A peaks-over-threshold analysis of extreme traffic load effects on bridges

Franziska Schmidt, Xiao Yi Zhou, François Toutlemonde

Abstract

For purposes of design of new structures or safety assessment of existing structures, extreme effects need to be assessed, for example through calculation of characteristic values of high quantile. Indeed, these values make it possible to determinate stresses for extreme events (extreme weather impacts like temperatures or rain, but also extreme traffic events). This paper focuses on modeling and estimating tail parameters of bridge traffic load effects by generalized Pareto distribution (GPD) to obtain these characteristic values. Its main advantage is that it makes use of all relevant data on the high load effect induced by the passage of traffic, not just the daily maxima used for generalized extreme value (GEV) distribution. However, two critical problems of parameter estimation and threshold selection complexify the use of GPD in practice. A Kolmogorov-Smirnov (KS) statistic based threshold selection method is used to choose the optimal threshold, and the maximum likelihood estimator is adopted to estimate the parameters in this study. Results show that the GPD method is a theoretically well supported technique for fitting a parametric distribution to the tail of an unknown underlying distribution.

1 INTRODUCTION

A bridge structure should have the capacity to withstand the load expected with a given probability over the design life: it is stated in Eurocode 1 that the structures should be able to resist to the characteristic values for 1000-year return period, which means events with a probability of exceedance of 5% in 50 years.

But it is obvious that it is impossible to collect enough data to determine the characteristic value of load effect expected over lifetime of loading. Even getting sufficient data for the short return period would require several samples of short period data in order to assess the corresponding probability, which is not currently available. Therefore, some form of statistical projection should be performed.

In development of live load model of AASHTO, Nowak [1] proposed a method that uses normal

probability paper. This method had and still has great influence on other American researchers [4]. Straight lines are fitted to the tails of the load effect distributions plotted on normal probability paper. The extreme value of load effect can be easily determined by the inversed normal distribution, and the cumulative distribution of maximum load effect is calculated by raising the fitted distribution to a certain power according to the extreme value theory. Although the method is extensively used, the choice of the tail fraction is empirical or subjective: Nowak [2] does not state how he selects the beginning of the tail and Sivakumar [5] uses the upper 5%.

In the background studies for developing the traffic load model of Eurocode 1, five methods have been proposed to predict extreme traffic load or traffic load effect [6]. These are:

i. a half-normal curve fitted to the end of the histogram,
ii. a Gumbel distribution fitted to the tail of the histogram,
iii. the asymptotic extreme distribution obtained with extreme value theory,
iv. Rice's formula for stationary Gaussian processes,
v. Monte-Carlo simulation of artificial traffic and Gumbel extrapolation.

Among these methods, Rice's formula is different from the others as it uses the full history of the load effect. It has been proved that if the process is Gaussian and stationary, the up-crossing intensity can be expressed by Rice's formula [7]. In practice, the load effect is a mixed Gaussian process, therefore the level crossing rate has a Gaussian tail. Cremona [9] proposed an automatic method to select the optimal threshold from which the tail can be expressed by Rice formula. The mathematical assumption of Gaussian stationary process is a strong requirement, it requires the length of influence line of interest is large compared to the length occupied by the single vehicles [10]. The approach may be not suitable to common load effect of short- to medium-span bridge as it has sharply varying nonzero influence area.

In recent years, the extreme value theory has been used for modeling traffic load effect. Many authors approach the problem by identifying the maxima load effect recorded during a loading event or in a reference period such as a day or a week, and then fit these maxima to an extreme value distribution. Bailey and Bez [11] determine that the Weibull distribution is the most appropriate to model the extreme traffic load effect. Caprani et al. [12] classify the loading events by the number of trucks involved in the maximal load effect, and then the maxima of each loading event are modeled by GEV distribution. This block maxima approach is wasteful of data as only one data in each block is

taken. The second highest value in one block may be larger than the highest of another block and this is not accounted for generally.

There exists a peaks-over-threshold (POT) approach which avoids this drawback, by using the Generalized Pareto distribution.

Indeed, due to the advances in extreme value theory, the GPD emerged as a natural family for modeling exceedances over a high threshold. The POT approach has shown its importance and success in a number of statistical analysis problems relating to finance, insurance, hydrology, geographical phenomena, and other domains [15]. However, despite the sound theoretical basis and wide applicability, fitting of this distribution in practice is not a trivial exercise. Two main factors, the choice of threshold and the choice of the parameter estimator, affect the accuracy of the estimations of the return values. An extensive discussion of the various parameter estimation methods has been given by Zea Bermudez and Kotz [18]. Scarrott and MacDonald [19] review the threshold estimation methods. While each methodology has its advantages and disadvantages, traditional parameter estimator such as maximum likelihood is undefined in some regions of the parameter space. Due to these difficult, the POT approach has not gotten much attention for modeling extreme bridge load effects [14].

The POT approach makes full use of all the data present in the tail as Nowak's approach, but it approximates the tail through extreme value distribution. Therefore, the POT approach is a natural family for modeling exceedances over a high threshold. The purpose of this paper is to introduce an efficient procedure for using POT approach to determine the extreme bridge traffic load effect. The rest of the article is organized as follows: In Section 2, we provide key distributional properties of the GPD. In Section 3, we fit the GPD to the bridge traffic load effects and summary the results. Discussions are carried out and conclusions are drawn in Section 4.

2 THE PEAKS-OVER-THRESHOLD METHOD AND GENERALIZED PARETO DISTRIBUTION

Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed random variables with common distribution function F, and let's M_n denote its maximum:

$$M_n = \max(X_1, \cdots, X_n).$$

In theory the distribution of M_n can be derived exactly for all values of n:

$$P\left\{\max_{i\le n} X_i \le x\right\} = P\left\{X_1 \le x, \cdots, X_n \le x\right\} = F^n(x)$$.

But in most cases, the distribution function F is unknown. The value of F^n can be approximated by extreme value distribution if the maximum, M_n, belongs to an maximum domain of attraction. In this case:

$$F^n(x) \approx G(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

where ξ, σ and μ are shape , scale and location parameters, respectively. The events that exceed some high threshold u can be regarded as extreme events also, the behavior of these extreme events is given by the conditional probability

$$P\left\{X > u+y | X > u\right\} = \frac{1-F(u+y)}{1-F(u)}$$

For large enough u, the distribution of $(X-u)$, conditional on $X>u$, is approximated with the extreme value theory by:

$$H(y) = 1-\left(1+\xi\frac{y}{\sigma}\right)^{-1/\xi}$$

In extreme value theory, the previous distribution belongs to the generalized Pareto family. The cumulative distribution function (CDF) of the GPD with shape, scale and location parameters ξ, σ and μ, respectively, is defined as

$$G(x) = \begin{cases} 1-\left[1+\dfrac{\xi(x-\mu)}{\sigma}\right]^{-1/\xi} & \xi \neq 0 \\ 1-\exp\left(-\dfrac{x-\mu}{\sigma}\right) & \xi = 0 \end{cases}$$

The probability density function of the GPD is given by

$$g(x) = \begin{cases} \dfrac{1}{\sigma}\left[1+\dfrac{\xi(x-\mu)}{\sigma}\right]^{-1/\xi-1} & \xi \neq 0 \\ \dfrac{1}{\sigma}\exp\left(-\dfrac{x-\mu}{\sigma}\right) & \xi = 0 \end{cases}$$

3 APPLICATION OF THE GPD TO THE ANALYSIS OF TRAFFIC LOAD EFFECTS

3.1 Bridge Traffic Load Effect

3.1.1 Traffic Data

Traffic data, taken from a piezo-ceramic weigh-in-motion system on the A9 motorway near Saint Jean de Vedas, in the South-East of France, is used to validate the POT method on the estimation of characteristic bridge traffic load effects. Weight and dimensional data were collected for trucks travelling in the slow and fast lanes in one direction of the 6-lane motorway from January 2010 to May 2010. The reliable loading data is useful for estimating extreme traffic load effect on bridge. The unreliable data (trucks with single axle, "0" axle weight and etc) can be eliminated from records with some chosen criteria [5]. For operational reasons, on some days the WIM system may not be active for some hours, it is important to exclude these days to ensure having continuous traffic.

581011 trucks for 86 days were drawn from the original data excluding data of error, weekends and system inactive days. The histogram of gross vehicle weight of the remaining trucks is presented in Figure 1: the quality of the records may be good as can be seen at the high percentage of trucks with GVW around the weight limit. The traffic composition plot shows that 5-axle trucks are the most frequently trucks used in France, and the characteristic of this truck affects the load effect on bridges.

Bridges supporting bi-directional 4 lanes are common in France. An artificial synthesis approach is used to create a sample with 4 lanes traffic from the records measured on 2-lane.

3.1.2 Influence lines

Previous studies have demonstrated the critical influence lines for developing load model load effects of bending moment at mid-span of a simply supported bridge, bending moment at first mid-span of two-span continuous bridge and hogging moment at middle support of the former. In this study, these three types of load effect are studied. In additional to these, three load effects for the three-span continuous bridge are studied as it is an usual type of bridge in France. A complete list of the influence lines are indicated in Table 1.

Thus three types of bridge structure are investigated: simply supported, two-span continuous with equal span length, and three-span continuous with span configuration of 0.75 L + L + 0.75 L. In the calculations 4 span lengths were considered: 20, 30, 40, and 50 m.

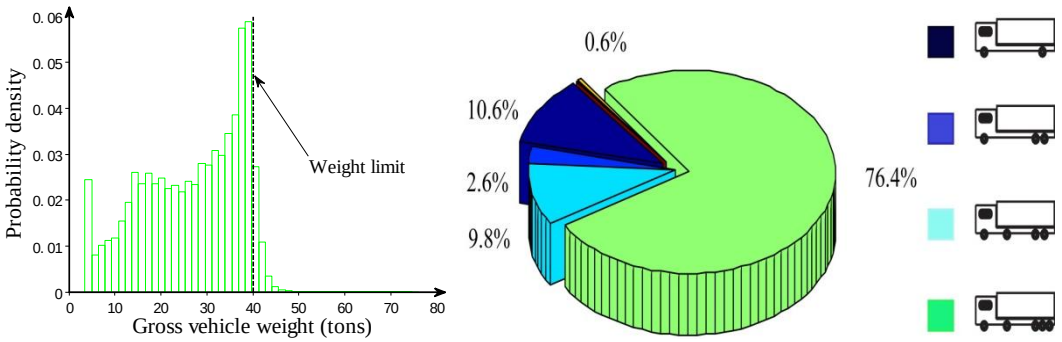


FIGURE 1 Traffic composition and histogram of GVW

TABLE 1 Influence lines used in calculation of load effect

Item	Description	Representation
I1	Bending moment at mid-span of a simply supported bridge	
I2	Bending moment at middle support of a two-span continuous bridge	
I3 & I4	Minimum and maximum bending moment at first mid-span of a two-span continuous bridge	
I5	Bending moment at first mid-span of a three span continuous bridge	
I6	Bending moment at second support of a three-span continuous bridge	
I7	Bending moment at second mid-span of a three span continuous bridge	

3.2 Preliminary study

In this section, we fit the GPD model to the bridge traffic load effects resulting from combination of traffic data and influence lines, which has been extensively studied in the literature by using other extrapolation methods (see, e.g. [9]). In the first stage, we detail the application of GPD to the moment induced by the 4 lanes of traffic at mid-span of a simply supported bridge with span of 50 m. The standard exponential Quantile-Quantile (Q-Q) plot in FIGURE 2(left) gives evidence of lighter than exponential (LTE) distribution nature of the hourly maximum traffic load effects as the curve has a convex shape. This becomes even more apparent in the Mean Excess (ME) plot given in FIGURE 2(right) with a negative slope ratio. The asymptotic linear tail on the ME plot indicates a Pareto type distribution can reasonably fit to the tail from the turn point.

3.2.1 Estimation of The GPD Parameters

The last decade has seen development of a number of approaches for parameter estimation in

GPD applications. The maximum likelihood (ML), the probability weighted moments (PWM) and the method of moments (MM) are often used. An extensive discussion of the various methods has been presented by de Zea Bermudez [18]. The main advantages and disadvantages of each method mentioned in the paper are thus known. Moharram et al [23], Singh and Ahmad [24], and Mackay et al [25] utilize Monte Carlo simulations to evaluate the performance of the estimators for GPD. All of the authors conclude that no one method is preferable to another for all simulated numerical samples. The ML estimator is asymptotically efficient, but for small samples the ML estimators does not always exist and when it does exist it has a larger bias and variance than other estimators. The PWM and MM estimators make straightforward computation property, but the existence of PWM estimator is restricted to shape parameter less than 0.5 and 1.0 for the MM estimator. Previous studies have demonstrated that the maxima of traffic load effect converge to Gumbel or Weibull distribution [11]. Therefore, the three extensively used estimators are available for traffic load effects.

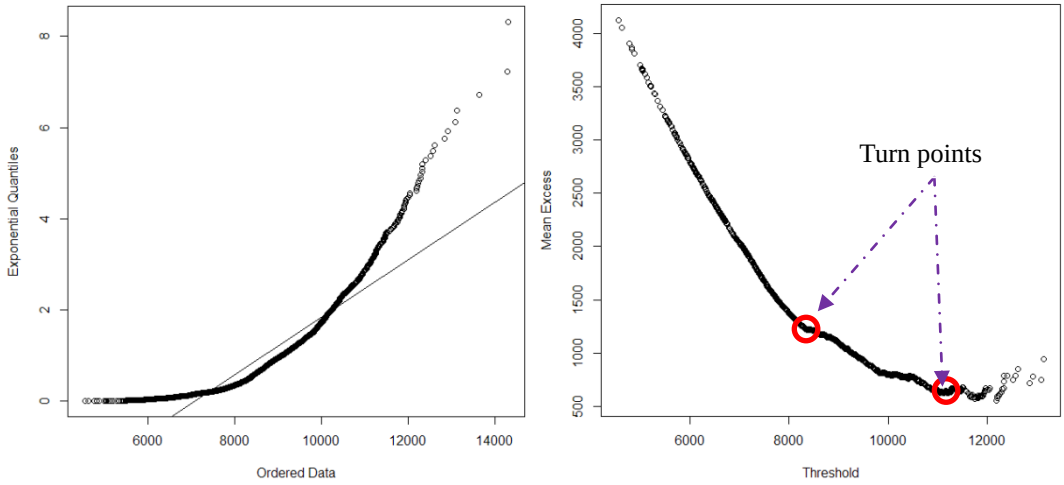


FIGURE 2 Standard Q-Q plot (left) and Mean Excess plot (right) for hourly maximum bending moment

3.2.2 Choice of The Optimal Threshold

The complex shape of the ME plot in FIGURE 2 indicates variation of the properties for excesses over high threshold. We need a refined method for choosing optimal threshold that can make trade-off between bias and variation. The determination of such a threshold still remains a delicate point of the whole set-up of GPD-fitting to distribution tails. A known recommendation to use for this purpose, the so-called Mean Excess Function cannot be considered as a satisfactory solution to this problem. The automatic threshold selection method proposed by Cremona [9] has shown its success in the optimal extrapolation for level crossing method. It should be a good solution for the choice of optimal

threshold for the use of POT approach. The Kolmogorov Distance (KD) between the tail histogram and the fitted GPD is used as a value for choice of optimal threshold. This statistical test compares two probability functions corresponding respectively to the reference and the tested distribution. If $S(x)$ and $F(x)$ are the probability functions, the K-test studies the statistics of the variation D :

$$D = \max_{-\infty < x < +\infty} \left(\left| S(x) - F(x) \right| \right)$$

For a value d, the probability $P(d>D)$ is approximated by the Kolmogorov function Q_K :

$$P\left(d > D\right) = Q_K\left(\sqrt{N}\, d\right) = 2\sum_{i=1}^{n} (-1)^{k-1} e^{-2k^2\left(\sqrt{N}d\right)^2}$$

FIGURE 3 shows the estimates of the shape parameter and P-value using the ML, MM and PWM estimators with threshold value. It is clear that the PWM estimator is very sensitive to the threshold value as shown in Figure 3. The ML estimator is slightly less sensitive that the PWM and MM estimators, displaying a slightly lower variation of shape parameter with threshold value. At a threshold of $u=11000$ the shape parameter appears to reach a plateau. The corresponding P-value of KS test keeping at a high level over 0.8 means the GPD model fits the excesses well. The fitting is quite sensitive to the endpoints, as the threshold closing to the tail, the shape parameter and P-value fluctuate display some irregular fluctuations. Therefore, the optimal threshold should be selected from the plateau around $u=11000$. Table 2 summarizes the results of the application. The results include the optimal thresholds selected for each parameter estimator, corresponding estimates of return level for 1000 years return period.

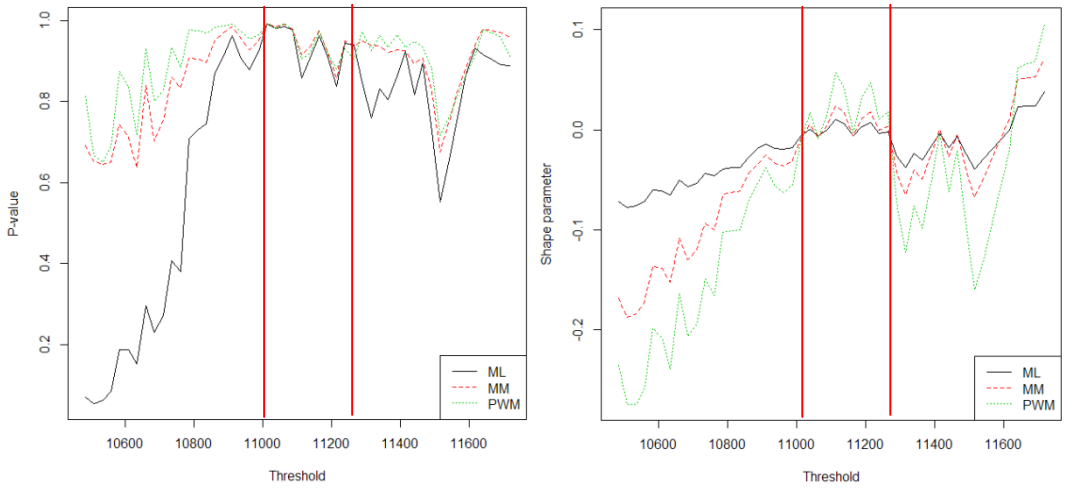


FIGURE 3 Threshold vs. P-value (left) and threshold vs. shape parameter

TABLE 2 ML, MM and PWM estimates

Method	Threshold	No. excesses	P-value	Shape	Return level
ML	11012.07	117	0.99046	-0.00087	22575.21

MM	11012.07	117	0.994234	0.003436	23004.72
PWM	11012.07	117	0.993113	0.012504	23960.33

3.3 Results

The availability of GPD has been illustrated in the previous preliminary study, and then we extend its application to other types of load effect and span lengths. The load effects are determined by combining 4-lane free flowing traffic and influence lines for span length from 20 to 50 m. Characteristic values of the load effects for 1000-year return period are extrapolated using the GPD. The results are compared with those found by conventional GEV, and also compared with the design values calculated from LM 1 of Eurocode 1. The GEV is fitted to daily maxima to avoid correlation, while the GPD is applied to hourly maxima ensuring more data can be involved. In order to reduce possible difference arising from parameter estimator, parameter estimator of ML is used to GPD and GEV.

A summary of results is presented in TABLE 3. Mostly, the predictions by using GEV and POT methods are less than design effects in all cases. Interestingly, the differences compared to design values for various type of load effect are around 40%. Also shown in TABLE 3 are predictions obtained by fitting the daily maxima to the GEV, which were slightly less than those obtained by the GPD approach. Comparisons of the POT results with the 1000-year GEV results are given in FIGURE 4.

4 DISCUSSION AND CONCLUSIONS

In this article, we have introduced the Peaks-over-threshold method for statistical analysis of traffic load effects on bridges. It is based on the excesses over high threshold can be modeled by the generalized Pareto distribution. Its main advantage is that it use of all relevant data on the high tail induced by the passage of traffic, not just daily or annual maxima. The two difficulties of threshold selection and parameter estimation for using this method have been discussed on the case of bridge traffic load effects. Some remarks can be made:

- The automatic approach based on KS test can efficiently identify the optimal threshold for the use of POT method.
- The GPD model has shown its success on predicting characteristic bridge traffic load effects as it captures the tail behavior very well.

- The almost same differences indicate the Load model 1 (LM1) of Eurocode 1 has a consistent safety margin for various types of load effect.
- Based upon the results it is reasonable to conclude that the LM1 of Eurocode is sufficient for modern free flowing traffic.

TABLE 3 Comparison of POT results

Load effect	Span length (m)	Characteristic load effect (kN.m)			Percentage difference (%)		
		LM1	GEV	POT	GEV[a]	POT[a]	POT[b]
I1	20	8115	4527	4481	-44.21	-44.79	-1.03
	30	14209	7801	8254	-45.10	-41.91	5.82
	40	21540	12267	11915	-43.05	-44.68	-2.86
	50	30109	17386	22575	-42.26	-25.02	29.84
I2	20	-4775	-2779	-5081	-41.80	6.40	82.83
	30	-9027	-4689	-6432	-48.06	-28.75	37.17
	40	-14514	-6960	-7589	-52.05	-47.71	9.05
	50	-21239	-9355	-9671	-55.95	-54.46	3.38
I3	20	-1769	-1101	-1249	-37.74	-29.40	13.41
	30	-3121	-1874	-1875	-39.97	-39.94	0.05
	40	-4782	-2853	-3361	-40.35	-29.72	17.81
	50	-6752	-3888	-3806	-42.42	-43.63	-2.10
I4	20	6402	3847	3771	-39.91	-41.09	-1.95
	30	11157	6140	6810	-44.96	-38.96	10.91
	40	16841	8735	9211	-48.14	-45.31	5.45
	50	23455	12477	13431	-46.80	-42.74	7.64
I5	20	6452	5226	3804	-19.00	-41.04	-27.21
	30	11325	4457	4051	-60.64	-64.23	-9.12
	40	17188	6815	7480	-60.35	-56.48	9.77
	50	24042	7386	7607	-69.28	-68.36	3.00
I6	20	-4763	-1951	-2098	-59.04	-55.95	7.54
	30	-8886	-3376	-5904	-62.01	-33.56	74.87
	40	-14162	-5446	-5738	-61.54	-59.48	5.36
	50	-20592	-7122	-8877	-65.41	-56.89	24.65
I7	20	5338	3482	3433	-34.76	-35.70	-1.43
	30	9290	4667	4762	-49.77	-48.74	2.04
	40	13986	7287	7641	-47.90	-45.37	4.87
	50	19426	9980	10333	-48.62	-46.81	3.54

a Related to load effects from LM1 of EC1.

b Related to GEV predictions.

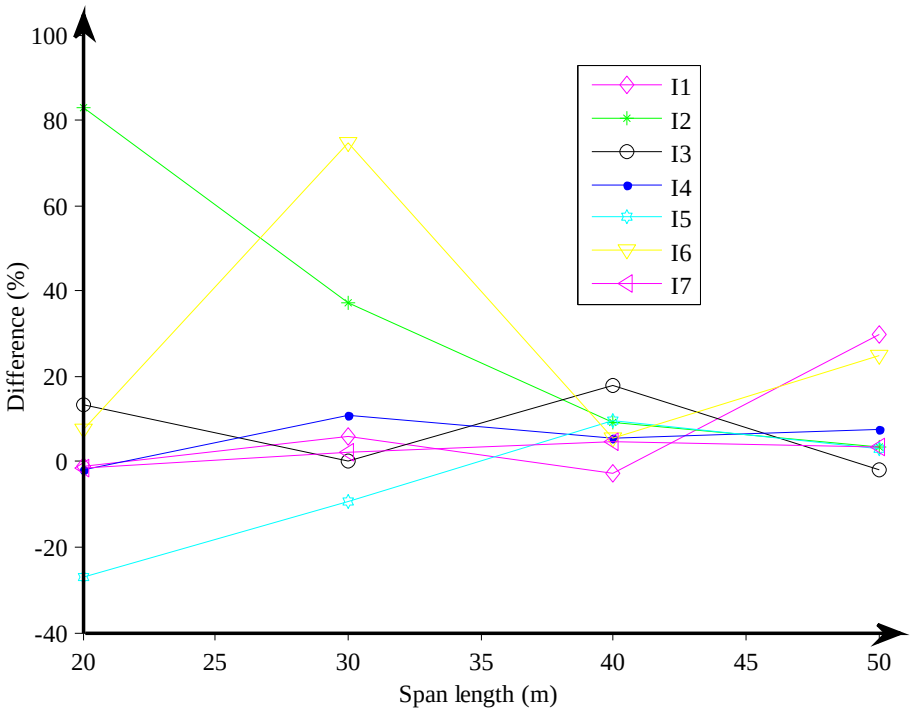


FIGURE 4 Difference

REFERENCES

[1] Nowak, A.S. and Hong, Y.-K. (1991) Bridge Live-Load Models. Journal of Structural Engineering 117, 2757-2767

[2] Nowak, A.S. (1993) Live load model for highway bridges. Structural Safety 13, 53-66

[3] Nowak, A.S. (1994) Load model for bridge design code. Canadian Journal of Civil Engineering 21, 36-49

[4] Kozikowski, M. (2009) WIM based live load model for bridge reliability. University of Nebraska

[5] Sivakumar, B., et al. (2011) Protocols for collecting and using traffic data in bridge design. In National Cooperative Highway Research Program

[6] Jacob, B. (1991) Methods for the prediction of extreme vehicular loads and load effects. In WG8 report, Eurocode 1: Traffic Loads on Bridges, LCPC

[7] Rice, S.O. (1944) Mathematical analysis of random noise. Bell Systems technical Journal 23, 282-332

[8] Rice, S.O. (1945) Mathematical analysis of random noise. Bell Systems technical Journal 24, 46-156

[9] Cremona, C. (2001) Optimal extrapolation of traffic load effects. Structural Safety 23, 31-46

[10] Ditlevsen, O. (1994) Traffic Loads on Large Bridges Modeled as White- Noise Fields. Journal of Engineering Mechanics 120, 681-694

[11] Bailey, S. and Bez, R. (1999) Site specific probability distribution of extreme traffic action effects. Probabilistic Engineering Mechanics, 19-26

[12] Caprani, C.C., et al. (2008) Characteristic traffic load effects from a mixture of loading events on short to medium span bridges. Structural Safety 30, 394-404

[13] Crespo-Minguillon, C. and Casas, J.R. (1997) A comprehensive traffic load model for bridge safety checking. Structural Safety 19, 339-359

[14] James, G. (2003) Analysis of Traffic Load Effects on Railway Bridge. In Structual Engineering Division, Royal Institute of Technology

[15] Holmes, J.D. and Moriarty, W.W. (1999) Application of the generalized Pareto distribution to extreme value analysis in wind engineering. Journal of Wind Engineering and Industrial Aerodynamics 83, 1-10

[16] Pisarenko, V.F. and Sornette, D. (2003) Characterization of the frequency of extreme earthquake events by the generalized Pareto distribution. Pure and Applied Geophysics 160, 2343-2364

[17] Brazauskas, V. and Kleefeld, A. (2009) Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view. Insurance: Mathematics and Economics 45, 424-435

[18] de Zea Bermudez, P. and Kotz, S. (2010) Parameter estimation of the generalized Pareto distribution - Part I & II. Journal of Statistical Planning and Inference 140, 1353-1388

[19] Scarrott, C. and MacDonald, A. (2012) A review of extreme value threshold estimation and uncertainty quantification. REVSTAT - Statistical Journal 10, 33-60

[20] Coles, S. (2001) An introduction to statistical modeling of extreme values. Springer-Verlag London Limited

[21] Enright, B. and O'Brien, E.J. (2011) Cleaning weigh-in-motion data: techniques and recommendations. International Society for Weigh-In-Motion

[22] Ferreira, A., et al. (2003) On optimising the estimation of high quantiles of a probability distribution. Journal of Theoretical and Applied Statistics 37, 401-434

[23] Moharram, S.H., et al. (1993) A comparative study for the estimators of the Generalized Pareto distribution. Journal of Hydrology 150, 169-185

[24] Singh, V.P. and Ahmad, M. (2004) A comparative evaluation of the estimators of the three-parameter generalized pareto distribution. Journal of Statistical Computation and Simulation 74, 91-106

[25] Mackay, E.B.L., et al. (2011) A comparison of estimators for the generalized Pareto distribution. Ocean Engineering 38, 1338-1346