A Bimodal Distribution Function with Fuzzy Regression for Predicting Truck Load Population including Overloads

Bora Jang[1*] and Jamshid Mohammadi[2]
[1]*Sharma & Associates, Inc., USA*
[2]*Department of Civil, Architectural and Environmental Engineering, Illinois Institute of Technology, USA*
**Corresponding author: bjang5@hawk.iit.edu*

Abstract: This paper presents a statistical model for truckload frequencies including those from overloads. A Bimodal distribution function with statistical parameters obtained using fuzzy regression is presented for predicting truck loads including the contribution from overloads. Overload trucks often appear as a sizeable portion of truck populations on highways. In applications when damage estimation of transportation facilities such as pavements and bridges is desired, theoretical models providing a reasonable representation of truck load populations including overloads will be useful. Load populations mostly exhibit inconsistent patterns - often with two or more distinct peaks. This is because of a combination of loaded and empty trucks as well as overloads in the population. And as such, simple statistical distribution models fail to portray a realistic representation of truck load populations. In this paper, several truck load populations for 5-axle vehicles were acquired from the States of Illinois and Michigan in US and used to demonstrate the suitability of Bimodal models in representing the data. The goodness-of-fit tests using both the traditional Kolmogorov-Smirnov (K-S) and more modern Anderson-Darling (A-D) methods were used to demonstrate the suitability of a particular model to represent the data. The results show that a combination of beta and lognormal distribution can conveniently be used as a suitable model to represent the truck load populations that were analyzed in this study. Using the Bimodal model, theoretical distributions are developed to (1) represent the entire truck load population with WIM data; and (2) predict the population with limited data. If no truck load population is available, the Bimodal distribution model can still be used, with certain assumptions, based on fuzzy regression using traffic pattern information available for the roadway.

Keywords: Truck overloads, statistical models, mixed distribution function, Kolmogorov-Smirnov (K-S) test, Anderson-Darling (A-D) test, fuzzy regression

1. Introduction

1.1. Truckload Distributions and Overloads

Truck load data play an important role in condition assessment and life-cycle management of transportation systems such as pavements and bridges. Methods used in condition assessment of these systems require a realistic estimation of wear and tear and damage accumulation as a result of the repeated application of truck loads. For a given system, damage assessment can be done by using truck load data directly in a discrete format (Jang and Mohammadi, 2018). However, such a method will only be specific to the system for which the truck load data are available. Furthermore, since the data is used in a discrete format, the frequencies of load occurrences can only be described with specific ranges. Each range will have an upper and lower limit;

and any other load value within the limits must be approximated with either the upper or the lower limit. A more versatile and realistic approach in damage assessment can be achieved by using a theoretical model that best represents the truck load data. The advantage of using a theoretical truck load data model is that (1) it affords the formulation of the damage assessment independent of the type of structure and in a rather theoretical format; and (2) the model allows the continuity in the load data and eliminates the issues with the discrete data ranges as described.

For a given class of trucks, the intensity and frequency of load occurrences in the population vary to a great extent. The load population does not often show a consistent pattern; and as such, a single probability distribution function is generally unable to represent the entire population. Often the distribution exhibits two or more distinct peaks. This is attributed to the variety of loads in the population; and in most cases, the appearance of two peaks indicates a combination of loaded and empty trucks in the mix (Jang and Mohammadi, 2017). Load data also contain overload frequencies that may be significant in some cases resulting in multiple peaks in the population. The term overload, as used here, refers to truck weights in excess of the 356 kN (80 kips) limit.

Studies on the effect of truck load on highway systems have shown that the load population often contains occurrences of overloads that in certain cases may be frequent (Mohammadi and Shah, 1992; Snyder et al., 1985; and Timm et al., 2005). According to a recent report from the National Cooperative Highway Research Program, overload occurrences increase over time based on the trend in the Federal Highway Administration (FHWA) data (CPCS, 2016). Observing dual peaks appearing in the load data for 5-axle trucks, Mohammadi and Shah (1992) used a theoretical mixed distribution model to represent the occurrence of the dual peaks in truck load populations. Timm et al. (2005) used a distribution model as a mix of multiple probability functions to represent axle load spectra. They offered a combination of lognormal and normal distributions, which was supported by some statistical analyses.

Available related studies on truckloads in general and truck overloads in particular have primarily focused on the significance of load effects in highway bridges and presenting load models that can be used in bridge design and evaluations. Among related studies addressing truck live loads, those by Nowak (1994) and Nowak and Hong (1991) are reported here. Nowak (1994) used load data from truck survey for the purpose of developing live load models for bridges in Ontario, Canada. Although no specific truckload model is introduced, the live load models from load effects such as bending moments and shear forces are developed and recommended for use in bridge design. The study by Nowak and Hong (1991) is probability-based and provides a statistical representation of the bridge live load and related parameters (bridge span length and number of lanes). Again, no specific reference to a model that can represent load truck by axles and especially overloads is provided. Miao and Chan (2002) conduct a similar study using WIM data from Hong Kong and extreme bridge girder load effects (such as estimates of daily occurrences of extreme bending moments and shear forces) and developed statistical models for bridge live loads. Specific to this study is the introduction of statistical models for gross vehicle weight estimation using the inverse normal probability values and Type I extreme value distribution. They compared their load models with those in AASHTO and noted a relatively close agreement. Their study indicates that using such models, the gross vehicle weight for the bridge design life is 41.8 tons with a probability of 0.95 and 60 tons with a probability of 0.98. These values translate respectively to 409 kN (92 kips) and 587 kN (132 kips). Both these values correspond to overload situations in in US; and the load models presented indicate that there are 0.05 and 0.02 probability that the truck weight will actually exceed 409 kN (92 kips) and 587 kN (132 kips), respectively.

Fu and Hag-Elsafi (2000) used the New York State WIM data in developing load models for bridges including overloads. They present samples of gross vehicle data, which includes a significant presence of overloads. Their model is primarily based on a lognormal distribution and focuses on providing equations applicable to load effects (such as bending moment) by treating the load source (i.e., gross vehicle weight) as a random variable. In presenting load distributions with overloads, they suggest combining probability values for normal load and overloads. Sample results presented for populations with significant overload presence clearly show the appearance of two peak values as expected and reported by others as well.

Tabatabai et al. (2017) use WIM data in Wisconsin for a period of one year to develop statistics for extreme truck loads for Class 9 category (which includes 5-axle single trailer trucks). For the data analyzed, the Class 9 constitutes about 62% of the entire data. The data indicates a 95% probability for a 467 kN (105 kips) with the maximum load recorded at 1,079 kN (242 kips). The study indicates that the load modeling for this class of trucks can be used to represent all other categories. Statistical analyses conducted considered the multi-modal effect of the data thus proposing to use mixed probability values. However, each probability function for a specific load range was determined around the peak value in that region. The model was developed by combining these different probability functions in which each function appears in the equation by a weight. The weights correspond to the number of sample data in different data ranges. The model indicates a significant amount of overload gross vehicle weight present in the load population. It was also compared with simulation data.

The main objective of this work is to use samples of truck load data and develop a mathematical model to represent a reasonable distribution for overload trucks. The use of such models is in applications where methods for transportation system damage assessment require information on the frequency and magnitude of loads, especially overloads. Specific transportation systems that can benefit from these models include bridges and pavements. In most part, real WIM data are used for damage assessment of these systems. However, in areas where WIM data is not available, models (such as those presented in this paper) will be helpful to estimate information on load frequency and magnitude.

1.2. Statistical Analysis

Specific to truck load data, available data is in discrete forms compiled primarily using a Weigh-in-Motion (WIM) system. As such data are compiled in only a limited number of stations across states, their use may become limited to the locations and states from which the data have been obtained. The use of mathematical models, that describe the truck load data through theoretical statistical distributions, may be an approach offering models that are more universally applicable to all locations. Since the discrete truck load distributions often show multiple peaks, conceivably, “mixed-type” distributions may be suitable models to represent them (Mohammadi and Shah, 1992). The overloads especially add more complexity to the shape of the load distribution, since they occur with less frequencies and constitute the area of the load population at the extreme upper tail of any theoretical distribution model.

A goodness-of-fit test is used to validate a theoretical distribution model when a particular probability distribution is specified to develop the model for a random set of data. Several studies have performed goodness-of-fit tests in addition to regression analyses to obtain suitable theoretical distributions to truckload data (Mohammadi and Shah, 1992; Timm et al., 2005). Such studies involve compiling actual data from the WIM systems and conduct statistical tests to determine the type of probability distribution models that would represent the load data. When mixed-type distribution models are involved, an extra parameter of the distribution model includes the common value that boundaries between any two functions.

Statistical test methods used in this paper are Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) tests (Ang and Tang, 2007; Jang, 2018). The A-D test has advantages over the K-S test; and it may be suitable for application to the WIM data, since the sample-size is very large. Since most overloads occur at the extreme end (tail) of the load population, the A-D test offers another advantage in the sense that it considers weights for these extreme values.

2. Weight-In-Motion Data used in Analysis

Nine different sets of WIM data are used in this study to find the characteristics of the truckload distribution and the truck overloads. The WIM data were collected from six different stations by the Illinois Department of Transportation (IDOT) and three different stations by Michigan Department of Transportation (MDOT) from June, 2014 to May, 2015 (Jang and Mohammadi, 2017). WIM data sets used in this study are calibrated by providers, IDOT and MDOT, to control any errors associated with the vehicle class, axle spacing and weight, speeds, etc.

A long enough period of time is needed to capture a realistic representation of truckload frequencies. In highway bridges, a two- or three-day period seems to be satisfactory (Hahin et al., 1993). Approximately, 4 weeks and a year of data represent 90% and 99% of the level of reliability, respectively. The data acquisition process may have to be conducted in several two- or three-day sessions at various times during the year to include any seasonal changes in the traffic pattern. Seasonal changes are usually caused by weather or other local factors. The effect of these changes is a periodic pattern in time that completes a cycle within a calendar year and is thus repeated on a yearly basis (Qu et al., 1997). The specific features of truck load population remain almost the same although the average yearly truck traffic changes from one highway to another (Mohammadi and Polepeddi, 2000). As such, a reasonable estimate of the population of overloads can be obtained based on the yearly truck traffic data available for the condition assessment a given system.

To better understand the distribution of truck weights in WIM data, it is necessary to analyze the data further by looking closely at the axle load and vehicle class data and specifically determining the presence of overloaded gross vehicle weight (GVW) in the data pool. The information on the frequency and distribution of the overloads in the data pool is especially important considering the fact that these loads, especially when occurred frequently, may accelerate the damage to pavements and bridges.

The load spectra from the WIM data are obtained in terms of the number of axles and the vehicle classifications in different data locations by developing a MATLAB program (MATLAB, 2016). Figure 1 shows the truck population of vehicle class 9 in the five-axle truck category for the combination of nine WIM data. In terms of the truck population according to the number of axles, the five-axle trucks constitute the bulk of the truck population except for the data from two of the nine stations.

Furthermore, it is noted that the truck population indicates that the vehicle class 9 is dominant in the five-axle category in the population except for the data from two of the nine stations. In addition to data from the five-axle and vehicle class 9 trucks; these two data sets appear to also include data from two-axle and the vehicle class 2 and 5 trucks (which does not include overloads). It is indicated that a significant portion of the overload data is in fact in the load data from the vehicle class 9 and the five-axle truck population. Furthermore, these types of trucks make up of a relatively large percentage of the overall truck populations on highways (as portrayed in WIM station data). Therefore, the WIM data for five-axle truck and vehicle class 9 were considered in this study, as the representative of a typical truck load data.

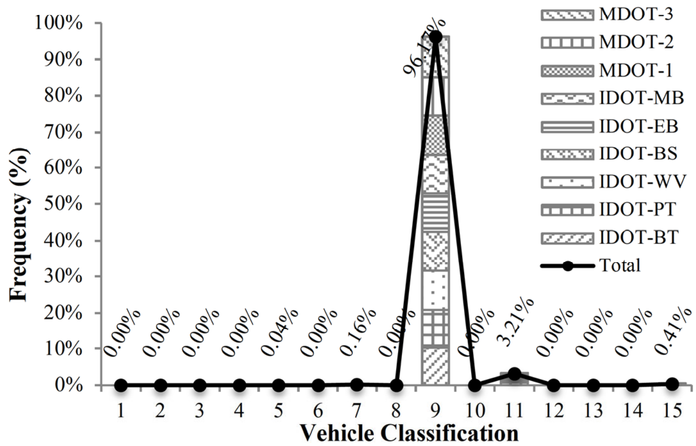


Figure 1. Truck population versus the vehicle class 9 in the five-axle truck category

3. A Bimodal Distribution Model for Overloads

The truckload distribution for the gross vehicle weight that exhibits two distinctive peaks is referred to a Bimodal distribution. A Bimodal distribution indicates a need for two different continuous probability distributions (f_1 and f_2) in which the overall truckload data is divided into two parts with a common load, S_L, representing the boundary between the two functions. As a simple method, the common load can be selected subjectively by visual examination of the data through trial and error. Alternatively, one may treat the common load as an extra parameter in the overall Bimodal distribution model, which will need to be estimated using statistical methods such as the method of maximum likelihood.

In developing the Bimodal distribution model, it was decided to use beta and lognormal distributions because of (1) the fact that they only accept positive values; and (2) others have also recommended them (e.g., Mohammadi and Shah 1992). Several combinations of beta and lognormal distributions were tried for functions, f_1 and f_2. The combinations considered included beta-beta, lognormal-lognormal, and beta-lognormal distributions (Jang, 2018). The beta distribution is among a few distributions that are appropriate for bounded random variables. The beta distribution, for a zero-lower bound, is (see for example, Ang and Tang, 2007)

$$f(s) = \frac{1}{B(q,r)} \frac{s^{q-1}(s_0 - s)^{r-1}}{{s_0}^{q+r-1}} \tag{1}$$

In which q and r are the shape parameters of the beta distribution, s_0 is the upper limit, and $B(q,r)$ is the beta function. Parameters, q and r, are positive and determine the shape of distribution. The beta function is written in the following form.

$$B(q,r) = \int_0^1 x^{q-1}(1-x)^{r-1}dx \tag{2}$$

The lognormal distribution may especially be suitable to represent overloads in the truckload frequency distribution with a one-sided tail effect. The function is written as follows (see for example, Ang and Tang, 2007):

$$f(s) = \frac{1}{\sqrt{2\pi}(\zeta s)} exp\left[-\frac{1}{2}\left(\frac{ln\, s - \lambda}{\zeta}\right)^2\right] \tag{3}$$

In which λ and ζ are the location and shape parameters of the lognormal distribution, respectively.
Figure 2 describes the concept of mixed density function with consideration for the common load.

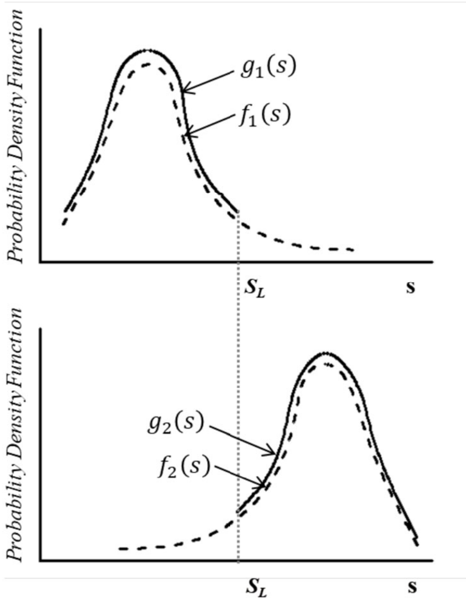


Figure 2. Schematic of mixed density function with consideration for the common load (S_L)

The probability density function for each part in the population will need to be adjusted since f_1 and f_2 are valid for $s < S_L$ and $s > S_L$, respectively (see Mohammadi and Shah, 1992). The adjusted density functions, g_1 and g_2, are therefore obtained from the proportion of each probability in the following forms:

$$g_1(s) = \frac{f_1(s)}{F_1(S_L)} \tag{4}$$

and

$$g_2(s) = \frac{f_2(s)}{1 - F_2(S_L)} \tag{5}$$

In which F_1 and F_2 are the probability distribution functions of f_1 and f_2. Using these adjusted probability density functions, the mixed density function is then written as

$$g(s) = \begin{cases} \alpha_1 g_1(s), & for\ s \le S_L \\ \alpha_2 g_2(s), & for\ s > S_L \end{cases} \tag{6}$$

where parameters α_1 and α_2 represent the proportions of entire load population that are below and above the common load, respectively. Note that $\alpha_1 + \alpha_2 = 1$.

4. Testing Goodness-of-fit of Load Distribution Models

As indicated earlier, to test the validity of a mix probability function to represent the data, statistical methods such as the Kolmogorov-Smirnov (K-S) test and the Anderson-Darling (A-D) test are useful (Ang and Tang, 2007; Anderson and Darling, 1954; Stephens, 1974). The WIM data were analyzed and tested against several types of statistical models when combined together. For each type, the aforementioned K-S and A-D tests were conducted in an effort to identify a Bimodal distribution model to represent the truck load data including the significance of the overloads that appear in the extreme right-hand tail of the model. The analysis shows that a combination of beta and lognormal distributions is especially useful in representing the data.

4.1. Kolmogorov-Smirnov (K-S) test

The details of the K-S test for similar engineering applications can be found for example in Khisty et al. (2012). The test is considered to be suitable for comparing the observed cumulative frequency with the cumulative distribution function of an assumed theoretical distribution. The maximum difference between the observed cumulative frequency function, $S_n(x)$, and the theoretical cumulative distribution function, $F_X(x)$, is the measure of discrepancy between the two. The maximum difference, D_n, is determined by

$$D_n = \max_x |F_X(x) - S_n(x)| \tag{7}$$

For a given significance level, the test passes if the value of the maximum difference is smaller than a critical value. We can consequently believe that the theoretical distribution model is appropriate to represent the observed data.

4.2. Anderson-Darling (A-D) test

Since the application of this test is less known in engineering, the details of the test are presented herein. The test is sensitive to discrepancies between the observed and theoretical frequencies; and as such, it places more weight at the tails of the distribution rather than near the median. This is certainly an advantage over the more traditional K-S test. The A-D statistic can be calculated by the following equation (see for example, Ang and Tang, 2007):

$$AD^2 = -\sum_{i=1}^{n}\left[\frac{(2i-1)\{lnF_X(x_i) + ln[1 - F_X(x_{n+1-i})]\}}{n}\right] - n \tag{8}$$

In which, AD^2 is the Anderson-Darling statistic, $F_X(x)$ is the theoretical cumulative distribution function, and n is the sample size. A small sample size is not valid for the A-D test and hence the sample size should be larger than 7.

The A-D statistic would receive more contributions from the tails of distribution, since it is expressed in terms of the logarithm of the probabilities. However, the critical values for the A-D test are dependent on the specific distributions being tested. Tables of critical values are currently not available exclusively for the beta distribution and for any combinations of multiple distribution functions. Nevertheless, the values of AD^2 are valuable for use as a comparable measure when various combinations of different distribution models are considered. For example, if AD^2 is relatively large, the reason may be attributed to the application of Bimodal distribution in representing the data with the appearance of a cluster of data triggering the larger value for AD^2.

5. Results and Discussions of Statistical Assessment

Each of the nine load spectra of the WIM data is modeled using several possible combinations of beta and lognormal distributions. The statistical tests require rearranging data into discrete ranges of the loads. After several trials, load ranges in 22.25 kN (5 kips) intervals (or less) are found to be suitable for the analysis. For the common load values lower than 200 kN (45 kips) and greater than 311 kN (70 kips), the Bimodal distribution model could not be established with distinct multiple peaks (which are the characteristics of the truck load populations). Thus, a suitable common load value is anywhere between 200 to 311 kN (45 to 70 kips) based on the frequencies of various load values in the entire truck load population.

Based on the aforementioned conditions, the proposed mixed probability distribution models were developed for various common loads corresponding to three kinds of the combination of beta and lognormal distributions as described earlier (i.e., beta-beta, lognormal-lognormal, and beta-lognormal). The goodness-of-fit tests are conducted to verify the validity of the proposed mixed probability distribution models for each case by the K-S and A-D methods. It is noted that the K-S and A-D tests are conducted to also determine the appropriate upper bound of beta distribution for the density function f_2 (see Eqs. 4-6) in the model made up of a combination of two beta functions. Common load values between 200 to 311 kN (45 to 70 kips) are considered in performing the K-S and A-D tests. Figure 3 illustrates the results of K-S and A-D tests for different values of the upper bound of beta distribution and different WIM stations. The minimum values of the test statistic (i.e., the maximum difference between theoretical and observed values in the K-S and AD^2 in the A-D tests) in terms of different values for the common loads are computed and plotted as shown in Figure 3. A theoretical distribution model, in which the upper bound is 556 kN (125 kips), results in the lower values of both tests for all WIM stations. Thus, the results using this upper bound value will be used to compare the three possibilities for the combination of beta and lognormal distributions, as described earlier.

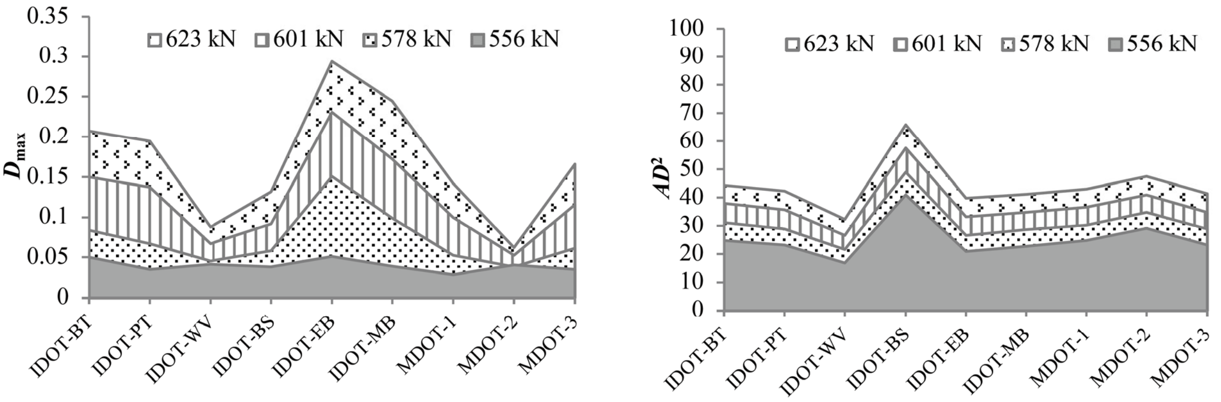


Figure 3. Test statistics (D_{max} in K-S and AD^2 in A-D tests) in terms of various values of the upper bound of beta distribution for nine WIM stations

Figure 4 represents the results of K-S and A-D tests for different values of the common load and combinations of the Bimodal distribution models for all WIM stations. As seen in the results, no test statistics value for the combination of two beta distributions is listed near the common load over 245 or 267 kN (55 or

60 kips). This is because the beta distribution is found to be unsuitable to present the distribution for load values larger than the common load.

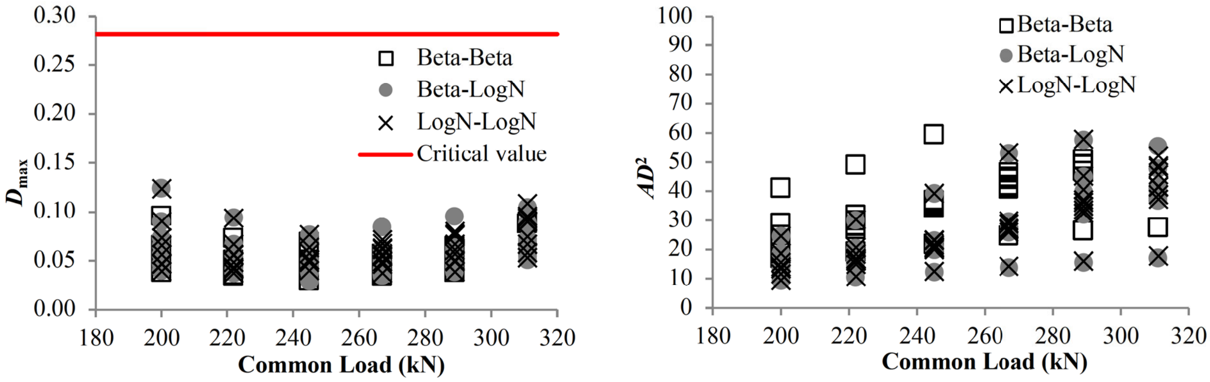


Figure 4. Results of K-S and A-D tests with various common loads and combinations of Bimodal distribution for the nine WIM stations

For the K-S test, all theoretical distributions for nine WIM stations are determined to be acceptable to model the load data at 5% significance level (critical value = 0.282). Considering various common load values, a combination of the two beta and beta-lognormal distributions with the common load at 245 kN (55 kips) is most suitable to represent the entire truck load population per results from the K-S test.

Since the critical value in the A-D test is not known, the method is used as a way of comparatively deciding the suitability of one model over the others. The A-D statistics mainly decrease as the common load decrease for most cases. The combination of two beta distributions results in greater values for the test statistics (AD^2 values) than the other kinds of Bimodal distribution for all common load cases. Furthermore, Bimodal distributions containing at least one lognormal distribution represent very similar values in the A-D statistics for all common load cases. This is attributed to the fact that the lognormal distribution tends to dominate the load population at load values larger than the common load. The overall results show that a combination of the beta and lognormal distribution yields a favorable model with the smallest A-D statistics. It is reasonable to believe that the lognormal distribution is more suitable in representing the data in the portion of the population where the second peak appears, which includes the overloads in the tail of the distribution model. The results indicate that the combination of beta-lognormal distribution model with the common load of 200 kN (45 kips) is perhaps well suited to represent the data.

The K-S test generally draws conclusions based on the middle portion of the data. And as such, the data at extremes (at tails) is not very well represented. On the other hand, the A-D test draws conclusions more on the extreme loads. Therefore, if only the overload distribution is of interest, the A-D test is perhaps more appropriate. However, in general because the entire population is needed for damage estimations, it makes sense to rely on both tests (i.e., K-S and A-D tests) in any given problem. For this reason, we suggest a range for the common load be specified (in this case, 200 to 245 kN (45 to 55 kips)), so that the significance of both the middle and the tail data effects is incorporated in damage estimation of an affected transportation system (pavement or bridge). Depending on the focus of analysis, however, one of three values (200, 222, or 245 kN (45, 50, or 55 kips)) may be used in practice. For analysis in this study, as a result of the two tests, and at least within the limited data analyzed for the Illinois and Michigan truckloads, it is reasonable to conclude that a combination of the beta and lognormal distributions with the common load of 200 kN (45 kips) is a suitable theoretical distribution for truckload data (including the overloads).

It is noted that for most situations, and especially for 5-axle in vehicle class 9, the range of common loads specified would be the representative of the data in a Bimodal distribution model. In fact in situations where the overloads make up a significant portion of the overall load population, and when overloads are at unusually large values, the common load may become larger than the range specified in this study. However, such conditions are very rare, especially considering the fact that very large overloads happen under special conditions and very infrequently.

6. Proposed Distribution Model when WIM Data is available

In situations when the truck load population is available, the theoretical distribution proposed in this study may be used to describe the entire truck load population. Figure 5 shows a flowchart on how the load population may be quantified for the case if WIM data is available.

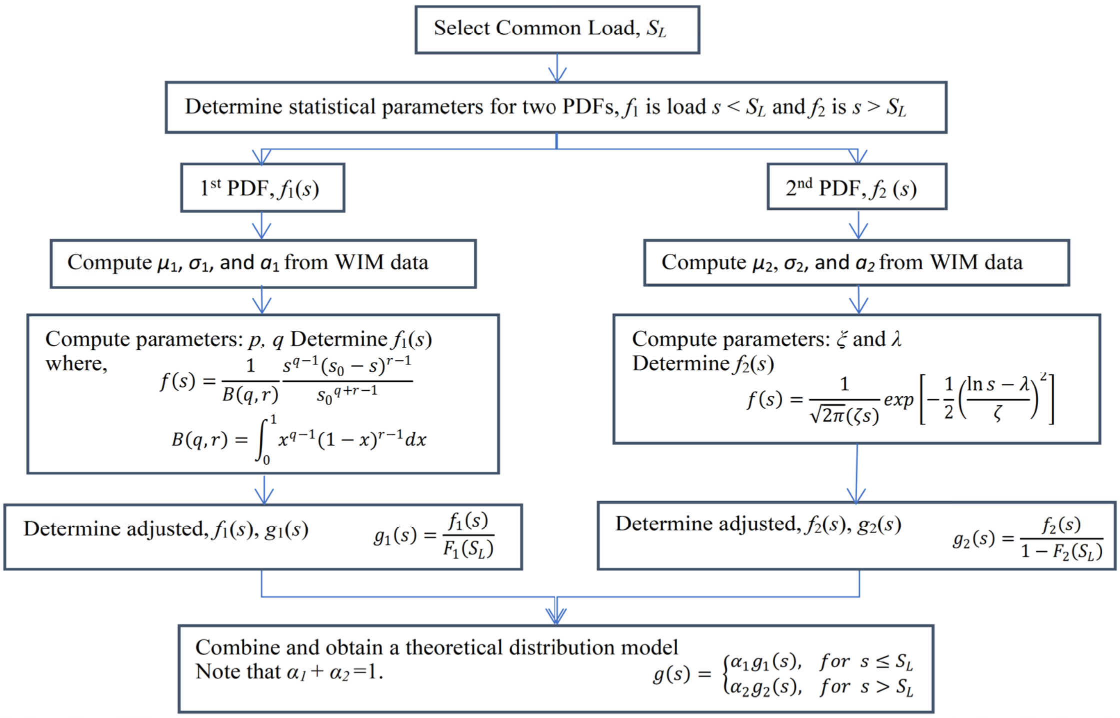
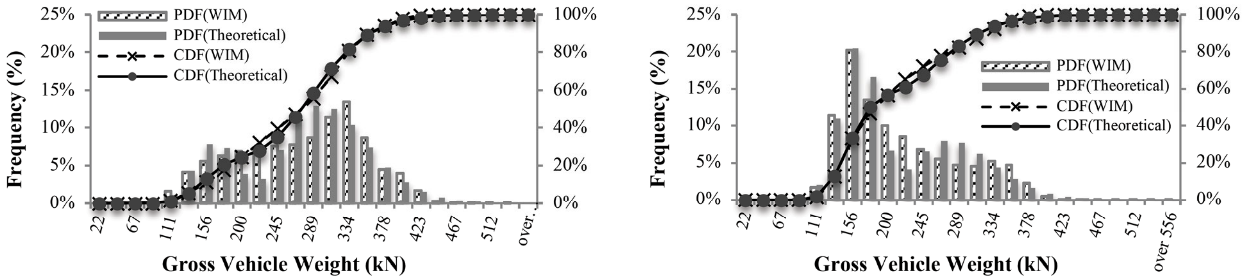


Figure 5. Results of K-S and A-D tests with various common loads and combinations of Bimodal distribution for the nine WIM stations

The parameters driving the Bimodal distribution modes are then obtained using the available WIM data. A common load, S_L, needs to be selected among possible values (i.e. 200, 222, or 245 kN (45, 50, or 55 kips)) for the truck load population. Then the mean and standard deviation (i.e. μ_1, σ_1, μ_2, and σ_2) for the two parts of the load population (i.e., the portion representing values of load lower than S_L and those representing values of load larger than S_L) are computed. Furthermore, the percentage of the load frequencies in each portion, i.e. α_1 and α_2, are computed, in which α_1 is the ratio of the sample points in s < S_L divided by the total sample points; while α_2 is the ratio of the sample points in s > S_L divided by the total sample points. The theoretical distribution can then be determined from these parameters.

As an example, the frequency and cumulative probability distribution for the observed WIM data and the theoretical model using the Bimodal distributions for four WIM stations are illustrated for the common load of 200 kN (45 kips) in Figure 6.



(a) IDOT-WV (b) MDOT-2

Figure 6. The frequency and cumulative probability distribution for the observed WIM data and the theoretical model for four WIM stations

7. Proposed Distribution Model when WIM Data is not Available

If no truck load population is available, one may still use the theoretical models proposed in this study with certain assumptions and based on the roadway traffic and other information. The average daily truck traffic (ADTT) and the number of overload permits that an agency issues for bridges along the roadway may be used as basic information in developing a model for the load distribution. As a first step, using the available information, the percentage of overloads in the entire truck load population can be determined as:

$$OL^* = \frac{the\ number\ of\ permits\ per\ year}{ADTT \times 365} \tag{9}$$

In which, OL^* is the minimum percentage of overloads in the entire truck load population and *ADTT* is the annual daily truck traffic. In the absence of WIM data, statistical parameters such as μ_1^*, σ_1^*, μ_2^*, and σ_2^* have to be estimated based on the available information using regression analysis – either through numerical or logical schemes.

7.1. Fuzzy Regression Analysis

In general, a regression analysis, especially, a linear regression model, is commonly used to analyze relationship and correlation between dependent and independent variables and to predict the trend in the data set (Ang and Tang, 2007). Usually, a large number of collected data is required. If the data set is of small size, the relation between variables is imprecise. In such cases, rather than the statistical linear regression model, logical regression analysis may be used as described in this paper. Statistical parameters (i.e., μ_1^*, σ_1^*,

μ_2^*, σ_2^*, etc.) depend on volume of traffic, which can be used as the basic information in deriving these parameters.

A fuzzy regression model was developed by Tanaka et al. (1982), based on the fuzzy logic to treat the imprecise and uncertain phenomenon in a data set. The main approach in this model is to minimize the data fuzziness (and thus uncertainty) using a linear programming method. Over decades, many researchers have conducted studies to improve the outcome of the fuzzy regression model (Tanaka et al., 1982; Lee and Tanaka, 1999; Wang and Tsaur, 2000; Chen and You, 2014). In this paper, the classical regression analysis with crisp-input (i.e., the percentage of overload, *OL*) and fuzzy-output (i.e., statistical parameters required in the Bimodal distribution for predicting the truck load population) is presented in evaluating the relationship between variables. The model includes a confidence interval in treating the fuzzy nature of the relationship between variables.

The general form of fuzzy linear output by Tanaka et al. (1982) is as follows:

$$\widetilde{Y_i} = \widetilde{A_0} + \sum_{j=1}^{m} \widetilde{A_j} x_{ij} \tag{10}$$

In which, $\widetilde{Y_i}$ is the fuzzy linear output with $i = 1, 2, \ldots, n$, $\widetilde{A_0} = (a_0, c_0)$, $\widetilde{A_j} = (a_j, c_j)$ are fuzzy coefficient with $j = 1, 2, \ldots, m$, $x_{ij} = [x_{1j}, x_{2j}, \ldots, x_{nj}]$ and is a vector of input variables with respect to j^{th} fuzzy coefficient. In the fuzzy coefficient (a_j, c_j), a_j is a mode and c_j is the spread in the membership function (MF). A symmetric triangular model is considered in this study; and the expression for MF is written as follows:

$$MF(a_j) = \begin{cases} max\left(1 - \dfrac{|a_j - c_j|}{c_j}\right), & a_j - c_j \leq a_j \leq a_j + c_j \\ 0, & otherwise \end{cases} \tag{11}$$

The shape of membership function can be selected based on the distribution of data and can be in an asymmetric triangular or trapezoidal form.

To obtain the result of fuzzy linear regression analysis with crisp-input and fuzzy-output, the total degree of fuzziness (spread) is minimized as follows:

$$\begin{aligned} &Minimize \\ &\qquad \sum_{i}^{n}\left(c_0 + \sum_{j=1}^{m} c_j |x_{ij}|\right) \\ &Minimize \\ &a_0 + \sum_{j=1}^{m} a_j x_{ij} + (1-h)\left[c_0 + \sum_{j=1}^{m} c_j |x_{ij}|\right] \geq y_i + (1-h)e_i \\ &a_0 + \sum_{j=1}^{m} a_j x_{ij} - (1-h)\left[c_0 + \sum_{j=1}^{m} c_j |x_{ij}|\right] \leq y_i - (1-h)e_i \\ &c_j \geq 0, \quad 0 \leq h \leq 1, \quad i = 1,2,\ldots,n, \quad j = 1,2,\ldots,m \end{aligned} \tag{12}$$

In which, in addition to the aforementioned variables, y_i is the mode and e_i is the spread in a symmetric triangular membership function for the fuzzy output, respectively, and h is a factor to control the level of support (spread) in the membership function. The support of the fuzzy regression model prediction shows the range of possible values. All data points should be included in the interval created by the h index. Increasing the value of h the possible values of output which is not presented in the data set will be considered in the interval. The value of h can be determined through an expert knowledge.

After estimating the fuzzy output number based on the fuzzy regression, the crisp value of output can be determined by the defuzzification procedure. Figure 7 presents a popular method, which involves the calculation of the centroid for the area under the membership function. It is suggested to select the defuzzification method based on the form of membership function used in the analysis.

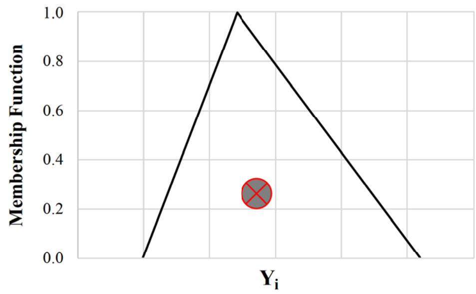


Figure 7. The fuzzy membership function (MF) and defuzzification: triangle shape and centroid calculation

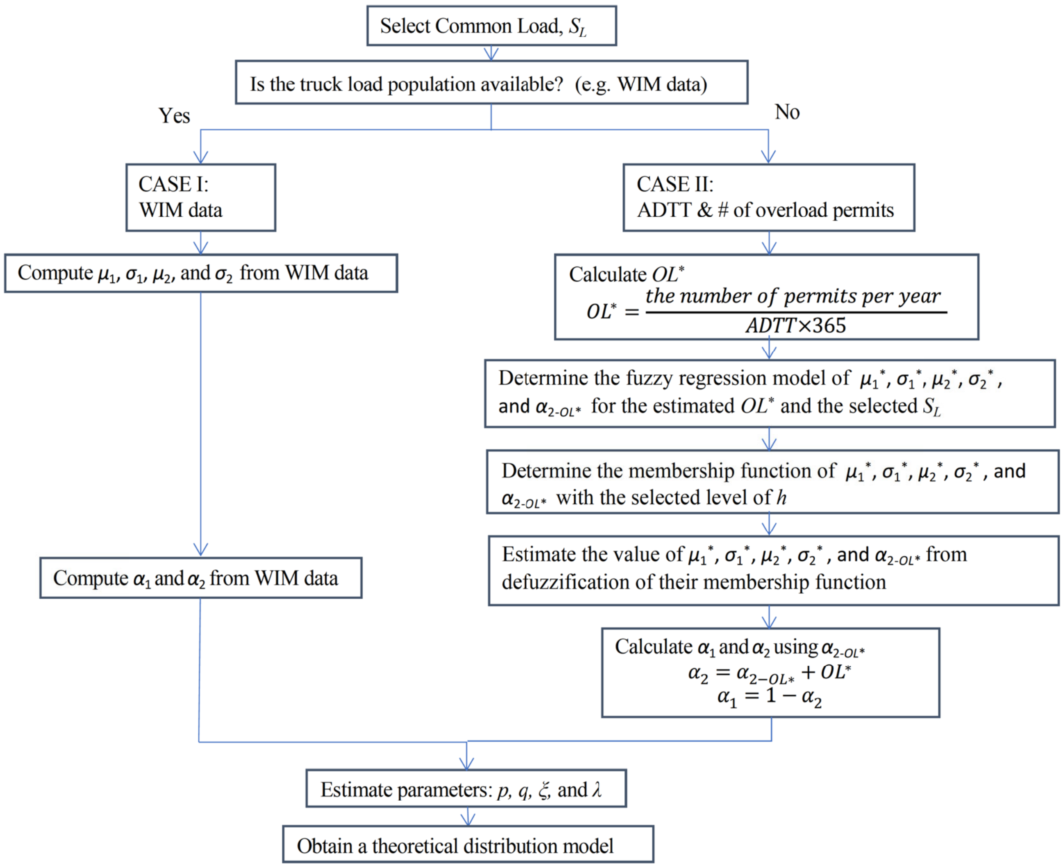


Figure 8. Flowchart for obtaining a suitable statistical truck load distribution model

Statistical parameters determined from the fuzzy regression model are then used to obtain the Bimodal distribution for the load population in the absence of any WIM data. Figure 8 shows a flowchart on how the load population may be estimated for the two cases of (1) WIM data is available; and (2) no WIM data is available.

7.2. Numerical Example and Discussions of Fuzzy Regression Model

Upon selecting the common load, i.e. 200, 222, or 245 kN (45, 50, or 55 kips), for the truck load population, approximate values for the mean and standard deviation of the two portions of the load population (i.e. μ_1^*, σ_1^*, μ_2^*, and σ_2^*) and the parameters α_1 and α_2 can be estimated. It is noted that we must first compute an auxiliary parameter (i.e. $\alpha_{2\text{-}OL}^*$) from OL^* in Eq. 9 and then computing α_1 and α_2 as shown in Eq. 13. The estimates for α_1 and α_2 are:

$$\alpha_2 = \alpha_{2-OL}(OL^*) + OL^* \ \ and\ then\ \ \alpha_1 = 1 - \alpha_2 \tag{13}$$

As an example, considering a common load, S_L = 200 kN (45 kips), the membership function of fuzzy output from the regression analysis is developed and presented in Figs. 8 and 9.

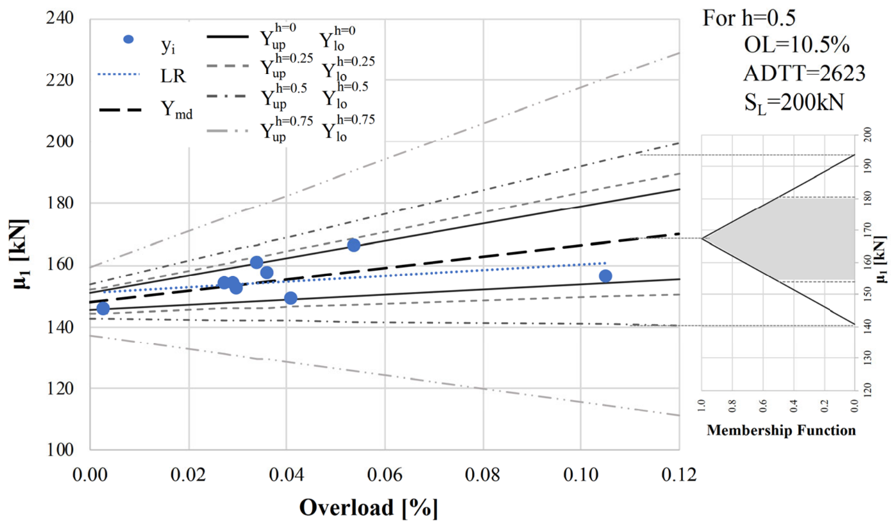


Figure 8. The fuzzy linear regression model for μ_1^* for S_L = 200 kN (45 kips) and a triangular MF

Using the fuzzy regression model given h, the values of the statistical parameters for μ_1^*, σ_1^*, μ_2^*, σ_2^*, and $\alpha_{2\text{-}OL}^*$ can be estimated. In Figure 8, the shaded area represents the observed fuzzy output without the consideration for h. The additional area (not shaded) expands the confidence interval and increase the probability to capture the data beyond the available set. It is further noted that $\alpha_{2\text{-}OL}^*$ is also bounded by 0 and 1 limits, i.e., $0 \leq \alpha_{2-OL}^* \leq 1$.

The fuzzy linear regression model results provided in Figures 8 and 9 are developed based on the data set collected. These results are reasonable predicted values of the parameters according to the fuzzy regression model with limited available data. When additional data becomes available, the shape of boundary and membership function may be calibrated for more robust estimates for the parameters of the truck load distribution.

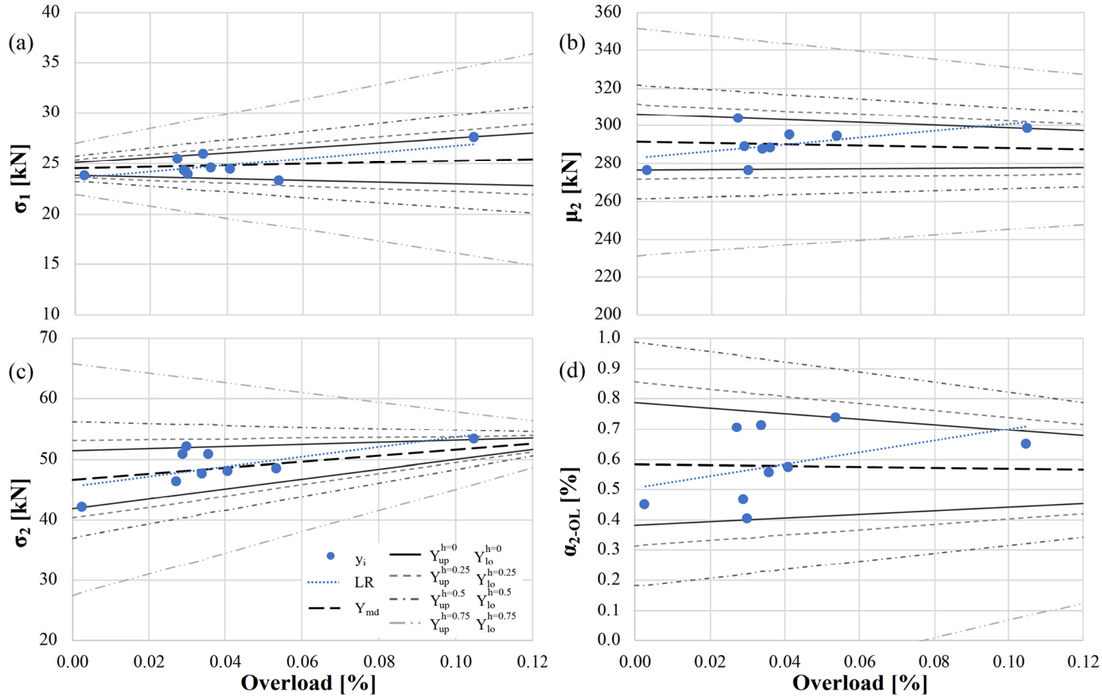


Figure 9. The fuzzy linear regression model for σ_1^*, μ_2^*, σ_2^*, and $\alpha_{2\text{-}OL}^*$ for S_L = 200 kN (45 kips)

8. Comparing Results from the Model with those from other Studies

Using the model presented in this paper, the gross vehicle weights were estimated for 95% and 98% probability values. Studies by Tabatabai et al. (2017) and Miao and Chan (2002) provide load levels corresponding to 95% and 98% values also. Fu and Hag-Elsafi (2000) do not specifically provide load estimates for these probability values, instead they present, in graphics form, the frequencies at different load levels including overloads. The graph can then be used to also read off the values corresponding to 95 and 98% probabilities. Table I provides a summary of the values from these references and those predicted using the models in this paper. Although the estimates from the current study show lower than those indicated by others, it is emphasized that there are marked differences in the data used in the three references listed in Table I.

The WIM data used in the three references listed in Table I all had a significantly higher GVW data as opposed to the data used in the current study. Furthermore, heavier truck weights were also present in their data. The overload content in the WIM data used in our study was at about 10% of the total loads with the maximum GVW at lower levels than those reported in the three references. Nevertheless, by knowing an estimate for the percentage of the overload and the largest load in the population, the model presented in this study provides a useful tool to predict load populations that can be used for damage assessment in payments and bridges, in the absence of having specific WIM data.

Table I. Gross vehicle weight estimates for 95 and 98% probability level in kN (kips)

Probability Level (%)	Model Values[1]	Tabatabai, Titi and Zhao (2017)[2]	Miao and Chan (2001)	Fu and Hag-Elsafi (2000)[3]
95	378.3 (85.0)	471.0 (104.9)	413.0 (92.0)	422.8 (95.0)
98	400.5 (90.0)	669.7 (150.5)	587.0 (132.0)	600.1 (135.0)

[1] Maximum values estimated are provided.
[2] Loads effects provided in this reference were used to arrive at GVW.
[3] Values were obtained from GVW graphs presented in this reference.

9. Limitations of Study

The model proposed in this study was intended for demonstrating how mathematical Bimodal distribution models can be used to represent truckload data. Available WIM data from two sources (in the US States of Illinois and Michigan) were used for this purpose. Having some type of mathematical model is useful in applications where truckload (including overloads) information is needed. The model presented in this paper can also be used in situations where WIM data is not readily available, as long as some preliminary information (for example the percentage of overloads in the truck load population) is known. There are several limitations to this study as explained below.

1. The model developed is exclusive to the sources from which WIM data was acquired. Since there are variations expected across truckload data from different states, the model is not intended to offer a universal method for estimating truckloads and overloads. WIM data from more states would be needed for development of a universal model, which also requires introducing additional parameters in it.

2. The study utilized the WIM data as received from the sources. State agencies in compiling WIM data use their own methods of quality assurance and control in maintaining the accuracy of their data. Thus, this study did not include any additional verification and/or validation on the original WIM data received.

3. A key parameter in the proposed Bimodal distribution model is the "common load." The suggested values for this parameter are based on the WIM data analyzed; and as such, they represent the 5-axle vehicle class 9 category. Although this class of vehicles is reported to make up a major percentage of the total population of truckloads, in rare cases when frequent occurrences of very heavy trucks are involved in the population, the common load will need to be studied further and revised.

4. In the fuzzy regression analysis, various type of membership functions may be employed. It is noted that more complex membership function forms, especially asymmetric ones, require an additional set of data for their validation. A simple triangular linear function is approximate, yet provides a convenient means for using the logical regression analysis method explained in this paper.

10. Conclusions

The following presents the main conclusions from this study.

1. When the entire population of truck load in a given WIM station is analyzed, the distribution often appears with two distinct peaks indicating that a Bimodal distribution function is perhaps needed to represent the data.

2. Based on the WIM data compiled, and statistical tests, it is shown that a Bimodal distribution comprised of a beta distribution and a lognormal distribution would provide a reasonable model for the entire truck load data, including the overloads.

3. The study further enhances the statistical load analysis process by introducing using Anderson-Darling test in addition to the more traditional Kolmogorov-Smirnov test.

4. Specific to the Illinois and Michigan truckload data, a combination of the beta and lognormal distribution function with a common load value at 245 kN (55 kips) appears to suitably represent the WIM truck load data including the overloads.

5. When no WIM data is available and the truck load distribution need to be developed, this study proposes the model development based on some preliminary roadway information (such as ADTT and the number of annual overload permit issued).

6. Using the percentage of overloads in the truck load population estimated and the fuzzy regression model for the required statistical parameters, the proposed Bimodal distribution can be developed for cases when no or limited WIM data are available.

Acknowledgements

The writers wish to acknowledge the Illinois and Michigan Department of Transportation for their cooperation and support of the research in providing the WIM data.

References

Anderson, T. W. and Darling, D. A. A Test of Goodness of Fit. *Journal of the American Statistical Association*, 49(268):765-769, 1954.

Ang, A. H. and Tang, W. H. *Probability Concepts in Engineering – Emphasis on Applications to Civil and Environmental Engineering*. John Wiley & Sons, Inc., 2007.

Chen, S. P. and You, S. S. A Fuzzy Linear Regression Model Based on Dissemblance Index. *International Journal of Mathematical and Computational Sciences*, 8(9):1278-1283, 2014.

CPCS, Perkins Motor Transport, Inc., and Portscape, Inc. Multi-State, Multimodal, Oversize/Overweight Transportation. NCHRP Report 830, Transportation Research Board, Washington, D.C., 2016.

Fu, G. and Hag-Elsafi, O. Vehicular overloads: Load model, bridge safety, and permit checking. *Journal of Bridge Engineering*, ASCE, 5(1):49-57, 2000.

Hahin. C., South. J. M., Mohammadi. J., and Polepeddi, R. Accurate and rapid determination of fatigue damage in steel bridges. *Journal of Structural Engineering*, ASCE, 119(1):150-168, 1993.

Jang, B. *Bridge Rating with Considerations for Fatigue Damage Effect from Truck Overloads*. PhD Dissertation, Illinois Institute of Technology, Chicago, IL, 2018. Civil Engineering PhD.

Jang, B. and Mohammadi, J. Bridge rating modification to incorporate fatigue damage from truck overloads. *Bridge Structures*, 13(2-3):01-107, 2017.

Jang, B. and Mohammadi, J. Significance of fatigue damage from overload trucks on bridge load rating. *MATEC Web of Conferences*, 165, 10005, 2018.

Khisty, C. J., Mohammadi, J., and Amedkudzi, A. *Systems Engineering with economics probability and statistics*. Ross Publishing company, Fort Lauderdale, FL., 2012.

Lee, H. and Tanaka, H. Fuzzy Approximations with Non-Symmetric Fuzzy Parameters in Fuzzy Regression Analysis. *Journal of the Operations Research*, Society of Japan, 42(1):98-112, 1999.

MATLAB and Statistics Toolbox Release 2016b. The MathWorks, Inc., Natick, Massachusetts, United States, 2016.

Miao, T. J. and Chan, T. H. Bridge live load models from WIM data. *Engineering Structures*, 24(8):1071-1084, 2002.

Mohammadi, J. and Shah, N. Statistical Evaluation of Truck overloads. *Journal of Transportation Engineering*, ASCE. 118(5):651-665, 1992.

Mohammadi, J. and Polepeddi, R. Bridge rating with consideration for fatigue damage from overloads. *Journal of Bridge Engineering*, ASCE, 5(3):259-265, 2000.

Nowak, A. S. and Hong, Y. K. Bridge live-load models. *Journal of Structural Engineerin*g, ASCE, 117(9):2757-2767, 1991.

Nowak, A. S. Load model for bridge design code. *Canadian Journal of Civil Engineering*, 21(1):36-49, 1994.

Qu, T., Lee, C. E., and Huang, F. Traffic-load Forecasting using Weigh-In-Motion Data. TX-99/987-6, Texas Department of Transportation, Austin, TX, 1997.

Snyder, R. E., Likins, G. E., and Moses, F. Loading spectrum experienced by bridge structures in the United States. Report RD-851012, Federal Highway Administration, Washington, D.C., 1985.

Stephens, M. A. EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, 69(347):730-737, 1974.

Tabatabai, H., Titi, H., and Zhao, J. WIM-based assessment of load effects on bridges due to various classes of heavy trucks. *Engineering Structures*, 140:189-198, 2017.

Tanaka, H., Uejima, S., and Asai, K. Linear Regression Analysis with Fuzzy Model. *IEEE Transactions on Systems, Man, And Cybernetics*, Smc-12(6):903-907, 1982.

Timm, D. H., Tisdale, S. M., and Turochy, R. E. Axle load spectra characterization by mixed distribution modeling. *Journal of Transportation Engineering*, 132(2):83-88, 2005.

Wang, H. F. and Tsaur, R. C. Bicriteria Variable Selection in a Fuzzy Regression Equation. *Computers and Mathematics with Applications*, 40:877-883, 2000.