A Polynomial Short-Term Traffic Flow Prediction Model

NICOLAE CIONT, RODICA DORINA CADAR, MIHAI ILIESCU, MELANIA ROZALIA BOITOR
Department of Railways, Roads and Bridges
Technical University of Cluj Napoca
str. Memorandumului nr. 28, 400114 Cluj Napoca
ROMANIA
nicolae.ciont@cfdp.utcluj.ro http://www.utcluj.ro

Abstract: In this paper, a polynomial short-term traffic flow prediction model is proposed. It consists of a cubic function built to forecast the short-term evolution of traffic volumes during working days morning peak hours. The traffic data was collected using a weigh-in-motion system installed on an important multilane highway in Romania. Information processing was carried out using a designed application software, which provided separate volumes for consecutive 15-minute intervals during the three-weeks study period in early 2015. The obtained results indicated good prediction accuracies, showing that the proposed polynomial model is a useful short-term traffic prediction tool.

Key-Words: short-term traffic prediction, polynomial model, intelligent transport systems, weigh-in-motion, prediction performance, forecast methodology.

1 Introduction

Intelligent Transport Systems (ITS) are advanced applications which provide the users information on different transport modes, thus offering the possibility to use the existing road network in a safe and efficient manner. They assist both users and engineers in the design, operational processes and maintenance of road networks [8]. ITS applications use software applications and information technology in order to improve the efficiency of road transportation systems [22].

Moretti et al. [28] claim that most decisions in transportation engineering are marked by imprecision and uncertainty. Traffic engineering is that transportation engineering branch that deals with the planning, design and operational aspects of both roads and their associated systems and networks. In many cases, the associated traffic problems are solved using soft computing methodologies [28]. According to Guo et al. [12], there are two main types of traffic management and control systems: reactive and proactive. The former are based on current traffic conditions, whereas the latter use traffic predictions. Accurate real-time data and short-term traffic forecasting are essential for the development of efficient and dynamic traffic management and control systems [44]. The evolution of ITS systems has provided essential traffic engineering tools, which are highly useful to users, administrators and engineers. A considerable percentage of these tools are based on short-term traffic forecasting models, which use past and current traffic data inputs. They provide the anticipatory traffic parameters required for the design and functionality of proactive systems [12].

Short-term traffic flow prediction represents one of the most important aspects of ITS [55]. As a part of traffic control and management [28], traffic forecasting constitutes the process of directly estimating conditions at a future time, based on continuous traffic information [22]. It is also used to build prediction intervals and evaluate the forecast uncertainty [12]. Traffic prediction is based on information collected using different types of sensors. They record real-time traffic data in order to predict likely changes in traffic flow [24].

Short-term traffic prediction is essential to the effectiveness and event detection accuracy of ITS [53]. At the same time, Zhang et al. [55] emphasise that the performance of ITS depends on the quality and accuracy of real-time traffic data. Components such as advanced traveler information systems (ATIS), online transportation management and advanced traffic management systems (ATMS) extensively use predicted traffic information [22][24][51][55]. Furthermore, traffic forecasting models are applied to route guidance systems, adaptive ramp metering, signal control, deployment of emergency management systems and variable message signs [14][35].

According to Lin et al. [24], the evolution of variables such as volume, speed or travel time is predicted over periods typically ranging from 5 to
60 minutes, based on current and historic measurements [14]. However, some researchers consider different time intervals, ranging from less than 5 minutes [53] to 30 minutes [35] or even a few hours [14][39][41].

The remainder of this paper is organised as follows: section 2 presents a brief literature review on short-term traffic forecasting, while section 3 is focused on the adopted methodology. The model development, implementation and performance are discussed in section 4, based on a case study conducted in Romania. The conclusions of this study and several future research directions are presented in section 5.

2 Brief literature review

Studies on traffic prediction have begun in the second part of the 20th century and have been developing ever since. In recent years, many studies have been published on the subject. One of the earliest short-term traffic prediction methods is spectral analysis, proposed by Nicholson and Swann in 1974 [29]. It provides the required framework to highlight cyclical patterns in the collected traffic data [55]. The performance of short-term traffic forecast depends on the recorded information accuracy, as well as on the applied short-term prediction method [43]. The growing use of ITS is a current feature of traffic management and control strategies. The procedures are carried out using flexible mathematical methods and computers [22].

According to Sun et al. [35], forecasting models can be classified into three groups: using only historical data, using only current data and using both datasets. Van Hinsbergen et al. [38], as well as van Lint and van Hinsbergen [39], divide the short-term traffic forecast methods into three categories: naïve, parametric and non-parametric. Approaches that combine these forecast methods are called hybrid methods. The paper by Habtemichael and Cetin [14] describes each of these categories.

Traditional forecasting methods predict the current value as a function of its past values [55]. Time series models such as the Autoregressive Integrated Moving Average (ARIMA) [3][9][15][23] and the Seasonal ARIMA (SARIMA) [13][26][36][46] have been used in many studies. The ARIMA model, which predicts the current value on regression of its past value, has become a reference for the newly developed models [55]. The SARIMA model was also implemented in a large number of studies. However, one of its limitations is that it assumes a linear correlation among the time series values and may not capture the non-linearity in a complex transportation system [24].

Non-parametric approaches, such as neural networks (NN) [40][50][52], k-nearest neighbor (k-NN) [24][33][53] or support vector machines (SVM) [2][44][48], do not make strong assumptions on the underlying model form [12]. K-NN has a small error ratio and good error distribution [20]. According to Karlaftis and Vlahogianni [21], NNs have a good forecast ability and modeling adaptability, offering good predictions, as SVMs do. However, the same researchers have previously noted that most methodologies face difficulties at modeling and predicting traffic volume, due to its oscillating nature [42].

Recent studies used traffic data collected using non-intrusive monitoring tools, such as Bluetooth [1][34] and GPS [16][25]. The effectiveness of these systems has continuously improved, but further development is still necessary [43].

Other prediction methods were used by researchers in their studies, each with its own advantages and disadvantages: time series models [27][47], fuzzy logic system [54], wavelet network models [49], artificial intelligence [17] etc. The local linear regression model proposed by Sun et al. [35] performed better than other non-parametric approaches, such as k-NN.

Filtering methods have been widely applied to short-term traffic forecast models. These include: Kalman filter [12][19][45], Bayesian dynamic linear models, particle filters [5], recursive least square (RLS), least mean square (LMS) filters etc. The adaptive Kalman filter proposed by Guo et al. [12] led to acceptable prediction intervals, even under volatile traffic conditions. Wang et al. [43] proposed a new Bayesian combination method (BCM), which improved the accuracy of the traditional BCM, originally put forward by Petridis et al. [30].

Statistical techniques such as simple smoothing [4][10][31] were also used.

3 Methodology

The main purpose of the forecast methodology proposed in this paper is to approximate the short-term evolution of traffic volumes during weekdays morning peak hours, based on past traffic data. The approach is based on traffic evolution in a peri-urban area intensely used by commuters. Traffic volumes are highest during the morning and afternoon peak hours. However, the prediction model may be applied to other settings as well, where traffic stream parameters variation is different.
The adopted methodology is based on polynomial regression, although many researchers have studied and emphasised the advantages of different other forecasting models. Fan and Gijbels [11] concluded that linear regression performs better than other models. Sun et al. [35] affirm that methods such as the k-NN [32] or kernel smoothing [10] may be included in the group of local polynomial regression models [11]. Literature includes studies based on a wide range of other prediction models.

3.1 Cubic function

The prediction method used in this paper is based on a polynomial function of degree 3, which is linear in the coefficients (eq.(1)):

\[ P(x) = A_3 \cdot x^3 + A_2 \cdot x^2 + A_1 \cdot x + A_0 \]  (1)

where \( A_0, A_1, A_2, A_3 \) – coefficients, \( A_3 \neq 0 \).

The employed cubic function was intended to approximate the variation of peak traffic volumes in the morning, during working days. The main purpose was to establish a suitable function, based on continuously collected traffic data, and to evaluate its performance as a short-term traffic prediction model for 15-minute time intervals. It is also to be noted that the concept of polynomial regression may be adopted for other time intervals as well, as long as the model performance is satisfactory.

The proposed methodology assumes traffic volumes are counted using a functional monitoring system. Furthermore, the recorded data needs to be processed and validated using an appropriate tool, which provides separate volumes for consecutive 15-minute intervals. This approach was considered suitable, as it was employed in other similar works. Researchers have also used sub-hourly intervals to build short-term traffic prediction models. Wang et al. [43] proposed their new BCM using data collected over 15-minute intervals. Kumar et al. [22] applied Artificial Neural Network (ANN) for short-term prediction using past traffic data collected during 15-minute intervals. Lin et al. [24] proposed their forecasting method called k-nearest neighbor based local linear wavelet neural network (kNN-LLWNN) based on five-minute traffic volumes.

4 Model development

4.1 Data collecting and processing

The traffic data used in this study was collected using a high-speed weigh-in-motion (WIM) system, installed on the European E60 road, in the peri-urban area of the city of Cluj Napoca, Romania [6]. The specified location is a major commuting route between the inner city and the suburban area nearby...
The system converts and exports the recorded traffic data into daily Microsoft Access files. The data processing tool used in this paper is a designed application software [6][8]. It represents an essential tool for the processing, analysis and reporting of the recorded traffic data.

The built polynomial forecast tool was trained based on traffic data collected on ten working days, between Monday, 23 February 2015 and Friday, 6 March 2015. The model was then tested using information recorded over the following five weekdays (9-13 March 2015).

The employed highly efficient traffic data processing tool provided the average hourly traffic repartition on working days for the studied period (Fig. 1, dotted line). As mentioned before, volumes are highest during the morning and afternoon rush hours (Fig. 1). The short-term traffic forecast model proposed in this paper is based on the morning peak volumes. This is supported by the results obtained by Ciont et al. [7], who showed that, in the studied area, 93% of morning trips are taken between 7:00 and 9:00 hrs, and that 8:00 – 8:15 hrs is the busiest quarter-hour in 51% of cases. Specifically, the procedure in this paper analysed data collected from 7:30 to 9:00 hrs (Fig. 1, continuous line).

### 4.2 Model implementation and performance

The studied time interval was divided into six 15-minute intervals, numbered 1 to 6 (Fig. 2). They correspond to intervals 7:30 – 7:45, 7:45 – 8:00 and so on, until 8:45 – 9:00 hrs, respectively. The designed application software provided the separate traffic volumes for the six consecutive 15-minute intervals, necessary to build the prediction model. The study focused on the highlighted area in Fig. 1, using the historical data obtained for the studied 15-minute intervals on the ten analysed weekdays (Fig. 2).

![Fig. 1 Average hourly traffic repartition](image1.png)

The prediction performance measures obtained in the building process indicated promising results (Table 1), especially for the 8:00 – 8:45 hrs interval. The resulted values were satisfactory enough to test the model using traffic data collected over the following five weekdays (Fig. 3). Considering the initial results, it was expected that the forecast would perform well.

<table>
<thead>
<tr>
<th>15-min interval</th>
<th>MAE</th>
<th>MAPE [%]</th>
<th>VAPE [%]</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:30 - 7:45</td>
<td>25</td>
<td>2.5</td>
<td>1.1</td>
<td>0.61</td>
</tr>
<tr>
<td>7:45 - 8:00</td>
<td>31</td>
<td>2.8</td>
<td>2.4</td>
<td>0.65</td>
</tr>
<tr>
<td>8:00 - 8:15</td>
<td>13</td>
<td>1.1</td>
<td>1.5</td>
<td>0.49</td>
</tr>
<tr>
<td>8:15 - 8:30</td>
<td>26</td>
<td>2.5</td>
<td>1.8</td>
<td>0.60</td>
</tr>
<tr>
<td>8:30 - 8:45</td>
<td>18</td>
<td>1.7</td>
<td>0.5</td>
<td>0.97</td>
</tr>
<tr>
<td>8:45 - 9:00</td>
<td>32</td>
<td>3.2</td>
<td>1.7</td>
<td>0.61</td>
</tr>
</tbody>
</table>

![Fig. 2 Prediction model on historical data](image2.png)

![Fig. 3 Prediction model on test data](image3.png)

Indeed, the testing phase showed that the overall accuracy of the proposed short-term traffic prediction model is very good (Table 2). An average MAE of 28 vehicles/15 minutes was obtained (compared to 24 in the building phase), showing a good prediction, under the given circumstances. MAE values as low as 11-12 vehicles/15 minutes.
resulted, indicating minimum MAPE values of 1.1 – 1.2 %. As in the building phase, the prediction accuracy was at its highest between 8:00 and 8:45 hrs, as well as between 7:30 and 7:45 hrs. However, it is to be noted that the performance measures obtained for the other quarter-hourly intervals (7:45 – 8:00 hrs, 8:45 – 9:00 hrs) still indicate good prediction accuracies, although the model performed better during other periods. Excellent values obtained for the VAPE and MASE measures were also encouraging.

Table 2. Prediction performance measures, testing phase

<table>
<thead>
<tr>
<th>15-min interval</th>
<th>MAE</th>
<th>MAPE [%]</th>
<th>VAPE [%]</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:30 - 7:45</td>
<td>12</td>
<td>1.2</td>
<td>0.9</td>
<td>0.79</td>
</tr>
<tr>
<td>7:45 - 8:00</td>
<td>42</td>
<td>4.0</td>
<td>2.7</td>
<td>0.82</td>
</tr>
<tr>
<td>8:00 - 8:15</td>
<td>32</td>
<td>3.0</td>
<td>3.2</td>
<td>0.61</td>
</tr>
<tr>
<td>8:15 - 8:30</td>
<td>11</td>
<td>1.1</td>
<td>1.6</td>
<td>0.58</td>
</tr>
<tr>
<td>8:30 - 8:45</td>
<td>24</td>
<td>2.4</td>
<td>1.7</td>
<td>0.69</td>
</tr>
<tr>
<td>8:45 - 9:00</td>
<td>46</td>
<td>5.0</td>
<td>6.1</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Although the proposed prediction model performs well, it also has several limitations. Its performance is affected on isolated days with low traffic volumes (i.e. national holidays), which may represent an unpredicted event during the building phase. In this case, other short-term forecasting tools might perform better. Other shortcomings may be encountered if such a model is applied during weekends or on locations where volume peaks appear irregularly (e.g. near concert halls or stadiums). However, given the conditions considered in this paper, the proposed polynomial model is a useful short-term traffic prediction tool. The obtained results are encouraging, showing that the polynomial model is performant. Although in this paper the proposed function was built based on peak volumes during the morning, similar models may be applied to other hour intervals as well.

5 Conclusions

This paper proposes a short-term traffic flow prediction model, based on polynomial regression. A cubic function was built to model the traffic volumes evolution during weekdays morning peak hours, based on past traffic data. The information used in this paper was collected using a WIM system installed on an important commuting route between the Cluj Napoca inner city and a major nearby suburban area, in Romania.

A designed application software was the essential data processing tool, employed to analyse and report the recorded traffic data. It provided separate traffic volumes for consecutive 15-minute intervals during the morning peak hours in February and March 2015.

The results provided by the employed performance measures, in both building and testing phases, indicated good prediction accuracies. Despite its several shortcomings, the proposed polynomial model is a useful short-term traffic prediction tool.

Further studies are planned to employ other prediction models as well, based on the continuously recorded traffic data. Results comparison should indicate the advantages and disadvantages of each model. Applying a suitable short-term prediction model may also be considered for other times of day or week. The obtained results could be used in correlation with other traffic systems, such as ATIS, ATMS, signal control or variable message signs, to provide real-time information to both users and administrators.

Acknowledgement:

This paper is supported by the Sectoral Operational Programme Human Resources Development POSDRU/159/1.5/S/137516 financed from the European Social Fund and by the Romanian Government.

References:


