Title	Modelling Extreme Traffic Loading on Bridges Using Kernal Density Estimators
Authors(s)	Leahy, Cathal, O'Brien, Eugene J., Enright, Bernard
Publication date	2011-10-13
Publication information	Leahy, Cathal, Eugene J. O'Brien, and Bernard Enright. "Modelling Extreme Traffic Loading on Bridges Using Kernal Density Estimators." Eugenides Foundation, 2011.
Conference details	Innovations on Bridges and Soil-Bridge Interaction (IBSBI 2011), Athens, Greece, October 13-15, 2011
Publisher	Eugenides Foundation
Item record/more information	http://hdl.handle.net/10197/4077

IBSBI 2011, October 13-15, 2011, Athens, Greece

MODELING EXTREME TRAFFIC LOADING ON BRIDGES USING KERNEL DENSITY ESTIMATORS

Cathal Leahy[1], Eugene J. OBrien[1] and Bernard Enright[2]

[1] UCD School of Civil, Structural & Environmental Engineering, Newstead, Dublin 4, Ireland
[2] Dublin Institute of Technology, Bolton Street, Dublin1, Ireland
e-mail: Cathal.Leahy@ucdconnect.ie, eugene.obrien@ucd.ie, bernard.enright@dit.ie

ABSTRACT: Kernel density estimators are a non-parametric method of estimating the probability density function of sample data. In this paper, the method is applied to find characteristic maximum daily truck weights on highway bridges. The results are then compared with the conventional approach.

KEY WORDS: Bridge; Characteristic; Traffic; Kernel Density Estimators; Loading.

1 INTRODUCTION

Bridge safety assessment involves a comparison of load effect (stress, etc.) and the capacity of the bridge to resist that effect. Probabilistic assessment requires the convolution of the probability density functions for load effects and resistances. However, a load and resistance factor approach is commonly applied which requires the calculation or estimation of characteristic levels of load effect and resistance. This paper concentrates on the loading side of this equation and describes a method of estimating characteristic load effects for road bridges using Weigh-in-Motion (WIM) traffic data.

The accurate estimation of characteristic load effect is critically dependent on the extreme upper tail of the load distribution. Relatively few measured values are available for this tail region, and some form of interpolation and extrapolation is required. A popular approach is to plot the measured data on Gumbel probability paper [1] and to fit a type III Generalised Extreme Value (Weibull) distribution to this data [2-4]. Characteristic load effects can be estimated using this fitted distribution. A return period of 1000 years is used in the Eurocode for the design of new bridges, based on a 5% probability of exceedance in 50 years. The U.S. AASHTO design code is based on the distribution of the 75-year maximum loading [5].Lesser periods have been used for assessment, typically in the 5 to 10 year range [6]. One of the disadvantages of this method is that it is assumed that the data comes from a Weibull distribution. Kernel density estimators (KDEs) are a non-parametric method of estimating the probability density function (PDF) of sample data. The PDF is built from the measured data without assuming that it comes from a certain theoretical distribution. As a result, the PDF is ‘more true’ to the original data as it does not force a theoretical distribution upon it.

1.1 Introduction to kernel density estimators

In the KDE method, each sample data point is replaced by a component density (kernel function), and these densities are then added to form the complete PDF. Rectangular, triangular and Normal kernel functions are common although any distribution can be used. Depending on the parameters of the kernel function used, the resultant PDF will have different characteristics. The KDE method is not then entirely non-parametric but may provide a compromise between a purely non-parametric approach and a parametric approach [7]. *Figs. 1 and 2* illustrate a simple example of how the method works. *Fig. 1* shows the histogram of 30 loads randomly sampled from a Normal distribution. It is clear that the histogram is not an accurate representation of the true distribution. Using the KDE method a normal kernel function of area 1/30 is created for each data point (shown at bottom of *Fig. 2*). These individual kernel functions are then added to create the PDF, which gives a much better approximation of the theoretical normal distribution than the histogram. This simple example uses just 30 data points but the more data points that are available, the more accurate the estimate of the true distribution.

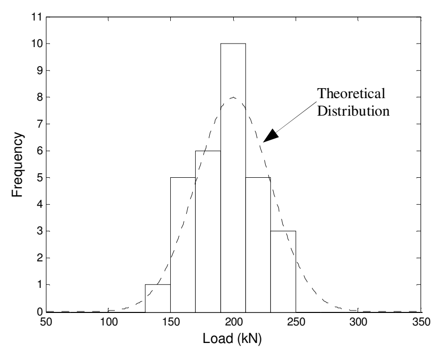


Figure 1. Histogram of 30 sample data points

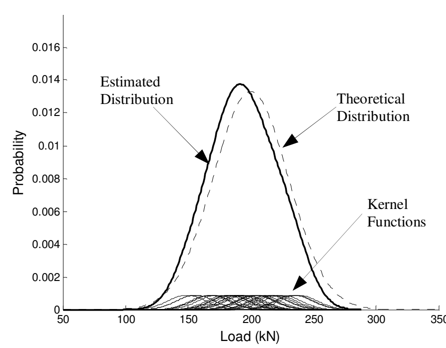


Figure 2. Kernel density estimate of PDF for 30 sample data points

The bandwidth of the kernel functions refers to the width of the individual distributions. In the case of *Fig. 2*, the bandwidth would refer to the standard deviation of the normal kernel functions. This is an important factor in the KDE method and has a significant influence on the smoothness of the estimated PDF. Smaller bandwidths result in fewer data points influencing the estimate at any one point, which gives a ‘bumpier’ estimate of the PDF [8]. As the bandwidth increases, a smoother PDF is achieved as there is more overlap between the individual kernel functions. A certain degree of smoothness is desirable but as the bandwidth increases, the estimated PDF becomes a poorer fit to the measured data. The optimal bandwidth is therefore a compromise between smoothness and achieving a good fit to the original data. This optimal bandwidth is often chosen by plotting a number of different bandwidths and subjectively picking the one which best fits one's expectations for the density. For many applications this method is sufficient but an objective approach is more appropriate for inexperienced users or if multiple data sets are to be analysed [9, 10].

Towards the peaks of the sample data, where there are many sample data points, small bandwidths are best but towards the tails of the data, where there are very few sample data points, larger bandwidths are required to obtain a smooth PDF and a more appropriate estimate of the density. Scott [11] suggests methods for picking bandwidths and for varying bandwidths based on the distribution of the data.

KDEs are known to work well for interpolation of measured data but less well for extrapolation beyond the measured data [7]. This paper aims to develop a method for improving the accuracy of the method for extrapolation in order to estimate characteristic load events.

2 APPLICATION OF KERNEL DENSITY ESTIMATORS TO LOAD DATA

Measured maximum daily load data is usually assumed to be consistent with the Generalized Extreme Value distribution. Type 1 Generalized Extreme Value (Gumbel) distributions are therefore used here as the kernel functions. In this paper the bandwidth of the kernel function refers to the scale parameter of the Gumbel kernel functions. *Figs. 3* and *4* show the histogram of 1000 truck weights, randomly sampled from a Gumbel distribution. *Fig. 3* uses a fixed bandwidth of 5 kN to estimate the PDF of the data while *Fig. 4* uses a larger fixed bandwidth of 40 kN. The difference in smoothness between the two estimates of the distribution is clear. The smaller bandwidth gives a PDF which follows the data very closely but fails to smooth the local peaks in the histogram (which appear due to the randomness in the limited data set). It is clear in *Fig. 4* that the larger bandwidth has over-smoothed the data set and produced a poor fit to histogram of the data. To achieve a good fit to the original data, while also smoothing the PDF sufficiently, a variable bandwidth can be used. A smaller bandwidth is used where there are high densities of data and the bandwidth increases as the data points become more sparse, i.e., towards the tails of the data.

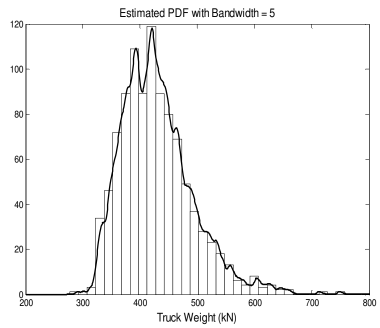


Figure 3. Kernel density estimate of PDF using bandwidth of 5 kN

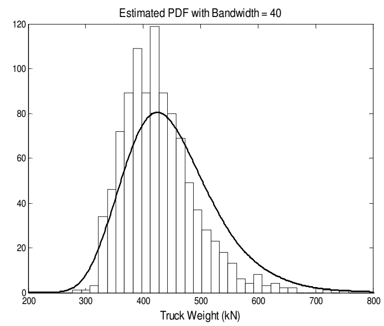


Figure 4. Kernel density estimate of PDF using bandwidth of 40 kN

To find an appropriate starting bandwidth at the mode of the data Eq. (1) is used [11]. This equation is for Normal kernel functions but works well with the Gumbel kernel functions used here.

$$h=1.06\sigma n^{-0.2} \qquad (1)$$

where: h is the bandwidth
σ is the standard deviation of the sample
n is the sample size

Eq. (2) is used to increase the bandwidth with increasing distance from the mode of the data. This approach was developed by Abramson [12] and cited in Scott [11].

$$h_{\mathrm{i}}=\frac{k}{\sqrt{f(x_i)}} \qquad (2)$$

where: $f(x_i)$ is the density function
k is a constant
h_i is the bandwidth used

Eq. (1) is used to obtain the starting bandwidth at the mode of the data. The constant k is then calculated by substituting this value into Eq. (2). *Fig. 5* shows that using Eq. (2), the bandwidth increases rapidly with increasing distance from the mode of the data. These large bandwidths resulted in overestimation of the value of the PDF in the extrapolation region of the tail. To address this, a bandwidth cap is required to prevent the bandwidth from increasing above a certain level.

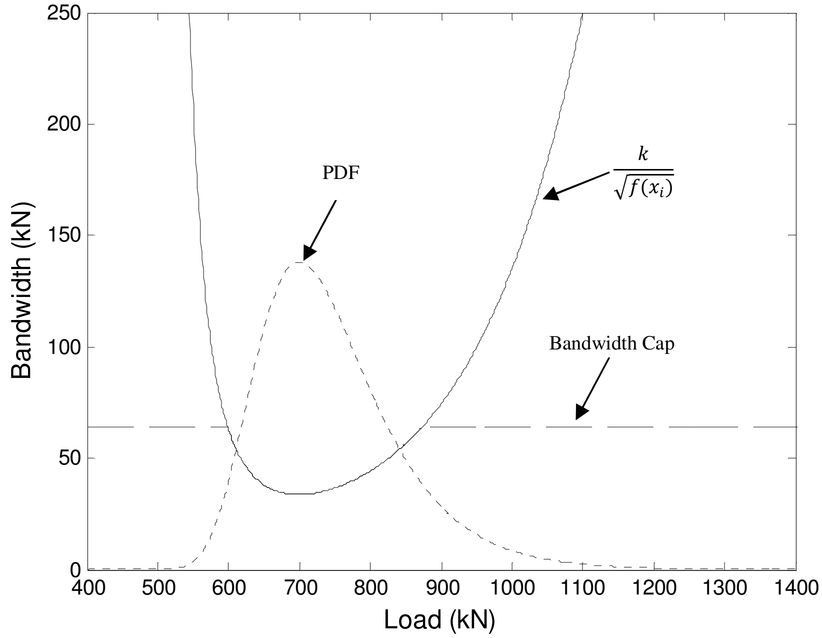


Figure 5. Variation in calculated bandwidth with respect to PDF

Different approaches for calculating the optimal bandwidth cap are investigated. The method which gave the best results is based on the scale parameter of the sample data. The scale parameter of a Gumbel distribution can be estimated using Eq. (3) [13].

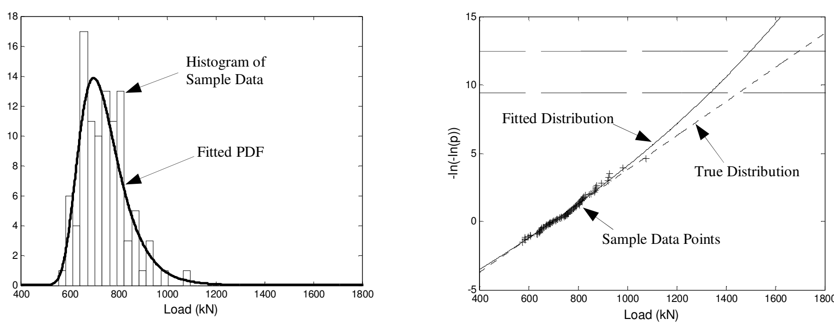
$$\sigma = \frac{S\sqrt{6}}{\pi} \tag{3}$$

where: S is the standard deviation of the data and
σ is the scale parameter of the data

To calibrate this method the KDE approach is applied to different data sets using caps of 70, 80 and 90% of the estimated scale parameter for the datasets. These sample data sets are generated from Gumbel distributions using 10 sets of location and scale parameters. Parameters which correspond to typical truck gross vehicle weights and individual, tandem and tridem axle weights are chosen. Different combinations are selected so as to give different ratios of one parameter to the other. For each set of Gumbel parameters, 10 sets of 100 data points are randomly generated and the KDE method applied to each data set for the three different caps. In all cases the cap of 80% of the estimated scale parameters gives the best estimates of the true distribution.

3 RESULTS

To compare the accuracy of the KDE method to current best practice, 10 datasets, each containing 100 maximum daily truck gross vehicle weights, are randomly generated from a Gumbel distribution with a location parameter of 700 kN and a scale parameter of 80 kN. Maximum daily weigh-in-motion (WIM) data is used as a guide for picking these parameters. Based on a five day week, this would represent 5 months of WIM data. These datasets are first analysed using a conventional approach, i.e., the truck weights are plotted on Gumbel probability paper and a Weibull distribution fitted to the data points. *Fig. 6* shows one of the data sets and the fitted Weibull distribution. The 50 year and 1000 year return period characteristic values are then calculated using the fitted distribution. The estimated return periods for the 10 sets of data are compared with their true theoretical values and a root mean square error (RMSE) calculated.

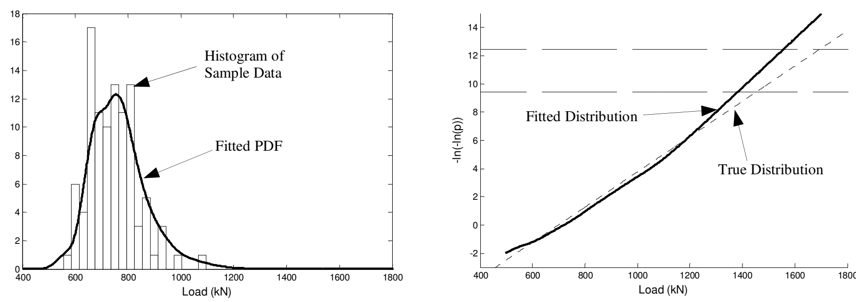


(a) Histogram and fitted PDF

(b) Gumbel probability paper with 50 and 1000 year return period levels shown (- -)

Figure 6. Weibull distribution fitted to data

The KDE method is then applied to the same 10 data sets and the 50 and 1000 year return period values were again estimated and the RMSE calculated. *Fig. 7* shows the KDE estimate of the distribution for the same data set as *Fig. 6*. A comparison of *Figs. 6(a)* and *7(a)* indicates a better fit with the KDE approach but a more realistic theoretical distribution from the conventional approach, as might be expected.



(a) Histogram and fitted PDF

(b) Gumbel probability paper with 50 and 1000 year return period level shown (- -)

Figure 7. Kernel density estimate of distribution of data

Table 1 shows the RMSEs for the two approaches. The KDE method achieves better estimates for both the 50 and 1000 year return period values with an overall reduction in the RMSE of 28%.

Table 1. Root mean squared error of estimated return periods for both methods

	50 yr RMSE	**1000 yr RMSE**
Fitted Weibull	6.12 %	8.51 %
Kernel Density Estimators	3.99 %	6.49 %

4 CONCLUSION

Accurate estimation of characteristic loading is critical for both bridge assessment and design. This estimation process is highly dependent on the extreme upper tail of the distribution of measured data where relatively few data points are available. The kernel density estimator method provides a non-parametric method of interpolating between and extrapolating beyond the data in this region. When applied to the 10 data sets of 100 randomly generated truck gross vehicle weights in this paper, it achieved a substantial reduction in error when estimating the 50 year and 1000 year load events.

ACKNOWLEDGMENTS

The financial support of the Irish National Roads Authority for this research is gratefully acknowledged.

REFERENCES

[1] Castillo, E., *Extreme Value Theory in Engineering*. Academic Press, San Diego, 1988.

[2] Caprani, C.C., OBrien, E.J., McLachlan, G.J., "Characteristic traffic load effects from a mixture of loading events on short to medium span bridges", *Structural Safety*, Vol. 30, No. 5, pp. 394-404, 2008.

[3] Bailey, S.F., Bez, R., "Site specific probability distribution of extreme traffic action effects", *Probabilistic Engineering Mechanics*, Vol. 14, No. 1-2, pp. 19-26, 1999.

[4] Gindy, M., Nassif, H.H., "Comparison Of Traffic Load Models Based On Simulation And Measured Data", *Proc. Joint International Conference on Computing and Decision Making in Civil and Building Engineering*, pp. 2497-2506, Montréal, 2006.

[5] Nowak, A.S., "Live load model for highway bridges". *Structural Safety*, Vol. 13, pp. 53-66, 1993.

[6] Nowak, A.S., Nassif, H.H. and DeFrain, L., "Effect Of Truck Loads On Bridges", *Journal of Transportation Engineering (ASCE)*, Vol. 119, No. 6, pp. 853-67, 1993.

[7] Kim, K.D., Heo, J.H., "Comparative study of flood quantiles estimation by nonparametric models", *Journal of Hydrology*, Vol. 260, No. 1-4, pp. 176-193, 2002.

[8] Moon, Y.I., Lall, U., "Kernel Quantile Function Estimator For Flood Frequency Analysis", *Water Resources Research*, Vol. 30, No. 11, pp. 3095-3103, 1994.

[9] Silverman, B.W., *Density Estimation for Statistics and Data Analysis*, Monographs on statistics and applied probability, Chapman & Hall, London, 1986.

[10] Jones, M.C., Marron, J.S., Sheather, S.J., "A Brief Survey Of Bandwidth Selection For Density Estimation", *Journal Of The American Statistical Association*, Vol. 91, No. 433, pp. 401-407, 1996

[11] Scott, D. W., *Multivariate density estimation : theory, practice, and visualization*, Wiley, New York, 1992.

[12] Abramson, I.S. "On bandwidth variation in kernel estimates - a square root law". *The Annals of Statistics*, Vol. 10, No. 4, pp. 1217-1223, 1982.

[13] Ang, A. H.-S. and Tang, W.H., *Probability concepts in engineering : emphasis on applications in civil & environmental engineering*, Second Edition, Wiley, New York, 2007.