PREDICTION OF PASSENGER CAR FUEL CONSUMPTION USING ARTIFICIAL NEURAL NETWORK: A CASE STUDY IN THE CITY OF NIŠ

UDC (629.4.016.2+004.032.26):531.39

Bratislav Predić¹, Miloš Madić², Miloš Roganović³, Marko Kovačević¹, Dragan Stojanović¹

¹University of Niš, Faculty of Electronic Engineering, Department of Computer Science, Niš, Republic of Serbia
²University of Niš, Faculty of Mechanical Engineering, Niš, Republic of Serbia
³University of Niš, Faculty of Occupational Safety, Niš, Republic of Serbia

Abstract. The reduction of CO₂ emission which is in direct relationship with fuel consumption is of prime importance for the future sustainable use of passenger cars. For the given passenger car, the fuel consumption in urban areas is mostly affected by the conditions related to traffic and driving behavior. In this paper, an artificial neural network model for the prediction of passenger car fuel consumption in the City of Niš was developed based on experimentally measured data recorded through on-board diagnostics equipment. Fuel consumption was assumed to be a function of car speed, a city zone, an hour of day and a day of week. A comprehensive preliminary investigation revealed that single hidden layer artificial neural network model having ten neurons can be efficiently trained with Levenberg-Marquardt algorithm to provide satisfactory prediction accuracy. Finally, the analysis of effects of the selected independent variables on the fuel consumption was discussed based on twelve 3D surface plots.

Key words: fuel consumption, OBD-II, modeling, prediction, artificial neural network

1. INTRODUCTION

Ever increasing population and the number of motor vehicles at a global level creates concerns about the transportation sustainability considering that this sector depends exclusively on fossil fuels, non-renewable energy sources, which have harmful impacts on both the environment and human health. Thus, with the ultimate aim to improve mobility, transportation planning must address and consider several issues including energy savings,
CO₂ emission reduction, unpredictable fuel price, social, economical, government’s policy objectives and regulations and other issues. For fulfillment of these goals and creation of strategies/policies for each transportation activity, i.e. it’s planning and optimization, a base and necessary step is to estimate and predict the fuel consumption and emissions.

In the literature, the estimation of fuel consumption and emissions is usually carried out by the application of mathematical models that consider various number of parameters related to vehicle engine, traffic, road, vehicle, weather and driver related categories [1, 2]. For a given vehicle type and weather conditions (i.e. temperature, humidity and wind speed), fuel consumption depends on a number of parameters which can be grouped in the following categories: (i) engine (type, volume, number of cylinders and valves,...), (ii) vehicle characteristics (weight, tire pattern and pressure, ...), (iii) road (grade, surface roughness,...), (iv) traffic (density, velocity, flow, number of vehicle stops,...), and (v) driving pattern (gear changing, speed, acceleration,...). Quantification of the effects of these parameters on fuel consumption is necessary to develop methods and strategies for fuel consumption prediction and fuel economy improvement [3]. It should be noted here that fuel consumption and emissions models are specific and valid in the geographical area where model data are gathered. As noted by Cartenì et al. [4], traffic flow (average speed, accelerations and complete mobility behavior in general), geometric infrastructure (width, radii of curvature, slopes, lateral disturbance index) and environment (average temperature, altitude, rainfall index, characteristics of the wind etc..) influence, in a non-negligible way, traffic-derived emission and fuel consumption factors.

Importance of fuel consumption estimation and prediction has attracted a number of researchers which have perceived this important topic from various aspects and in different context. Previous researches encompassed analyses of available data, fuel consumption data collection based on questionnaires, simulation and measurement and development of different mathematical models (instantaneous, modal, average speed, and other) for fuel consumption prediction. Ahn et al. [1] developed regression based mathematical models to predict vehicle fuel consumption and emissions using instantaneous speed and acceleration as explanatory variables. Wang et al. [3] investigated the influence of driving patterns on fuel consumption using a portable emissions measurement system on passenger cars. It has been observed that the modeled and measured fuel consumption rates for vehicles are in good agreement, and most of the differences between them are within 20%. He et al. [5] analyzed the current and future oil consumption and CO₂ emissions from China’s road transport sector. Treiber et al. [6] developed a model for the instantaneous fuel consumption estimation that includes vehicle properties, engine properties, and gear-selection schemes and implemented it for different passenger car types representing the vehicle fleet under consideration. Rahimi-Ajdadi and Abbaspour-Gilandeh [7] applied artificial neural network (ANN) and multiple regression models for prediction of fuel consumption of tractors. In the presented research, in order to obtain the best possible mathematical model, six ANN training algorithms and different ANN topologies were tested, as well as a stepwise procedure in the case of multiple regression analysis. Kara Togun and Baysec [8] developed explicit ANN model for the prediction of torque and brake specific fuel consumption (BSFC) of a gasoline engine in terms of spark advance, throttle position and engine speed. The experimental data from totally 81 test runs was used to train and test the ANN model. The authors concluded that the proposed ANN approach is quite accurate, fast and practical. The same authors in another study [9] revealed that ANN showed slightly better performance
than genetic programming in prediction of torque and BSFC. Yusaf et al. [10] investigated the use of ANN modeling to predict brake power, torque, BSFC, and exhaust emissions of a diesel engine modified to operate with a combination of both compressed natural gas and diesel fuels. The conducted analysis of the experimental data by the ANN revealed that there was good correlation between the ANN predicted results and the measured data, with correlation coefficient values ranging between 0.92 and 0.99. In order to improve the ANN based modeling, several topologies and combinations of activation functions in hidden and output layer were evaluated and trained using the experimental data. Siami-Irdemoosa and Dindarloo [11] used ANN approach to predict heavy mining dump trucks’ fuel consumption per cycle based on cyclic haulage activities. Mean absolute percentage error (MAPE) of 10% demonstrated applicability of ANN in prediction of the fuel consumption. Masikos et al. [12] proposed robust ANN based model for the prediction of energy consumption of a fully electric vehicle (FEV) by using twelve input variables related to vehicle, traffic, road and weather context as well as the driver profile. In order to accelerate, the typically slow rate of convergence experienced with the method of gradient descent for the purpose of ANN training, the use of the conjugate gradient descent method was proposed. Based on the achieved estimation error of 12.36% on average, the authors concluded that the proposed ANN model is quite reliable for predicting vehicle energy consumption. Wu and Liu [13] developed an ANN model for the prediction of car fuel consumption by taking into account the following predictors: make of car, engine type, weight of car, vehicle type and transmission system. The authors concluded that a conventional ANN model trained with conventional backpropagation (BP) algorithm has a learning capability for fuel consumption prediction. In a connected research, however, the authors showed that the results of training and testing from all databases showed that the radial basis function (RBF) ANN was better and faster than the BP ANN model [14]. El-Shawarby et al. [15] analyzed the impact of vehicle speed and acceleration levels on vehicle fuel consumption and emission rates using field data gathered under real world driving conditions. In addition, a comparison between the on-road fuel consumption and emission measurements and the Virginia Tech microscopic (VT-Micro) model estimates was presented to demonstrate the validity of the VT-Micromodel for the analysis of vehicle cruising and acceleration behavior. The VT-Micro model is a nonlinear regression model that utilizes a multi-dimensional polynomial model structure. This multiple regression model (VT-Micro) related the dependent variables (instantaneous fuel consumption and emission estimates) to a set of quantitative independent variables: namely, instantaneous speed and acceleration levels. Recently, Martínez-Morales et al. [16] applied ANN based modeling approach to predict the fuel consumption and NOx emission of a four stroke spark ignition engine. Engine speed, angle of the admission throttle valve, engine load, injection time ignition angle, and the intake manifold absolute pressure were selected as input variables. In the proposed approach, the multi-objective particle swarm optimization was used to determine weights of RBF ANN model.

Road transportation, as the dominant transportation mode in the Republic of Serbia, is the main source of fuel consumption and CO₂ emissions which are directly proportional to the fuel consumption. Although regarding fuel consumption and CO₂ emission hybrid vehicles and electromobiles can offer significant reduction, they are still too expensive for the average buyer. Moreover, other issues such as lack of fuel/charging stations, relatively short driving ranges, etc. resulted in the fact that today the fossil fuel cars remain predominant
in road transportation. Therefore, estimation of fuel consumption is an essential part of any attempt to reduce costs and the environmental impacts associated with fuel consumption. In recent years the fuel consumption of passenger cars has become one of the most important issues in car design but also one of the major concerns for consumers which use them in everyday urban and highway transport. In order to model complexities, nonlinearities and interactions between different variables and fuel consumption that inherently exist, this paper proposes an artificial intelligence approach based on the use of ANNs. Although it has been seen from literature review that ANNs have been recognized as a powerful mathematical tool for fuel consumption prediction, this paper proposed the development of instantaneous (microscopic) ANN based mathematical model that estimates the passenger car fuel consumption in the city of Niš (Serbia). More specifically, based on experimentally measured data recorded through on-board equipment, an ANN model which considers traffic conditions and the time history of average car speed as input variables is developed.

2. EXPERIMENTAL PROCEDURE AND DATA COLLECTION

The research methodology used in this study consists of four main stages. In the first stage experimental data measured in passenger car about fuel consumption in the city of Niš were collected. In the second stage data were statistically analyzed and randomly divided into two sets, one for ANN model training and one for ANN model validation. In the third stage, by using Levenberg-Marquardt algorithm and experimentally obtained data, the ANN mathematical model for the prediction of passenger car fuel consumption is developed. In the final stage, based on the developed model an analysis of fuel consumption by using twelve 3D surface plots showing interactions between average car speed and hour for each day of week and each city zone is performed.

For the purpose of this study vehicle running parameters were collected from vehicle’s engine control unit using on-board diagnostics (OBD-II) scanner via controller area network bus interface. Conveniently, integrated, all-in-one vehicle tracking and monitoring device XT-2000 produced by XirgoTech was used. This device integrates OBD-II interface, global positioning system receiver and cellular network general packet radio service modem. Using OBD-II interface, this device is capable of collecting standard vehicle operating parameters including ignition status, engine RPM, current fuel consumption, vehicle battery voltage and standard fault codes. Each data tuple is tagged with a timestamp and geographic coordinates by the device and such data packet is sent to configured data collection service via packet-data transfer service of a mobile telco operator. The data collection service is a specialized TCP message parsing and storage Windows network service developed in-house specifically for this study. The data collection service uses geospatial database with detailed road network data to map-match each received data tuple to specific road network segment to allow data classification and segmentation. Road network segments data includes road type (motorway, primary, secondary, tertiary, footway etc.), direction (one-way/two-way), number of lanes etc.

Fig. 1 shows schematic view of the experimental setup used for data collection.
The passenger car used in the present study is a second generation (make 2010 with 40000 km traveled) Toyota Yaris, a subcompact car produced by Toyota since 2005. The car is powered by a 1.33 l petrol engine with four cylinders, rated at 74 kW at 4500 rpm with manual six-speed transmission. The vehicle was driven by only one person following the same driving pattern in accordance with traffic conditions.

In order to collect data from a wide-variety of traffic conditions and the various road types, the vehicle was driven in the randomly chosen routes in the city of Niš, Serbia and in different timings. The Niš city area occupies 597 km$^2$ with 391 km long road network [17]. The road structure is made of main, regional and local roads, and for the purpose of the analysis the entire city area was divided into four city road zones: narrow city center (zone 1 - red), broad city center (zone 2 - yellow), inner city zone (zone 3 - blue), wider city zone (zone 4 - green) (Fig. 2).
Over the period of 30 days (from 1 to 31 January, 2015) data were collected from the vehicle using the on-board equipment. Due to different traffic conditions that occur during the day, a typical driving trip in urban areas consists of varying amounts of vehicle idling (at traffic lights, stop signs, crosswalks, etc.), acceleration, deceleration and cruising modes. Also, different city areas have different road conditions with varying amount of traffic and traffic speed, whereby city centers are generally heavily congested.

As fuel consumption on given route considerably differ considering traffic conditions, for the prediction of fuel consumption the following independent (explanatory) variables were selected: day of week, hour of day (h), city zone and average car speed (km/h). Here it should be noted that fuel consumption drastically varies from individual to individual driving style, i.e. aggressive, calm (eco-driving). From a physical point of view, the driving behavior is represented by the speed-time and acceleration-time diagrams. In a more aggressive driving pattern the acceleration rate is higher than in the non-aggressive one and emissions are increased [18]. It has also been reported that there is up to 40% difference in fuel consumption between a calm driver and an aggressive one [19], whereby under an average speed of 70 km/h the driver’s behavior has more opportunity to differ from an ideal fuel efficient behavior, i.e. the driver has a significant influence on car fuel consumption [20].

Since outliers can have deleterious effects on statistical analyses and result in subsequent misleading interpretations of the results, statistical graphical analyses (box and scatter plots) were used to eliminate outliers from the initially prepared set of data containing 250 sets. Table 1 gives the descriptive statistics of the collected data used in ANN model development and validation. Variables day of week and city zone are categorical (nominal) variables, whereas hour of day (h) is represented in discrete domain from 00:00 to 24:00 h with interval of 1 h. Thus in Table 1 only descriptive statistics of the car speed and fuel consumption are given.

**Table 1** Descriptive statistics of the obtained data

<table>
<thead>
<tr>
<th>Variables</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car speed (km/h)</td>
<td>17</td>
<td>60.8</td>
<td>29.5</td>
</tr>
<tr>
<td>Fuel consumption (l/100 km)</td>
<td>5.9</td>
<td>13.3</td>
<td>8.7</td>
</tr>
</tbody>
</table>

For the purpose of ANN training and validation, entire set of obtained data was randomly divided into two sets: (i) training set containing 75% of data, and (ii) testing set containing 25% of data which was used for testing the generalization ability of the ANN model.

### 3. Mathematical Modeling

Among different mathematical modeling techniques, multiple regression analysis and ANNs are two most popular mathematical modeling techniques that are widely used for model development using empirically obtained data. Suitability and effectiveness of using ANNs for the prediction of fuel consumption has been well emphasized in previous researches. On the other hand, although development of multiple regression models follows the well-defined procedure, it requires far less time and effort and eases interpretation of results in comparison to ANN models, as shown by Ahn et al. [1], there may be situations in which negative values of fuel consumption are predicted using regression models. As noted by the authors these inconveniences give rise to find proper data transformations.
In accordance with the previous discussion for the purpose of mathematical modeling of the relationships between independent variables (day of week, hour of day, city zone and car speed) and fuel consumption of the passenger car in the city of Niš, feed-forward single hidden layer ANN was used. This type of ANN process information from $n$ input neurons through $m$ neurons (in hidden layer) to one neuron (in output layer) according to the following relation:

$$ y = f_2 \left[ \sum_n W \cdot f_1 \left( \sum_n V \cdot X + B \right) + b \right], $$

where $V$ and $W$ are matrices of weight connections between input and hidden layer and hidden and output layer, respectively, $B$ is bias matrix of weight connections for hidden neurons, $b$ is bias of output neuron, $X$ is vector of input parameters, $f_1$ and $f_2$ are transfer (activation) functions in hidden and output layers, respectively, and $y$ is ANN output.

The ANN model development is not a trivial task since one must consider a number of issues, including initialization of connection weights, number of hidden neurons, selection of training algorithm, selection of transfer functions etc. As a result of a comprehensive preliminary investigation it has been observed that 4-8-1 ANN architecture with hyperbolic tangent sigmoid and linear activation functions in hidden layer and output layer, respectively, provides acceptable results. These activation functions were used because it was assumed that there exists a nonlinear relationship between independent variables and fuel consumption. In order to increase the convergence speed, the Levenberg-Marquardt algorithm was chosen for the purpose of ANN training. Initialization of weight connections and biases was done according to the Nguyen-Widrow method. Here it should be noted that for the same ANN architecture, in order to overcome the problem of convergence to local minima, different initialization schemes were tested.

After the ANN training process was finished, i.e. near optimal values of weight coefficients and biases of the ANN are determined, the developed ANN model must be tested for generalization capability. In order to estimate the generalization capability of the developed ANN model, the test data which were not used in model development stage were used. The statistical methods of root mean square error (RMSE) and MAPE were used to measure ANN model performance. RMSE and MAPE are calculated by the following equations:

$$ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( t_i - p_i \right)^2}, $$

$$ MAPE = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{|t_i - p_i|}{t_i} \times 100\% \right), $$

where $t$ is the $i$-th target (experimental) value, $p$ is the $i$-th predicted value by the ANN model and the $n$ is the number of sample data.

Table 2 illustrates the performance of the developed ANN model using both training and testing data in terms of RMSE and MAPE. As it can be seen, these values are fairly reasonable, hence the developed model can be used to provide estimation of the fuel consumption for an arbitrary combination set of independent variable values within considered range of independent variable values. It can be argued that MAPE of around 12% is acceptable considering that the model didn’t include other variables which are known to affect the fuel consumption, such as weather conditions for example.
Table 2 Statistical assessment of the developed ANN model

<table>
<thead>
<tr>
<th>Statistical method</th>
<th>Training data</th>
<th>Testing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>1.31</td>
<td>1.42</td>
</tr>
<tr>
<td>MAPE</td>
<td>12.3</td>
<td>13.3</td>
</tr>
</tbody>
</table>

4. RESULTS AND DISCUSSION

Analysis and discussion of results was performed by means of twelve 3D surface plots showing the interactions effects of hour of day and car speed for different days of week and different city zones (Figs. 3, 4 and 5). Because of space restrictions 3D surface plots are given only for Monday (as a representative working day), and two weekend days, Saturday and Sunday.

From Figs. 3, 4 and 5 one can observe that for a given day the fuel consumption decreases from zone 1 to zone 4. This is because of less traffic upon which less crowd is created. However, this effect on fuel consumption is less pronounced during the weekend. This can be explained by the increased intensity of traffic both in the central part of the city and on the access roads to the city located in zones 3 and 4.

![Fig. 3 Car fuel consumption for Monday](image-url)
Generally, for a given city zone and time of day with an increase in the car speed fuel consumption decreases. But for evening hours of weekends one can notice that there is an optimal speed range for a given zone, where fuel consumption is minimal (Figs. 4b,c, 5b). This can be explained by the possible impact of vehicle stops on traffic lights that occurred, as it is clear that in early evening hours one may expect low intensity traffic. It is well known that any change in speed requires acceleration (or deceleration), whereas during acceleration, the fuel to air ratio is higher than optimal, thus ability to maintain constant speed at certain section assures less fuel consumption. Regardless of the city zone considered the fuel consumption is increased in the afternoon and evening hours. Also, one may observe in Figs. 4a, 4b, 5a, 5b that there exists an increase in fuel consumption in city zones 1 and 2 also in the early morning hours during the weekend. This can be explained by the increased traffic intensity in some parts of the city. However as given in Fig. 3c, 3d, if there is a possibility to maintain average speed higher than 35 km/h on weekdays in the afternoon hours savings in the fuel consumption can be achieved.
Based on given 3D surface plots one can notice that fuel consumption less than 8 l/100 km can be achieved in different city zones, but at different hours of day and days in the week. Average fuel consumption by city zones for working and weekend days are given in Table 3. As clearly indicated in Table 3 fuel consumption gradually decreases from zone 1 to zone 4 and this is more pronounced in the weekend days.

**Table 3** Average fuel consumption by city zones

<table>
<thead>
<tr>
<th>Days in the week</th>
<th>Zone 1</th>
<th>Zone 2</th>
<th>Zone 3</th>
<th>Zone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working days</td>
<td>8.42</td>
<td>8.16</td>
<td>7.9</td>
<td>7.78</td>
</tr>
<tr>
<td>Weekend days</td>
<td>8.71</td>
<td>8.44</td>
<td>8.22</td>
<td>8.1</td>
</tr>
</tbody>
</table>

4. **CONCLUSION**

The prediction of fuel consumption and the associated CO₂ emissions and estimation of travel time in urban areas are very important tasks in transportation planning in order to improve traffic flow as well as to conform to environmental concerns aiming to reduce air pollution and achieve energy savings. In the present study, instantaneous ANN based mathematical model for prediction of the passenger car fuel consumption during winter period in the city of Niš was developed. Based on experimentally measured data recorded
through on-board equipment, the proposed mathematical model was aimed to relate car speed, city zone, day of week and hour of day, as independent variables, and passenger car fuel consumption. It has been observed that single hidden layer ANN model trained with Levenberg-Marquardt algorithm has learning capability for fuel consumption estimation at satisfactory level. From the analysis of obtained results it is revealed that the passenger car fuel consumption in the city of Niš is predominantly affected by the time of day, ability to maintain a constant average speed in the given city zone and finally by the day of week. Further, an increase of about 4% in fuel consumption for the weekend days irrespective of the city zone is also observed. It may be concluded that the introduction of preventive measures to reduce crowds and congestions such as the regulation of traffic by the police in most critical periods of the day, for each day singularly, within city zones can contribute to reduction of fuel consumption and associated CO₂ emissions.

The practical implications of the conducted research and developed model are multiple and are reflected in the following. Within the Niš city zone for the given route and day in the week and time, one can predict average travel time based on the average car speed. Also, predicted fuel consumption rates can be used for the estimation of travel costs as well as associated CO₂ emissions.

As is the case with any other empirical models, the developed ANN prediction model is valid for particular passenger car (production year, mileage and driving pattern) and the Niš city area where the data were collected. This does not necessarily restrict the relevance of the results of other similar compact car types which are very common in the present vehicle fleet, however, one can expect only rough prediction estimation. In order to improve robustness of the model as well as prediction accuracy of the developed ANN model, incorporation of additional variables such as weather related factors, acceleration levels and number of stops is future research scope.

Acknowledgement: This paper was realized as a part of the projects “Studying climate change and its influence on the environment: impacts, adaptation and mitigation” (III 43007) financed by the Ministry of Education and Science of the Republic of Serbia within the framework of integrated and interdisciplinary research for the period 2011-2015.

REFERENCES


