Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 83 (2016) 774 – 781
Procedia Computer Science

The 6th International Conference on Sustainable Energy Information Technology (SEIT 2016)

# Fuel Consumption Models Applied to Automobiles Using Real-Time Data: A Comparison of Statistical Models

Ahmet Gürcan Çapraz[a],Pınar Özel[b], Mehmet Şevkli[c]*,Ömer Faruk Beyca[d]

[a,b,c] *Fatih University Industrial Engineering Department, 34500, Buyukcekmece, Istanbul,Turkey*
[d] *Istanbul Technical Üniversity Industrial Engineering Department, Maçka 34367, Istanbul,Turkey*

**Abstract**

Even though the number and variety of fuel consumption models projected in the literature are common, studies on their validation using real-life data is not only limited but also does not fit well with the real-time data. In this paper, three statistical models namely Support Vector Machine (SVM), Artificial Neural Network and Multiple Linear Regression are used in term of prediction of total and instant fuel consumption. The models are compared against data collected in real-time from three different passenger vehicles on three routes by causal drive, using a mobile phone application. Our outcomes reveal that, the results obtained by the models vary depending on the total consumption and instant consumption correlation. Support Vector Machine model of fuel consumption expose comparatively better correlation than the other statistical fuel consumption models.

Peer-review under responsibility of the Conference Program Chairs

*Keywords:* fuel consumption models, support vector machine, artificial neural network, linear regression, machine learning.

## 1. Introduction

Over one third of global energy usage is due to transportation[1, 2], the majority of which is obtained through petroleum products. Consumption of fuel by vehicle engines results in greenhouse gas (GHG) emissions, the most prominent of which is $CO_2$, all of which have detrimental effects on the environment. The second largest source of $CO_2$ emissions is the combustion of gasoline and diesel in vehicles used in transportation[3]. To reduce environmental

* Corresponding author. Tel.: +90 212 866 33 00; fax: +902128663412.
*E-mail address:* msevkli@fatih.edu.tr


Peer-review under responsibility of the Conference Program Chairs
doi:10.1016/j.procs.2016.04.166

externalities, it is imperative that fuel efficiency of vehicles is improved. To this end, whilst renewable fuel sources and alternative fuel vehicles might be a solution in the long-term, reducing emissions and fuel consumption of vehicles operating under existing technologies should be the goal in the short to medium term.

The amount of $CO_2$ emissions from a vehicle is proportional to the amount of fuel consumed by the engine. Fuel efficiency depends on the driving mode, which in turn is dependent on several factors. There are many different ways to estimate emissions using the information on fuel consumption. However, the choice of the nature of emission functions becomes crucial for when accurate estimates are needed in planning for transportation, be it at operational, tactical or strategic levels. The literature is relatively rich in describing models to estimate fuel consumption of vehicles, but numerical validation results are limited, are described only for some of these models, and in isolation. Further empirical studies using real data are needed not only to gauge the quality of the estimations suggested by such models, but also to be able to provide comparison results across the various models.

This is precisely the aim of the present paper. In particular, the main contribution of this study is to test the performance and comparison of three statistical models namely Support Vector Machine (SVM), Artificial Neural Network and Multiple Linear Regression against real-time data that is collected from a passenger car using existing on-board diagnostics (OBD2), a Bluetooth interface and a smartphone. Using the results, we provide a numerical comparison of three statistical models.

The rest of the paper is structured as follows. In Section 2, we provide a brief overview of the statistical models of fuel consumption and relevant literature. In Section 3, experimental set-up and process of data collection is described. Numerical results are presented in Section 4 and conclusions are given in Section 5.

## 2. Statistical models of Fuel consumption

### *2.1. Multiple Linear Regression*

Linear Regression is the most widely used statistical model to construct a linear relationship between the variables. Linear regression models assume that response (dependent) variable is a linear function of model parameter (independent variable). In multiple linear regression there are more than one model parameters in order to predict the response variable as shown in the following equation.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{1}$$

where, $\beta_0$ is the constant term, ($\beta_1, \ldots, \beta_n$) are the dependent variables' ($x_1, \ldots, x_n$) coefficients and y is the response variable. The Multiple Linear Regression was applied to fuel consumption models by N. Fumo and M.A. R. Biswas[4]. Stepwise multiple regression method of greenhouse gas emission modeling in the energy sector in Poland proposed by A.K. Wiecek[5].

In this study the response variable, fuel consumption, is predicted by utilizing Speed, Acceleration, Engine Revolutions per Minute (RPM), Volumetric Efficiency, exhaust gas recirculation (EGR) Commanded and Slope for a specific vehicle. When all the variables are used in analysis of linear regression, the correlation between observed values and predicted values is quite high. Some of the variables (Engine RPM, Volumetric Efficiency and, EGR Commanded) are related to machine and cannot be controlled by the drivers. Hence, Speed, Acceleration and Slope which can be controlled are chosen to be used in analysis for further analysis.

### *2.2. Artificial Neural Network*

Artificial Neural Network (ANN) is a supervised learning technique that mimics the biological neural networks. Complex structure of neural networks allows ANN to capture the non-linear structure of data accurately. In ANN input vectors are propagated through layers of neurons in order to estimate corresponding output as shown in Fig.1.

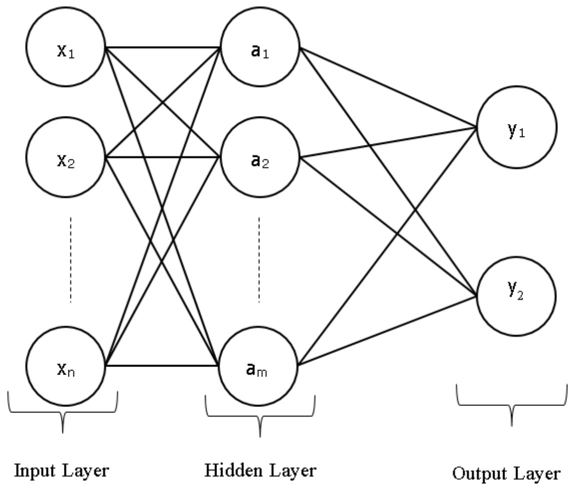


Fig. 1. Representation of Typical 3 Layer Neural Network Architecture

Fig.1 represents the ANN architecture with *n*-dimensional input vector and *m* neurons in the hidden layer. Every element of the input vector has a relationship with every neuron in the hidden layer. The strength of the relationship between input $x_i$ and neuron $a_j$ is determined by weight $w_{ij}^1$. The relationships between neurons in the hidden layer and the output is associated with the weights in the second layer which is denoted by $w_{ij}^2$. The output vector y is predicted by forward propagating the input vector through layers of the network as shown in the following equation for one hidden layer neural network architecture.

$$a = f^1(w^1x) \tag{2}$$

$$y = f^2(w^2a) \tag{3}$$

where x is the vector of inputs and $w^i$ is the weight matrix for the layer *i*, $f^i$ is the activation function of the layer *i*. In this study, sigmoid function is selected as activation function for the first layer and linear function is utilized for the second layer.

In recent years, ANN has been applied various disciplines including energy management, especially forecasting fuel consumption and $CO_2$ emission. ANN approach has been used by K. Ermis et.al[6], in world green energy consumption, in prediction of tractor fuel consumption F.R. Ajdadi and Y.A. Gilandeh[7], in a predictive system for car fuel consumption J.D. Wu and J.C. Liu[8], in prediction of torque and specific fuel consumption of a gasoline engine N.K. Togun and S.Baysec[9].

### 2.3. Support Vector Machine

Support Vector Regression is a version of Support Vector Machine which was put forward in 1996 by Vapnik, V. et. al.[10]. Given a set of input-output pairs $(x_1,x_2\ldots,x_{(n)},)$ $y_1,y_2\ldots,y_n)$ Support Vector Regression (SVR) methods aims to approximate the following function

$$f(x) = wx + b \tag{4}$$

by minimizing the following objective function

$$\frac{1}{2}\,\|w\|^2 + C\,\frac{1}{n}\sum_{i=1}^{n} L(x_i, y_i) \tag{5}$$

where ‖ w ‖ is the regularization term, $L(x_i, y_i)$ is the loss function and C is the trade-off between model complexity and error on training dataset. The graphical representation of SVR can be seen in Figure 2. The advantage of SVR is to present convex solution space resulting in unique solution.

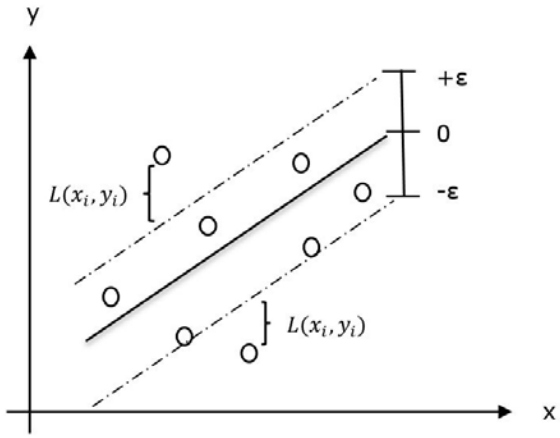


Fig. 2. The graph of Support Vector Regression

The data points are not always in a linear classification; the kernel functions enable us to transform the nonlinear dataset into a linear separation format. Fig. 3 shows the transformation of nonlinear dataset to linear dataset by using kernel functions.

$$y = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*). K(x_i, x) + b \tag{6}$$

Where y is output, $\alpha_i$ and $\alpha^*$ are lagrange multipliers, $x_i$ is input vector, $K(x_i,x)$ is kernel function, and b is bias.

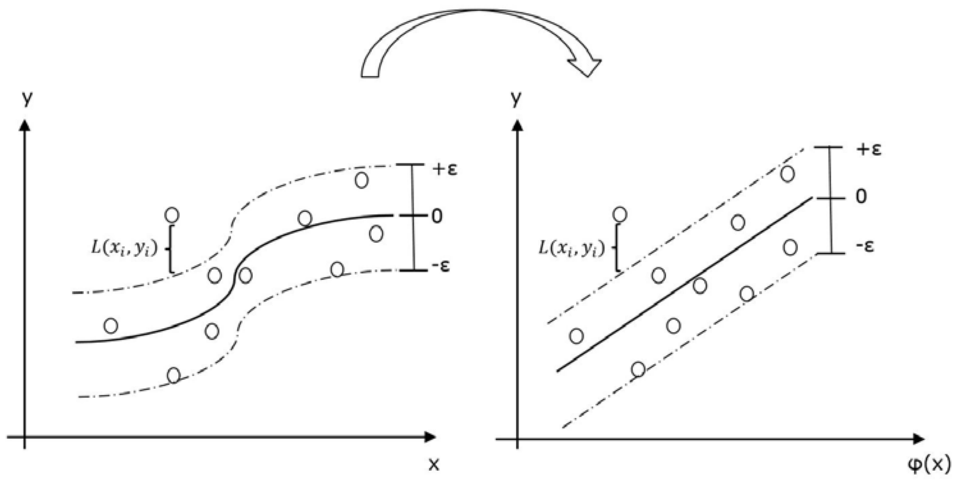


Fig. 3. The transformation of nonlinear dataset to linear dataset by using Kernel functions

## 3. Real time data collection

Numerical experiments are conducted using different vehicles which is shown in Table 1. The data were collected when the drivers were in the normal course.

Table 1 Vehicle properties

| Vehicle Model | Fuel Type | Engine Displacement (ml) | Frontal Surface Area(m2) | Vehicle Weight (kg) | Trip Distance (km) | Average Speed (km/h) |
|---|---|---|---|---|---|---|
| Toyota Corolla | Diesel | 1396 | 2.09 | 1425 | 139,23 | 140 |
| Kia Ceed | Diesel | 1582 | 2.25 | 1800 | 432,42 | 100 |
| BMW X3 | Diesel | 1980 | 2.66 | 2610 | 361,81 | 120 |

In all the experiments, we assume three different vehicles driven on three different routes which are listed in Table 2. First route is from Ankara to Bolu, Turkey, of 139.23 km length. Second route is from Ankara to Kayseri, Turkey, of 432.42 km length. Third route is from Istanbul to Gerede, Turkey, of 361.81 km length.
The latitude and longitude coordinates of the starting and ending points of the test routes are given in Table 2.

Table 2 Routes Coordinates

| | Starting Points | | Ending Points | |
|---|---|---|---|---|
| | longitude | latitude | longitude | latitude |
| Route 1 by Toyota Corolla | 32,60486 | 40,11768 | 31,64131 | 40,76668 |
| Route 2 by Kia Ceed | 32,25257 | 40,60962 | 35,54927 | 38,75007 |
| Route 3 by BMW X3) | 28,734311 | 41,05889 | 32,25309 | 40,60868 |

Data from the vehicle has been collected in real-time using existing on-board diagnostics (OBD2), a Bluetooth interface and a smartphone. An OBD2 protocol allows accessing the vehicle's Engine Control Unit (ECU) easily through a Bluetooth OBD2 connector (see Fig. 4). Instant data are obtained via Torque Pro[11] that uses the OBD2 connector. Many recently manufactured vehicles support the use of OBD2. The diagnostic socket should be fitted to the vehicle using an adapter plug prior to instant data collection.

Torque Pro is a tool for performance and diagnosis, and runs on the Android phones of any kind. The tool keeps the driver informed via several sensors located within the Engine Management System of the vehicle. Through OBD2, it is possible to access data from the ECU and OBD2 itself becomes an extremely useful source of information for screening and solving problems with the vehicle.

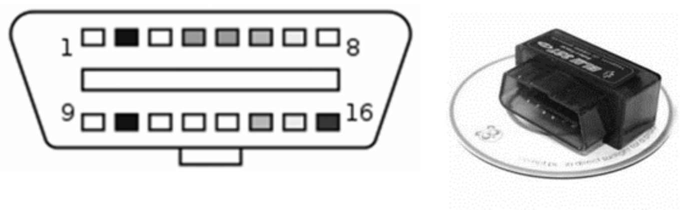


Fig. 4. The OBD2 connector and the Bluetooth adaptor

In order to capture the parameters needed by the fuel consumption models via Torque Pro, a Bluetooth adaptor is plugged into the OBD2 socket of the vehicle, through which parameters such as the vehicle weight and engine displacement are input into the Torque Pro application on the smart phone.

## 4. Results

The collected data is analyzed using Multiple Linear regression, Artificial Neural networks and Support Vector machine models. These models are compared within themselves and each other.

In order to predict the fuel consumptions with regression models. Speed, Acceleration, Engine RPM, Volumetric Efficiency, EGR Commanded and Slope of the road are used as input variables. When all the variables are used in analysis, the correlation between real data and predicted data is quite high. Although all variables are used in the analysis, not all of them are controllable. Speed and acceleration are controllable by driver and the slope is dependent on the route selection. In order to see impact of speed, acceleration and slope, further analysis are conducted.

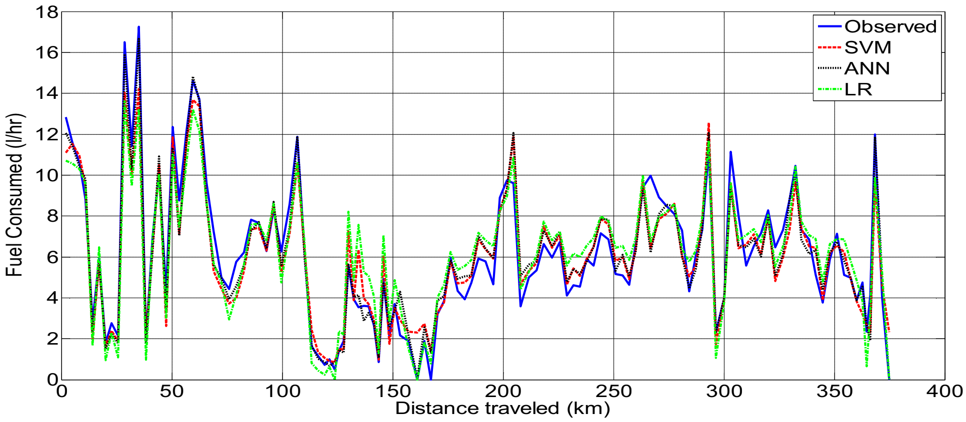


Fig. 5. Fuel consumption rate prediction of the three models

Fig 5 illustrates the fuel consumption rate prediction of the three models comparing with the observed data. X-axis show the distance traveled in km while the y-axis represents the fuel consumption rate in l/hr. Solid line is observed fuel production rate, dashed line is the prediction with SVM, dotted line is prediction with ANN and the dash-dot line is the prediction with linear regression. Fig 5 also reveals that SVM and ANN makes very close predictions to observed data outperforming the linear regression.

Random subsampling is used for testing the performance of statistical models. Data is divided into 80% of training and 20% of testing randomly for each run. The procedure is repeated for 1000 times. Average and standard deviation (in parenthesis) values are given in Table 3 for each statistical model.

In all statistical methods data is first normalized then statistical method are applied. Least square method is used to find parameters for linear regression. In neural network one hidden layer neural network architecture is utilized with 10 neurons in the hidden layer. Sigmoid activation function is used for the hidden layer and linear activation function is used for the output layer. Levenberg–Marquardt algorithm is implemented to find the weights which will minimize the least square error. For support vector regression epsilon-SVR is applied with radial basis kernel function. Both the epsilon value in the loss function and gamma value in radial basis kernel function is specified as 0.1 and the cost parameter C is set as 1.

As seen from the Table 3, SVM provides better result for all automobile kinds comparing ANN and linear regression. SVM outperforms other models for both accuracy and consistency looking at average and standard deviation of R values. Even when we just use speed, acceleration and slope as input values SVM delivers better results than other statistical models.

In the next analysis we will present more generalized results instead of constructing a model for each kind of automobile. More generalized model is created by adding additional input variables such as Frontal Area, Engine Displacements, and Mass of Vehicle. The new model can be applied all kinds of automobiles. Table 4 shows result of the new model for each statistical model.

Table 3 Results of Statistical Models

| | | All Variables | | | Speed, Acceleration and Slope | | |
|---|---|---|---|---|---|---|---|
| Models | | Linear Regression | Neural Network | Support Vector Machine | Linear Regression | Neural Network | Support Vector Machine |
| Toyota-Corolla | Training | 0,89 (0,005) | 0,96 (0,007) | 0,97 (0,001) | 0,85 (0,006) | 0,93 (0,007) | 0,93 (0,002) |
| | Testing | 0,89 (0,019) | 0,95 (0,021) | 0,96 (0,007) | 0,85 (0,022) | 0,91 (0,029) | 0,92 (0,013) |
| Kia-Ceed | Training | 0,78 (0,014) | 0.92 (0,014) | 0,94 (0,002) | 0,75 (0,015) | 0,90 (0,013) | 0,91 (0,004) |
| | Testing | 0,76 (0,056) | 0,89 (0,048) | 0,92 (0,014) | 0,74 (0,057) | 0,88 (0,033) | 0,90 (0,018) |
| BMW-X3 | Training | 0,90 (0,004) | 0,93 (0,010) | 0,95 (0,002) | 0,76 (0,011) | 0,80 (0,019) | 0,80 (0,006) |
| | Testing | 0,90 (0,015) | 0,91 (0,037) | 0,92 (0,011) | 0,75 (0,043) | 0,78 (0,044) | 0,79 (0,026) |

Table 4 Results of a new statistical model for all automobiles

| | All Variables | | | Controllable Variables | | |
|---|---|---|---|---|---|---|
| | Linear Regression | Neural Network | Support Vector Machine | Linear Regression | Neural Network | Support Vector Machine |
| Training | 0,87 (0,004) | 0,95 (0,005) | 0,96 (0,001) | 0,81 (0,006) | 0,90 (0,006) | 0,90 (0,002) |
| Testing | 0,86 (0,016) | 0,94 (0,019) | 0,95 (0,005) | 0,80 (0,022) | 0,88 (0,024) | 0,90 (0,009) |

Table 4 shows that SVM outperforms other methods in both cases where in the first case all variables are used as input and in the second in addition to automobile properties (Frontal Area, Vehicle Weight and Engine Displacement) just the slope, instantaneous acceleration and speed is used. Neural network also provides competitive results however support vector regression gives more robust results especially for testing data which can be seen from standard deviations of R values.

## 5. Conclusion

In this study several methods including statistical methods are exploited in order to obtain the best prediction method for instantaneous fuel consumption. Three statistical methods (LR, ANN, and SVR) are applied for the data that is generated from three kind of automobiles. Amongst the statistical models SVR offer best results. We consider the generalized statistical models, SVR gives 0.90 R value on the testing data with only 0.009 standard deviation after thousands runs.

This research demonstrates that SVR can be applied to prediction of fuel consumption models efficiently and outperforms the other statistical models in terms of R values.

The broader goal of this study is to find the best prediction method to estimate the fuel consumption for long distance in terms of using the advantage of the slope of the road.

## 6. Acknowledgements

This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Project no. 114E421.

## References

1. Ahn K, Rakha H. The effects of route choice decisions on vehicle energy consumption and emissions, Transport Res D-Tr E 2008; 13:151-167.

2. Rakha H, Ahn K, Moran K, Saerens B, & Bulck E. Virginia tech comprehensive power-based fuel consumption model: Model development and testing. Transport Res D-Tr E 2011; 16(7):492–503.

3. Epa.gov. Carbon Dioxide Emissions-Climate Change-US EPA; 2015. <http://www3.epa.gov/climatechange/ghgemissions/gases/co2.html> [accessed 20.01.2016].

4. N. Fumo and M.A.R.Biswas. Regression analysis for prediction of residential energy consumption. Renewable and Sustainable Energy Reviews 47(2015)332–343

5. A.K. Wiecek. Stepwise multiple regression method of greenhouse gas emission modeling in the energy sector in Poland. Journal of Environmental Sciences 30(2015) 47– 54

6. K. Ermis, A. Midillib, I. Dincerc, M.A. Rosenc. Artificial neural network analysis of world green energy use. Energy Policy 35 (2007) 1731–1743

7. F.R. Ajdadi, Y.A. Gilandeh. Artificial Neural Network and stepwise multiple range regression methods for prediction of tractor fuel consumption. Measurement 44 (2011) 2104–2111

8. J.D. Wu, J.C. Liu. A forecasting system for car fuel consumption using a radial basis function neural network.

Expert Systems with Applications 39 (2012) 1883–1888

9. N.K.Togun and S. Baysec. Prediction of torque and specific fuel consumption of a gasoline engine by using artificial neural networks. Applied Energy 87 (2010) 349–355

10.Drucker, Harris; Burges, Christopher J. C.; Kaufman, Linda; Smola, Alexander J.; and Vapnik, Vladimir N. (1997); "Support Vector Regression Machines", in Advances in Neural Information Processing Systems 9, NIPS 1996, 155–161, MIT Press

11.Torque Pro (OBD 2 & Car), (2015), <https://play.google.com/store/apps/details?id=org.prowl.torque&hl=en> [accessed 20.01.2015]