A Non-parametric Bayesian Network for multivariate probabilistic modelling of Weigh-in-Motion System Data

Mendoza-Lugo, Miguel Angel; Morales-Nápoles, Oswaldo; Delgado-Hernández, David Joaquín

DOI
10.1016/j.trip.2022.100552

Publication date
2022

Document Version
Final published version

Published in
Transportation Research Interdisciplinary Perspectives

Transportation Research Interdisciplinary Perspectives 13 (2022) 100552

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.elsevier.com/locate/trip

A Non-parametric Bayesian Network for multivariate probabilistic modelling of Weigh-in-Motion System Data

Miguel Angel Mendoza-Lugo [a,*], Oswaldo Morales-Nápoles [a], David Joaquín Delgado-Hernández [b]

[a] Delft University of Technology, Stevinweg 1, 2628CN Delft, The Netherlands
[b] Autonomous University of the State of Mexico, Cerro de Coatepec, Ciudad Universitaria, 50100, Toluca, Mexico

ARTICLE INFO

Keywords:
Weigh in motion
Bayesian Network
Vehicle loads
Simulation

ABSTRACT

Weigh-in-motion (WIM) systems help to collect data such as vehicular loads, individual axle loads, vehicle type, and number of axles. This is relevant in engineering because traffic load performs an essential function in the design of new bridges and in the reliability assessment of existing ones, in traffic analysis and other areas of engineering. Therefore, when WIM data is not available, computing synthetic WIM observations that adequately approximate the statistical dependence between variables is important. In this paper, WIM measurements from the Netherlands and Brazil were analysed, and a set of non-parametric Bayesian Networks (NPBNs) is presented. This paper significantly improves on previous results by allowing observations of inter-axial distance to be generated, by allowing several sources of data to be used in the modelling and by making software available to researchers and practitioners interested for generating synthetic observations based on the distribution of vehicle type. In particular, statistical models to describe the weight and length of different vehicle types are derived. Three NPBNs were quantified using data from: (i) six WIM locations of the motorway network of the Netherlands, (ii) one WIM location in one city route of Rotterdam, The Netherlands, and (iii) one WIM location of one highway in Araranguá city located in the south of Brazil. Additionally, a Graphical User Interface (GUI) for the six Dutch WIM motorways locations was developed. To illustrate a possible use of the model when WIM data is not available. The GUI was used to compute synthetic WIM observations using data collected through traffic counters gathered in Toluca city in central Mexico, as input. This paper shows that the methodology here presented is widely applicable and depends only on the assessment of vehicle type configuration.

1. Introduction

Vehicle load investigation is essential for the reliability assessment of existing road infrastructure because the exceedance of legal weight limits may cause serious threats to road transport operation and road infrastructure. For example, by increasing the risk of deterioration of pavements and bridges. One approach for describing the traffic flow characteristics is data gathered through Weigh-in-Motion (WIM) systems. WIM is a technology that allows measuring vehicle attributes while the vehicle is travelling at full highway speed. Hence, significant amounts of data such as axle loads, vehicle type and inter-axle distance are collected.

Applications of the data gathered by the Weigh-in-Motion system include reliability assessment of bridges, pavements and, monitoring of overloaded traffic. Consequently, WIM systems are widely used around the world. For instance, countries such as Slovenia, Ireland, and England, operate 216 WIM sites installed in their national road networks (Kreslin et al., 2016). However, WIM systems are expensive in terms of initial capital costs and life cycle maintenance costs. As a result, a large number of countries or regions around the world do not operate any WIM system at present. They are mostly using common techniques such as pneumatic road tubes, some disadvantages of this technique include inaccurate axle counting when trucks volumes are high and absence of vehicle weighing (Crespo-Minguillón and Casas, 1997).

Therefore, computing good artificial WIM data is relevant in road infrastructure safety assessment because an accurate model enhances the reliability estimation. To accurately simulate WIM observations, statistical correlations between the variables need to be modelled. Generally, this can be carried out through empirical factors, linear correlations, and copulas (Crespo-Minguillón and Casas, 1997; Federal Highway Administration's Intelligent Transportation Systems Program Office, 2007; S. et al., 2006; Kim and Song, 2019). Nevertheless, these

* Corresponding author.
E-mail addresses: m.a.mendozalugo@tudelft.nl (M.A. Mendoza-Lugo), o.moralesnapoles@tudelft.nl (O. Morales-Nápoles), david.delgado@uaemex.mx (D.J. Delgado-Hernández).

https://doi.org/10.1016/j.trip.2022.100552
Received 9 September 2021; Received in revised form 10 December 2021; Accepted 29 January 2022
Available online 22 February 2022

studies are focused only on axle loads or provide fixed inter-axle distances.

The aim of this research is to construct a WIM data model of heavy vehicles (total weight above 34 kN) from data obtained in the available WIM stations. In order to generate synthetic data statistically representative of the real observations. The model is based on a type of Bayesian Network (BN) called non-parametric Bayesian Network (NPBN). Unlike previous studies, it provides data that describes all the main variables of a WIM data set i.e. vehicle type, gross vehicle weight, individual axle loads, total vehicle length, and inter-axle distances. Furthermore, to make use of the model more convenient, we developed a graphical user interface (GUI) of the Dutch available motorway WIM locations.

In Section 2, the main definitions and concepts of NPBNs used in the work are presented. In Section 3, we present the framework for generating synthetic WIM data. Observations of six Dutch WIM locations by clustering the vehicle types by vehicle configuration are used. Section 4 present the main results and the validation of the NPBN models. Next, in Section 5, two case studies of the framework are presented: an NPBN quantified with WIM observations collected in Rotterdam city in the Netherlands by classifying vehicle types according to the number of axles, and an NPBN quantified with data collected in Araranguá city in Brazil by clustering vehicle types by classification WIM codes. As will be seen later, the validation of the methodology is performed with the Dutch case (Section 3.1) while the cases in Section 5 are presented as examples of the use of the methodology. Section 6 presents the developed GUI together with one possible application of the model when WIM data is not available. Finally, the conclusions are drawn in Section 7.

2. Non-parametric Bayesian networks

When modelling multivariate probability distributions, Bayesian Networks (BNs) are considered effective tools (Pearl, 1988). BNs are directed acyclic graphs (DAG), consisting of nodes and arcs. The nodes of BNs represent random variables and the arcs represent the probabilistic relations between these variables (Neil et al., 2000). A BN encodes the probability density or mass function on a set of variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ by specifying a set of conditional independence statements in the DAG associated with a set of conditional probability functions (Oswaldo and M., 2015). A major overview of applications of BNs may be found in Marcot (2017), Marcot and Penman (2019) and Weber et al. (2012).

One type of BN which has the advantage of managing hundreds of variables in a rapid manner is the non-parametric Bayesian network (NPBN). The NPBN methodology proposes a technique for dealing with any one-dimensional marginal distribution (as long as it is invertible) while inducing the dependence structure given by the DAG of the BN using one parameter copulas for which zero correlation entails independence (Hanea et al., 2015). The theory of NPBNs introduced in Kurowicka and Cooke (2005) is based around bivariate copulas. Distributions are assigned to the nodes and one (conditional) parameter copulas to the arcs of the DAG (Joe, 1997). Copulas are a class of bivariate distributions whose marginals are uniform (Genest and MacKay, 1986). The copula of two continuous random variables X_i and X_j with $i \neq j$ is the function C such that their joint distribution can be written according to Eq. (1) (Oswaldo and M., 2015). For one parameter copula families, the vector of parameters $\boldsymbol{\theta}$ in Eq. (1) provides a relationship between the copula and measures of association namely rank correlation (r) which assess the strength of the monotonic relationship between variables.

$$F_{X_i,X_j}(X_i, X_j) = C_\theta \left[F_{X_i}(X_i), F_{X_j}(X_j) \right] \tag{1}$$

In an NPBN, different bivariate copulas parameterized by rank correlation so that zero correlation implies independence, can in principle be used for different arcs. For large models, the usage of the Gaussian copula grants computational advantages. In fact, this is the main advantage of the choice of Gaussian copulas in our framework. In Morales-Nápoles and Steenbergen (2014) the authors have provided evidence for accepting the Gaussian copula as a valid underlying model for WIM data when compared with the Gumbel and Clayton copulas. As will be seen later in Section 4, the choice of the Gaussian copula still renders valid results for applications of the proposed model. Hence, to simplify and reduce the joint distribution sampling, the Gaussian copula is assumed. The Gaussian copula, with ρ as a parameter is given by Eq. (2). Where $\boldsymbol{\Phi}_\rho$ is the bivariate standard normal cumulative distribution function with product-moment correlation ρ and $\boldsymbol{\Phi}^{-1}$ the inverse of the one dimensional (1D) standard normal distribution function. A correlation equal to 0 implies independence for the Gaussian copula.

$$C_\rho(u, v) = \boldsymbol{\Phi}_\rho[\boldsymbol{\Phi}^{-1}(u), \boldsymbol{\Phi}^{-1}(v)]; (u, v) \in [0, 1]^2 \tag{2}$$

The dependence measure of interest is the conditional rank correlation due to its relationship with conditional copulas. The conditional rank correlations of X_i and X_j given $X_k = x_k, \ldots, X_z = x_z$ are (Oswaldo and M., 2015):

$$r(X_i, X_j \mid X_k = x_k, \ldots, X_z = x_z) = r(\widetilde{X_i}, \widetilde{X_j}) \tag{3}$$

where $\widetilde{X_i}$ and $\widetilde{X_j}$ have the conditional distribution of $X_i, X_j \mid X_k = x_k, \ldots, X_z = x_z$. Mathematical details can be found in Hanea et al. (2006).

The relationship between the parameter ρ and the Gaussian copula correlation r is given by Eq. (4) (Pearson, 1907). Partial correlations can be computed from correlations with the recursive Eq. (5) (Yule and Kendall, 1951). Partial correlations are equal to conditional correlations for the joint normal distribution. To compute the correlation matrix of the standard normal transformation of $\mathbf{X}$ by using Eq. (5) and the conditional independence statements of the BN. Rank correlation could be converted to partial correlations with Eq. (4).

$$\rho(X, Y) = 2sin\left(\frac{\pi}{6} r(X, Y)\right) \tag{4}$$

$$\rho_{1,2;3,\ldots,m} = \frac{\rho_{1,2;4,\ldots,m} - (\rho_{1,3;4,\ldots,m})(\rho_{2,3;4,\ldots,m})}{\sqrt{(1 - \rho^2_{1,3;4,\ldots,m})(1 - \rho^2_{2,3;4,\ldots,m})}} \tag{5}$$

In an NPBN, the immediate precursor of a node X_i are called parents denoted by $pa(X_i)$. For each variable X_i with m-parents, $X_1 = pa_1(X_i), \ldots, X_k = pa_m(X_i)$, associate the arc $pa_j(X_i) \to X_i$ with the rank correlation. The assignment is empty if $pa(X_i) = \varnothing$.

$$\begin{aligned} &r[X_i, pa_j(X_i)], j = 1 \\ &r[X_i, pa_j(X_i) | pa_1(X_i), \ldots, pa_{j-1}(X_i)], j = 2, \ldots, m \end{aligned} \tag{6}$$

In general, Hanea et al. (2015) show that given: (i) a DAG with m nodes specifying conditional independence relationships in an NPBN; (ii) m random variables $X_1, \ldots, X_m$, assigned to the nodes, with invertible distribution functions $F_1, \ldots, F_m$; (iii) the (non-unique) specification Eq. (6) of conditional rank correlations on the arcs of the NPBN; (iv) a copula realizing all correlations $\in (-1, 1)$ for which correlation 0 entails independence; the joint distribution of the m variables is uniquely determined. The joint distribution satisfies the characteristic factorization Eq. (7) and the conditional rank correlations in Eq. (6) are algebraically independent.

$$f(X_i, \ldots, X_m) = \prod_{i=1}^{n} f_{X_i | pa(X_i)} \tag{7}$$

The following criterion are established in order to read conditional independence statements of the graph: (i) X_1 is not marginally independent of X_3 ($X_1 \not\perp X_3$) i.e. $(X_1) \to (X_2) \to (X_3)$, (ii) X_1 and X_3 are conditionally independent given X_2 ($X_1 \perp X_3 \mid X_2$) i.e. $(X_1) \leftarrow (X_2) \to (X_3)$ and (iii) X_1 and X_3 are not conditionally independent given X_2 ($X_1 \not\perp X_3 \mid X_2$) i.e $(X_1) \to (X_2) \leftarrow (X_3)$. In Pearl (1988), Hanea et al.

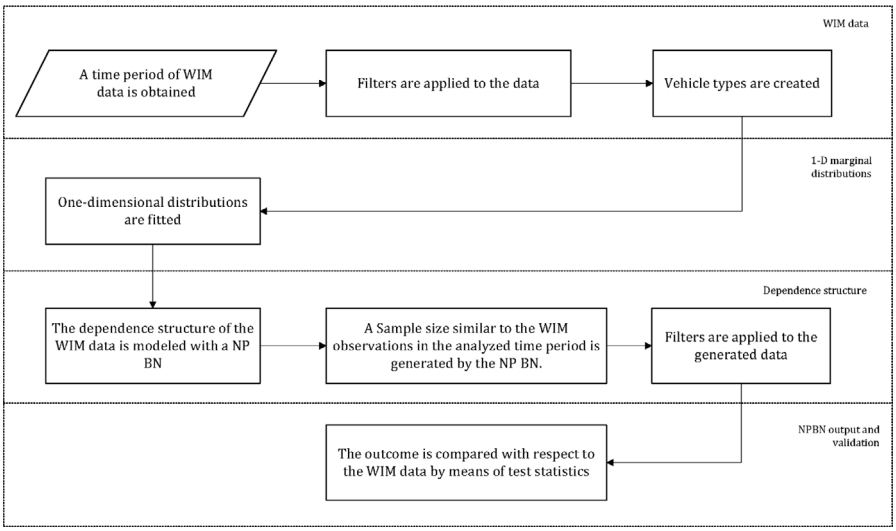

Fig. 1. Framework for build the NPBN and generate WIM samples.

(2006) more details and a complete discussion of the semantics of a BN is presented.

In order to find a given conditional distribution, the conditional distribution on the standard normal transformation of **X** is calculated. Then, by using the inverse distribution function of each one-dimensional (1-D) margin, the margins are transformed back to their original units. For a complete treatment of the NPBN framework the reader is referred to Hanea et al. (2015) and references therein. Now that the concepts of the NPBN used in this work have been introduced, the steps for building the network of interest will be presented.

3. Framework for generating synthetic WIM data

In this section, the methodology to generate synthetic WIM data through Non-parametric Bayesian Networks is presented. Fig. 1 shows the whole framework divided into four categories: WIM data, One-dimensional marginal distributions, dependence structure and NPBN output and validation.

First, a period of WIM observations is obtained. The data collected through WIM system includes type of vehicle, road and direction, vehicle length [cm], total vehicular weight [kN], one observation per axle load [kN], distance from vehicle front to first axle [cm], and one observation per inter-axle distance [cm]. Filters are applied to the data in order to detect and exclude unreasonable measurements or possible errors of the WIM system's vehicle classification algorithm. Then, vehicle types are created either by clustering vehicles based on the number of axles, body configuration (trailer, semi-trailer, bus, single unit, etc.) or vehicle classification according to the classification code of the WIM system (this will be further discussed in Section 3.1). A 1-D probability distribution is fitted for each axle load and inter-axle distances of the created vehicle types (Section 3.2). Next, the dependence structure of the WIM data is modelled through a NPBN in which the nodes represent the individual axle loads and individual inter-axle distances and the arcs represent probabilistic dependence between connected nodes (Section 3.3). Once the model has been adequately quantified, a sample of a size similar to the WIM observations in the analysed period is generated by the model. Then, filters are applied to the samples to leave out non-heavy vehicles and unreasonable values. Finally, the outcome is compared statistically with respect to the WIM data (Section 4). Each step within the framework is detailed next.

3.1. Weigh In Motion data

According to Vrouwenvelder and Waarts (1993), a measurement period of about one month outside public holidays in the Netherlands is representative of the traffic load distribution in highways. Thus, WIM data corresponding to April 2013 for three Dutch locations in both the right (-R) and the left (-L) driving directions, were used. The measurements were taken in highways A12 (km 42) Woerden, A15 (km 92) Gorinchem and A16 (km 41) Gravendeel. Thus, in this paper, when referring for example to data from the A15 in the right direction we will write A15-R and similarly for all other data sources.

The number of observations available ranged from 124 347 in the A15-L to 220 840 in the A16-L. Moreover, previous research shows that WIM observations have an amount of incorrect measurements (Gillmann, 1992; Kentucky Trasportation Center, 2013; Quinley, 2010). In this paper, twenty seven filter criteria were applied as presented in Table A.1, described in Kreslin et al. (2016) and And and Vervuurt (2015). Furthermore, Table 1 shows an overview of the number of WIM registrations that were removed based on the filtering criteria.

Once the filtering procedure is completed, a total of 264 different vehicle codes were observed in the April 2013 WIM data with five main body configurations: Buses (B), Tractor–Semitrailer–Trailer (R), Tractor–Semitrailer (T), Single-unit multi-axle vehicle and/or Single unit multi-axle vehicle–Semitrailer (V) and Others vehicles (O). Fig. A.1 shows an overview of the vehicle codes observed in the WIM measurements. The codes used in the WIM system consist of letters and digits that define the sequence of axle groups. For example, a seven axle vehicle with the configuration Tractor–Semitrailer with one axle at the front of the cabin and two at the rear and the rear semitrailer with a quad is coded as T12O4. Notice that the letter O in the T12O4 code represent the semitrailer unit (*oplegger* in Dutch). Because of the database size, it is not feasible to investigate the complete configuration of axle loads for each vehicle. Therefore, 26 vehicle types were created. These are presented in Table 2, grouped per vehicle configuration and number of axles. Letters represent the vehicle configuration and digits correspond to the number of axles. The complete table with all vehicles categories in the WIM system is presented in Table A.2. It may be observed that in the data corresponding to the A12-L highway all vehicle types are present. However, for the other data sets, two categories were excluded. This is the case of category O6 (according to the notation in the second column of Table 2), for data sets A12-R, A16-L, and A16-L and category O11 for data sets A15-L, A15-R, and A16-R. Also, the resulting filtered data does not include vehicles with more than eleven axles.

Table 1
Percentage of filtered vehicles.

	A12-L	A12-R	A15-L	A15-R	A16-L	A16-R
Vehicles in the database	165 373	161 589	124 449	124 789	220 920	203 598
Vehicles after filter criteria	161 519	152 537	120 216	121 179	215 530	199 751
% of filtered vehicles	2,3%	5,6%	3,4%	2,9%	2,4%	1,9%

Table 2
Created vehicle types from most observed WIM observations.

Vehicle (i)	Type	No. axles (n_i)	Code					
1	B2	2	B11					
2	B3	3	B111	B12				
3	O3	3	O3					
4	O4	4	O4					
5	O5	5	O5					
6	O6	6	O6					
7	O8	8	O8					
8	O9	9	O9					
9	O10	10	O:					
10	O11	11	O>					
11	R5	5	R11111					
12	R6	6	R111111	R11112	R11211	R1122		
13	R7	7	R111121	R11113	R11221	R1123	R1222	
14	R8	8	R112121	R12221	R1223			
15	R9	9	R121221					
16	T3	3	T11O1					
17	T4	4	T11O4	T11O11	T12O1			
18	T5	5	T11O3	T11O21	T11O111	T12O2	T12O11	T21O11
19	T6	6	T11O4	T11O1111	T12O3	T12O21	T12O111	
20	T7	7	T12O4	T12O1111				
21	V2	2	V11					
22	V3	3	V11A1	V12	V21	V111		
23	V4	4	V11A2	V11A11	V13	V22	V211	V1111
24	V5	5	V11A12	V12A2	V12A11			
25	V6	6	V12A12	V22A2	V22V11			
26	V7	7	V22A12					

3.2. One-dimensional marginal distributions for individual axle load

Once the data are filtered and the vehicle types are created, for each axle load, a one-dimensional marginal distribution is approximated with a Gaussian Mixture (GM) (McNicholas and Murphy, 2008). Gaussian Mixture distribution has been used in previous studies (Kim and Song, 2019; Enright and O'Brien, 2013; Kim and Song, 2021) because of its ability to approximate multi-modal distributions. These are typical for individual axle loads in WIM data. A GM is a weighted sum of G Gaussian densities (each one referred to as a component) expressed as follows:

$$f(x) = \sum_{g=1}^{G} \pi_g \phi\left(x \mid \mu_g, \sigma_g\right) \tag{8}$$

where $g = 1, \ldots, G$, $\sum_g \pi_g = 1$ are the mixture weights and $\phi\left(x \mid \mu_g, \sigma_g\right)$ are components of Gaussian densities with parameters μ_g and σ_g.

The expectation maximization (EM) algorithm (McLachlan, G. J. and Peel, 2000) is used to fit the Gaussian Mixtures to the individual axle load data per vehicle type. The relative goodness-of-fit Akaike information criterion (AIC) (Akaike, 1974) is used to choose the best-fitting Gaussian mixture distribution that describes individual axle loads per vehicle type. The AIC score rewards models with high log-likelihood while accounting for the number of parameters to prevent over-fitting. The model with the lower AIC score is expected to have a balance between its ability to fit the data and avoid over-fitting. As a result, the number of components of the fitted distributions ranged from four to seven. As observed in And and Vervuurt (2015), by adding more components, the tail of the fitted distribution function is increasingly dominated by individual observations in the tail of the distribution which would result in overestimation of individual axle loads. A total of 837 distributions were fitted for all locations. As an example, Fig. 2 shows the fitted GM distribution function for the vehicle type B3 in highway A15-L and its parameters are presented in Table 3.

Although fitting a GM distribution to axle loads is an appropriate approach, this is not the case for the inter-axle distances because, differently to axle loads, there are a finite number of or vehicle lengths according to vehicle category. Therefore, for each inter-axle distance and vehicle length, we use the empirical cumulative distribution function (ECDF) defined as:

$$\hat{F}_n(x) = \frac{1}{N+1} \sum_{i=1}^{N} \mathbf{I}\left\{X_i \leq x\right\} \tag{9}$$

where $\mathbf{I}$ is the indicator function, namely $\mathbf{I}\left\{X_i \leq x\right\}$ is one if $X_i \leq x$ and zero otherwise. Now that the data to be used in our framework and the 1-D marginal distributions have been described, the characterization of probabilistic dependence is presented.

3.3. Dependence: WIM NPBN

After the 1-D marginal distributions for the random variables were selected, the dependence structure of the WIM observations was modelled with a non-parametric Bayesian Network. A total of 26 submodels, that correspond to each vehicle type are built from each of the six WIM databases. Let $i = \{1, 2, 3, \ldots, 26\}$ be a set of indices corresponding to the 26 vehicle types, previously presented in Table 2, and n a set whose elements represent the number of axles of each vehicle type (see Table 2). We use, $X_{i,j}$ to denote the random variable representing the jth axle load. Notice that $j = \{1, \ldots, n_i\}$ according to the ith vehicle type. X_{i,n_i+1} denotes total vehicle length. Inter-axle distance per vehicle type i is denoted as X_{i,n_i+1+j}. The first inter-axle distance per vehicle type X_{i,n_i+2} corresponds to the distance between the front of the vehicle and the first axle. Our data does not report the distance between the last axle and the end of the vehicle and hence it is not modelled.

Notice that the model assumes that within axle load and inter-axle distance nodes $X_{i,j} \perp X_{i,j-2} | X_{i,j-1}$ for $\forall j$, where $A \perp B|C$ means that A and B are conditionally independent given C. This assumption

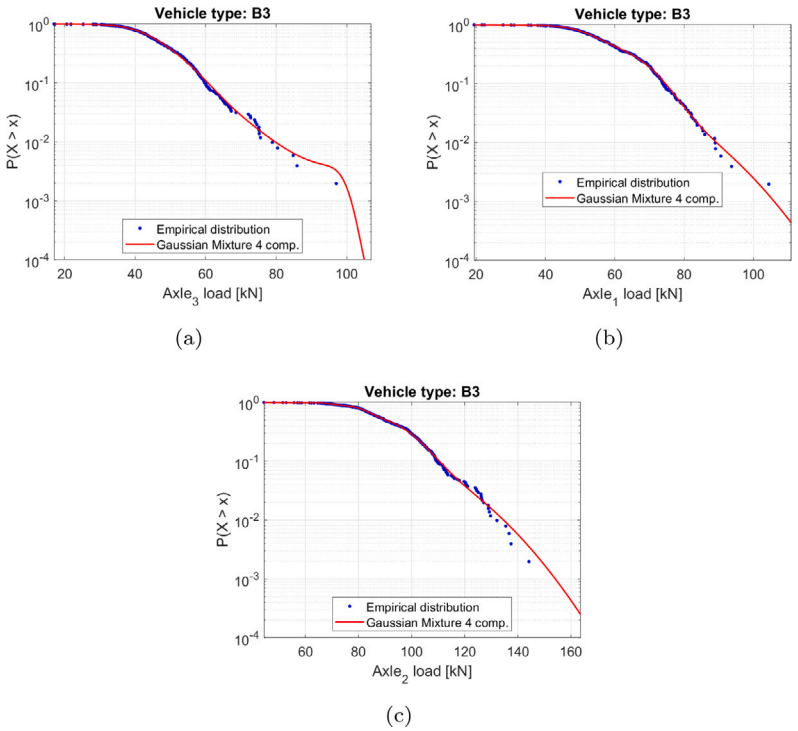

Fig. 2. Gaussian mixture of 4 components fitted distribution function for vehicle type B3 in highway A15-L: (a) Axle 1 load [kN]; (b) Axle 2 load [kN]; (c) Axle 3 load [kN].

Table 3
Gausian mixture components for vehicle type B3 in highway A15-L.

Axle 1 load [kN]			
Component	μ	σ	π
1	52.40	5.89	0.506
2	66.13	16.76	0.115
3	68.37	7.01	0.347
4	39.79	12.26	0.032

Axle 2 load [kN]			
Component	μ	σ	π
1	98.67	8.52	0.400
2	70.51	2.92	0.074
3	82.98	3.92	0.260
4	95.89	21.78	0.266

Axle 3 load [kN]			
Component	μ	σ	π
1	51.25	15.47	0.197
2	42.08	5.90	0.358
3	50.66	7.67	0.441
4	99.61	2.45	0.003

indicates, for example, that for a particular intermediate axle (or intermediate inter-axle distance), if the values of the loads (or distances) of two axles adjacent to it were known, knowing the value of any other axle load (or inter-axle distance) not adjacent to it would not update the distribution of the particular intermediate axle load (or inter-axle distance).

The total vehicle weight per vehicle type (W_i) is given by Eq. (10). The total vehicle length of the vehicle type i is $L_i = X_{i,n_i+1}$. Notice that W_i and L_i are deterministic functions of vehicle type i. For example, if $i = 2$ (vehicle type B3 according to Table 2), then the total vehicle weight of B3 is the sum of its individual axle loads, i.e. $W_2 = X_{2,1} + X_{2,2} + X_{2,3}$ and its total length is $L_2 = X_{2,4}$. Therefore, W is the random variable representing vehicle weight regardless of vehicle type. The distribution of total vehicle weight is often used when investigating bridge reliability for certain elements of the bridge. Similarly as for W, L is a random variable representing vehicle length regardless of the vehicle type. If the distance between the last axle and the end of the vehicle were to be required, this could be computed per vehicle type as: $X_{i,n_{i+1}} - \sum_i X_{i,n_{i+1+j}}$.

$$W_i = \sum_{j=1}^{n_i} X_{i,j} \tag{10}$$

As an example, a representation of the NPBN for highway A15-L is shown in Fig. 3. In this case, $i = \{1, 2, 3, \ldots, 25\}$. The model consists in 301 nodes and 436 arcs. The arcs represent correlations between axle loads, between axle loads and total vehicle length and between inter axle distance per vehicle type. The NPBN is implemented in MATLAB 2019b with the toolbox BANSHEE (Paprotny et al., 2020).

The unconditional rank correlations $r(X_{i,j}, X_{i,j-1})$, for all i and for $j > 1$, between individual axle loads and inter-axle distances per vehicle type for the NPBN A15-L model can be found in Tables B.1 and B.2. Notice that the correlation of the last axle load (X_{i,n_i}) to total vehicle length ($X_{i,n_{i+1}}$) is also presented. No clear pattern regarding its size or direction is observed across vehicle types. For example, as can be seen in Table B.1, the correlation ranges between (roughly) −0.3 and 0.79. A possible reason for this dependence is the design length of the vehicle according to its purpose. For example longer vehicles might be able to carry heavier loads in the last axle which could explain a correlation as high as 0.79. As mentioned earlier this pattern is not clear across vehicles types. Further investigation of this dependence could be a way to improve the model here presented.

The corresponding conditional rank correlations matrices between the random variables, as colour maps, for the six WIM locations can be found in Figs. B.1 to B.3. After the detailing of three categories of the framework, in the successive section, the results and validation of the NPBN will be presented

4. Results

4.1. NPBN model output

The output of the model is a data set, similar to the WIM measurements, with 26 columns. The first column corresponds to vehicle type according to the notation in the second column of Table 2. The second corresponds to total vehicle weight (W) in kN. Columns 3 to 13 to individual axle loads (A) in kN. The 14th column to total vehicle

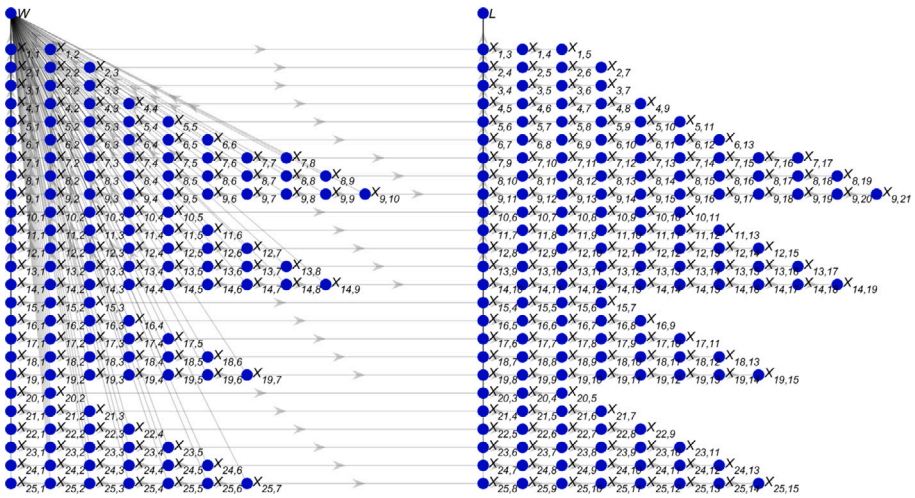

Fig. 3. BN model for A15-L highway of the WIM system in the Netherlands. The left side of the network represents the $X_{i,j}$ axle loads. The right side represents the vehicle length $X_{i,n_{i+1}}$ and the inter-axle distances $X_{i,n_{i+1+j}}$.

Table 4
BN model data set output (1st to 13th column).

Type	W	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
V5	303.97	64.74	93.89	48.20	43.53	53.60	NaN	NaN	NaN	NaN	NaN	NaN
T4	222.66	58.27	59.71	51.35	53.33	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
T3	176.42	59.62	51.29	65.51	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
T6	283.02	58.14	57.4	50.20	35.99	43.29	38.01	NaN	NaN	NaN	NaN	NaN

Table 5
Table 4 (continued, 14th to last column).

L	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
18.95	1.73	5.21	1.37	6.51	1.42	NaN	NaN	NaN	NaN	NaN	NaN
13.19	12.80	30.40	52.20	1.44	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...
16.77	20.40	43.50	68.30	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
17.16	16.80	32.10	14.10	5.96	1.36	1.39	NaN	NaN	NaN	NaN	NaN

length (L) in m, the 15th to the distance from vehicle front to the first axle ($D1$) in m and columns 16 to 25 to individual inter-axle distance in m, i.e. distance from axle 1 to axle 2 ($D2$), distance from axle 2 to axle 3 ($D3$) and so on. Not a number (NaN) strings are placed in fields where no data is computed. An example of the output table is presented in Tables 4 and 5. Note that the total vehicle weight is the sum of the individual axle load observations.

4.2. Validation

To validate the model, one month period of random samples per WIM location were generated from the corresponding NPBN model. As an example, the observed and simulated W and L for the A15-L highway (165 000 random samples) are presented in Fig. 4. Selected test statistics, the Nash–Sutcliffe model Efficiency Coefficient (NSE) (McCuen et al., 2006) and the Mean Absolute Error (MAE) (Willmott and Matsuura, 2005) are summarized in Table 6. The NSE takes values in $(-\infty, 1]$, where 1 indicates a perfect fit while the MAE takes values within $[0, \infty)$, where zero indicates a perfect fit.

To graphically assess if the generated synthetic data represent the real WIM observations we use the scatter plot presented in Fig. 4. The scatter plot was created by plotting the sorted WIM observations against the sorted synthetic observations for a single simulation. If both datasets are the same, one would expect to see the points lying in a straight line. As can be seen, Fig. 4(a) shows that the model slightly underestimates W (for this particular realization) in the interval between 900 kN and 1100 kN and Fig. 4(b) shows a slight overestimation of L in the interval between 12 m to 17 m. In general, as expected, for axle loads, more deviations with respect to the original data are observed in the tail of the distribution of total weight. Nevertheless, the model provides an overall good fit for the data. This can also be observed in the results of Table 6, where the values of NSE and MAE for W and L are close to 1 and 0 respectively. For example, for the A15-L highway, the NSE and MAE values for W are 0.99 and 0.46 respectively, while for L the corresponding values are 0.96 and 0.35.

Table 6
Test statistics of the model for different WIM locations on The Netherlands.

Highway	Total vehicle weight		Total vehicle length	
	NSE	MAE	NSE	MAE
A12-L	0.99	0.30	0.98	0.24
A12-R	0.99	0.37	0.98	0.25
A15-L	0.99	0.46	0.96	0.35
A15-R	0.99	0.48	0.97	0.30
A16-L	0.99	0.25	0.97	0.26
A16-R	0.99	0.32	0.97	0.25

Additionally, Figs. 5 and 6 shows the correlation matrix plot of axle loads and inter-axle distances for both, the observed and simulated data. The figures show the results of the most observed vehicle type (T5) at the A15-L highway. Notice that, the correlations of the simulated data between each axle load and the correlation between each inter-axle distance differ slightly compared to the WIM data.

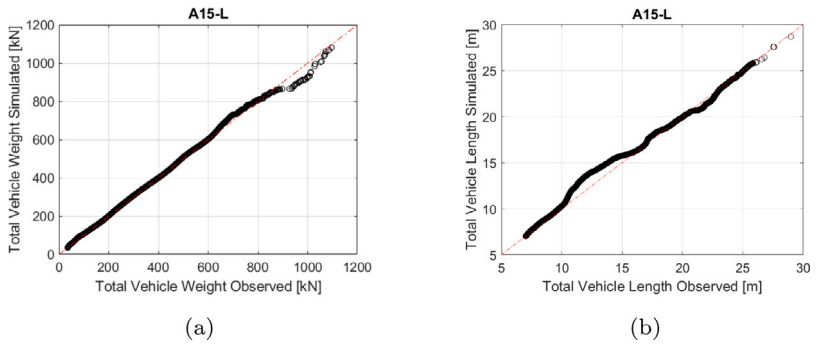

Fig. 4. Observed and simulated: (a) total vehicle weight and total vehicle length (b) with a one month period.

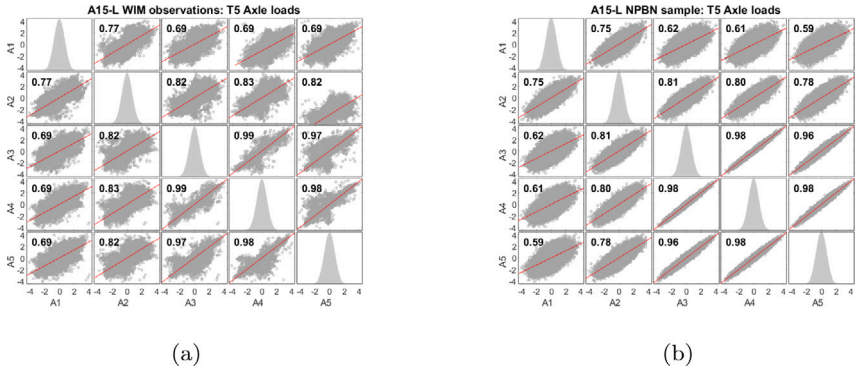

Fig. 5. Comparison between axle loads generated by the NPBN model and the WIM data of the T5 vehicle type in highway A15-L: (a) Correlation matrix for the empirical data standard transformed; (b) Correlation matrix for the synthetic data standard transformed.

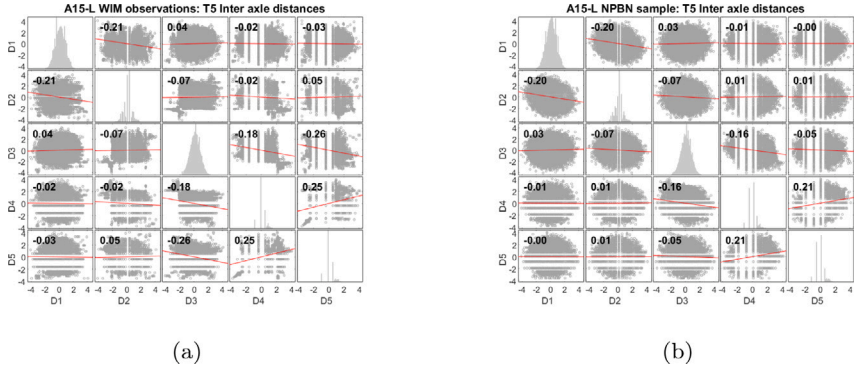

Fig. 6. Comparison between inter axle distances generated by the NPBN model and the WIM data of the T5 vehicle type in highway A15-L: (a) Correlation matrix for the empirical data standard transformed; (b) Correlation matrix for the synthetic data standard transformed.

This means that the NPBN model, under the assumption of the normal copula, correctly approximates the dependence structure of the empirical data. Furthermore, Fig. 7 shows the exceedance probability plots comparison of the observed and simulated variables, total vehicle weight, and total vehicle length at the A15-L. As can be seen in Fig. 7, the synthetic data shows similar behaviour as the observed data at the tail of the distributions. Figs. B.4 to B.6 show same comparisons plots for the other locations under study. Notice that this corresponds to a single realization of our NPBN. Perfect agreement between the observed data and every simulation is neither expected nor desired. Rather, approximating the general traffic configuration is the aim of the model especially for later use in investigation of infrastructure reliability.

Diverse engineering applications are often interested in determining the extremes observations of the phenomena. Thus, accurate synthetic data should be able to approximate them. Table 7 shows the load corresponding to the heaviest (Max W) and longest vehicles (Max L) in the original data. It shows their corresponding exceedance probability and the value observed in one simulation of the NPBN for the same probability of exceedance $P(X > x) = 1 - P(X \leq x)$. This is shown for the six locations under investigation. For example, the heaviest vehicle observed at A15-L has a total weight of 1085.89 kN with a probability of exceedance of $P(X > x) \approx 8.52E{-}06$ (as can be seen in Fig. 7(a)). The total vehicle weight in the simulation with probability of exceedance of $8.52E{-}06$ has a total weight of 1067.94 kN. This means that the difference between observed and simulated Max W at A15-L is around 1.62%. Similarly, the longest vehicle observed at A15-L has a total length of 27.58 m with $P(X > x) \approx 8.52E{-}06$ (see Fig. 7(b)). The total vehicle length in the simulation with probability of exceedance of $8.52E{-}06$ has a total length of 27.57 m. The resulting difference between observed and simulated Max L at A15-L is around 0.4%.

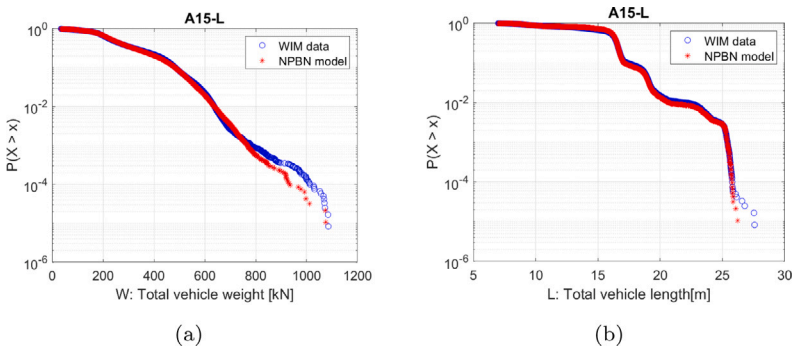

Fig. 7. Exceedance probability plots comparison between variables of interest (W and L) generated by the NPBN model and the WIM data in highway A15-L: (a) Total vehicle weigh distribution [kN] for original data and data generated with NPBN model. (b) Total vehicle length distribution [m] for original data and data generated with NPBN model.

Table 7
Comparison of the simulated heaviest and largest vehicles.

Highway	$P(X > x)$	Observed	Simulated	Difference %	Observed	Simulated	Difference %
		Max W [kN]	Max W [kN]		Max L [m]	Max L [m]	
A12-L	6.19E−06	1080.50	1090.01	0.88	29.62	29.05	1.92
A12-R	6.54E−06	1096.87	1068.43	2.59	29.36	28.31	3.58
A15-L	8.52E−06	1085.50	1067.94	1.62	27.58	27.57	0.04
A15-R	8.34E−06	1085.89	1075.20	0.98	29.11	28.93	0.62
A16-L	4.67E−06	1016.26	1007.44	0.87	28.36	28.26	0.35
A16-R	5.03E−06	1044.21	1044.80	0.06	27.23	27.47	0.88

According to Table 7, highway A12-R has the biggest difference for Max W and Max L with 2.59% and 3.58%, respectively. The smallest difference can be found in highway A16-R with 0.06% for Max W and 0.88% for Max L. Overall, the differences between the estimations from the observed and simulated data are minor for all study locations. This shows that the performance of the method and NPBN models can be considered effective. In order to further investigate model validity we perform split-sample analysis. In Appendix B we present a comparison between synthetic data generated by the model quantified with the training data set (80%) and the observations of the test data set (20%). This is a commonly used split fraction in traffic analysis (see Rutherford and McNeill, 2011; Ravilla et al., 2021; Shahid et al., 2021, for example). The observations correspond to the A16-R highway. From these results, it is clear that the performance of the model is in agreement with the results of the NPBN model quantified with the unsplit data set. In order to profit from all data available, further all our analysis is performed with models quantified with the full data set.

Next, to show the adaptability of the framework, two case studies are presented in Section 5: (i) a NPBN quantified using data of a Dutch city road by classifying vehicle types according to the number of axles, and (ii) a NPBN using data of a Brazilian highway, classifying vehicle types according to the automatically generated WIM vehicle codes.

5. Case studies

5.1. NPBN from Rotterdam city WIM data

One WIM station in the municipality of Rotterdam (South Holland, The Netherlands) is considered for this case study. The WIM system was located in the bridge *Beukelsbrug* in the S115 city route. The observations correspond to May 2013. In total, 14 different heavy vehicle codes were observed in the S115 WIM data set. The WIM recorded vehicle codes are only numbers that define the sequence of axle groups. These codes differ from the ones found in the WIM system of the Dutch motorways described in Section 3.1. Hence, a similar classification as the one presented in Table 2 is not possible.

Moreover, Fig. 8 shows a comparison of the exceedance probability plots of W and L between the Dutch WIM locations (A16-L with the lowest heavier W observation and A12-R with the highest heavier W measurement) and the S115 location. As can be seen in Fig. 8, as expected, heavier and longer vehicles circulate on the Dutch motorways compared to those circulating on regional roads. Consequently, the NPBN purposed in Section 3.2 does not meet the requirements to compute similar S115 city route WIM observations. Hence as a consequence, a new model has to be quantified with data from Rotterdam. By modifying the vehicle classification and then applying the framework described in Section 3, the synthetic WIM data can be generated. For this case study, vehicle types were created grouping them by the number of axles. Therefore, four vehicle types were generated as presented in Table 8.

Table 8
Created vehicle types WIM vehicle codes grouped by number of axles.

Vehicle (i)	Type	No. axles (n_i)	Codes				
1	2 axles vehicle	2	2	11			
2	3 axles vehicle	3	12	21	111		
3	4 axles vehicle	4	22	31	211	1111	
4	5 axles vehicle	5	212	221	311	1211	2111

A total of 14 Gaussian Mixtures (9 Gaussian mixtures of 4 components, 3 GM of 5 components and 2 GM of 6 components) were fitted to the axle loads choosing the best fit according to the AIC criterion. Fig. 9 shows the NPBN model for the S115 WIM location with $i = \{1, 2, \ldots, 4\}$. The DAG consists of 34 nodes and 46 arcs. A random sample with a size similar to the WIM observations was generated from the NPBN model. The comparison between the S115 location WIM data and that generated by the model is presented in Fig. 10. The figure was generated with 6000 samples. As can be seen, Fig. 10(a) shows that the model slightly underestimates W in the interval between 400 kN to 600 kN. In Fig. 10(b) no important deviations are presented. Overall, the model approximates well the measured total vehicle weight and total vehicle length.

5.2. NPBN from Araranguá, Brazil WIM data

For this case study, we analysed WIM data provided by the Transport and Logistics Laboratory (LabTrans) of the Santa Catarina Federal

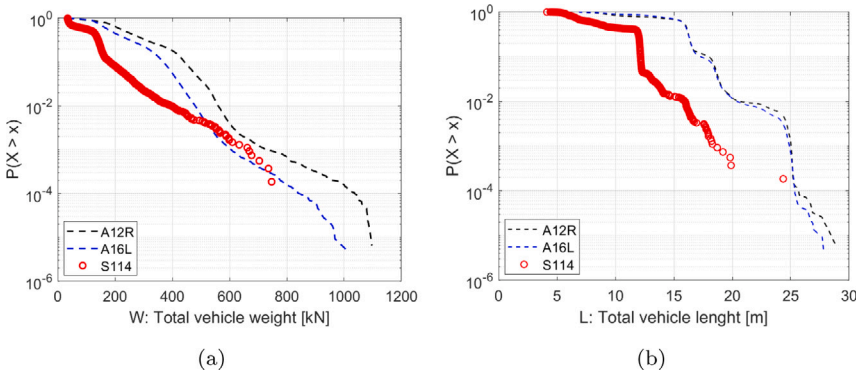

Fig. 8. Exceedance probability plots of: (a) Total vehicle weight [kN] and (b) Total vehicle length [m] for the locations: A12-R, A16-R, and S115.

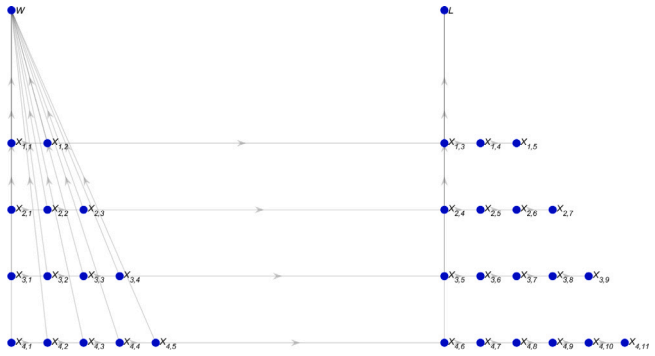

Fig. 9. NPBN model for S115 road WIM location. The left side of the network represents the $X_{i,j}$ axle loads. The right side represents the vehicle length $X_{i,n_{i+1}}$ and the inter axle distances $X_{i,n_{i+1+j}}$.

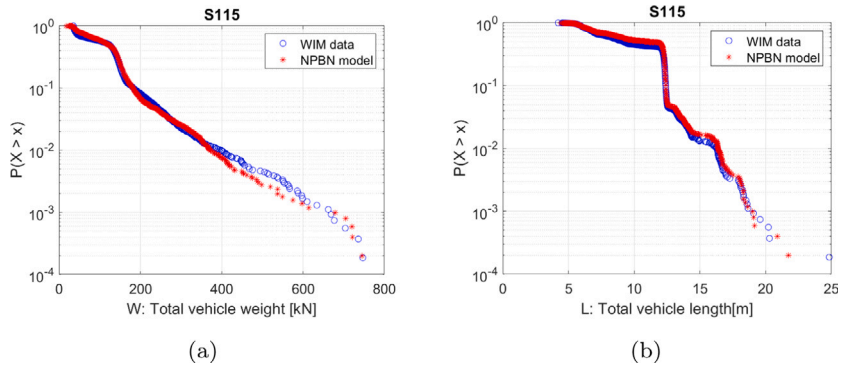

Fig. 10. Comparison between variables of interest generated by the NPBN model and the WIM data in S115 location: (a) Total vehicle weigh distribution [kN] for original data and data generated with the NPBN model, (b) Total vehicle length distribution [m] for original data and data generated with the NPBN model.

University (UFSC for its acronym in Portuguese). We use the automatically determined WIM codes to assign vehicle types. The data was gathered at the Federal Highway route BR-101 km 418, located in the city of Araranguá in the South of Brazil. The measurements were taken in April 2014. The data set has a total of 66 820 vehicles with total vehicle weight above 34 kN.

Following the framework for generating WIM synthetic data, described in Section 3, the WIM database is analysed. About 6.2% of the observations were excluded and 45 different classification codes were found with vehicles from 2 up to 8 axles. The codes were automatically determined by the WIM system according to the Brazilian National Department of Transport Infrastructure (DNIT for its acronym in Portuguese) (Departamento Nacional De Infra-Estrutura De Transportes, 2012). As stated in Departamento Nacional De Infra-Estrutura De Transportes (2012), the numbers in the code corresponding to the number of axles and the letters to vehicle classification. To name a few: C is a Tractor–Trailer, S a Tractor–Semitrailer, and I is a Trailer–Semitrailer with an inter-axle distance of more than 2.4 m. Letter D corresponds to a tandem (group of 2 axles), T to a tridem (group of 3 axles), and Q to a quad (group of 4 axles). For example, a 2C2 vehicle represents a Tractor with 2 axles plus a 2 axles Trailer. The 45 classification codes (or vehicle types) are presented in Table 9.

A total of 230 Gaussian Mixtures were fitted to one-dimensional axle loads choosing the best fit according to the AIC criterion. Next, the dependence structure can be constructed. Fig. 11 shows the NPBN model for the BR-101 highway with $i = \{1, 2, \ldots, 45\}$. The network consists of 507 nodes and 735 arcs. Finally, a random sample with a size similar to the WIM observations was generated from the NPBN model. The comparison between the BR-101 WIM data and that generated by

Table 9
Registered vehicle classification codes in the WIM data set of BR-101 highway. April 2014.

Vehicle (i)	Type	No. axles (n_i)	Vehicle (i)	Type	No. axles (n_i)	Vehicle (i)	Type	No. axles (n_i)
1	2C	2	16	3C	3	31	3S2	5
2	2C2	4	17	3C2	5	32	3S3	6
3	2CB	2	18	3C3	6	33	3T4	7
4	2D4	6	19	3D4	7	34	3V5	8
5	2DL	4	20	3DB	3	35	4C	4
6	2I1	5	21	3DL	5	36	4CD	4
7	2I2	4	22	3I1	6	37	4DB	4
8	2I3	5	23	3I2	5	38	4DT	7
9	2LD	5	24	3I3	6	39	4R2	6
10	2N3	5	25	3JD	6	40	Unknown 3 axles	3
11	2N4	6	26	3LD	6	41	Unknown 4 axles	4
12	2S1	3	27	3N3	6	42	Unknown 5 axles	5
13	2S2	4	28	3P5	8	43	Unknown 6 axles	6
14	2S3	5	29	3QD	7	44	Unknown 7 axles	7
15	3BC	3	30	3S1	4	45	Unknown 8 axles	8

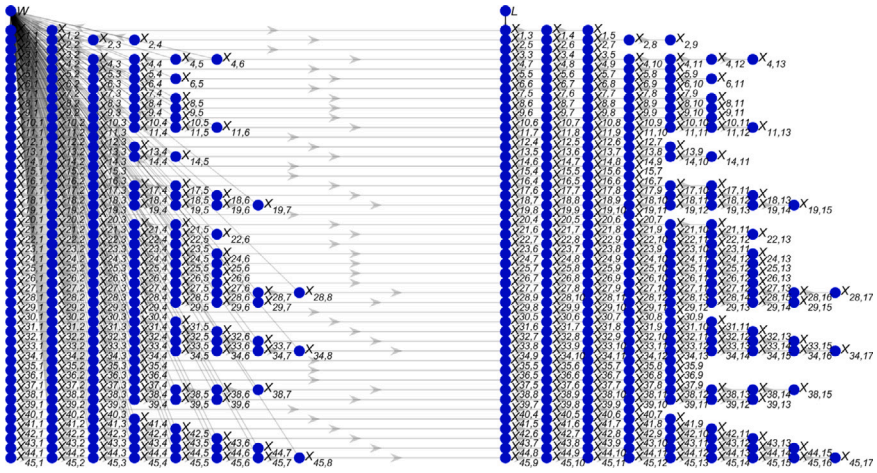

Fig. 11. NPBN model for the 45 vehicle types of the WIM system in BR-101 highway located in Araranguá, Brazil. The left side of the network represents the $X_{i,j}$ axle loads. The right side represents the vehicle length $X_{i,n_{i+1}}$ and the inter axle distances $X_{i,n_{i+1+j}}$.

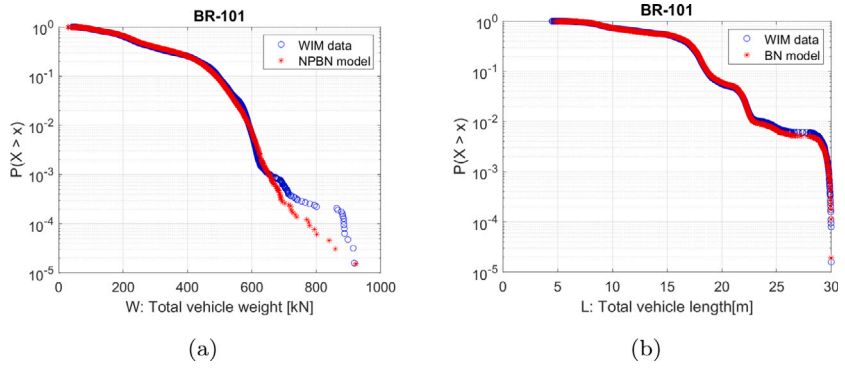

Fig. 12. Comparison between variables of interest generated by the NPBN model and the WIM data in BR-101 highway: (a) Total vehicle weight distribution [kN] for original data and data generated with the NPBN model; (b) Total vehicle length distribution [m] for original data and data generated with the NPBN model.

the BN model is presented in Fig. 12. The figure was generated with 71 000 samples as described in Section 3.3.

As can be seen in Fig. 12(a), the exceedance probability computed with synthetic observations shows a deviation in the interval of 750 kN to 850 kN with respect to the one computed with WIM measurements. Nevertheless, the model still approximates the maximum values of the empirical distribution. The computed values for L are quite similar to the observed ones (see Fig. 12(b)). As has been noted, in this case, the methodology for computing synthetic WIM observations also approximates the data. The results are consistent for all studied WIM locations. In the following section, a Graphical User Interface (GUI) for the six Dutch WIM motorways locations will be presented. The GUI illustrates a possible use of the model when WIM data is not available. The GUI is used to compute synthetic WIM observations using data collected through (for example) traffic counters as input.

6. Graphical User Interface application

When WIM data is not available due to the scarcity or absence of proper equipment, the most utilized devices to collect traffic data are the pneumatic road tubes. The change in air pressure within the tube is measured as the vehicle's wheels pass over the tube (Beyer,

Table 10
Selected corresponding vehicle types.

SCT	WIM NPBN	No. vehicles	Proportion
B2	B2	11 041	0.68000
C2	V2	3 576	0.22020
C3	V3	887	0.05462
C4	V4	2	0.00012
T2S2	T4	7	0.00043
T3S2	T5	523	0.03220
T3S3	T6	184	0.01133
T2S1R2	R5	2	0.00012
T2S2R2	R6	10	0.00062
T3S2R2	R7	3	0.00018
O9	O9	5	0.00031
Total		16 240	1

2015). Pneumatic road tubes are mainly used for vehicle classifications, estimation of average daily traffic and estimation of direction of travel. Consequently, the information regarding axle loads and inter-axle distances cannot be collected. Therefore, to have an insight over possible axle loads and axle inter-distances we can use the NPBN described in Section 3 with data obtained from pneumatic road tubes as input i.e. vehicle classification and proportion of vehicle types.

We developed a simple stand-alone Graphical User Interface (GUI) of the six Dutch motorways WIM locations model described in Sections 3.1 to 3.3. The GUI was implemented in Python, using the PyQt5 package for the Windows operative system. Notice that no fitting or goodness of fit procedures are implemented in the GUI. Rather, the results from the fitting and goodness-of-fit procedures and the created vehicles types introduced in Section 3.2 are the core of the engine for generating synthetic observations using the GUI.

To exemplify the use of the GUI, we use data from pneumatic road tubes gathered in the low-speed lane of the Paseo Tollocan avenue (Mexican federal highway 15) located in the city of Toluca Mexico. The data was obtained in the last week of August 2014. The information was provided by the Autonomous University of the State of Mexico (UAEMex). Table 10 shows the number of observed vehicles per vehicle type according to the vehicle classification of the Ministry of Communications and Transport (SCT for its acronym in Spanish) of Mexico (SCT, 2008). Table 10 shows also the NPBN equivalent vehicle types (matched with the Mexican types according to their silhouette) and the proportions of the registered vehicles by the pneumatic road tubes .

Now, it is possible to enter the information in Table 10 into the GUI (see Fig. 13) to compute the axle loads and inter-axle distances together with total vehicle weight and total vehicle length. Thus, the information obtained by the pneumatic road tubes is extended. The output of the GUI are histograms (Fig. 14(a)) and exceedance cumulative plots (Fig. 14(b)) of W and L. Additionally, the computed observations can be stored in a comma-separated value (CSV) file. For this example, to compute the observations, we use the so-called "Hypothetical highway" which is the mixture of the six available highways in the GUI. This means that each computed observation comes from one randomly selected NPBN Dutch motorway model. A quick user guide for the NPBN WIM graphical user interface can be found in Appendix C.

As can be seen in Fig. 14, the maximum W is around 600 kN and the maximum L is approximately 25 m. The most frequent vehicles are the ones with 100 kN to 150 kN and those with a total length of 11 m to 12.5 m. Which correspond to Buses (B2) and single-unit two-axle vehicles (V2) representing around 90% of the total observed vehicles. Information to download the Graphical User Interface executable application can be found in the supplementary material related to this article.

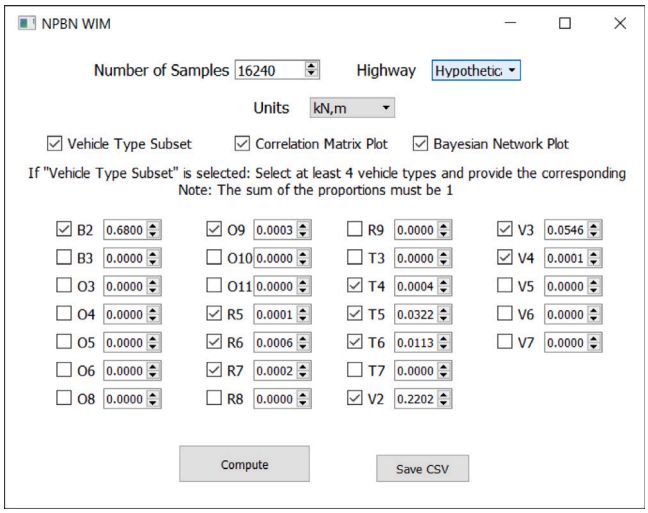

Fig. 13. GUI Main window.

7. Conclusions

In this paper, we presented an improved methodology to compute synthetic WIM observations of heavy vehicles through Non-parametric Bayesian Networks (NPBNs). The model provides data that describes: vehicle type, total (gross) vehicle weight, individual axle loads, total vehicle length, and inter-axle distances. In total eight high dimensional NPBNs with up to 507 nodes and 735 arcs were quantified. The first six NPBNs were quantified with data corresponding to six WIM locations of the motorway network of the Netherlands. 26 vehicle types were created grouping the registered vehicles in the WIM data set per vehicle configuration and per number of axles. The seventh NPBN was quantified using WIM observations collected in a city route in Rotterdam city, The Netherlands. 4 vehicle types were created classifying vehicle types by the number of axles. Data collected in Araranguá city in Brazil were used to quantify the last NPBN. For this model, 45 vehicle types were created according to the vehicle classification WIM codes.

The NPBN models properly simulate the dependence structure of the empirical data. The correlations of the simulated data between axle loads and inter-axle distances slightly deviate from the ones computed with the WIM observations. Small differences in the exceedance probabilities of the empirical and simulated data can be observed for the total vehicle weight (W) in the range of 800 kN to 1000 kN. For the total vehicle length (L), the models show a minor overestimation of the exceedance probabilities of lengths in the interval between 12 m and 17 m. Nevertheless, the models provide a good fit for the data. Values of NSE for W and L are in the interval of 0.97 to 0.99 and, while for MAE are in the range of 0.24 to 0.48. The difference between the estimation of the tail of the distributions, i.e., the heaviest vehicle ($MaxW$) and longest vehicle($MaxL$) for the same probability of exceedance $P(X > x)$ are around 0.04% and 3.6%. In general, it can be noted that all presented NPBN models correctly approximate the WIM observations. Additionally, to make use of the NPBN model more convenient, we developed a Graphical User Interface (GUI) for the models of the six Dutch WIM motorways. A potential application of the model when WIM data is not available was presented by using the GUI to compute synthetic WIM observations. We used data collected through traffic counters, gathered in Toluca city in central Mexico, as input.

The framework for computing synthetic WIM observations here presented can be applied in any WIM location, using site-specific WIM records and site-specific vehicle types. The key aspect of the methodology is a good assessment of the vehicle types when constructing the model. Usually, previous studies have constructed vehicle types by grouping per number of axles. This vehicle classification works.

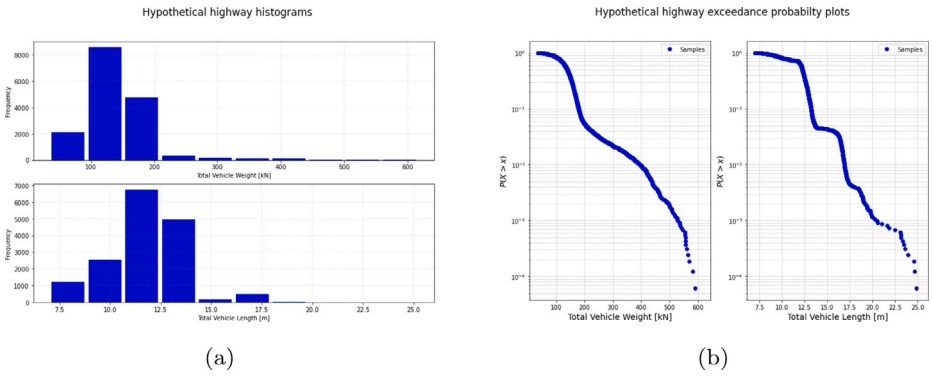

Fig. 14. NPBN WIM GUI output: (a) W and L histograms. (b) W and L exceedance probability plots.

Table A.1
Filtered criteria described in Kreslin et al. (2016) and And and Vervuurt (2015).

	Filter
1	Wheelbase length less than 1 m.
2	Wheelbase less than 30 m and first or last spacing above 10 m.
3	Wheelbase larger than 40 m.
4	Trucks with axle load below or equal to 0 tons.
5	Any axle weight larger than 40 tons.
6	Any axle weight larger than 15 tons and above 85% of gross vehicle weight.
7	Trucks with gross weight below or equal to 0 tons.
8	Sum of axle loads not within 50 kg of gross vehicle weight.
9	Truck with closely spaced, i.e. less or equal to 2 m first two axles, one of which is larger than 10 tons and over 2.5 times heavier than other axle.
10	First spacing larger than 15 m.
11	Any spacing less than 0,4 m.
12	Miss match between number of axle spacings and number of axle loads.
13	Sum of axle spacings not within 50 mm of wheelbase.
14	Number of axles below or equal to 1.
15	First axle spacing in the interval of 10 m–15 m.
16	Each spacing in range of 0,4 m–0,7 m.
17	Each spacing in range of 0,7 m–1,0 m.
18	Each axle load in the interval of 25 tons–40 tons.
19	Each axle load below 0,5 tons.
20	Vehicles with same WIM identification number (ID).
21	Vehicles with a gross vehicle weight below 3,56 tons.
22	Vehicles with a gross vehicle weight above 112 tons.
23	Vehicles with a speed greater than 120 km/h.
24	The vehicles with gross vehicle weight larger than 71.3 tons and or length bumper-to-bumper above than 25,5 m and axle spacing above 12,5 m (data related to a combination of two vehicles).
25	Vehicles with inter axle distances less than 75 cm.
26	Duplicate records.

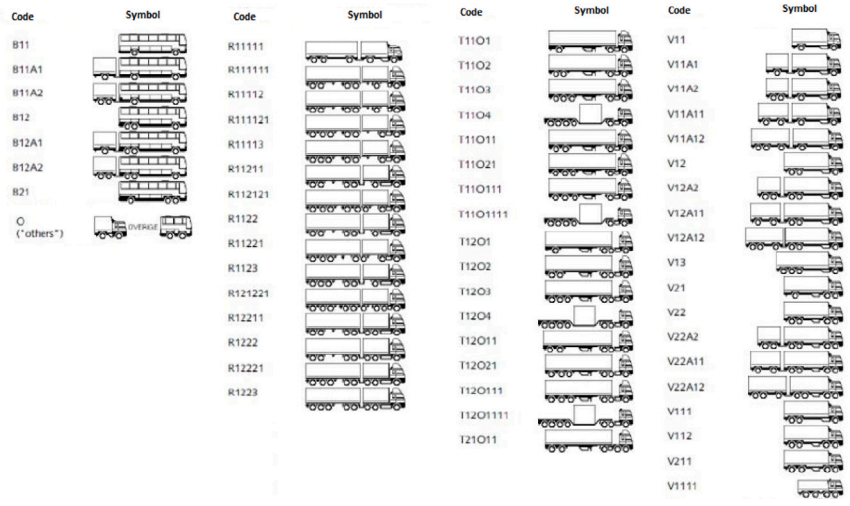

Fig. A.1. Most observed vehicle codes in the Dutch WIM measurements (April 2013).

Table A.2
Created vehicle types from all observed Dutch WIM codes (April 2013).

Item	Class	Code											
1	B2	B11	B2										
2	B3	B111	B12	B3									
3	O3	O3											
4	O4	O4											
5	O5	O5											
6	O6	O6											
7	O8	O8											
8	O9	O9											
9	O10	O=											
10	O11	O>											
11	R5	R11111	R1112	R1211	R122								
12	R6	R111111	R11112	R11121	R1113	R11211	R1122	R12111					
		R1212	R123	R1311	R132								
13	R7	R1111111	R111112	R111121	R11113	R111211	R11122	R112111					
		R124	R13111	R133	R2221	R223	R11311	R1123					
		R1132	R115	R121111	R12112	R1213	R12211	R1222					
		R11212	R11221										
14	R8	R11111111	R1111112	R1111121	R111113	R111122	R1112111	R111212	R111221	R11123	R1121111	R112112	R112121
		R11213	R112211	R11222	R1124	R113111	R11312	R11321	R1133	R1211111	R121112	R121121	R12113
		R121211	R12122	R1214	R122111	R12212	R12221	R1223	R12311	R1232	R125	R131111	R1313
		R13211	R1322	R134	R2123	R2213	R2222	R224					
15	R9	R1112121	R1112211	R11124	R1121121	R112113	R1122111	R112221	R11223	R1125	R1134	R12111111	R1211112
		R1211121	R121113	R1212111	R121212	R121221	R12123	R1221111	R122112	R122121	R12213	R1224	R123111
		R12321	R1233	R126	R1314	R132111	R13221	R1323	R1332	R1341	R135	R1413	R144
		R2214	R2223	R225	R234	R3312	R54						
16	T3	T11O1											
17	T4	T111O1	T11O11	T11O2	T12O1	T21O1	T2O2						
18	T5	T111O11	T111O2	T11O111	T11O12	T11O21	T11O3						
		T12O11	T12O2	T21O11	T21O2	T2O21	T2O3	T3O2					
19	T6	T111O111	T111O12	T111O21	T111O3	T11O1111	T11O112	T11O121	T11O13	T11O211	T11O22	T11O31	T11O4
		T12O111	T12O12	T12O21	T12O3	T21O111	T21O12	T21O21	T21O3	T2O22	T2O4	T3O3	
20	T7	T111O112	T111O121	T111O13	T111O22	T111O31	T111O4	T12O1111	T12O112	T12O121	T12O13	T12O211	T12O22
		T12O31	T12O4	T21O211	T21O22	T21O4	T3O4						
21	V2	V11											
22	V3	V111	V11A1	V12	V21	V3							
23	V4	V1111	V112	V11A11	V11A2	V121	V13	V211	V22	V4			
24	V5	V111A11	V111A2	V11A111	V11A12	V12A11	V12A2	V21A11	V21A2				
25	V6	V1111A11	V1111A2	V111A111	V111A12	V112A11	V112A2	V121A11	V12A111	V12A12	V12A21	V12A3	V13A11
		V13A2	V211A11	V211A2	V21A12	V22A11	V22A2						
26	V7	V1111A111	V1111A12	V1111A3	V112A111	V112A12	V112A21	V112A3	V121A111	V121A12	V121A3	V13A111	V13A12
		V13A21	V13A3	V211A12	V211A3	V22A111	V22A12	V22A21	V22A3	V4A12			

Table B.1
Rank correlation per vehicle type between axles A15-L location.

Vehicle (i)	Type	No axles (n_i)	$X_{i,1}, X_{i,2}$	$X_{i,2}, X_{i,3}$	$X_{i,3}, X_{i,4}$	$X_{i,4}, X_{i,5}$	$X_{i,5}, X_{i,6}$	$X_{i,6}, X_{i,7}$	$X_{i,7}, X_{i,8}$	$X_{i,8}, X_{i,9}$	$X_{i,9}, X_{i,10}$	$X_{i,10}, X_{i,11}$	$X_{i,n_i}, X_{i,n_{i+1}}$
1	B2	2	0.87										0.2
2	B3	3	0.81	0.85									0.31
3	O3	3	0.62	−0.1									0.09
4	O4	4	0.61	0.03	0.73								0.26
5	O5	5	0.64	0.06	0.66	0.73							0.04
6	O6	6	0.32	0.08	0.95	0.63	0.99						−0.28
7	O8	8	0.75	0.81	0.79	0.65	0.95	0.93	0.97				−0.19
8	O9	9	0.75	0.61	0.85	0.49	0.83	0.99	0.99	0.98			0.27
9	O10	10	0.73	0.4	0.49	0.71	0.71	0.98	0.99	0.99	0.98		−0.79
10	R5	5	0.7	0.22	0.6	0.92							−0.04
11	R6	6	0.01	0.16	0.94	0.47	0.96						0.43
12	R7	7	0.42	0.12	0.89	0.84	0.89	0.99					0.08
13	R8	8	0.48	0.41	0.66	0.76	0.8	0.87	0.94				−0.06
14	R9	9	0.48	0.66	0.62	0.37	0.98	0.85	0.99	0.99			−0.16
15	T3	3	0.45	0.54									0.36
16	T4	4	0.58	0.74	0.96								0.4
17	T5	5	0.77	0.82	0.99	0.98							0.05
18	T6	6	0.44	0.28	0.65	0.99	0.99						−0.2
19	T7	7	0.51	0.51	0.63	0.95	0.96	0.93					−0.21
20	V2	2	0.81										0.54
21	V3	3	0.59	0.6									−0.21
22	V4	4	0.42	0.42	0.9								−0.1
23	V5	5	0.56	0.59	0.59	0.93							0.12
24	V6	6	0.57	0.45	0.55	0.72	0.95						−0.05
25	V7	7	0.49	0.36	0.7	0.46	0.83	0.99					−0.15

Table B.2
Rank correlation per vehicle type between inter axle distances for A15-L location.

Vehicle (i)	Type	No axles (n_i)	$X_{i,n_{i+1}}$, $X_{i,n_{i+1+1}}$	$X_{i,n_{i+1+1}}$, $X_{i,n_{i+1+2}}$	$X_{i,n_{i+1+2}}$, $X_{i,n_{i+1+3}}$	$X_{i,n_{i+1+3}}$, $X_{i,n_{i+1+4}}$	$X_{i,n_{i+1+4}}$, $X_{i,n_{i+1+5}}$	$X_{i,n_{i+1+5}}$, $X_{i,n_{i+1+6}}$	$X_{i,n_{i+1+6}}$, $X_{i,n_{i+1+7}}$	$X_{i,n_{i+1+7}}$, $X_{i,n_{i+1+8}}$	$X_{i,n_{i+1+8}}$, $X_{i,n_{i+1+9}}$	$X_{i,n_{i+1+9}}$, $X_{i,n_{i+1+10}}$
1	B2	2	0.63	0.11								
2	B3	3	0.28	−0.03	0.03							
3	O3	3	0.04	0.11	−0.33							
4	O4	4	0.06	0.15	0.37	0.13						
5	O5	5	0.16	0.34	0.54	−0.39	−0.22					
6	O6	6	0.48	0.18	−0.09	−0.28	−0.42	0.01				
7	O8	8	0.09	0.14	−0.84	−0.49	−0.82	−0.39	−0.52	−0.23		
8	O9	9	0.44	0.27	−0.77	−0.79	−0.82	−0.39	−0.28	−0.05	0.73	
9	O10	10	0.6	0.77	0.18	0.32	−0.64	−0.58	−0.22	0.46	0.35	0.01
10	R5	5	0.15	−0.07	0.51	0.19	0.05					
11	R6	6	−0.1	−0.23	−0.27	−0.31	−0.07	0.07				
12	R7	7	0.16	0.03	−0.11	−0.68	−0.09	−0.75	0.33			
13	R8	8	0.15	0	0.07	−0.54	−0.19	−0.71	−0.68	0.25		
14	R9	9	0.23	0.18	−0.35	−0.49	−0.64	−0.49	−0.55	−0.41	0.32	
15	T3	3	0.22	−0.16	−0.06							
16	T4	4	0.27	−0.13	0.02	−0.25						
17	T5	5	0.28	−0.21	−0.07	−0.18	0.25					
18	T6	6	0.21	0.15	0.15	−0.22	−0.27	0.59				
19	T7	7	−0.08	−0.16	0.41	−0.16	−0.26	−0.27	−0.06			
20	V2	2	0.52	0.31								
21	V3	3	0.18	0.07	−0.21							
22	V4	4	0.28	0.17	0.75	−0.1						
23	V5	5	0.41	0.15	−0.25	−0.39	−0.38					
24	V6	6	0.29	0.03	−0.48	−0.41	−0.22	−0.15				
25	V7	7	0.08	0.07	−0.71	−0.04	−0.06	−0.41	−0.19			

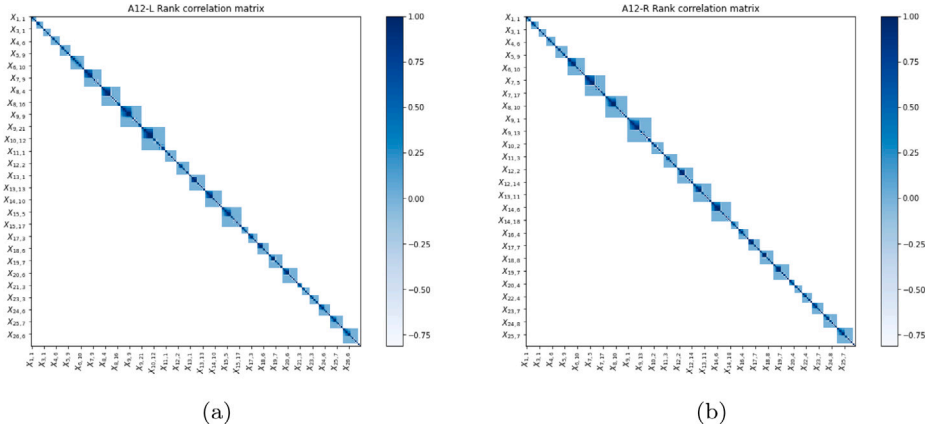

Fig. B.1. Bayesian Network rank correlation matrix corresponding to the Dutch A12 highway. (a) left lane and (b) right lane.

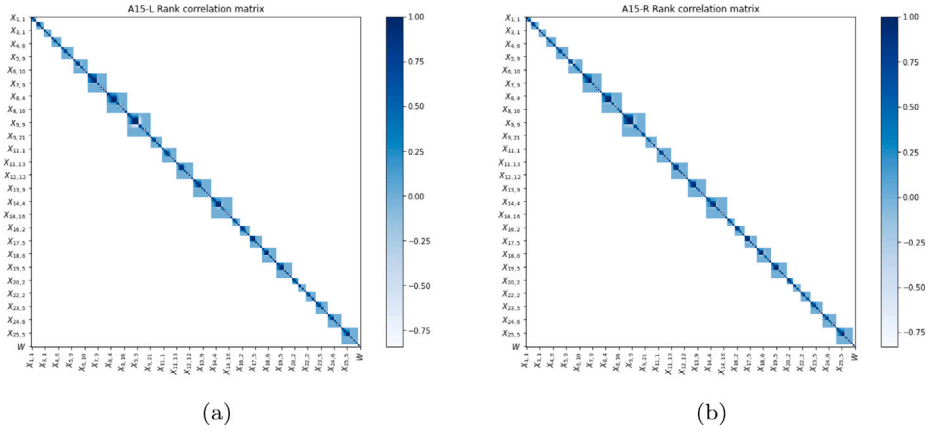

Fig. B.2. Bayesian Network rank correlation matrix corresponding to the Dutch A15 highway. (a) left lane and (b) right lane.

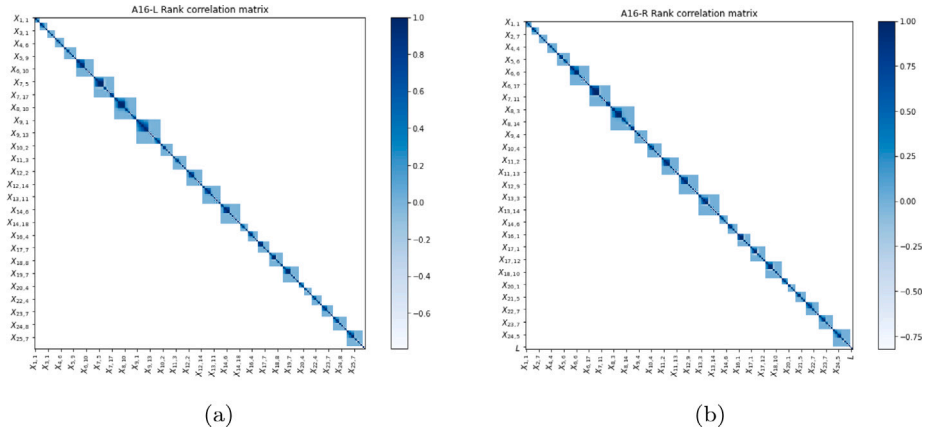

Fig. B.3. Bayesian Network rank correlation matrix corresponding to the Dutch A16 highway. (a) left lane and (b) right lane.

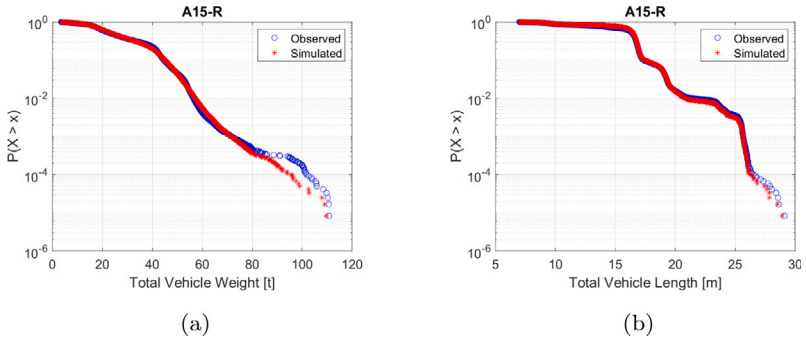

Fig. B.4. Comparison between variables of interest generated by the BN model and the WIM data in highway A15: (a) Total vehicle weight [kg] comparison right lane; (b) Total vehicle length [cm] comparison right lane.

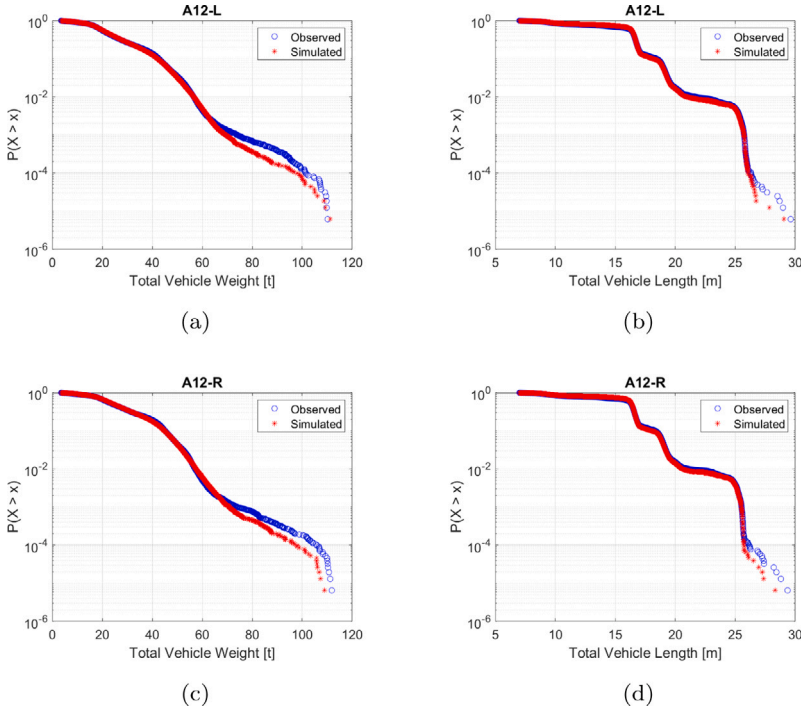

Fig. B.5. Comparison between variables of interest generated by the BN model and the WIM data in both driving directions of highway A12: (a) Total vehicle weight [t] comparison left lane; (b) Total vehicle length [m] comparison left lane; (c) Total vehicle weight [t] comparison right lane; (d) Total vehicle length [m] comparison right lane.

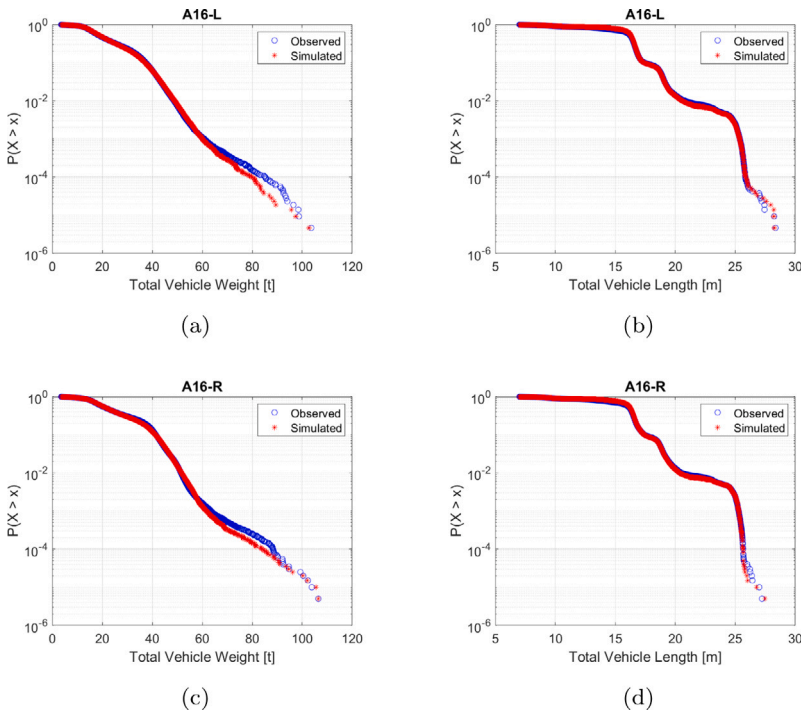

Fig. B.6. Comparison between variables of interest generated by the BN model and the WIM data in both driving directions of highway A16: (a) Total vehicle weight [t] comparison left lane; (b) Total vehicle length [m] comparison left lane; (c) Total vehicle weight [t] comparison right lane; (d) Total vehicle length [t] comparison right lane.

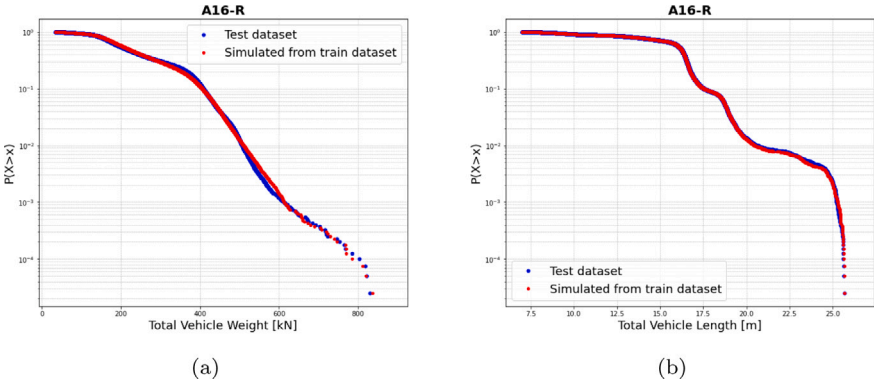

Fig. B.7. Comparison between data generated by the trained model and test data set for (a) total vehicle weight and (b) total vehicle length (all vehicle types).

However, often more insight into the different vehicle types is required. In these cases, a better classification can be performed, for example, by using the WIM automatically generated codes or grouping vehicles by body configuration. Our methodology allows the use of these classifications. Thus, more accurate data can be simulated. The next steps in our research correspond to the application of the models here presented in risk and reliability of individual infrastructure (bridges for example) and road networks including multiple infrastructures.

CRediT authorship contribution statement

Miguel Angel Mendoza-Lugo: Data curation, Writing – original draft, Methodology, Software, Writing – review & editing, Investigation, Validation, Formal analysis, Visualization. **Oswaldo Morales-Nápoles:** Supervision, Project administration, Resources, Conceptualization, Writing – review & editing, Methodology, Investigation, Formal analysis. **David Joaquín Delgado-Hernández:** Data curation, Resources, Investigation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Mexican National Council for Science and Technology (CONACYT), Mexico, for the financial support given through the scholarship 2019-000021-01EXTF-00564, CVU 784544, to carry out this research. The authors also acknowledge Gustavo Garcia Otto and Rafael Aleixo de Souza (Santa Catarina Federal University) for providing Brazilian WIM data. We thank Dr. ir. C.B.M. Blom (TU Delft/ Rotterdam Municipality) and Ir D.G. Schaafsma (Rijkswaterstaat) for providing Dutch WIM data. Additionally, the authors acknowledge the Faculty of Engineering of the Autonomous University of the State of Mexico for providing Mexican traffic counter data.

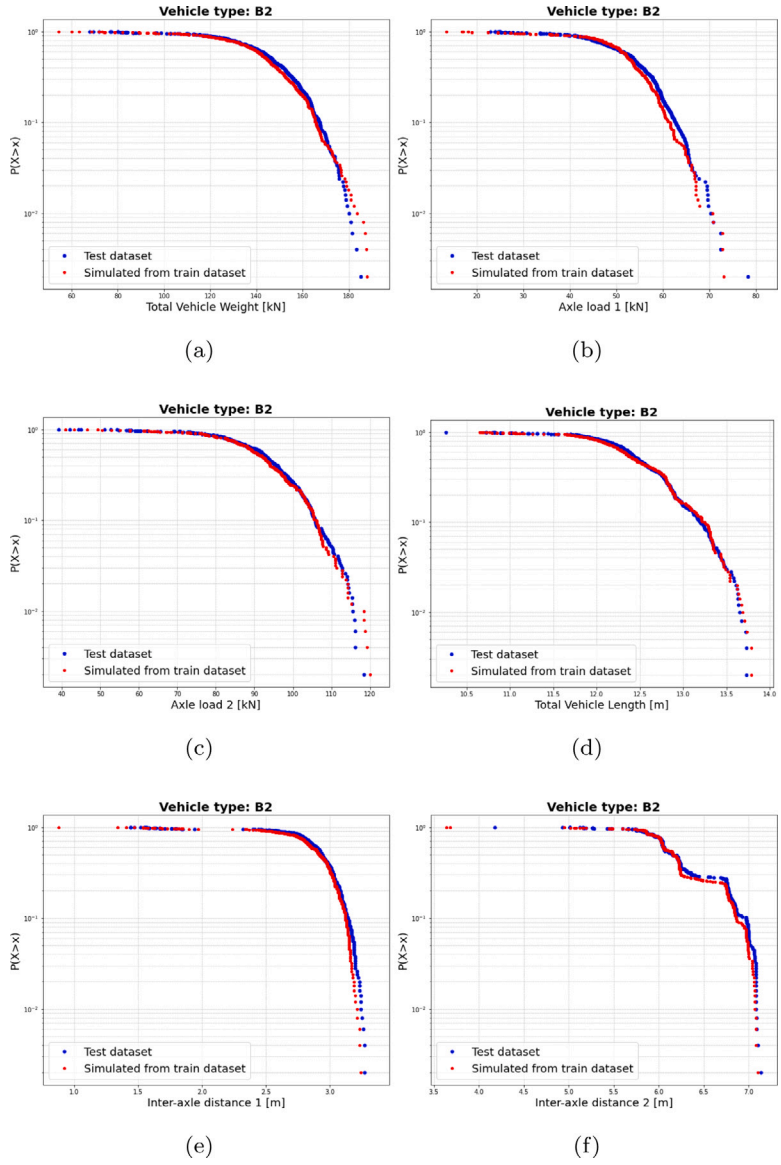

Fig. B.8. Comparison between data generated by the trained model and test data set: (a) total vehicle weight, (b) and (c) axle loads, (d) total vehicle length and (e) and (f) inter-axle distances of the vehicle type B2.

Vehicle	Type	Silhouette
1	B2	
2	B3	
3	O0	
4	O3	
5	O4	
6	O5	
7	O6	
8	O7	
9	O8	
10	O9	
11	R5	
12	R6	
13	R7	
14	R8	
15	R9	

Vehicle	Type	Silhouette
16	T3	
17	T4	
18	T5	
19	T6	
20	T7	
21	V2	
22	V3	
23	V4	
24	V5	
25	V6	
26	V7	

Fig. C.1. Available vehicle types.

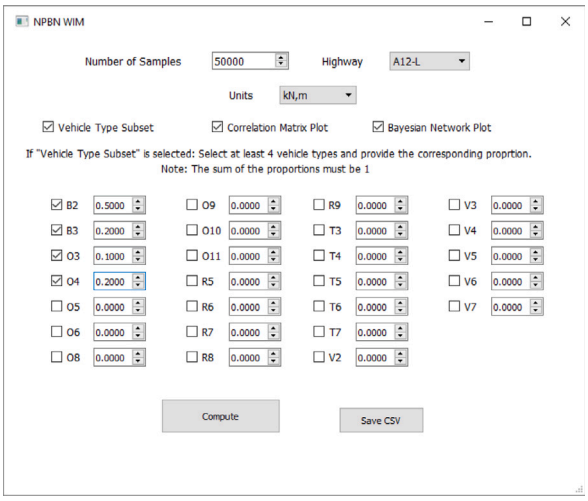

Fig. C.2. GUI main window.

Appendix A. Traffic data

See Tables A.1 and A.2 and Fig. A.1.

Appendix B. NPBN model

Table B.1 shows the unconditional rank correlations $r(X_{i,j}, X_{i,j-1})$, for all i and $j > 1$ (according to the notation in Section 3.3), between individual axle loads per vehicle type for the NPBN A15-L model. The first two columns indicate the vehicle type i and the name, the third column, the corresponding number of axles (n_i). Every entry in the next 10 columns indicates the rank correlations between the first and second axle ($X_{i,1}, X_{i,2}$), second and third ($X_{i,2}, X_{i,3}$), and so on until the 10th and 11th axle ($X_{i,10}, X_{i,11}$) per vehicle type. The last column indicates the rank correlations between the last vehicle axle and the total vehicle length ($X_{i,n_i}, X_{i,n_{i+1}}$). As the number of axles increases, the correlations are stronger in the axles close to the end of the vehicle. In contrast, the correlation between X_{i,n_i} and $X_{i,n_{i+1}}$ and, as is expected, is low.

Similarly, Table B.2 shows the rank correlations between individual inter-axle distances per vehicle type. The first two columns indicate the vehicle type (i) and the corresponding name, the third column, the number of axles per vehicle type (n_i). The fourth column indicates the rank correlation between the total vehicle length and the first axle distance ($X_{i,n_{i+1}}, X_{i,n_{i+1+1}}$). The next 10 columns indicate the rank correlations between the first and second axle distance ($X_{i,n_{i+1+1}}, X_{i,n_{i+1+2}}$), second and third ($X_{i,n_{i+1+2}}, X_{i,n_{i+1+3}}$), and so on until the 10th and 11th axle ($X_{i,n_{i+1+10}}, X_{i,n_{i+1+11}}$) per vehicle type. Notice that the correlations almost zero is observed in some cases. The corresponding rank correlations matrices between the random variables, as colour maps, for the six WIM locations can be found in Figs. B.1 to B.3. After the detailing of three categories of the framework, in the successive section the results and validation of the NPBN will be discussed.

Figs. B.1 to B.3 shows the corresponding rank correlations matrices between the random variables, as colour maps, for the six studied dutch WIM locations described in Section 3.1.

Split data analysis

To further investigate model validity, we split the data set into an 80:20 ratio (80% of the data set goes into the training set and 20% of the data set goes into the testing set). Fig. B.7 shows a comparison between synthetic data generated by the model quantified with the training data set and the observations of the test data set. The observations correspond to the A16-R highway. Additionally, Fig. B.8 shows the same comparison for the two-axle vehicle B2 of total vehicle weight, total vehicle length, axle loads and inter-axle distances. As can be seen in Figs. B.7 and B.8 the model is able to represent the data points in the training data set to a good degree. Same type of analysis was performed with data corresponding to other highways and vehicle types in our database. The patterns are similar to those briefly presented here. The model captures the main complexities of the data set.

Appendix C. GUI quick user guide

The Graphical User Interface NPBN WIM computes synthetic WIM data. The observations correspond to April 2013 for three Dutch locations in both the right (R) and the left (L) driving directions. The measurements were taken in highways A12 (km 42) Woerden, A15 (km 92) Gorinchem, and A16 (km 41) Gravendeel. Additionally, a hypothetical highway was created which is a combination of all six available WIM locations in the model. Thus, each simulated vehicle randomly chooses one of the locations to compute the synthetic data. The 26 codes (vehicle types) used in the GUI WIM consist of a letter and a number that define the number of axles. The letter represents the vehicle configurations: Buses (B), Tractor–Semitrailer–Trailer (R), Tractor–Semitrailer (T), Single-unit multi-axle vehicle and/or Single unit multi-axle vehicle–Semitrailer (V) and Others vehicles (O). For example, a seven-axle vehicle with the configuration Tractor–Semitrailer is coded as T7. The vehicle types and the silhouette are presented in Fig. C.1.

To compute the desired amount of WIM observations, in the main window of the GUI (see Fig. C.2), the user can choose between the 26 vehicle types and the seven locations (A12-L, A12-R, A15-L, A15-R, A16-L, A16-R, and Hypothetical). The option for choosing the desired units is also available. Additionally, there are three main checkboxes: (i) vehicle type subset, (ii) correlation matrix plot, and (iii) Bayesian network plot. Which actions are described next:

(i) If the “Vehicle Type Subset" check box is selected. The user needs to select at least 4 of the available 26 vehicle types and provide their corresponding proportions. Otherwise, 26 vehicle types will be used to generate synthetic observations.
(ii) If the “Correlation Matrix Plot" check box is selected. The rank correlation matrix plot will be shown as colour map (Fig. C.3(a))
(iii) If the “Bayesian Network Plot" check box is selected. The Non-Parametric Bayesian Network (NPBN) direct acyclic graph will be shown (Fig. C.3(b)).

If the user chooses a Hypothetical highway no rank correlation matrix plot nor NPBN direct acyclic graph will be shown. Once all the values are set, by pressing the button “Compute” the synthetic WIM observations will be generated. Plots of histograms and exceedance probability plots of total weight (W) and total vehicle length (L) will be generated automatically (see Fig. C.4). Finally, the computed data can be stored into a comma-separated values (CSV) file by pressing the button “Save CSV”.

Appendix D. Graphical user interface executable application

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.trip.2022.100552.

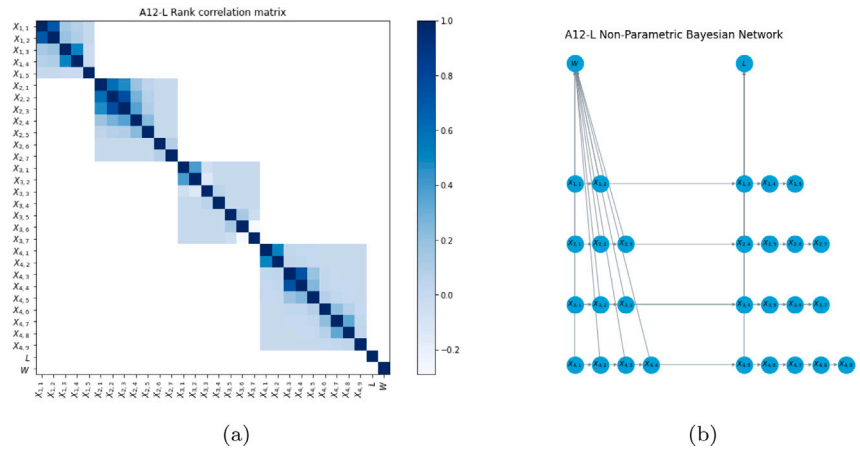

Fig. C.3. (a) Rank correlation matrix colour map and (b) NPBN direct acyclic graph.

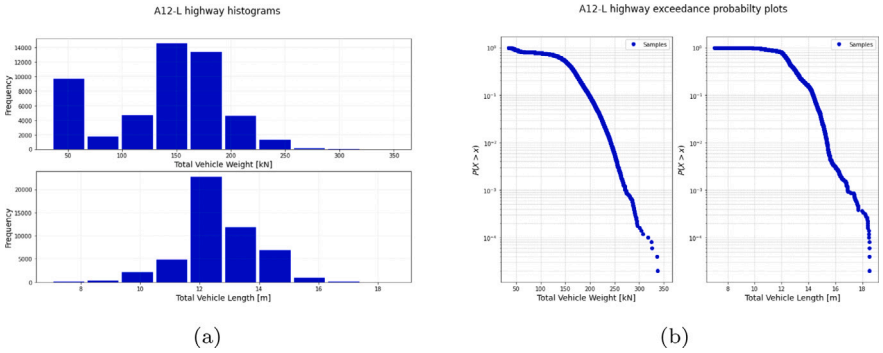

Fig. C.4. (a) Computed W and L histograms (b) Computed W and L exceedance probability plots.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19 (6), 716–723. http://dx.doi.org/10.1109/TAC.1974.1100705.

And, D.A., Vervuurt, O.M.N.A., 2015. Traffic load model for short span bridges. Tech.rep., Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek, Delft.

Beyer, P., 2015. Non-intrusive detection, the way forward. In: Southern African Transport Conference.

Crespo-Minguillón, C., Casas, J.R., 1997. A comprehensive traffic load model for bridge safety checking. Struct. Saf. 19 (4), 339–359. http://dx.doi.org/10.1016/S0167-4730(97)00016-7, URL http://www.sciencedirect.com/science/article/pii/S0167473097000167.

Departamento Nacional De Infra-Estrutura De Transportes, 2012. Quadro De Fabricantes De Veículos. Coordenação Geral de Operações Rodoviárias.

Enright, B., O'Brien, E.J., 2013. Monte Carlo simulation of extreme traffic loading on short and medium span bridges. Struct. Infrastructure Eng. 9 (12), 1267–1282. http://dx.doi.org/10.1080/15732479.2012.688753.

Federal Highway Administration's Intelligent Transportation Systems Program Office, 2007. A Summary of Vehicle Detection and Surveillance Technologies used in Intelligent Trasportation Systems. Tech. rep.

Genest, C., MacKay, J., 1986. The joy of copulas: Bivariate distributions with uniform marginals. Amer. Statist. 40 (4), 280–283. http://dx.doi.org/10.2307/2684602, URL http://www.jstor.org/stable/2684602.

Gillmann, R., 1992. Calibration and adjustment of weigh-in-motion data. Transp. Res. Rec..

Hanea, A.M., Kurowicka, D., Cooke, R.M., 2006. Hybrid method for quantifying and analyzing Bayesian belief nets. Qual. Reliab. Eng. Int. 22 (6), 709–729. http://dx.doi.org/10.1002/qre.808.

Hanea, A., Morales Napoles, O., Ababei, D., 2015. Non-parametric Bayesian networks: Improving theory and reviewing applications. Reliab. Eng. 'I' Syst. Saf. 144, 265–284. http://dx.doi.org/10.1016/j.ress.2015.07.027, URL https://www.sciencedirect.com/science/article/pii/S0951832015002331.

Joe, H., 1997. Multivariate Models And Multivariate Dependence Concepts. CRC Press.

Kentucky Trasportation Center, 2013. Wim Data Collection and Analysis. Tech.rep., University of Kentucky, Kentucky.

Kim, J., Song, J., 2019. A comprehensive probabilistic model of traffic loads based on weigh-in-motion data for applications to bridge structures. KSCE J. Civ. Eng. 23 (8), 3628–3643. http://dx.doi.org/10.1007/s12205-019-2432-9.

Kim, J., Song, J., 2021. Bayesian updating methodology for probabilistic model of bridge traffic loads using in-service data of traffic environment. Struct. Infrastructure Eng. 1–16. http://dx.doi.org/10.1080/15732479.2021.1924797.

Kreslin, A.Ž., Maja, Leahy, C., OBrien, E., Schmidt, F., Pedersen, C., 2016. Guidelines on collecting WIM data and forecasting of traffic load effects on bridges. Tech. rep, Conference of European Directors of Road.

Kurowicka, D., Cooke, R.M., 2005. Distribution-free continuous Bayesian belief nets. In: Modern Statistical And Mathematical Methods In Reliability. In: Series on Quality, Reliability and Engineering Statistics, vol. 10, World Scientific, pp. 309–322. http://dx.doi.org/10.1142/9789812703378_0022.

Marcot, B.G., 2017. Common quandaries and their practical solutions in Bayesian network modeling. Ecol. Model. 358, 1–9. http://dx.doi.org/10.1016/J.ECOLMODEL.2017.05.011, URL https://www.sciencedirect.com/science/article/pii/S0304380016308134?via{%}3Dihub.

Marcot, B.G., Penman, T.D., 2019. Advances in Bayesian network modelling: Integration of modelling technologies. Environ. Model. Softw. 111, 386–393. http://dx.doi.org/10.1016/J.ENVSOFT.2018.09.016, URL https://www.sciencedirect.com/science/article/pii/S1364815218302937.

McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the Nash–sutcliffe efficiency index. J. Hydrol. Eng. 11 (6), 597–602.

McLachlan, G. J. and Peel, D., 2000. Finite Mixture Models. Wiley Series in Probability and Statistics, New York.

McNicholas, P.D., Murphy, T.B., 2008. Parsimonious Gaussian mixture models. Stat. Comput. 18 (3), 285–296. http://dx.doi.org/10.1007/s11222-008-9056-0.

Morales-Nápoles, O., Steenbergen, R.D., 2014. Analysis of axle and vehicle load properties through Bayesian networks based on weigh-in-motion data. Reliab. Eng. Syst. Saf. 125, 153–164. http://dx.doi.org/10.1016/j.ress.2014.01.018, URL https://www.sciencedirect.com/science/article/pii/S0951832014000283 Special issue of selected articles from ESREL 2012.

Neil, M., Fenton, N., Nielson, L., 2000. Building large-scale Bayesian networks. Knowl. Eng. Rev. 15 (3), 257–284. http://dx.doi.org/10.1017/S0269888900003039, URL https://www.cambridge.org/core/product/identifier/S0269888900003039/type/journal{_}article.

Oswaldo, M.-N., M., S.R.D.J., 2015. Large-scale hybrid Bayesian network for traffic load modeling from weigh-in-motion system data. J. Bridge Eng. 20 (1), 4014059. http://dx.doi.org/10.1061/(ASCE)BE.1943-5592.0000636.

Paprotny, D., Morales-Nápoles, O., Worm, D.T.H., Ragno, E., 2020. BANSHEE–A MATLAB toolbox for non-parametric Bayesian networks. SoftwareX 12, 100588. http://dx.doi.org/10.1016/j.softx.2020.100588, URL http://www.sciencedirect.com/science/article/pii/S2352711020303010.

Pearl, J., 1988. Chapter 2 - Bayesian inference. In: Pearl, J.B.T. (Ed.), Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, San Francisco (CA), pp. 29–75. http://dx.doi.org/10.1016/B978-0-08-051489-5.50008-4, URL http://www.sciencedirect.com/science/article/pii/B9780080514895500084.

Pearson, K., 1907. On Further Methods Of Determining Correlation. In: Department of applied mathematics. University college, University of London. Drapers'company research memoirs, (vol. 16), Dulau and Company, URL https://books.google.nl/books?id=TKYSAAAAYAAJ.

Quinley, R., 2010. WIM Data Analyst's Manual. Tech.rep., FHWA Office of Pavement Technology.

Ravilla, B.S., Bein, W., Martinez, Y.E., 2021. Analysis of traffic based on signals using different feature inputs. In: Latifi, S. (Ed.), ITNG 2021 18th International Conference On Information Technology-New Generations. Springer International Publishing, Cham, pp. 239–243.

Rutherford, G., McNeill, D.K., 2011. Statistical vehicle classification methods derived from girder strains in bridges. Can. J. Civ. Eng. 38 (2), 200–209. http://dx.doi.org/10.1139/L10-128.

S., S., Devdas, M., A., M.P., 2006. Multivariate simulation and multimodal dependence modeling of vehicle axle weights with copulas. J. Transp. Eng. 132 (12), 945–955. http://dx.doi.org/10.1061/(ASCE)0733-947X(2006)132:12(945).

SCT, 2008. NOM-012-SCT-2-2008, Sobre el peso y dimensiones máximas con los que pueden circular los vehículos de autotransporte que transitan en las vías generales de comunicación de jurisdicción federal. Diario Oficial de la Federacion, p. 36.

Shahid, N., Shah, M.A., Khan, A., Maple, C., Jeon, G., 2021. Towards greener smart cities and road traffic forecasting using air pollution data. Sustain. Cities Soc. 72, 103062. http://dx.doi.org/10.1016/j.scs.2021.103062, URL https://www.sciencedirect.com/science/article/pii/S2210670721003462.

Vrouwenvelder, A., Waarts, P.H., 1993. Traffic loads on bridges. Struct. Eng. Int. 3 (3), 169–177. http://dx.doi.org/10.2749/101686693780607796.

Weber, P., Medina-Oliva, G., Simon, C., Iung, B., 2012. Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. Eng. Appl. Artif. Intell. 25 (4), 671–682. http://dx.doi.org/10.1016/J.ENGAPPAI.2010.06.002, URL https://www.sciencedirect.com/science/article/pii/S095219761000117X.

Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. 30 (1), 79–82.

Yule, G.U., Kendall, M.G., 1951. An introduction to the theory of statistics. J. Symb. Log. 16 (1), 51. http://dx.doi.org/10.2307/2268667.