model the frequency of design load overcoming events. These models appear to be the dominant mathematical tools for modelling non-negative discrete response variables as in the case of such events, owing to their solid statistical properties. However, exposure variables refer to a variable, whose frequency of events must be zero when it is equal to zero. Therefore, the frequency model of design load overcoming events was estimated by a generalised linear regression model (GLRM) with a negative binomial regression (NBR) error structure, as applied in other engineering fields (Barabino et al., 2021, 2023; Liu et al., 2005).

More formally:
- $TR \subset S$ and $TE \subset S$ are the training and testing subsets, respectively.
- $\widetilde{H}(s)$ (events/timeslot) is the predicted value for the observed $H(s)$.

- α, β, γ_{f_i} (predictor specific) are the coefficients of the frequency model.
- $E(s)$ (veh/h) is the exposure factor observed during $s \in S$.
- p_H (−) is the p-value associated with the model, i.e., global statistical significance of the model according to the Chi-square test.
- p_{f_i} (−) is the p-value associated with the frequency predictor f_i, i.e., the significance of every single predictor according to the t-test.
- dr (−) is the deviance ratio, i.e., the ratio between the regression deviance and the degree of freedom.

Hence, the frequency was predicted according to the following functional form Eq. (1).

$$\widetilde{H}(s) = \alpha E(s)^{\beta} e^{\sum_{f_i \in F, f_i \neq E} \gamma_{f_i} f_i(S)} \quad \forall\ s \in S \tag{1}$$

To implement a model validation technique, which relies on unbiased out-of-sample evaluations, the dataset was randomly split into training (TR) and testing (TE) subsets before performing the fitting procedure. Hence, the fitting procedure started with the identification of the predictors (f_i) to be included in the model. As for the exposure factor ($E(s)$), the hourly flow was chosen to impose null frequency predictions for those timeslots with a null vehicular flow. As for the other predictors, these were extracted from F by applying a straightforward selection method. First, a preliminary filtering procedure was performed to prevent multicollinearity issues by removing high-correlated predictors, as follows.

1) Compute the correlation matrix related to each predictor $f_i \in F$ and the response variable $H(s)$.
2) Individuate the couples of high correlated predictors (i.e., correlation index greater than 0.8, according to Shrestha (2020)).
3) Eliminate the predictor which had the lower correlation index with the response variable $H(s)$, for each couple of high correlate predictors.

Next, an automated stepwise technique based on forward selection and backward elimination was adopted to individuate the best set of predictors from the list obtained through the preliminary filtering procedure. According to the highest dr and p_H of the estimated model, either a backward or a forward stepwise technique was preferred. Once the model was fitted, the sign of each coefficient and the significance of each predictor (p_{f_i}) were assessed to understand the influence that each factor would have on the predicted frequency.

2.3. Machine learning approach

The frequency prediction model was also built in a more innovative way by using a machine learning approach. According to the literature, an ANNM was chosen above other machine learning algorithms.

More formally:
- $TR \subset S$, $VA \subset S$ and $TE \subset S$ are the training, validation, and test subsets, respectively.

ARTICLE IN PRESS

6   J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

- $IN \in \mathbb{R}^{|S|,|F|}$ (predictor specific) is the input matrix for the ANNM fitting process, i.e., the matrix of the frequency predictors (i.e., F) observed during each $T(s) \in S$.
- $TG \in \mathbb{R}^{|S|}$ (events/timeslot) is the target vector for the ANNM fitting process, i.e., the vector of the frequencies of design load overcoming events (i.e., $H(s)$) observed during each $T(s) \in S$.
- $\widetilde{TG} \in \mathbb{R}^{|S|}$ (events/timeslot) is the predicted value for TG.
- ω is a mathematical function, which relates the input matrix IN to the target vector TG.
- $\widetilde{\omega}$ is an approximation of the function ω.
- P is the set of ANNM hidden layer perceptrons and $p \in P$ is the generic perceptron. The perceptron is the basic structure of the network and is a simplified model of a biological neuron.
- $\theta \in \mathbb{R}^{|F||P|+2|P|+1}$ (predictor specific) is the generic vector containing the parameters of the ANNM and $\theta_0 \in \mathbb{R}^{|F||P|+2|P|+1}$ is the specific vector obtained through the learning phase.
- PFI_{f_i} (events/timeslot) is the permutation feature importance associated to the predictor $f_i \in F$.
- $\epsilon \in \mathbb{R}^{|S|}$ (events/timeslot) is the vector of residual values, i.e., the difference between the target vector TG and the predicted one $\widetilde{TG}$.
- $MSE(\theta)_{VA}$ (events/timeslot) is the mean squared error computed on the validation subset (VA) as a function of the parameter vector (θ).

Then, assume the existence of a mathematical function ω, which relates the input matrix IN to target vector TG, such that $TG = \omega(IN)$. If this function exists, it can be interpreted as a computation model that links the causes (IN) with their observed effects (TG). The ANNM defines a mapping $TG = \widetilde{\omega}(IN, \theta) + \epsilon$ and learns the value of the parameter vector θ that leads to the best approximation (Ian et al., 2016).

A two-layer feed-forward network was selected to perform this mapping, being the more straightforward among the ANNM architectures (Schmidhuber, 2015). In this architecture, the information moves only forward direction from the input nodes, across the hidden nodes and towards the output nodes. This network has two layers of perceptrons: a hidden layer of $|P|$ perceptrons and an output layer on one perceptron. A hyperbolic tangent sigmoid activation function was selected for the hidden layer, which was recommended due to its antisymmetric shape (Romero Reyes et al., 2013; Sramka et al., 2019). A linear activation function was chosen for the output layer since the response was an integer variable.

Before performing the training phase, the set of timeslot observations (i.e., S) was randomly divided into three subsets: training, validation, and test. The training set (i.e., TR) was presented to the network during training, and the network parameters were calibrated to fit the training data. The validation set (i.e., VA) was used to measure network generalisation to prevent overfitting and halting training when generalisation stopped improving. The testing set (i.e., TE) did not affect training and so provided an independent measure of network performance during and after training. During the training phase, θ was adjusted to minimise a certain cost function computed on the validation subset (Brownlee, 2019). Particularly, the $MSE(\theta)_{VA}$ was selected as cost function, being

the most popular for regression problems where a quantity has been predicted (Reed and Marks II, 1999). Thus, the training procedure was formalized as follow in Eq. (2).

$$\theta_0 = \arg \min_{\theta} MSE(\theta)_{VA} \tag{2}$$

A backpropagation algorithm was selected being one of the most widely used for training feedforward neural networks (Ian et al., 2016). At the end of the training phase, the best parameter vector θ_0 was determined and, thus, the functional form of the ANNM for frequency prediction was obtained, which is shown in Eq. (3).

$$\widetilde{H}(s) = \widetilde{\omega}\big((f_i(s) \in F), \theta_0\big) \quad \forall s \in S \tag{3}$$

It is noteworthy that training multiple times generated different results due to diverse initial conditions and random splitting. Thus, several training instances were issued, and the best model was selected according to the lower MSE.

Next, the significance of each predictor was evaluated to understand its influence on the forecasted frequency. To perform this task, the permutation feature importance (i.e., PFI_{f_i}) was adopted because it is recognised as an indicator that produces reliable results, especially for non-linear or opaque estimators (Strobl et al., 2007). The PFI_{f_i} is defined as a decrease in a model score when a single factor value is randomly shuffled (Breiman, 2001). The technique destroys the link between the factor and the target, and the decrease in the model score reflects how much the model depends on the factor. This method has the advantage of being model-independent and can be calculated many times with various permutations of the characteristic. In this paper, the test dataset rather than the training dataset was used to fit the indicator, since this way, the significance of features in an out-of-sample prediction can be emphasized even more.

- N is the set of random permutation instances and $n \in N$ is the generic instance.
- $IN_{nf_i} \in \mathbb{R}^{|S|,|F|}$ (predictor specific) is the corrupted version of the input matrix at the permutation $n \in N$, which was obtained by randomly shuffling the column associated to $f_i \in F$ of the uncorrupted input matrix IN.
- $MSE(\theta_0)_{TE}$ (events/timeslot) is the MSE of the trained ANN model, calculated on the test dataset TE, considering the uncorrupted input matrix IN.
- $MSE(\theta_0)_{TE nf_i}$ (events/timeslot) is the MSE of the trained ANN model, calculated on the test dataset TE, considering the corrupted input matrix IN_{nf_i}.
- $s = -MSE(\theta_0)_{TE}$ (events/timeslot) is the reference score of the trained ANN model on the uncorrupted input matrix IN.
- $s_{nf_i} = -MSE(\theta_0)_{TE nf_i}$ (events/timeslot) is the score of the trained ANN model on the corrupted input matrix IN_{nf_i}.

Then, multiple permutation instances ($|N|$) were performed for each explanatory factor $f_i \in F$ and the associated permutation feature importance was calculated by considering the average of the scores relating to the different corrupted datasets shown in Eq. (4).

ARTICLE IN PRESS

J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

7

$$\text{PFI}_{f_i} = s - \frac{1}{|N|} \sum_{n \in N} s_{nf_i} \quad \forall \, f_i \in F \tag{4}$$

2.4. Comparison between econometric and machine learning performances

Once the GLRM and ANNM were estimated, they were compared to find the best model in fitting and predicting the frequency of design overloading events. To make this choice, three comparison strategies were employed as follows.

   I Regression plots (RPs).
  II Some measures of errors (MoEs).
 III The ratio among the number of observed vs predicted events.

As for I, RPs were a useful way to visually assess the fitting and forecasting performance of a model. These were graphs showing the pairs' target frequency-vs-predicted frequency on a Cartesian plane. More formally, for a specific subset (i.e., TR, VA and TE) and model (i.e., GLRM and ANNM), each RP was obtained by putting the pairs $(H(s), \widetilde{H}(s))$ on a Cartesian plane and by determining the correlation coefficient (denoted as R) between the target variable (i.e., $H(s)$) and the predicted variable (i.e., $\widetilde{H}(s)$). Next, RPs associated with GLRM and ANNM were compared considering the following items. 1) The closer to the first quadrant bisector are the $(H(s), \widetilde{H}(s))$ points, the greater the performance will be, because a point on the first quadrant bisector corresponds to a perfect fitting or prediction (i.e., $\widetilde{H}(s) = H(s)$). 2) The closer to 1 that R is, the greater the model performance will be, since when $R = 1$ it indicates a perfect correlation between observed and predicted values.

As for II, some MoEs were evaluated, being common indicators to measure the differences between the values predicted by a model and the corresponding values observed in the real world. MoEs include: the mean absolute error (MAE), the root mean squared error (RMSE), and the coefficient of variation (CoV) among them. The MAE measures the average magnitude of the errors in a set of predictions, without regard to their direction. It is defined as the arithmetic mean of the absolute errors. Since the MAE is a linear score, it gives equal weight to each individual error in the sample. The RMSE is a quadratic scoring rule that measures the average magnitude of error. It is computed by squaring the difference between the prediction and the corresponding observed value, and then averaging these values over the sample. Next, the square root is calculated. Since the RMSE is a nonlinear score, it gives relatively great weight to large errors. This is because errors are squared before they are averaged. This means that the RMSE is particularly helpful when large errors are unwanted. The CoV describes the model fit in terms of the relative sizes of the squared errors and the values taken by the response variable. It is defined as the ratio of the RMSE to the mean of the response variable. Hence, it can be interpreted as a normalized version of the RMSE indicator. More formally, let $\epsilon(s)$ be the residual of the prediction during $T(s)$, i.e., the component of the vector ϵ associated to $T(s)$, which is shown in Eq. (5).

$$\epsilon_s = H(s) - \widetilde{H}(s) \quad \forall \, s \in S \tag{5}$$

Therefore, as for the training subset (TR), MoEs were calculated according to Eqs. (6)–(8).

$$\text{MAE} = \frac{\sum\limits_{s \in \text{TR}} |\epsilon(s)|}{|\text{TR}|} \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{\sum\limits_{s \in \text{TR}} \epsilon(s)^2}{|\text{TR}|}} \tag{7}$$

$$\text{CoV} = \frac{\text{RMSE}}{\frac{\sum\limits_{s \in \text{TR}} |H(s)|}{|\text{TR}|}} \tag{8}$$

As for VA and TE subsets, MoEs were computed according to the previous equations where TR was replaced with VA and TE, respectively. Next, the MoEs associated with GLRM and ANNM were compared considering that lower values are better.

As for III, the sum of the observed and predicted number of design load overcoming events (denoted as SUM and $\widetilde{\text{SUM}}$) was first computed throughout the entire monitoring period. Next, the ratio between $\widetilde{\text{SUM}}$ and SUM (denoted as π) was determined. More formally, as for the TR, these quantities were calculated according to Eqs. (9)–(11).

$$\text{SUM} = \sum_{s \in \text{TR}} H(s) \tag{9}$$

$$\widetilde{\text{SUM}} = \sum_{s \in \text{TR}} \widetilde{H}(s) \tag{10}$$

$$\pi = \frac{\widetilde{\text{SUM}}}{\text{SUM}} \tag{11}$$

As for the VA and TE subsets, SUM, $\widetilde{\text{SUM}}$ and SUM and π were computed according to the same previous equations where TR was replaced with VA and TE, respectively. Next, π ratios were compared considering that the closer to 1 that the π ratio is, the greater the fitting performance will be, because $\pi = 1$ means that the total number of predicted events coincides with the observed number.

3. Results and discussion

3.1. Descriptive statistics

A total of 2 million WIM raw records were collected during a four-month observation period (T). After performing the pre-processing procedure, 0.9 million vehicular records (around 45% of the entire passing traffic flow) were maintained on the validated traffic dataset (Table 1). Although a high percentage of data was excluded, these mainly referred to light and fast vehicles (i.e., cars and vans) for which the WIM was unable to measure certain parameters. The reason for this was that WIM systems have been primarily conceived for weighing

ARTICLE IN PRESS

8   J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

Table 1 – Details about the WIM raw data collected during the monitoring period.

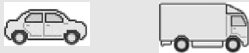| Description	Detail
Start date	1st January 2022
End date	30th April 2022
Number of raw vehicular records	2,002,320
Number of validated vehicular records (after FP and QCA)	904,980
Duration of each timeslot (s)	3600
Number of timeslots	2696

heavier vehicles rather than lighter ones. Hence, the reliability of the consequent frequency analyses was not impaired, since lighter vehicles give only a slight contribution in determining the load acting on a bridge.

Seven vehicular classes were recognized by the WIM system (Table 2). In Table 2, the fraction of vehicles inside each class refers to the validated traffic database. Although cars and vans were shown to be the most frequent vehicle types, as expected, a relatively high percentage (7.22%) of commercial vehicles (i.e., classes from 2 to 6) was observed, confirming the high quota of such vehicles in the South Ring Road. Notably, a relevant fraction of vehicles was found to have been overloaded, with class 5 (i.e., vehicles with more than 6 axles) and class 2 (i.e., single unit trucks and buses) being in the greater proportion (about 90% and 83%, respectively). To visualize the GVM distribution associated with each vehicular class, probability density functions (PDFs) were graphed (Fig. 3(a)–(g)). Multimodal shapes were also observed for several vehicular classes, suggesting the presence of vehicles with different load ratios (e.g., empty, partially loaded, or fully loaded), which was an expected outcome (Lu et al., 2018; Ren et al., 2019).

Since frequency modelling was performed monthly, the timeslot duration was set to one hour. Thus, a set of 2656 hourly timeslots (S) was obtained (Table 1). Next, the set of frequency predictors (F) was computed for each timeslot. The list of these predictors, including their definitions and

Table 2 – Vehicular classes recognized by the WIM device.

Class ID	Class name	Symbolic illustration	Limit mass according to the Italian TC (kg)	Fraction of vehicles inside the class (%)*	Fraction of vehicles exceeding the limit mass (%)**
1	Cars and vans		3500	92.889	0.277
2	Single-unit trucks and buses		18,000 (2 axles) 25,000 (3 or more axles)	0.597	82.588
3	Articulated trucks up to 6 axles		30,000 (2 axles) 40,000 (4 axles) 44,000 (5 or more axles)	5.733	44.978
4	Road trains up to 6 axles		24,000 (3 axles) 40,000 (4 axles) 44,000 (5 or more axles)	0.469	42.072
5	More than 6 axle vehicles		44,000	0.309	89.989
6	Isolated trailers		6000 (1 axle) 22,000 (2 axles) 26,000 (3 or more axles)	0.003	4.545
7	Unknown vehicles	?	44,000	<0.001	0.000

Note: * refers to the ratio of the number of vehicles in the specific class to the total number of vehicles in the validated traffic database. ** refers to the ratio of the number of vehicles in the specific class with a GVM above the TC limit to the total number of vehicles in the same class.
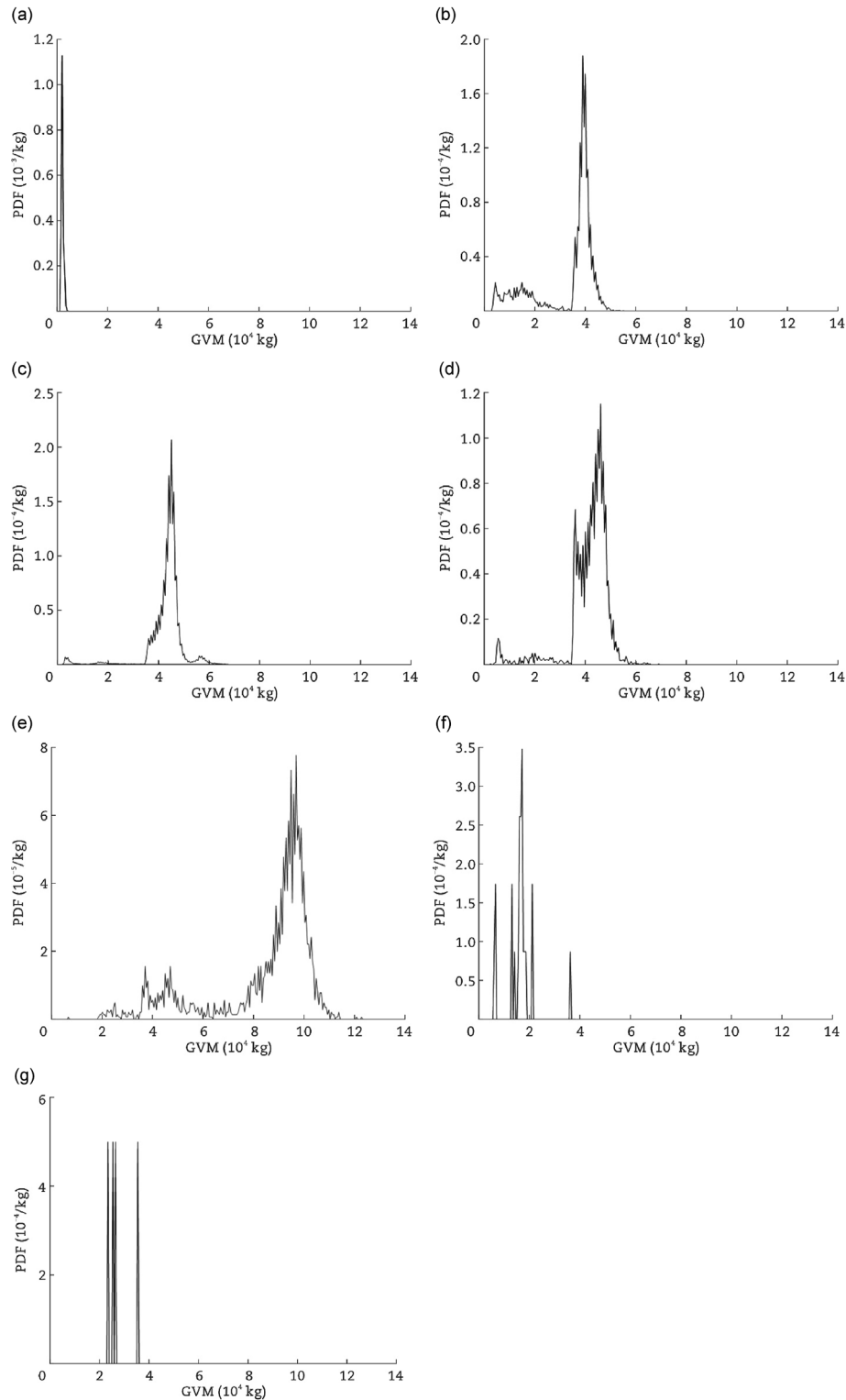
ARTICLE IN PRESS

J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

9

Fig. 3 — **PDFs of GVM associated to seven classes. (a) Class 1. (b) Class 2. (c) Class 3. (d) Class 4. (e) Class 5. (f) Class 6. (g) Class 7.**

some descriptive statistics, is given in Table A1 in the Appendix, which is otherwise self-explanatory.

Notably, although lane width and span length were constant terms in the bridge at hand, they were considered as predictors for completeness. Indeed, from a theoretical viewpoint, if additional bridges were included in the modelling, these factors could have different values and, thus, could be significant predictors. Moreover, several temporal context factors (e.g., type of day, type of hour, and lighting) might appear to not be directly related to traffic overloading.

ARTICLE IN PRESS

10   J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

Therefore, it could be argued that if these predictors were excluded from the modelling, the results could still be true. Nevertheless, traffic on the bridge is expected to change over the time. Hence, temporal context factors could have some influence on the frequency of traffic load overcoming events.

By comparing the overall vehicular load on the monitored bridge lane during each instant with the design load thresholds prescribed by Eurocode 1, the frequency of design overloading events occurred during each $T(s)$, i.e., the response variable H_s, was determined. Some descriptive statistics on $H(s)$ are given in Table A1 in the Appendix.

3.2. Generalised linear models

Once the WIM raw data had been processed, the frequency prediction model was first built by adopting the Econometric approach, as indicated in Eq. (1). A splitting ratio of 80% and 20% was employed to identify TR and TE, respectively (Fertlitsch, 2021). The coefficient estimates and significances for each component of the best-fit model are shown in Table 3, whereas Table 4 reports some summary statistics.

From a general perspective, this model fits the data well. Indeed, the statistical χ^2 test on dr produced a small p_{GL}-value for goodness-of-fit (<0.001). Therefore, the null hypothesis that all the regression coefficients were zero could be rejected. As for predictors, the results confirmed that most of these were highly significant (i.e., $p_{f_i} < 0.05$), which showed a strong regression effect. Focusing on each highly significant predictor separately, the following considerations arose.

As for the exposure risk factor, the results showed that the exponent (i.e., β) is positive as expected. Therefore, a 1 veh/h increase in the hourly flow would increase the number of overloading events as well, while keeping all other variables constant at their means. This result confirmed previous research because it has been widely recognised that the bridges subjected to the greatest traffic flows have the highest

Table 4 – Results of the best fit GLR frequency prediction model (econometric approach): summary statistics.

Source	Degree of freedom	Mean deviance	Parameter	Value
Regression	17	253.130	dr	253.130
Residual	2139	0.266	χ^2	0.001
Total	2156	2.260		

risk (Fiorillo and Ghosn, 2022; Ministero delle Infrastrutture e dei Trasporti, 2020).

As for temporal context factors, the negative coefficient associated with the type of day factor (i.e., γ_{f_4}) indicated that the frequency was lower over the weekend. This was a consequence of the fewer commercial activities during weekends. Nonetheless, it was a novel outcome because the type of day factor had not been investigated in previous safety analyses. Moreover, the strong significance of this factor ($p_{f_4} < 0.001$), suggested that excluding temporal context factors from the list of frequency predictors would have worsened the model fit.

As for the traffic flow characteristics factors, the speed factor was found to be significant. Particularly, a 1 km/h increase in the mean speed reduced the number of overloading events, whilst it kept all other predictors constant at their means. Therefore, the frequency decreased as the mean speed increased, as indicated by the negative coefficient (i.e., $\gamma_{f_{14}}$). This result could be justified as follows. First, the faster vehicles were generally characterized by a lower GVM, having a greater probability of belonging to lighter categories (Ventura et al., 2023a). Second, an increase in speed led to a reduction in the permanence time of the vehicles on the deck and, therefore, a lower probability of multiple presences on the bridge. Consequently, this suggested that imposing low-speed limits on bridge decks might have a negative safety impact. This was a noteworthy result, considering that RAs

Table 3 – Results of the best fit GLR frequency prediction model (econometric approach): coefficient estimates.

Parameter	Related predictor	Est.
$\log(\alpha)$	—	−1.920
β	Hourly flow (exposure factor)	0.400***
γ_{f_4}	Type of day-weekend wrt weekday	−0.510***
γ_{f_8}	Vehicular fraction-class 2	−8.390
$\gamma_{f_{11}}$	Vehicular fraction-class 5	6.950
$\gamma_{f_{14}}$	Speed-mean	−0.042***
$\gamma_{f_{29}}$	Length-minimum	−2.094**
$\gamma_{f_{40}}$	Axle imbalance ratio-mean	−1.426**
$\gamma_{f_{43}}$	Axle imbalance ratio-maximum	0.438**
$\gamma_{f_{44}}$	Interaxle-mean	1.084*
$\gamma_{f_{55}}$	GVM-length ratio-maximum	0.00043***
$\gamma_{f_{60}}$	GVM limit ratio-mean-class 1	3.910**
$\gamma_{f_{68}}$	GVM limit ratio-maximum-class 2	−0.151
$\gamma_{f_{70}}$	GVM limit ratio-maximum-class 4	0.403***
$\gamma_{f_{77}}$	Overloaded vehicles fraction-class 3	1.711***
$\gamma_{f_{78}}$	Overloaded vehicles fraction-class 4	0.271**
$\gamma_{f_{82}}$	Extremely loaded vehicles following one another	0.150**
$\gamma_{f_{87}}$	Overloaded axles fraction	19.260***

Note: * indicates that the variable is significant at 0.10 level or lower. ** indicates that the variable is significant at 0.05 level or lower. *** indicates that the variable is significant at 0.001 level or lower. wrt stands for with respect to.

ARTICLE IN PRESS

J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx   **11**

often prescribe very low speed limits for overloaded trucks with special permits moving on bridge decks (Ventura et al., 2020).

As for vehicular characteristics factors, three predictors were found to be highly significant: one related to the entire vehicle (i.e., length) and two to single axles (i.e., axle imbalance ratio and interaxle). Particularly, a 1 m increase in the minimum length of vehicle observed during a timeslot, reduced the number of overloading events, whilst keeping all other variables constant at their means (as shown by coefficient $\gamma_{f_{29}} < 0$). Perhaps, the interposition of short and light vehicles prevented multiple heavy vehicles from being on the bridge simultaneously. Therefore, an increase in the length of short vehicles interposed between two heavy vehicles following one another kept them away from each other and reduced the likelihood of concurrent presence on the bridge. According to what the authors were able to find in the literature, this was an original finding not pointed out in any earlier bridge safety analyses. Conversely, it appears that the axle imbalance factor had a contrasting effect. Indeed, the frequency of overloading events seemed to be reduced as the mean axle imbalance ratio of a unit ($\gamma_{f_{40}} < 0$) increased, whilst an increase in the maximum axle imbalance ratio induced the opposite result ($\gamma_{f_{43}} > 0$). These opposing behaviours could indicate that globally axle imbalance had only a slight influence on frequency.

As for the interaction between vehicular and bridge characteristics factors, the GVM length ratio was shown to be highly significant, whilst the coefficient was found to be positive as intuitively expected. Therefore, with a 1 kg/m increase in the GVM length ratio, the number of overloading events increased whilst keeping all other predictors constant at their means. Hence, this outcome confirmed that when short and heavily loaded vehicles were detected, the probability of observing a high traffic load on the bridge would increase, since more vehicular mass can simultaneously impinge upon on the entire deck length. Despite being evident, this aspect was not directly emphasized in past bridge risk evaluations, as far as the authors know.

As for compliance with TC prescription factors, this is the subgroup that has the greater influence on frequency, since the six factors were found to be extremely significant: five related to the vehicle and one related to single axles. Precisely, the frequency increased as the GVM limit ratio for class 1 and class 4 vehicles rose ($\gamma_{f_{60}} > 0$ and $\gamma_{f_{70}} > 0$). Likewise, a greater percentage of overloaded class 3 and class 4 vehicles led to a higher frequency of overcoming events ($\gamma_{f_{77}} > 0$ and $\gamma_{f_{78}} > 0$). These outcomes imply that vehicles exceeding the mass limits prescribed by the TC can induce a higher threat on the bridge than those that are compliant. Moreover, since only coefficients of class 1, class 3, and class 4 emerged as significant, illegally overloaded vehicles appeared to be more detrimental than those with permits, which are more likely to belong to class 5. This could be explained considering that illegally overloaded vehicles are generally shorter than those with permits, as the latter are conceived with more axles to reduce their mass linear density. These findings endorse Fiorillo and Ghosn (2018), who showed that the probability of failure increased with the percentage of illegally overweighted trucks.

Besides, a positive coefficient was found for the factor related to the detection of extremely loaded vehicles (GVM higher than 44,000 kg) following one another (i.e., $\gamma_{f_{82}}$). This is a consistent outcome because following vehicles are more likely to simultaneously act on the bridge deck than distant ones, thus inducing a lane load greater than design thresholds. This result confirms Fiorillo and Ghosn (2019) and Fiorillo and Ghosn (2022) who showed that the probability that overloaded trucks crossing the bridge simultaneously could have a negative effect on bridge reliability. As for single axle factors, a 1% increases in the overloaded axle fractions strongly increases the number of overloading events, as shown by the relatively high coefficient (i.e., $\gamma_{f_{87}}$). This result denotes that frequency increases significantly when the percentage of overloaded axles increases as well, because more of the vehicular load will act on the bridge lane. On the one hand, the outcome is novel since the overloaded axle fractions had never been explicitly accounted for in previous models. On the other hand, this finding was expected because it is well-known that overloaded axles lead to adverse impacts on infrastructure, such as a rise in bridge deterioration rates (Lou et al., 2016) and a decrease in road pavement life (Zhao et al., 2021).

3.3. Artificial neural network models

Next, the frequency model was fitted by adopting the machine learning approach based on an ANNM, according to Eqs. (2) and (3). A splitting ratio of 70%, 15% and 15% was employed to identify TR, VA and TE, respectively (Flach, 2012). Numerous attempts to gain improvement were undertaken by varying the number of neurons in the hidden layer and the training algorithm. Then, a network with 10 perceptrons in the hidden layer trained with the Levenberg-Marquardt algorithm was chosen because it returned the best data fitting. The training procedure was halted at epoch 9, corresponding to the minimum cost function computed on the validation subset $(\text{MSE}(\theta)_{\text{VA}})$, as indicated in Fig. 4. In epoch 9, the best validation performance was attained.
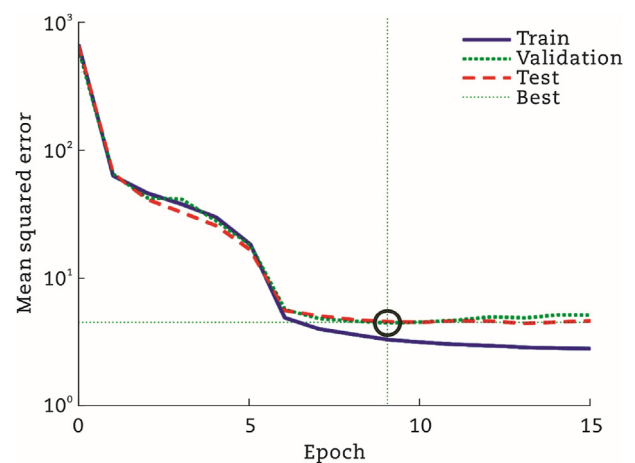

Fig. 4 – **Cost function for the ANN frequency model as a function of training epoch (machine Learning approach).**

ARTICLE IN PRESS

12   J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

Table 5 – Permutation feature importance of each predictor for the ANN frequency model (machine learning approach).

Frequency predictor	Symbol	PFI[*,**]	Position[**]	Frequency predictor	Symbol	PFI[*,**]	Position[**]
Hourly flow (exposure factor)	$f_1(s) = E(s)$	32.88	15	Interaxle-std. deviation	$f_{45}(s)$	12.64	22
Lane width	$f_2(s)$	NA	NA	Interaxle-minimum	$f_{46}(s)$	0.04	76
Span length	$f_3(s)$	NA	NA	Interaxle-maximum	$f_{47}(s)$	1.39	47
Type of day–weekend wrt weekday	$f_4(s)$	7.14	35	Normalized GVM-mean	$f_{48}(s)$	200.49	1
Lighting–daytime wrt night-time	$f_5(s)$	1.49	46	Normalized GVM-std. deviation	$f_{49}(s)$	10.67	25
Type of hour–peak hour wrt off-peak hour	$f_6(s)$	0.17	71	Normalized GVM-minimum	$f_{50}(s)$	−0.02	81
Vehicular fraction-class 1	$f_7(s)$	45.09	13	Normalized GVM-maximum	$f_{51}(s)$	57.19	11
Vehicular fraction-class 2	$f_8(s)$	0.87	52	GVM-length ratio-mean	$f_{52}(s)$	65.82	8
Vehicular fraction-class 3	$f_9(s)$	27.70	16	GVM-length ratio-std. deviation	$f_{53}(s)$	14.97	20
Vehicular fraction-class 4	$f_{10}(s)$	1.28	48	GVM-length ratio-minimum	$f_{54}(s)$	0.01	79
Vehicular fraction-class 5	$f_{11}(s)$	1.56	45	GVM-length ratio-maximum	$f_{55}(s)$	12.02	23
Vehicular fraction-class 6	$f_{12}(s)$	−0.43	84	GVM limit ratio-mean-all classes	$f_{56}(s)$	69.13	6
Vehicular fraction-class 7	$f_{13}(s)$	0.23	66	GVM limit ratio-std. deviation-all classes	$f_{57}(s)$	26.58	17
Speed-mean	$f_{14}(s)$	1.74	44	GVM limit ratio-minimum-all classes	$f_{58}(s)$	0.48	59
Speed-std. deviation	$f_{15}(s)$	0.24	65	GVM limit ratio-maximum-all classes	$f_{59}(s)$	0.54	57
Speed-minimum	$f_{16}(s)$	1.06	50	GVM limit ratio-mean-class 1	$f_{60}(s)$	7.76	33
Speed-maximum	$f_{17}(s)$	−0.13	82	GVM limit ratio-mean-class 2	$f_{61}(s)$	10.80	24
Headway-mean	$f_{18}(s)$	10.13	28	GVM limit ratio-mean-class 3	$f_{62}(s)$	3.62	41
Headway-std. deviation	$f_{19}(s)$	0.35	61	GVM limit ratio-mean-class 4	$f_{63}(s)$	14.47	21
Headway-minimum	$f_{20}(s)$	0.13	74	GVM limit ratio-mean-class 5	$f_{64}(s)$	8.67	31
Headway-maximum	$f_{21}(s)$	0.28	64	GVM limit ratio-mean-class 6	$f_{65}(s)$	0.06	75
Congestion condition-free flow	$f_{22}(s)$	0.18	70	GVM limit ratio-mean-class 7	$f_{66}(s)$	0.20	67
GVM-mean	$f_{23}(s)$	193.78	2	GVM limit ratio-maximum-class 1	$f_{67}(s)$	0.33	62
GVM-std. deviation	$f_{24}(s)$	95.16	4	GVM limit ratio-maximum-class 2	$f_{68}(s)$	6.70	36
GVM-minimum	$f_{25}(s)$	0.04	78	GVM limit ratio-maximum-class 3	$f_{69}(s)$	3.99	38
GVM-maximum	$f_{26}(s)$	92.16	5	GVM limit ratio-maximum-class 4	$f_{70}(s)$	10.66	26
Length-mean	$f_{27}(s)$	119.80	3	GVM limit ratio-maximum-class 5	$f_{71}(s)$	17.09	19
Length-std. deviation	$f_{28}(s)$	65.85	7	GVM limit ratio-maximum-class 6	$f_{72}(s)$	0.19	68
Length-minimum	$f_{29}(s)$	0.90	51	GVM limit ratio-maximum-class 7	$f_{73}(s)$	0.15	73
Length-maximum	$f_{30}(s)$	0.53	58	Overloaded vehicles fraction-all classes	$f_{74}(s)$	6.21	37
Axle number-mean	$f_{31}(s)$	50.27	12	Overloaded vehicles fraction-class 1	$f_{75}(s)$	−0.21	83
Axle number-std. deviation	$f_{32}(s)$	44.79	14	Overloaded vehicles fraction-class 2	$f_{76}(s)$	8.95	30
Axle number-minimum	$f_{33}(s)$	0.00	80	Overloaded vehicles fraction-class 3	$f_{77}(s)$	22.58	18
Axle number-maximum	$f_{34}(s)$	3.98	39	Overloaded vehicles fraction-class 4	$f_{78}(s)$	1.10	49
Axle typology-double wheels axle fraction	$f_{35}(s)$	8.66	32	Overloaded vehicles fraction-class 5	$f_{79}(s)$	0.72	53
Axle mass-mean	$f_{36}(s)$	9.58	29	Overloaded vehicles fraction-class 6	$f_{80}(s)$	−1.18	85
Axle mass-std. deviation	$f_{37}(s)$	65.18	9	Overloaded vehicles fraction-class 7	$f_{81}(s)$	NA	NA
Axle mass-minimum	$f_{38}(s)$	0.04	77	Extremely loaded vehicles following one another	$f_{82}(s)$	0.31	63
Axle mass-maximum	$f_{39}(s)$	0.42	60	Axle mass limit ratio–mean	$f_{83}(s)$	10.36	27
Axle imbalance ratio-mean	$f_{40}(s)$	1.97	43	Axle mass limit ratio-std. deviation	$f_{84}(s)$	63.35	10
Axle imbalance ratio-std. deviation	$f_{41}(s)$	0.71	54	Axle mass limit ratio-minimum	$f_{85}(s)$	0.19	69
Axle imbalance ratio-minimum	$f_{42}(s)$	0.63	55	Axle mass limit ratio-maximum	$f_{86}(s)$	2.55	42
Axle imbalance ratio-maximum	$f_{43}(s)$	0.59	56	Overloaded axles fraction	$f_{87}(s)$	3.66	40
Interaxle-mean	$f_{44}(s)$	7.16	34	Overloaded adjacent axles fraction	$f_{88}(s)$	0.16	72

Note: * the colour of the background is graduated from the most important predictor (red) to the least (white), ** NA indicates that the PFI cannot be calculated because the factor is constant across timeslots.

ARTICLE IN PRESS

J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

13

The permutation feature importance of each frequency predictor (PFI_{f_i}) in the best fit model is shown in Table 5.

As a general rule, factors related to vehicular characteristics, the interaction between vehicular and bridge characteristics, and compliance with TC make the greatest contributions to explaining the frequency of design overloading events. Indeed, several predictors belonging to these groups showed a high permutation feature importance.

Focusing on the 10 most important predictors separately, the following considerations were deemed important.

Normalised GVM-mean (f_{48}) is the predictor that gives the greatest contribution to explaining the frequency of design load overcoming events. This predictor (which belongs to the interaction between vehicular and bridge characteristics factors) surpasses three factors of the simple GVM of each passing vehicle (i.e., GVM-mean (f_{23}), GVM-std. deviation (f_{24})
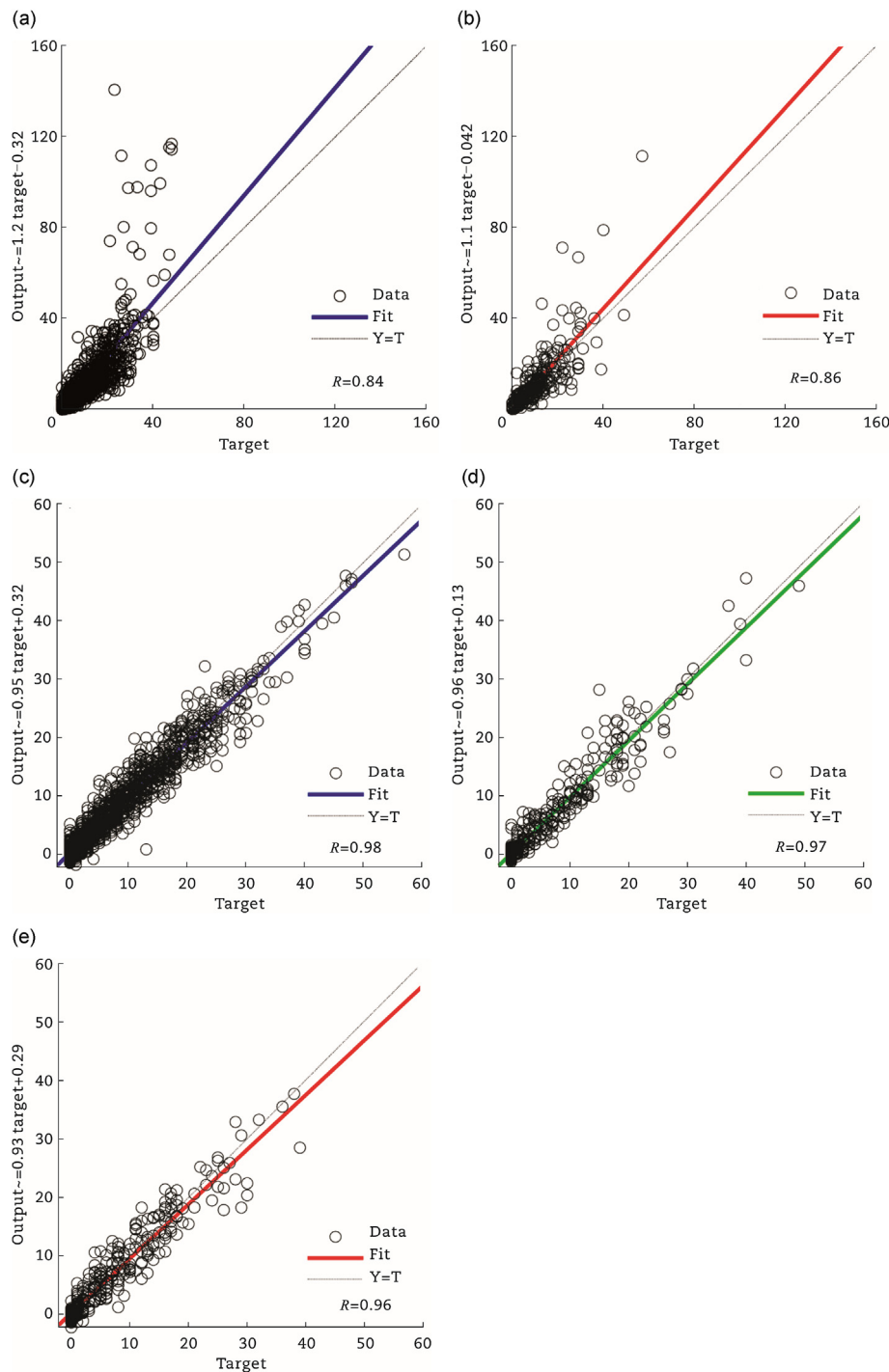

Fig. 5 – Regression plot (RP). (a) GLRM on TR dataset. (b) GLRM on TE dataset. (c) ANNM on TR dataset. (d) ANNM on VA dataset. (e) ANNM on TE dataset.

ARTICLE IN PRESS

14   J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

and GVM-maximum (f_{26})), ranked second, fourth and fifth, respectively. This is an interesting outcome since it suggests that the load acting on the bridge deck is more closely related to the interaction between the GVM and the bridge span length-vehicle length ratio rather than to the simple GVM (which belongs to vehicular characteristics factors). It can be explained considering that if the vehicular length is less than the bridge span, there will be at least a temporal frame in which the overall GVM will act on the deck. Conversely, only a fraction of the GVM will be on the deck at every timestamp. Whereas if the simplifying hypothesis of a GVM uniformly distributed along the vehicular length is assumed, this fraction will therefore be proportional to the bridge span length-vehicular length ratio.

Nevertheless, the significance of vehicular characteristic factors in frequency prediction is endorsed by 1) the third and seventh positions achieved by the two factors related to vehicular length parameters (i.e., length-mean (f_{27}) and length-std. deviation (f_{28}), respectively), and 2) the ninth position reached by a factor related to mass acting on each vehicular axle (i.e., axle mass-std. deviation (f_{37})).

Finally, factors associated with compliance with TC prescriptions also have a relevant impact on the frequency of design overloading events: a predictor of the presence of overloaded vehicles (i.e., GVM limit ratio-mean-all classes (f_{56})) and one of the detections of overladed axles (i.e., axle mass limit ratio-std. deviation (f_{84})) ranked at the sixth and tenth position, respectively.

3.4. Performance evaluation between GLRM and ANNM

Once the GLRM and ANNM were estimated, their fitting and prediction performances were compared to find the best modelling approach. As for the I comparison strategy, the RPs associated with the models were graphed (Fig. 5) and the correlation coefficient (R) among target and predicted frequency was computed (Table 6). In Fig. 5, the graphs are arranged by putting the models in the rows and the datasets in the columns. GLRM has only two datasets, therefore, only two graphs are presented.

Fig. 5 clearly indicates that the ANNM had a greater fitting and predictive capacity than the GLRM. Specifically, several relatively high residuals corresponding to the greater target frequencies were observed on both TR and TE subsets for the GLRM. Indeed, many ($H(s), \tilde{H}(s)$) points fell far away from the first quadrant bisector. Conversely, in the ANNM the ($H(s),$ $\tilde{H}(s)$) points feel closer to the first quadrant bisector on all TR, VA and TE subsets, implying relatively lower residuals. The greater performance of the ANNM was confirmed by the values obtained for the correlation coefficient as shown in Table 6. Indeed, the R metric of the GLRM was always under 0.9 for all subsets (TR and TE). Conversely, the R coefficient of the ANNM was close to 1 on all subsets (TR, VAL and TE). Specifically, a high R on the test dataset would be a fundamental result because TE had no effect on the training process and then provided an independent measure of the ANNM predictive performance.

As for the II comparison strategy, MoEs were computed for both GLRM and ANNM and on all TR, VA and TE subsets. The results in Table 6 indicate that the ANNM outperformed the GLRM both on fitting and forecasting tasks. Indeed, as for the fitting task, the ANNM performed better than the GLRM at fitting training data as evidenced by the fewer errors discovered on TR. Similarly, for the forecasting task, the ANNM's reduced errors on TE showed that it could more accurately anticipate the frequency of upcoming design load overcoming events than the GLRM.

The π ratio was computed for both GLRM and ANNM and on all TR, VA and TE subsets for the III comparison strategy. The findings in Table 6 demonstrated that, depending on whether TE or TR subset was considered, the GLRM overstated the overall number of overloading events by a percentage ranging from roughly 10% to 13%. Conversely, a contrasting effect was observed for the ANNM. Indeed, the ANNM slightly overestimated the number of overloading events by about 1% on TR subset, while it slightly underestimated the number of overcoming events by about 1% and 2% on VA and TE subsets, respectively. Nevertheless, these outcomes suggested that globally, the ANNM had a better fitting and predictive performance than the GLRM: although a moderate overappraisal in the number of overcoming events can be considered as conservative, an excessive π ratio such as those observed for the GLRM should be avoided to prevent unnecessary (and costly) traffic management actions.

4. Conclusions

Truck weights frequently exceed legal load limits, occasionally leading to road bridge failures. Hence, setting models to forecast the frequency of the events in which the design traffic

Parameter	GLRM		ANNM		
	Training	Test	Training	Validation	Test
I Comparison strategy					
Correlation coefficient related to RPs (R)	0.84	0.86	0.98	0.97	0.96
II Comparison strategy					
Mean absolute error (MAE) (events/timeslot)	2.23	2.24	1.13	1.31	1.37
Root mean squared error (RMSE) (events/timeslot)	6.52	5.56	1.82	2.11	2.13
Coefficient of variation (CoV)	1.03	0.91	0.32	0.33	0.39
III Comparison strategy					
Ratio among the total number of predicted and observed design load overcoming events (π)	1.13	1.10	1.01	0.99	0.98

Table 6 – Comparison of the fitting and prediction performances of GLRM and ANNM.

ARTICLE IN PRESS

J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx   **15**

loads of bridges are exceeded is an addition to RAM safety management. Indeed, the frequency of design overloading events is a key component of the methodologies aimed at assessing the risk related to traffic load hazard on bridges. While econometric models (EMs) based on generalised linear regressions (GLRMs) were seldomly employed to predict the frequency of such events, according to the best authors' knowledge, machine learning models (MLMs) based on artificial neural networks (ANNMs) had never been investigated in this context.

This study contributes to the literature as follows.

- It introduces a frequency-based metric of traffic overloading on road bridges that could be a driver of bridge failure probability. This is an addition to the current literature, which has investigated this probability using RAMs.
- It specifies, calibrates, and validates a GLRM and an ANNM, for the prediction of the frequency of traffic overloading on road bridges, according to a list of exposure, bridge side, temporal context, and traffic load hazard frequency predictors. Next, it shows the related effects and importance of both models.
- It compares the performance of the models by determining regression plots (RPs), some measures of errors (MoEs) and the ratio among the observed vs predicted number of events.

A real-world application was performed considering a large dataset of 2 million records collected by a weigh-in-motion (WIM) station on a bridge along a main road near Brescia (Italy). The results showed that vehicular characteristics, interaction between vehicular and bridge characteristics, and compliance with traffic code (TC) prescriptions have the greatest importance in explaining the frequency of overloading events. Moreover, the results indicated that the ANNM outperformed the GLRM. Indeed, a greater correlation coefficient among target and predicted frequencies, lower MoEs and a closer to unity ratio between the sum of the observed and predicted number of design load overcoming events were found for the ANNM.

The relevant implications of this research are shown as follows.

- The availability of frequency models with a greater predictive capacity than the current state-of-the-art could enhance the consistency of predictions. For instance, as frequency is an important risk component, the higher its predictive capacity, the more effective the risk prediction.
- Road authorities can implement more efficient traffic management strategies to improve the safety of bridges against traffic load hazards. For instance, with additional technologies (e.g., cloud computing platforms, traffic lights, and variable message signals), strategies could reroute groups of vehicles that pose a high risk of overloading a specific bridge. Particularly, since vehicles that break TC mass limits have been shown to increase the overloading frequency, the simultaneous presence of non-compliant vehicles on a bridge should be avoided by requiring them to leave the road at an upstream exit.

This study also suggests new directions for further investigation. First, the effect of a greater ability to predict frequency over existing methodologies for assessing traffic load hazards on bridges should be investigated (Ventura et al., 2024). Second, incorporating new variables based on data collected by additional sensors (e.g., accelerometers, strain gauges, intelligent traffic cameras) could further boost the performance of the frequency prediction models. Finally, this study used the load value as a straightforward criterion to evaluate the occurrence of overloading events as a driver of bridge failure probability. Since the traffic load effect directly influences bridge safety, a specific assessment of bridge safety involving an accurate evaluation of the structural response should be required. This is because for the same total applied load, the demand that this induces in terms of internal actions on the structural elements can be very dissimilar depending on the different load patterns. This specific analysis will help better individuate limit states of overloading events and can be applied in future studies.

Conflict of interest

The authors do not have any conflict of interest with other entities or researchers.

Appendix

Table A1 − Frequency predictors and response variable. List, definition, unit of measure, type, and some descriptive statistics. An earlier (less complete) version of this table was proposed in Ventura et al. (2023a).

Name of the frequency predictor	Symbol	Definition	Unit of measure	Type	Minimum*	Maximum*	Mean*	Standard. deviation*
Hourly flow (exposure factor)	$f_1(s) = E(s)$	Ratio among the number of vehicles passing on the bridge during $T(s)$ and the duration of $T(s)$ expressed in hours	veh/h	Continuous	1.00	953.00	335.68	240.16
Lane width	$f_2(s)$	Width of the monitored bridge lane	m	Continuous	3.00	3.00	3.00	0.00
Span length	$f_3(s)$	Length of the monitored bridge span	m	Continuous	23.50	23.50	23.50	0.00
Type of day-weekend wrt weekday	$f_4(s)$	1 if $T(s)$ falls on Saturday or Sunday; 0 otherwise	−	Binary	0.00	1.000	0.280	0.449
Lighting-daytime wrt night-time	$f_5(s)$	1 if $T(s)$ falls from 7:00 to 19:00; 0 otherwise	−	Binary	0.00	1.000	0.547	0.498
Type of hour-peak hour wrt off-peak hour	$f_6(s)$	1 if $T(s)$ falls from 7:00 to 9:00 or from 16:00 to 18:00; 0 otherwise	−	Binary	0.00	1.000	0.168	0.374
Vehicular fraction-class 1	$f_7(s)$	Fraction of vehicles in class 1 computed on $T(s)$	−	Continuous	38.27%	100.00%	91.89%	23.43%
Vehicular fraction-class 2	$f_8(s)$	Fraction of vehicles in class 2 computed on $T(s)$	−	Continuous	0	9.85%	0.42%	0.72%
Vehicular fraction-class 3	$f_9(s)$	Fraction of vehicles in class 3 computed on $T(s)$	−	Continuous	0	58.21%	6.75%	9.21%
Vehicular fraction-class 4	$f_{10}(s)$	Fraction of vehicles in class 4 computed on $T(s)$	−	Continuous	0	16.67%	0.65%	1.46%
Vehicular fraction-class 5	$f_{11}(s)$	Fraction of vehicles in class 5 computed on $T(s)$	−	Continuous	0	16.67%	0.28%	0.70%
Vehicular fraction-class 6	$f_{12}(s)$	Fraction of vehicles in class 6 computed on $T(s)$	−	Continuous	0	0.49%	0	0.02%
Vehicular fraction-class 7	$f_{13}(s)$	Fraction of vehicles in class 7 computed on $T(s)$	−	Continuous	0	12.50%	0	0.24%
Speed-mean	$f_{14}(s)$	Mean of the vehicular speed computed on $T(s)$	km/h	Continuous	12.0	81.6	73.4	4.6
Speed-std. deviation	$f_{15}(s)$	Std. deviation of the vehicular speed computed on $T(s)$	km/h	Continuous	0.0	25.1	7.0	1.1
Speed-minimum	$f_{16}(s)$	Minimum of the vehicular speed computed on $T(s)$	km/h	Continuous	5.0	74.0	54.1	8.1
Speed-maximum	$f_{17}(s)$	Maximum of the vehicular speed computed on $T(s)$	km/h	Continuous	12.0	90.0	88.9	2.2
Headway-mean	$f_{18}(s)$	Mean of the temporal distance between consecutive vehicles computed on $T(s)$	s	Continuous	3.8	4443.9	55.4	181.6
Headway-std. deviation	$f_{19}(s)$	Std. deviation of the temporal distance between consecutive vehicles computed on $T(s)$	s	Continuous	0.0	26,100.1	90.5	809.6
Headway-minimum	$f_{20}(s)$	Minimum of the temporal distance between consecutive vehicles computed on $T(s)$	s	Continuous	0.5	754.6	3.4	20.3
Headway-maximum	$f_{21}(s)$	Maximum of the temporal distance between consecutive vehicles computed on $T(s)$	s	Continuous	19.6	165,095.4**	437.9	4490.2
Congestion condition-free flow	$f_{22}(s)$	1 if no headway lower than 5 s occurs during $T(s)$; 0 otherwise	−	Binary	0.000	1.000	0.944	0.229
GVM-mean	$f_{23}(s)$	Mean of the GVM computed on $T(s)$	kg	Continuous	1588	28,378	5339	4490
GVM-std. deviation	$f_{24}(s)$	Std. deviation of the GVM computed on $T(s)$	kg	Continuous	0	31,018	8855	6823
GVM-minimum	$f_{25}(s)$	Minimum of the GVM computed on $T(s)$	kg	Continuous	1000	2400	1228	87
GVM-maximum	$f_{26}(s)$	Maximum of the GVM computed on $T(s)$	kg	Continuous	2000	122,600	52,992	36,049
Length-mean	$f_{27}(s)$	Mean of the vehicular length computed on $T(s)$	m	Continuous	2.52	8.39	3.38	0.94
Length-std. deviation	$f_{28}(s)$	Std. deviation of the vehicular length computed on $T(s)$	m	Continuous	0.00	6.40	1.86	1.30
Length-minimum	$f_{29}(s)$	Minimum of the vehicular length computed on $T(s)$	m	Continuous	2.20	3.35	2.27	0.07
Length-maximum	$f_{30}(s)$	Maximum of the vehicular length computed on $T(s)$	m	Continuous	2.67	35.95	12.91	6.10

ARTICLE IN PRESS

J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

Axle number-mean	$f_{31}(s)$	Mean of the axle number computed on $T(s)$	—	Continuous	2.00	3.93	2.25	0.32
Axle number-std. deviation	$f_{32}(s)$	Std. deviation of the axle number during $T(s)$	—	Continuous	0.00	2.51	0.63	0.49
Axle number-minimum	$f_{33}(s)$	Minimum of the axle number computed on $T(s)$	—	Discrete	2	2	2	0
Axle number-maximum	$f_{34}(s)$	Maximum of the axle number computed on $T(s)$	—	Discrete	2.00	10.00	5.64	2.44
Axle typology-double wheels axle fraction	$f_{35}(s)$	Fraction of axles with double wheels computed on $T(s)$	—	Continuous	0	47.62%	5.42%	5.73%
Axle mass-mean	$f_{36}(s)$	Mean of the axle mass computed on $T(s)$	kg	Continuous	809	7544	2184	1379
Axle mass-std. deviation	$f_{37}(s)$	Std. deviation of the axle mass computed on $T(s)$	kg	Continuous	90	5006	2235	1476
Axle mass-minimum	$f_{38}(s)$	Minimum of the axle mass computed on $T(s)$	kg	Continuous	500	1093	515	32
Axle mass-maximum	$f_{39}(s)$	Maximum of the axle mass computed on $T(s)$	kg	Continuous	1077	19,783	11,226	5691
Axle imbalance ratio-mean	$f_{40}(s)$	Mean of the ratio between the masses acting on left and right wheels computed on $T(s)$	—	Continuous	0.764	1.279	1.029	0.076
Axle imbalance ratio-std. deviation	$f_{41}(s)$	Std. deviation of the ratio between the masses acting on left and right wheels computed on $T(s)$	—	Continuous	0.072	0.297	0.148	0.026
Axle imbalance ratio-minimum	$f_{42}(s)$	Minimum of the ratio between the masses acting on the left and right wheels computed on $T(s)$	—	Continuous	0.500	1.015	0.673	0.103
Axle imbalance ratio-maximum	$f_{43}(s)$	Maximum of the ratio between the masses acting on the left and right wheels computed on $T(s)$	—	Continuous	0.918	2.000	1.686	0.182
Interaxle-mean	$f_{44}(s)$	Mean of the interaxes computed on $T(s)$	m	Continuous	2.47	3.59	2.70	0.08
Interaxle-std. deviation	$f_{45}(s)$	Std. deviation of the interaxes computed on $T(s)$	m	Continuous	0.00	1.66	0.69	0.37
Interaxle-minimum	$f_{46}(s)$	Minimum of the interaxes computed on $T(s)$	m	Continuous	0.44	3.35	1.48	0.49
Interaxle-maximum	$f_{47}(s)$	Maximum of the interaxes computed on $T(s)$	m	Continuous	2.67	18.39	5.93	1.84
Normalized GVM-mean	$f_{48}(s)$	Mean of the GVM, normalized with the ratio of bridge span length to vehicle length computed on $T(s)$	kg	Continuous	1588	28,378	5339	4490
Normalized GVM-std. deviation	$f_{49}(s)$	Std. deviation of the GVM, normalized with the ratio of bridge span length to vehicle length computed on $T(s)$	kg	Continuous	0	31,018	8854	623
Normalized GVM-minimum	$f_{50}(s)$	Minimum of the GVM, normalized with the ratio of bridge span length to vehicle length computed on $T(s)$	kg	Continuous	1000.00	2400.00	1227.60	86.55
Normalized GVM-maximum	$f_{51}(s)$	Maximum of the GVM, normalized with the ratio of bridge span length to vehicle length computed on $T(s)$	kg	Continuous	2000.00	122,600.00	52,976.49	36,044.29
GVM-length ratio-mean	$f_{52}(s)$	Mean of the ratio of GVM to vehicle length computed on $T(s)$	km/h	Continuous	602.09	2822.23	978.71	351.13
GVM-length ratio-std. deviation	$f_{53}(s)$	Std. deviation of the ratio of GVM to vehicle length computed on computed on $T(s)$	km/h	Continuous	0.00	2063.99	747.52	545.88
GVM-length ratio-minimum	$f_{54}(s)$	Minimum of the ratio of GVM to vehicle length computed on computed on $T(s)$	km/h	Continuous	149.53	791.37	476.41	59.93
GVM-length ratio-maximum	$f_{55}(s)$	Maximum of the ratio of GVM to vehicle length computed on computed on $T(s)$	km/h	Continuous	716.42	9957.63	4855.51	2634.51
GVM limit ratio-mean-all classes	$f_{56}(s)$	Mean of the ratio of GVM to TC limit computed on $T(s)$-all classes	—	Continuous	0.454	0.881	0.585	0.061
GVM limit ratio-std. deviation-all classes	$f_{57}(s)$	Std. deviation of the ratio of GVM to TC limit computed on $T(s)$-all classes	—	Continuous	0.000	0.501	0.173	0.061
GVM limit ratio-minimum-all classes	$f_{58}(s)$	Minimum of the ratio of GVM to TC limit computed on $T(s)$-all classes	—	Continuous	0.067	0.686	0.289	0.106

(continued on next page)

Table A1 — *(continued)*

Name of the frequency predictor	Symbol	Definition	Unit of measure	Type	Minimum*	Maximum*	Mean*	Standard. deviation*
GVM limit ratio-maximum-all classes	$f_{59}(s)$	Maximum of the ratio of GVM to TC limit computed on $T(s)$-all classes	–	Continuous	0.571	4.457	1.444	0.596
GVM limit ratio-mean-class 1	$f_{60}(s)$	Mean of the ratio of GVM to TC limit computed on $T(s)$-class 1	–	Continuous	0.444	0.771	0.547	0.030
GVM limit ratio-mean-class 2	$f_{61}(s)$	Mean of the ratio of GVM to TC limit computed on $T(s)$-class 2	–	Continuous	0.144	1.872	1.108	0.247
GVM limit ratio-mean-class 3	$f_{62}(s)$	Mean of the ratio of GVM to TC limit computed on $T(s)$-class 3	–	Continuous	0.073	1.339	0.919	0.210
GVM limit ratio-mean-class 4	$f_{63}(s)$	Mean of the ratio of GVM to TC limit computed on $T(s)$-class 4	–	Continuous	0.109	1.386	0.934	0.159
GVM limit ratio-mean-class 5	$f_{64}(s)$	Mean of the ratio of GVM to TC limit computed on $T(s)$-class 5	–	Continuous	0.491	2.559	1.868	0.400
GVM limit ratio-mean-class 6	$f_{65}(s)$	Mean of the ratio of GVM to TC limit computed on $T(s)$-class 6	–	Continuous	0.223	1.369	0.612	0.232
GVM limit ratio-mean-class 7	$f_{66}(s)$	Mean of the ratio of GVM to TC limit computed on $T(s)$-class 7	–	Continuous	0.532	0.800	0.626	0.119
GVM limit ratio-maximum-class 1	$f_{67}(s)$	Maximum of the ratio of GVM to TC limit computed on $T(s)$-class 1	–	Continuous	0.514	4.457	1.029	0.326
GVM limit ratio-maximum-class 2	$f_{68}(s)$	Maximum of the ratio of GVM to TC limit computed on $T(s)$-class 2	–	Continuous	0.144	2.000	1.245	0.238
GVM limit ratio-maximum-class 3	$f_{69}(s)$	Maximum of the ratio of GVM to TC limit computed on $T(s)$-class 3	–	Continuous	0.073	1.727	1.095	0.276
GVM limit ratio-maximum-class 4	$f_{70}(s)$	Maximum of the ratio of GVM to TC limit computed on $T(s)$-class 4	–	Continuous	0.109	1.564	1.022	0.165
GVM limit ratio-maximum-class 5	$f_{71}(s)$	Maximum of the ratio of GVM to TC limit computed on $T(s)$-class 5	–	Continuous	0.491	2.786	2.057	0.405
GVM limit ratio-maximum-class 6	$f_{72}(s)$	Maximum of the ratio of GVM to TC limit computed on $T(s)$-class 6	–	Continuous	0.223	1.369	0.613	0.232
GVM limit ratio-maximum-class 7	$f_{73}(s)$	Maximum of the ratio of GVM to TC limit computed on $T(s)$-class 7	–	Continuous	0.532	0.800	0.626	0.119
Overloaded vehicles fraction-all classes	$f_{74}(s)$	Fraction of vehicles with a GVM above TC limit computed on $T(s)$-all classes	–	Continuous	0	54.10%	4.73%	6.68%
Overloaded vehicles fraction-class 1	$f_{75}(s)$	Fraction of vehicles with a GVM above TC limit computed on $T(s)$-class 1	–	Continuous	0	33.33%	0.28%	1.08%
Overloaded vehicles fraction-class 2	$f_{76}(s)$	Fraction of vehicles with a GVM above TC limit computed on $T(s)$-class 2	–	Continuous	0	100.00%	82.59%	29.56%
Overloaded vehicles fraction-class 3	$f_{77}(s)$	Fraction of vehicles with a GVM above TC limit computed on $T(s)$-class 3	–	Continuous	0	100.00%	44.98%	28.83%
Overloaded vehicles fraction-class 4	$f_{78}(s)$	Fraction of vehicles with a GVM above TC limit computed on $T(s)$-class 4	–	Continuous	0	100.00%	42.07%	35.38%
Overloaded vehicles fraction-class 5	$f_{79}(s)$	Fraction of vehicles with a GVM above TC limit computed on $T(s)$-class 5	–	Continuous	0	100.00%	89.99%	23.48%
Overloaded vehicles fraction-class 6	$f_{80}(s)$	Fraction of vehicles with a GVM above TC limit computed on $T(s)$-class 6	–	Continuous	0	100.00%	4.55%	21.32%

ARTICLE IN PRESS

J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

19

Description	Symbol		Type				
Overloaded vehicles fraction-class 7	$f_{81}(s)$	—	Continuous	Fraction of vehicles with a GVM above TC limit computed on T(s)-class 7	0	0	0	0
Extremely loaded vehicles following one another	$f_{82}(s)$	—	Binary	1 if at least two vehicles with a GVM above 44,000 kg, following one another with a headway lower than 5 s, occurs during T(s); 0 otherwise	0.440	0.262	1.000	0.000
Axle mass limit ratio-mean	$f_{83}(s)$	—	Continuous	Mean of the ratio of axle mass to TC limit computed on T(s)	0.115	0.182	0.629	0.067
Axle mass limit ratio-std. deviation	$f_{84}(s)$	—	Continuous	Std. dev. of the ratio of axle mass to TC limit computed on T(s)	0.123	0.186	0.417	0.007
Axle mass limit ratio-minimum	$f_{85}(s)$	—	Continuous	Minimum of the ratio of axle mass to TC limit computed on T(s)	0.003	0.043	0.091	0.042
Axle mass limit ratio-maximum	$f_{86}(s)$	—	Continuous	Maximum of the ratio of axle mass to TC limit computed on T(s)	0.474	0.936	1.649	0.090
Overloaded axles fraction	$f_{87}(s)$	—	Continuous	Fraction of axles with a mass above TC limit computed on T(s)	2.39%	1.99%	13.78%	0
Overloaded adjacent axles fraction	$f_{88}(s)$	—	Continuous	Fraction of adjacent pairs of axles with a total mass above TC limit computed on T(s)	18.48%	19.06%	100.00%	0
Frequency of design load overcoming events	$H(s)$	—	Continuous	Number of times the traffic load exceeded one or more design load thresholds during T(s)	8.067	5.303	57.000	0.000

Note: * The italic font denotes that the indicator was calculated only during timeslots when at least one vehicle of the considered class was observed. ** This result is biased since the WIM system was down for around 46 h. Hence, the first vehicle passing following the WIM device restart was allocated a headway of nearly 46 h.

REFERENCES

Alqatawna, A., Rivas Alvarez, A.M., Garcia-Moreno, S.S.C., 2021. Comparison of multivariate regression models and artificial neural networks for prediction highway traffic accidents in Spain: a case study. Transportation Research Procedia 58, 277–284.

American Society for Testing and Materials (ASTM), 2017. Standard Specification for Highway Weigh-in-Motion (WIM) Systems with User Requirements and Test Methods. ASTM E1318-09. ASTM International, West Conshohocken.

Barabino, B., Bonera, M., Maternini, G., et al., 2021. Bus crash risk evaluation: an adjusted framework and its application in a real network. Accident Analysis and Prevention 159, 106258.

Barabino, B., Di Francesco, M., Ventura, R., 2023. Evaluating fare evasion risk in bus transit networks. Transportation Research Interdisciplinary Perspectives 20, 100854.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Brownlee, J., 2019. Loss and Loss Functions for Training Deep Learning Neural Networks (WWW Document). Available at: https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/, (Accessed 18 June 2022).

Chang, L., 2005. Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. Safety Science 43 (8), 541–557.

De Aloe, M., Ventura, R., Bonera, M., et al., 2022. Applying cost–benefit analysis to the economic evaluation of a tram-train system: evidence from Brescia (Italy). Research in Transportation Business and Management 47, 100916.

European Committee for Standardization (CEN), 2002. Eurocode–Basis of Structural Design. European Committee for Standardization, Brussels.

European Committee for Standardization (CEN), 2003. Eurocode 1: Actions on Structures–Part 2: Traffic Loads on Bridges. European Committee for Standardization, Brussels.

Faccin, C., Zavanella, L., Savoldi, E., et al., 2011. Piano del Traffico della Viabilitá Extraurbana (PTVE). Available: https://www.provincia.brescia.it/cittadino/viabilita-e-strade/piano-del-traffico-della-viabilita-extraurbana-ptve, (Accessed 18 July 2022).

Fertlitsch, A., 2021. Deep Learning Patterns and Practices. Manning Publications, Greenwich.

Fiorillo, G., Ghosn, M., 2014. Procedure for statistical categorization of overweight vehicles in a wim database. Journal of Transportation Engineering 140 (5), 1–12.

Fiorillo, G., Ghosn, M., 2018. Fragility analysis of bridges due to overweight traffic load. Structure and Infrastructure Engineering 14 (5), 619–633.

Fiorillo, G., Ghosn, M., 2019. Risk-based importance factors for bridge networks under highway traffic loads. Structure and Infrastructure Engineering 15 (1/3), 113–126.

Fiorillo, G., Ghosn, M., 2022. Risk-based life-cycle analysis of highway bridge networks under budget constraints. Structure and Infrastructure Engineering 18 (10/12), 1457–1471.

Flach, P., 2012. Machine Learning: the Art and Science of Algorithms that Make Sense of Data. Cambridge University Press, Cambridge.

Guo, L., Wei, J., Cui, S., 2011. Study of load spectra of the Jiaozhou Bay Expressway based on MEPDG. Applied Mechanics and Materials 97–98, 402–407.

Hegde, J., Rokseth, B., 2020. Applications of machine learning methods for engineering risk assessment–a review. Safety Science 122, 104492.

Hernandez, S.V., Tok, A., Ritchie, S.G., 2016. Integration of Weigh-in-Motion (WIM) and inductive signature data for truck body

ARTICLE IN PRESS

20   J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

classification. Transportation Research Part C: Emerging Technologies 68, 1—21.

Huang, P., Wang, J., Xu, X., et al., 2022. Improved multi-lane traffic flow simulation based on weigh-in-motion data. Measurement (Lond) 188, 110408.

Hwang, E.S., Kim, D.Y., 2019. Live load model for long span steel cable bridges considering traffic congestion scenarios. International Journal of Steel Structures 19, 1996—2009.

Ian, G., Yoshua, B., Courville, A., 2016. Deep Learning. The MIT Press, Cambridge.

Iatsko, O., Nowak, A.S., 2021. Revisited live load for simple-span bridges. Journal of Bridge Engineering 26 (1), 04020114.

Kim, J., Song, J., 2019. A comprehensive probabilistic model of traffic loads based on weigh-in-motion data for applications to bridge structures. KSCE Journal of Civil Engineering 23, 3628—3643.

Kim, J., Song, J., 2021. Bayesian updating methodology for probabilistic model of bridge traffic loads using in-service data of traffic environment. Structure and Infrastructure Engineering 19 (1), 77—92.

Liao, C., 2014. Generating reliable freight performance measures with truck GPS data. Transportation Research Record 2410, 21—30.

Liu, H., Davidson, R.A., Rosowsky, D.V., et al., 2005. Negative binomial regression of electric power outages in hurricanes. Journal of Infrastructure Systems 11 (4), 258—267.

Lou, P., Nassif, H., Su, D., et al., 2016. Effect of overweight trucks on bridge deck deterioration based on weigh-in-motion data. Transportation Research Record 2592, 86—97.

Lu, N., Liu, Y., Beer, M., 2018. Extrapolation of extreme traffic load effects on a cable-stayed bridge based on weigh-in-motion measurements. International Journal of Reliability and Safety 12, 69—85.

Lu, N., Ma, Y., Liu, Y., 2019. Evaluating probabilistic traffic load effects on large bridges using long-term traffic monitoring data. Sensors 19, 5056.

Maljaars, J., 2020. Evaluation of traffic load models for fatigue verification of European road bridges. Engineering Structures 225, 111326.

Mandić Ivanković, A., Skokandić, D., Žnidarič, A., et al., 2019. Bridge performance indicators based on traffic load monitoring. Structure and Infrastructure Engineering 15, 899—911.

Martinelli, V., Ventura, R., Bonera, M., et al., 2022. Effects of urban road environment on vehicular speed. Evidence from Brescia (Italy). Transportation Research Procedia 60, 592—599.

Marucci-Wellman, H.R., Corns, H.L., Lehto, M.R., 2017. Classifying injury narratives of large administrative databases for surveillance—a practical approach combining machine learning ensembles and human review. Accident Analysis Prevention 98, 359—371.

Micu, E.A., Malekjafarian, A., OBrien, E.J., et al., 2019. Evaluation of the extreme traffic load effects on the Forth Road Bridge using image analysis of traffic data. Advances in Engineering Software 137, 102711.

Ministero delle Infrastrutture e dei Trasporti, 2020. Linee Guida per la Classificazione e Gestione del Rischio, la Valutazione della Sicurezza ed il Monitoraggio dei Ponti Esistenti. Available at: https://www.mit.gov.it/sites/default/files/media/notizia/2020-05/1_Testo_Linee_Guida_ponti.pdf (Accessed 15 June 2022).

OIML, 2006. OIML R 134-1: Automatic Instruments for Weighing Road Vehicles in Motion and Measuring Axle Loads Part 1: Metrological and Technical Requirements—Tests 2006, 1—81. Available at: https://www.oiml.org/en/files/pdf_r/r134-1-e06.pdf (Accessed 16 June 2022).

Reed, R., Marks II, R.J., 1999. Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks. The MIT Press, Cambridge.

Ren, J., Thompson, R.G., Zhang, L., 2019. Impact of payload spectra of heavy vehicles on pavement based on weigh-in-motion data. Journal of Transportation Engineering Part B: Pavements 145 (2), 04019005.

Road Traffic Safety Management Systems, 2012. ISO 39001: Road Traffic Safety (RTS) Management Systems: Requirements with Guidance for Use. Road Traffic Safety Management Systems, Beijing.

Roh, H.-J., Sahu, P.K., Sharma, S., et al., 2016. Statistical investigations of snowfall and temperature interaction with passenger car and truck traffic on primary highways in Canada. Journal of Cold Regions Engineering 30 (2), 40150061.

Romero Reyes, I.V., Fedyushkina, I.V., Skvortsov, V.S., et al., 2013. Prediction of progesterone receptor inhibition by high-performance neural network algorithm. International Journal of Mathematical Models and Methods in Applied Sciences 7 (3), 303—310.

Salazar, F., Toledo, M.Á., González, J.M., et al., 2017. Early detection of anomalies in dam performance: a methodology based on boosted regression trees. Structural Control Health Monitoring 24 (11), e2012.

Schmidhuber, J., 2015. Deep Learning in Neural Networks: an Overview. University of Lugano & SUPSI, Galleria.

Schmidt, F., Jacob, B., Domprobst, F., 2016. Investigation of truck weights and dimensions using WIM data. Transportation Research Procedia 14, 811—819.

Shrestha, N., 2020. Detecting multicollinearity in regression analysis. American Journal of Applied Mathematics and Statistics 8 (2), 39—42.

Sramka, M., Slovak, M., Tuckova, J., et al., 2019. Improving clinical refractive results of cataract surgery by machine learning. PeerJ 7, 7202.

Strobl, C., Boulesteix, A.L., Zeileis, A., et al., 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 8, 1471.

Sujon, M., Dai, F., 2021. Application of weigh-in-motion technologies for pavement and bridge response monitoring: state-of-the-art review. Automation in Construction 130, 103844.

Susmita, Ray, 2019. A quick review of machine learning algorithms. In: International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, Faridabad, 2019.

Ventura, R., Barabino, B., Vetturi, D., et al., 2020. Bridge safety analysis based on the function of exceptional vehicle transit speed. The Open Transportation Journal 14, 222—236.

Ventura, R., Barabino, B., Vetturi, D., et al., 2023a. Bridge's vehicular loads characterization through weight-in-motion (WIM) systems. The case study of Brescia. European Transport/Trasporti Europei, https://doi.org/10.48295/et.2023.90.6.

Ventura, R., Barabino, B., Vetturi, D., et al., 2023b. Monitoring vehicles with permits and that are illegally overweight on bridges using weigh-in-motion (WIM) devices: a case study from Brescia. Case Study on Transport Policy 13, 101023.

Ventura, R., Barabino, B., Maternini, G., 2024. Traffic hazard on main road's bridges: real-time managing the risk of design load overcoming events. IEEE Transactions on Intelligent Transportation Systems, https://doi.org/10.1109/TITS.2024.3371265.

Wen, X., Xie, Y., Jiang, L., et al., 2021. Applications of machine learning methods in traffic crash severity modelling: current status and future directions. Transport Review 41 (6), 855—879.

Yannis, G., Antoniou, C., 2005. Integration of weigh-in-motion technologies in road infrastructure management. ITE Journal (Institute of Transportation Engineers) 75 (1), 39—43.

Zeng, Q., Huang, H., Pei, X., et al., 2016. Rule extraction from an optimized neural network for traffic crash frequency modeling. Accident Analysis and Prevention 97, 87—95.

ARTICLE IN PRESS

J. Traffic Transp. Eng. (Engl. Ed.) xxxx; xxx (xxx): xxx

21

Zhang, P., 2010. Advanced Industrial Control Technology. Elsevier, Amsterdam.

Zhang, G., Liu, Y., Liu, J., et al., 2022. Causes and statistical characteristics of bridge failures: a review. Journal of Traffic and Transportation Engineering (English Edition) 9 (3), 388—406.

Zhao, J., Wang, H., Lu, P., 2021. Impact analysis of traffic loading on pavement performance using support vector regression model. International Journal of Pavement Engineering 23 (11), 3716—3728.

Zhou, J., Caprani, C.C., Zhang, L., 2021. On the structural safety of long-span bridges under traffic loadings caused by maintenance works. Engineering Structures 240, 112407.

Zhou, J., Chen, Z., Yi, J., et al., 2020a. Investigation of multi-lane factor models for bridge traffic load effects using multiple lane traffic data. Structures 24, 444—455.

Zhou, J., Hu, C., Zhang, J., et al., 2022. Reliability assessment of existing concrete bridges under the passage of heavy trucks considering bending-shear interaction. Structure and Infrastructure Engineering 218, 108137.

Zhou, J., Liu, Y., Yi, J., 2020b. Effect of uneven multilane truck loading of multigirder bridges on component reliability. Structural Concrete 21 (4), 1644—1661.

Roberto Ventura is an assistant professor in transportation engineering at the University of Brescia (Italy). He received a master's degree in civil engineering in the 2019. In 2023, he earned a PhD in civil and environmental engineering, international cooperation and mathematics (curriculum of urban planning and mobility) at the University of Brescia (Italy). His primary research interests include the safety of the infrastructures subjected to heavy vehicle transit, risk analyses on transportation systems, and micromobility vehicles.
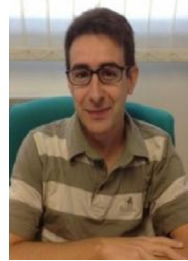
Benedetto Barabino is an associate professor of transportation engineering at the University of Brescia (Italy). He earned his PhD in transportation technology and economics from the University of Palermo in 2007. From the 2019 to the 2022, he was an assistant professor (tenure track) of transportation engineering at the University of Brescia. From 2017 to 2019, he was the head of the Department of Study and Research at CTM in Cagliari (Italy). His research interests include intelligent transportation systems, public transport planning, operations, service quality, bus safety and fare evasion.

Giulio Maternini is a full professor of transportation engineering at the University of Brescia (Italy). Since 2004 he is rector's deputy as a mobility manager at the University of Brescia. From 2010 to 2014, he was the Italian president of Italian Association of Traffic and Transportation (AIIT). Since 2012, he is a remote referee for Italian Ministry of University and Research (MIUR) funding calls. Since 2018, he is an editor of European Transport. Since 2019, he is the director of the Friendly City Study Center (CeSCAM). His research includes transport infrastructures, road design, and safety in transport.