A ROAD TRAFFIC PREDICTION STUDY BASED ON WEIGH-IN-MOTION DATA

Nicolae CIONT1, Rodica Dorina CADAR1, Dalia Sabina CÎMPEAN2

1 Technical University of Cluj Napoca, Department of Railways, Roads and Bridges, 28th Memorandumului Street, Cluj-Napoca, 400114, Romania
2 Technical University of Cluj Napoca, Department of Mathematics, 28th Memorandumului Street, Cluj-Napoca, 400114, Romania

Corresponding author: Dalia Sabina CÎMPEAN, Dalia.Cimpean@math.utcluj.ro

Abstract. An Advanced Traffic Management System (ATMS) represents one of the most important Intelligent Transportation Systems (ITS) subdomains. The main objectives of systems based on such technology are focused on improving road traffic parameters and traffic safety. This paper presents a weigh-in-motion (WIM) system used to record traffic data, while also proposing an application software to process the information. Two parametric models which describe the heavy vehicles traffic data were established by using a local regression smoothing. A statistical validation of the obtained fitting functions of the models was carried out and discussed. A number of goodness-of-fit measures were used to evaluate the overall performance of these simulation models. The results and discussion are clearly pointed out and compared in graphs, in order to obtain the best traffic prediction during a considered time period.

Key words: traffic data processing, weigh-in-motion, application software, mathematical models, prediction in time, parametric fit.

1. INTRODUCTION

A major problem of the last decades is the dramatic increase of traffic volume, due to mobility demands which become more acute every day in most of the world’s cities. This problem affects the urban communities both physically and mentally (Hennessy et al. [20]). Specifically, heavy vehicles have a negative impact on road structures and traffic flow.

Several techniques, dedicated to this problem, were applied and reported recently, such as visualisation techniques (Kosara et al. [26]). Various traffic sensors are available to measure traffic parameters, such as volume, speed and travel time (Jaume Barceló et al. [24]). Mathematical models are used to describe the registered data (Jin and Zhang [25], Bierlaire et al. [8], Li [28]).

Various methods to evaluate the existing traffic conditions and predict the ongoing state of traffic parameters have been developed (Comert and Bezuglov [14]). Parametric models include time-series and Kalman filtering. Nonparametric ones contain regression methods and neural networks. Chen et al. [11] compared different traffic prediction models, based on collected data, by using a loop detector located on the SR99 freeway in California. Vlahogianni et al. [35] reviewed in detail the progress of short-term traffic forecasting since 2004, by focusing on 10 challenging issues.

Intelligent Transportation Systems (ITS) represent advanced applications that provide innovative traffic management services. These systems are based on using information technology and software applications to design, use and maintain transportation systems. The main advantage of ITS is that the users are able to use the transportation network efficiently and safely. Officially introduced in 1991 (Abdulhai and Kattan [7]), ITS systems have known a quick development (Roess et al. [31]). An Advanced Traffic Management System (ATMS) represents one of the most important ITS subdomains. Its main objectives involve using technology to improve traffic flow and increase traffic safety at the same time.

Weigh-in-motion (WIM) represents the process of measuring the dynamic tyre forces of a moving vehicle and estimating the corresponding tyre loads of the equivalent static vehicle (ASTM E1318 [6]). High-speed weigh-in-motion (HS-WIM) systems are useful and unobtrusive traffic engineering tools (Slavik [32]), used to continuously measure different traffic parameters without stopping vehicles, thus improving
the monitoring efficiency. Most WIM systems include piezoelectric sensors and inductive loops, as well as electronic data processing equipment. This intrusive traffic monitoring technology (Bottero et al. [9]) is used worldwide (HPMS [22]; SHA Maryland DoT [4]; AND 584 [3]; Intertraffic world [1]; ITS International [2]) for applications such as: pavement design and research, bridge design and monitoring, traffic engineering and enforcement (Wang and Wu [36]).

WIM data can be used to evaluate traffic volume on a studied road sector in order to establish real axle loads. Therefore, a rationally-dimensioned road structure could be adopted. Traffic planning may also be based on WIM data. Building new roads, expanding existing ones and traffic management strategies are measures which are all implemented based on present and future estimated traffic data. Traffic modelling is based on measured volumes, speeds and delays. Furthermore, traffic counts are the main source of information used to establish traffic flow patterns (Georgia DoT [17]).

WIM data processing is of main importance to both managers and users. However, post-recording processing is often complicated, implying considerable efforts. This is one of the main reasons why users are unable to access relevant traffic information. Independent software applications for WIM data recording and processing have been developed in the last 15 to 20 years. In Europe, the first efforts were focused on creating a common database for WIM data storing and reporting (Henny [21]). The LTPP database established in the U.S.A. allowed the users to access the recorded traffic information (Hallenbeck [18]). Other applications developed at the end of the 90s were Telstar’s Roadways (Harding [19]) and ARRB’s WIMLINK (Cropley, [15]; Austroads [5]) and KAJI (Sunggiardi and Putranto [33]).

This paper gives a qualitative and a quantitative approach to heavy vehicles traffic on an East European carriageway. Its goals are to propose a WIM data processing software and to analyze the heavy vehicles traffic data collected using a HS-WIM system from a statistical point of view. Two mathematical models are obtained and compared to complete the study, giving a prediction of traffic flow congestion over a time period. The best proposed fitting function of the data is found with very good accuracy. Also, the registered data and the predicted models are in very good agreement for similar periods of time. The results, completed by a statistical validation, are plotted compared and discussed. The study is motivated by the practical decisions which could be taken to improve traffic flow, by following the proposed prediction model.

2. DATA COLLECTING AND PROCESSING

2.1. TRAFFIC DATA COLLECTING SYSTEM

The data collecting system used in this study is a HS-WIM device installed on the European E60 road (DN1 National Road), at the Western edge of the city of Cluj Napoca, Romania. (Fig. 1, Fig. 2). The studied road sector consists of a single carriageway with 4 (four) traffic lanes. The used WIM system has a piezo-loop-piezo (PLP) configuration, suitable for the selected location (NCHRP [30]).

![Fig. 1 – WIM location (map source: Google Maps).](image1)

![Fig. 2 – Cluj Napoca WIM system.](image2)

The recorded information is stored on a memory unit. The user may download this information either on site or through an online connection. The system’s electronic data processing equipment automatically
converts the traffic data into daily Microsoft Access (MS Access) files, which are then stored in a WIM database. The traffic data used in this paper was collected between April 25, 2013 and April 24, 2014.

2.2. DATA PROCESSING APPLICATION SOFTWARE

Each database file includes a considerable amount of detailed traffic information, such as: date, time, speed, lane number, direction, category, number of axles, gross vehicle weight etc. However, such large amount of data is difficult to handle, especially for long monitoring periods. The size of each daily file exported by the WIM software exceeds 20 MB, which is quite large. For a more familiar approach, data could be exported into MS Excel format, but, in that case, the data processing possibilities are rather limited. This is due to worksheet limits, implying time-consuming processes and high computer memory usage.

In order to enhance the WIM traffic data processing and analysis capability, application software was designed, based on the Visual Basic for Applications (VBA) programming language. Its main purpose is to provide a useful tool for user-related traffic engineering applications. It also includes a statistical module which evaluates basic functions (e.g. mean, median, mode etc.). Based on the processed information, further statistical analysis may be done.

The input files are analyzed and processed using the MS Access 2010 database management system. The basic application flowchart (Fig. 3) starts by loading the WIM database files for the desired study period. In some isolated cases, the WIM data recording equipment provides erroneous information, such as nonexistent lanes for example. This is why several validity checks are performed repeatedly, until the database is cleared of invalid recordings. As soon as the checks are passed, the application processes the input data, according to the user request. The application efficiency is enhanced by a selective database browsing algorithm, which scans the database files only for the lines and columns needed for the requested operation, and by other time and memory saving measures. By default, the results are displayed in MS Access format. However, the application also provides the option to export them to MS Excel.

![Fig. 3 – Application software flowchart.](image)

The designed application software evaluates several essential traffic engineering functions, providing important information on the following: total and average traffic volumes (hourly, daily, weekly, monthly); annual average daily traffic (AADT); lane distribution factors; peak hour studies, using 5 and 15-minute rate of flow intervals; spot speed studies, using 1, 2 or 5 km/h intervals; traffic density; speed-flow-density functions; equivalent axle loads, used for pavement design; capacity evaluation.

Using the designed application, the effects of different road network changes on traffic conditions may be studied. Ciont et al. [13] analyzed two possible traffic flow improvement scenarios for the studied area (Fig. 1), using software modeling and WIM measured data during the morning and afternoon peak hours. Designing road structures is also based on traffic evaluation. The processed WIM traffic data was used to comparatively study the design of road structures (Iliescu et al. [23]; Cadar et al. [10]). The results were calculated based on traffic parameters provided by standards, as well as using WIM data, showing that...
different work hypotheses may lead to technical, economical and environmental advantages. The studied area (Fig. 1) is intensely used by commuters. Traffic is rather congested during the morning and afternoon peak hours. Using the designed application, Ciont et al. [12] conducted a study on the peak hour factor (PHF), concluding that the maximum average PHF for working days on the studied sector is 0.950.

3. THE TRAFFIC SIMULATION MODELS

The traffic data used in this study was recorded by the presented HS-WIM system and processed using the designed application software. The approach on the traffic flow issue is based on empirical analysis, i.e., observation and mathematical curve fitting. To perform this analysis, the MatLab software was used.

The heavy vehicles (gross combination mass > 3500 kg) average weekday traffic, collected and processed by using the WIM measurements, were considered over one year time period. The WIM data processing software is showing that, for the studied period, heavy vehicles represent only 2.3% of traffic volume. This is mainly due to the fact that most lorries are using the Cluj-Napoca bypass (A3 motorway section, Fig. 1), thus avoiding entering the urban area. Even so, heavy vehicles passing through the city pose a problem to roads and traffic stream parameters, especially during peak hours.

The initial data were filtered and smoothed using a local regression smoothing. A weighed linear least-squares regression was performed (Casella and Berger [16]; Larson and Farber [27]; Moore and McCabe [29]) for two smoothing methods “lowess” and “loess”, respectively (see Fig. 4). The “lowess” uses a first degree polynomial since the “loess” uses a second degree polynomial. The smoothed value is given by the weighted regression at the predictor value of interest. The weights are given by

\[ u_i = \left(1 - \frac{x-x_i}{d(x)} \right)^3 \]

Here, \( x \) is the predictor value associated to the response value, \( x_i \) are the nearest neighbors of \( x \), as defined by the span, \( d(x) \) is the distance from \( x \) to the most distant predictor value within the span.

3.1. THE FITTING FUNCTIONS

Two parametric models were found to predict the real data in time. For the smoothed data by using “lowess”, the best fitting function is sinusoidal and was obtained in the form:

\[
f(x) = a_1 \sin(b_1 x + c_1) + a_2 \sin(b_2 x + c_2) + a_3 \sin(b_3 x + c_3) + a_4 \sin(b_4 x + c_4) + a_5 \sin(b_5 x + c_5) + a_6 \sin(b_6 x + c_6) + a_7 \sin(b_7 x + c_7) + a_8 \sin(b_8 x + c_8)
\]  

(1)

The coefficients (found with 95% confidence bounds) are:

- \( a_1 = 2084 \), \( b_1 = 0.05923 \), \( c_1 = 0.1272 \), \( a_2 = 1090 \), \( b_2 = 0.116 \), \( c_2 = 2.193 \), \( a_3 = 497.8 \), \( b_3 = 0.1498 \),
- \( a_4 = 5.302 \), \( a_4 = 56.97 \), \( b_4 = 0.4679 \), \( c_4 = 1.796 \), \( a_5 = 18.85 \), \( b_5 = 0.6194 \), \( c_5 = -0.02418 \), \( a_6 = 7.42 \),
- \( b_6 = 0.7264 \), \( c_6 = -0.6182 \), \( a_7 = 39.96 \), \( b_7 = 1.026 \), \( c_7 = -0.1524 \), \( a_8 = 37.99 \), \( b_8 = 1.038 \), \( c_8 = 2.571 \).

Also, the “loess” smoothing of the data was performed and a Fourier model was obtained for the fitting function, as seen below:

\[
g(x) = a_0 + a_1 \cos(xw) + b_1 \sin(xw) + a_2 \cos(2xw) + b_2 \sin(2xw) + a_3 \cos(3xw) + b_3 \sin(3xw) + a_4 \cos(4xw) + b_4 \sin(4xw) + a_5 \cos(5xw) + b_5 \sin(5xw) + a_6 \cos(6xw) + b_6 \sin(6xw) + a_7 \cos(7xw) + b_7 \sin(7xw) + a_8 \cos(8xw) + b_8 \sin(8xw)
\]  

(2)

Here, the coefficients (with 95% confidence bounds) are:

- \( a_0 = 1283 \), \( a_1 = -37.42 \), \( b_1 = 24.3 \), \( a_2 = -191.2 \), \( b_2 = -74.09 \), \( a_3 = -45.45 \), \( b_3 = -6.536 \),
- \( a_4 = 55.25 \), \( b_4 = -67.14 \), \( a_5 = 23.69 \), \( b_5 = 26.87 \), \( a_6 = -3.895 \), \( b_6 = 26.99 \),
- \( a_7 = -33.12 \), \( b_7 = -19.68 \), \( a_8 = 15.09 \), \( b_8 = 19.4 \), \( w = 0.1111 \).
3.2. STATISTICAL VALIDATION

The general simulation literature includes a large number of methods for the statistical validation of simulation models. For the present paper, a number of goodness-of-fit measures are used to evaluate the overall performance of simulation models. These are explained in works by Moore and McCabe [29] and Toledo and Koutsopoulos [34]. They will only be enumerated below and the values, important to obtain the best fit, will be pointed out.

The R-square is defined as the ratio of the sum of squares of the regression (SSR) and the total sum of squares (SST) and can take a value between 0 and 1. A value closer to 1 indicates that a greater proportion of variance is accounted for by the model. Then, for our two models given by (1) and (2), the R-square value is 0.99 that means the fit explains 99% of the total variation in the data about the average (precisely 0.9997 for (1) and 0.9967 for (2)). The sum of squares due to error (SSE) for which a value closer to 0 indicates that the model has a smaller error component (for example, the model (1)). The adjusted R-square is the best indicator of the fit quality when two models are compared and again a value closer to 1 (0.99 for our both proposed models) indicates a better fit. Finally, the root mean squared error (RMSE) known as the fit standard error is an estimate of the standard deviation of the random component in the data, with a value closer to 0 (3.2 for the best given model) indicating a fit that is more useful for prediction.

Considering the above matching conditions, our two proposed models approximate the real data with a very good accuracy and the best fitting function is given by (1), since the model described by (2) is also a very good choice for prediction.

4. RESULTS AND DISCUSSION

Figure 4 plots the initial data (the weights average of the vehicles traffic flow per week) for a year time period with blue dots. In the same graph, the smoothed values obtained by the “lowess” are compared with those obtained by “loess”. It should be mentioned that, the “lowess smooth” is a better option for the given problem, but, the “loess” it is also good and will give a good approximation using a Fourier fitting function. In the next figure, the best fitting function, given by (1), is represented, together with the initial and smoothed data (see Fig. 5). In the Fig. 6, the confidence and the prediction bounds are seen. The proposed fit is observed to give a very good approximation of the fitted data. Figure 7 represents the plot of the Fourier model, fit (2), given for smoothed “loess” data set and compared with the original data. It also gives a very good approximation of the studied problem. Figure 8, gives the comparison of the fitting functions (1) and (2) obtained with the two smoothing methods and reveals the advantage of comparing them to take the best statistic decision. It is worth to mention that, an analysis of the goodness to fit was obtained. The graphical methods (Figs. 4–8) are completed with the study of the goodness-to-fit statistics for the parametric models (Moore and McCabe [29]; Toledo and Koutsopoulos [34]). According to this, the best performances were found for the proposed model (1), represented in Fig. 5.

5. CONCLUSIONS

This paper presents a problem based on traffic data processing software, used to provide essential information based on database recordings collected using a HS-WIM system. Designed in VBA, its main purpose is to process, analyze and report the WIM traffic data, according to the user function request. The software provides important results and information on various parameters, used in both traffic studies and pavement design. The application functionality is crucial to effectively process the recorded WIM traffic data. The database is basically useless unless the designed software is applied. It may handle one or several WIM systems, providing essential information to both administrators and general users. In this study, the heavy vehicles traffic data was analysed, not only from a traffic engineering perspective, but also from a statistical point of view. Two mathematical models were found and compared to give a prediction of the traffic flow congestion in a time period. A statistical validation of these models was performed and reveals a very accurate fitting of the both functions with the smoothed data. The best fitting function of the data is found to be represented by (1). Also, the registered data and the predicted models are found to be in very
good agreement. The results are plotted and discussed together with a comparison of the proposed models and methods. The study could predict and prevent in a time period, the urban traffic congestion and problems which appear in similar given conditions.

![Figure 4](image1.png)

**Fig. 4** – A comparison of the original data (dots) and smoothed data (“lowess” and “loess”) versus time in weeks ($x$).

![Figure 5](image2.png)

**Fig. 5** – Fitting sinusoidal function $f$ versus time $x$, plotted together with the original and “lowess” smoothed data.

![Figure 6](image3.png)

**Fig. 6** – Analysis of the fitting function $f$, (1), for smoothed “lowess” dataset.
Fig. 7 – Analysis of the fitting function $g$, (2), given for smoothed “loess” data set.

Fig. 8 – The comparison of the fitting functions obtained with the two smoothing methods.

ACKNOWLEDGEMENTS

This research was supported by the Sectorial Operational Programme Human Resources Development POSDRU/159/1.5/S/137516 financed from the European Social Fund and by the Romanian Government. The WIM system was supported by the project “Development and support multidisciplinary postdoctoral programs in primordial technical areas of national strategy of the research – development – innovation”4DPOSTDOC, POSDRU/89/1.5/S/52603, project co-funded from European Social Fund through Sectorial Operational Program Human Resources 2007-2013”.

REFERENCES

1. INTERTRAFFIC WORLD, A lighter load, UKIP Media& Events Ltd., Abinger House, Church Street, Dorking, Surrey, RH4 1DF, United Kingdom, pp. 106-107, 2014.
3. *** Normativul pentru determinarea traficului rutier pentru proiectarea drumurilor din punct de vedere al capacității portante șii al capacității de circulație, AND 584/2012.