Research on vehicle distribution model based on POT theory

Junping Zhang¹,², Haiyun Huang², * and Yonghe Li²,⁵

¹School of Civil Engineering, Guangzhou University, Guangzhou 510006, China
²Guangdong Communication Polytechnic, Guangzhou 510650, China

*Corresponding author: Huanghy@gzu.edu.cn

Abstract. The vehicle overload phenomenon in expressway has become increasingly serious in China. Research on the extreme value of the vehicle load with its development trend is an urgent problem in the toll-by-weight mode. On the basis of statistical analysis of WIM data in Beijing-Zhuhai Expressway, by the extreme value theory and POT theory, a POT model had been developed. Vehicle weight distribution in the tail of the distribution function was obtained, which can predict the extreme value of vehicle load in any return period. Results showed that more heavy vehicles will appear in the future. It provided a useful reference for effective forecast of overweight vehicles.

1. Introduction

With the continuous development of China's auto-mobile industry, the overload transportation is in-creasing, and the road traffic load has produced significant variation, which has become one of the hot topics in the new research. In order to effectively limit the overload situation, toll-by-weight mode had been adopted. Vehicle weight generally takes dynamic Weight in Motion (WIM) technology, which had been successfully applied in many high-ways in our country.

Using WIM data to simulate and develop the distribution model of Operating vehicle load, some researchers at home and abroad had carried out study and analysis[1~6], but most of them had failed to properly describe the distribution of operating vehicle load in the tail. In fact, the extreme value of the operating the vehicle load is the key and core.

In this paper, on the basis of WIM data measured in Beijing-Zhuhai Expressway, by the extreme value theory and POT theory[7], a POT distribution model of operating vehicle load had been developed. Further, the current situation and trend of the operating vehicle load were discussed.

2. Vehicle load statistics

Beijing-Zhuhai Expressway is a National highway, and 80181 WIM data of vehicle load in a heavy truck Lane in April 2012 had been collected in one of the toll Station of this expressway.

The WIM data are divided into five types in accordance with "Guangdong Provincial Expressway toll vehicle classification standard". According to statistics, the vehicles number measured from one to five types were: 39533, 1602, 10826, 2268 and 25952, with the relative proportions of 49%, 2%, 15%, 3% and 32%.
The type 3 vehicle is including medium-sized passenger car, large passenger cars, medium-sized truck-based, with the average weight 11.2 tons. The weight distribution is quite dispersed, with 3.5 tons and 14.5 tons of two weight distribution peaks.

The type 5 vehicle is including heavy vehicles, heavy trailers car, 40 feet container trucks and other vehicles, with the average vehicle weight 48.8 tons. Vehicle weight distribution varying from 10 to 110 ton had two peaks of 20 tons and 55 tons with the maximum vehicle weight 107.9 tons.

The vehicle weight histogram of Type 3 and type 5 vehicles were shown as Figure 1~2.

Figure 1. Weight frequency polygon of type 3 vehicle

Figure 2. Weight frequency polygon of type 5 vehicle

3. POT model based on the extreme value analysis theory

At present, most researchers at home and abroad commonly used the hypothesis test, analysis of the parameters of the technical route \[8\] in the study of vehicle load model. This trial and error analysis method has strong subjectivity with in-accurate description the tail distribution of operating vehicle load. Based on this, this paper intends to use the extreme value analysis theory to develop the vehicle load model.

3.1. POT model

\(F(x)\) is an arbitrary distribution function of random variable sequence \(\{X_i\}\), which is assumed to have independent identically distributed random variables, \(F(x)\) support at the top of the point \(x^*\). There is a value of \(x^*<\mu<x^*\), called Threshold (Threshold), \(X_i\) is a sequence of random variables, \((X_i-\mu)\) for out of sequence. \(F_\mu(y)\) is defined as the conditional distribution function of the random variable, which can be expressed as formula (1).

\[
F_\mu(y) = P(X - \mu \leq y | X > \mu) = \frac{F(\mu+y) - F(\mu)}{1-F(\mu)} = \frac{F(x) - F(\mu)}{1-F(\mu)}
\]

So

\[
F(x) = F_\mu(y)(1-F(\mu)) + F(\mu)
\]  (2)

For sufficiently large threshold \(\mu\), the majority of unknown distribution function \(F(x)\) beyond the volume distribution function \(F_\mu(y)\) available GPD distribution \(G(y;\mu,\sigma,\xi)\) approximation is \(F_\mu(y)\approx G(y;\mu,\sigma,\xi)\), the type of substitution:

\[
F(x) = G(y;\mu,\sigma,\xi)(1-F(\mu)) + F(\mu)
\]  (3)

After the determination of the \(\mu\), you can get a large number of \(N_\mu\) than the threshold \(\mu\) in \(\{X_i\}\), according to the formula (3) with the frequency \((1-N_\mu/n)\) instead of the \(F(\mu)\), you can get the expression of \(F(x)\).
Through parameter estimation can be estimated \( \hat{\sigma} \) and \( \hat{\xi} \), the formula (4) turn to formula (5).

\[
F(x) = 1 - \frac{N(1 + \frac{\xi}{\sigma}(x - \mu))^{\frac{1}{\xi}}}{n} \tag{5}
\]

The key of using the threshold model to estimate the distribution of the tail of the vehicle load is the choice of threshold (\( \mu \)). Threshold \( \mu \) is too large, there will be only a small amount of excess sample, the estimated amount of variance will be high; Threshold \( \mu \) is too small, then the amount of Y and GPD distribution difference is larger, the estimated amount to be biased estimate.

3.2. Threshold selection

The method of selecting threshold \( \mu \) is mainly illustrated by graphic method and calculation method. Graphic method is based on the average of the amount of e(\( \mu \)) linear change or to determine the threshold of change caused by the parameter estimation of the change in the selection of threshold. Although the application of graphical method is widely used, there is the existence of strong subjectivity. Therefore, the kurtosis method for threshold selection, calculation steps are as follows:

1. Calculate the sample kurtosis \( K_n \):
   \[
   K_n = \frac{E(X_i - \overline{X}_n)^4}{[E(X_i - \overline{X}_n)^2]^2}, i = 1, 2, \ldots, n
   \]

2. To judge if the kurtosis, \( K_n \geq 3 \), the maximum value is selected to make \( X_i \) in \( N(\mu) \), which was removed from the sample;

3. Repeat the first step, second step, until the kurtosis is less than 3;

4. Select the largest \( X_i \) in the left sample point, which is the threshold.

3.3. Parameter estimation

The parameter estimation is based on the existing vehicle load data to estimate the unknown parameters of the POT model, which mainly include Maximum Likelihood Estimation, Moment Estimation, Bayes Estimation and so on\[9\]. After comparative analysis, the selection of the maximum likelihood method is used to estimate the parameters.

Assumed observed vehicle load sample sequence \( X_1, X_2, \ldots, X_n \), method of hypothesis and inference on the parameters of \( \xi \) and \( \sigma \) with Maximum Likelihood Estimation. By the formula (6) on both sides of the derivative can be GPD distribution density function.

\[
g(x) = \frac{1}{\sigma}(1 + \frac{\xi}{\sigma}x)^{-\frac{1}{\xi}-1} \tag{6}
\]

The log maximum likelihood as formula (7) is the estimation function of the GPD distribution for the upper and the natural logarithm.

\[
L(\sigma, \xi) = -n \ln \sigma - (1 + \frac{1}{\xi}) \sum_{i=1}^{n} \ln(1 + \frac{\xi}{\sigma}x_i) \tag{7}
\]

Make the partial derivation of \( \sigma \) and \( \xi \) separately, then make it equal to 0. The \( \sigma \) and \( \xi \)'s maximum likelihood estimates is \( \hat{\sigma} \) and \( \hat{\xi} \), formula (8) can be obtained.

\[
\begin{align*}
\frac{1}{\xi^2} \sum_{i=1}^{n} \ln(1 + \frac{\xi}{\sigma}x_i) - (1 + \frac{1}{\xi}) \sum_{i=1}^{n} \frac{x_i}{\sigma + \xi x_i} &= 0 \\
-n + (1 + \xi) \sum_{i=1}^{n} \frac{x_i}{\sigma + \xi x_i} &= 0
\end{align*} \tag{8}
\]

4. Example
The 25952 WIM data of type 5 vehicles were selected for example analysis. The kurtosis method was used to select threshold. Threshold $\mu=76.8$ can be obtained by the Matlab program\textsuperscript{[10]}. 

| Threshold | Kurtosis | $N_{\mu}$ | $N_{\mu}/n$ | Suitable |
|-----------|----------|-----------|-------------|----------|
| 76.8      | 2.9938   | 921       | 3.55%       | Yes      |

Then, according to the formula (10) and Matlab program on GPD distribution, parameters $\hat{\sigma}$ and $\hat{\xi}$ to can be estimated. The estimated results are shown in table 2.

| Threshold $\mu$ | Kurtosis | $\hat{\sigma}$ | $\hat{\xi}$ | JB test (P) value |
|----------------|----------|----------------|-------------|------------------|
| 76.8           | 6.87     | 6.28           | -0.036      | 0.001            |

Then based on GPD distribution, POT model can use to fit the tail distribution of the data. According to the formula (7), the tail of any the type 5 vehicle’s weight can be gotten as formula (9):

$$F_{w5}(x) = 1 - 0.0355(1 - (x - 76.8) / 6.28)^{10.036}$$ \hspace{1cm} (9)

The tail of types 5 vehicle weight distribution is plotted in Figure 3. It showed that the tail of distribution of the fitting results and the measured value were consistent with the better.

![Figure 3. Fitting results of vehicle weight distribution of type 5](image)

For the estimation of the vehicle load, the distribution function\textsuperscript{[11]} of any heavy vehicle $w$ can be gotten by the whole probability formula $F_w(x)$.

$$F_w(x) = a_3F_{w3}(x) + a_4F_{w4}(x) + a_5F_{w5}(x)$$ \hspace{1cm} (10)

Based on the theory of extreme value analysis of multi modal variables, the formula (10) is approximated by equation (11).

$$F_w(x) \approx [F_{w5}(x)]^{a_5}$$ \hspace{1cm} (11)

According to the measured data of recent years, type 5 vehicle accounted 66.5% for all heavy vehicles flow, and therefore $a_5=0.665$.

If the total flow of the reference period $T$ is $n$, then the return level of vehicle load is $U(T)$, equaled with probability value of vehicle weight distribution ($p=1-1/n$). And in the reference period $T$, the maximum value of vehicle load in (guaranteed rate was 95%) is the maximum weight distribution of 0.95 quartile ($W0.95$). The level $U(T)$ and the maximum value ($W0.95$) of vehicle load in different reference period $T$ were calculated and summarized in Table 3.

| Time  | Total flow | Type5 vehicle flow | $\mu=76.8$ |
|-------|------------|-------------------|-------------|
|       |            |                   |             |
| Time   | U(T)       | W0.95 (T) |
|--------|------------|-----------|
| 1 month| 39046      | 114.8     |
|        | 25952      | 128.6     |
| 1 year | 468552     | 126.5     |
|        | 311424     | 139.1     |
| 5 years| 2342760    | 133.5     |
|        | 1557120    | 145.4     |
| 10 years| 4685520    | 136.4     |
|        | 3114240    | 148.1     |
| 50 years| 23427600   | 142.9     |
|        | 15571200   | 153.9     |

From the estimated results, more than 114.8 tons of heavy vehicles will appear once a month, more than 126.5 tons of heavy vehicles on average once a year. It showed that the weighing charges policy of overweight vehicle had a certain inhibitory effect for aspects as gross weight and number, but still could not effectively control the variation of overweight vehicles. On the other hand, the measured operation vehicle load was far beyond the design load standard, which was a serious threat to the safety of highway and bridge. Therefore, it is necessary to continue to improve the weighing charges policy.

5. Conclusions
In this paper, based on the measured WIM data, The POT model of operating vehicle load in Beijing-Zhuhai Expressway was developed by using the extreme value POT theory. On this basis, the development and variation trend of the operation vehicle load was predicted. The results were shown as below.

(1) POT model of can be used to stimulate the tail distribution of vehicle load, and also can accurately predict vehicle load extreme value in any return period.

(2) Overweight vehicles are not accidental. There may be more heavy vehicles in the future, which is a serious threat to the safety of existing Highways and bridges.

(3) It is necessary to continue to improve the existed weighing charges policy.

Acknowledgments
These works was financially supported by the National Science Foundation of China(NO.51278134) and the Science and technology planning project of Guangdong Provincial Department of Transportation in 2014(NO.201402022).

References
[1] Gang M, Quan Q, Daojin L 2003 Bi-modal probabilistic model of highway and bridge vehicle loads Joural of Tsinghua University 43(10)
[2] Tao H 2010 Investigation Statistics and Simulation for Random Traffic Loading of Expressway Bridge Doctor thesis of Chang’an University
[3] Tao W, Wanshui H and Pingming H 2010 Research status and prospect on traffic loading for highway bridge Journal of Architecture and Civil Engineering 4 p31–37
[4] Jushu S, Chunfeng Y , Yuanming D 2009 Study on the Vehicle Load of Bridge Design on Heavy Traffic Highway Basic on Traffic Investigation Journal of Hebei University of Technology 1 p91–95
[5] Jinxin G and Wenjie L 2010 Highway bridge vehicle load probability model research (I) - a gauge system, weight fees and compulsory for super region Highway traffic science and technology 27 (6) p 40–45
[6] Jinxin G and Wenjie L 2010 Highway bridge vehicle load probability model research (II) - a gauge system, weight fees and compulsory for super region Highway traffic science and technology 27 (7) p 57–60
[7] Daoqi S 2005 Practical extremism statistical method (Tianjin Science and Technology Press )
[8] Shisong M, Jinglong W, Xiaolong P 2006 Advanced Mathematical Statistics (Beijing: Higher Education Press)
[9] Juan G 2008 Research on Parameter Estimation Methods of Generalized Extreme Value Distribution *Doctor thesis of Beijing North China Electric Power University*

[10] Haibin Z 2012 *Application of MATLAB* (Beijing: Tsinghua University Press)

[11] Shisong M, Jinglong W, Xiaolong P 2006 *Advanced Mathematical Statistics* (Beijing: Higher Education Press)