A machine learning approach for the estimation of fuel consumption related to road pavement rolling resistance for large fleets of trucks

F. Perrotta, T. Parry & L.C. Neves
University of Nottingham, Nottingham, United Kingdom

M. Mesgarpour
Microlise Ltd, Nottingham, United Kingdom

ABSTRACT: There remains a level of uncertainty concerning the methodological assumptions and parameters to consider in the estimation of road vehicle fuel consumption due to the condition of road pavements. In fact, recent studies highlighted how existing models can lead to very different results and that because of this, they are not fully ready to be implemented as standard in the life-cycle assessment (LCA) framework. This study presents an innovative approach, based on the application of the Boruta algorithm (BA) and neural networks (NN), for the assessment and calculation of the fuel consumption of a large fleet of truck, which can be used to estimate the use phase emissions of road pavements. The study shows that neural networks are suitable to analyse the large quantities of data, coming from fleet and road asset management databases, effectively and that the developed NN model is able to estimate the impact of rolling resistance-related parameters (pavement roughness and macrotexture) on truck fuel consumption.

1 INTRODUCTION

In the past, several studies showed that road pavement in poor condition can have significant effects on vehicle fuel consumption. These are for example the studies of Beuving et al. (2004), Sandberg et al. (2011), Karlsson et al. (2011), and Chatti & Zaabar (2012) that claimed that roughness and macrotexture can influence rolling resistance and therefore fuel consumption and greenhouse gas (GHG) emissions from road vehicles. This implies that a review of pavement maintenance strategies reduce overall GHG emissions.

However, existing models, that aim to estimate the impact of the road infrastructure on vehicle fuel consumption, give clear differences results (Trupia et al. 2016) and lack accuracy (Perrotta et al. 2018) that means road managers cannot make decisions with confidence.

These models need validation and continuous calibration, for example, HDM-4 (Kerali et al. 2006).

However, calibration and validation of these models can be expensive and time consuming as this requires experiments to be performed and even then may not be representative of real driving conditions.

Modern trucks are fitted with standard sensors (SAE International 2016) that keep track of various parameters (including fuel consumption) and are used by fleet managers to monitor the performance of their vehicles and optimize the operational costs of their fleets The same data, in combination with information from the road management system, can be used to develop accurate fuel consumption models using advanced regression techniques (Perrotta et al. 2017a; Perrotta et al. 2017b).

The paper applies Boruta Algorithm (BA) (Kursa & Rudnicki 2010) for the identification of the most significant variables and, following the procedure explained in Perrotta et al. (2017a), Artificial Neural Networks (NN) for the development of a fuel consumption model for a large fleet of trucks equipped with ~11 litres euro 6 engines. Since this is based on real data from the field it is expected that the developed model will give a high level of accuracy, higher than existing tools, offering a reliable solution to engineers in the field of pavement LCA for estimating the impact of the road surface conditions on vehicle fuel consumption during the use phase of the life cycle of the pavement.

2 DATA

The available dataset comes anonymized from the fleet management database of Microlise Ltd. Microlise is a company based in Nottingham that collects, processes, stores and analyses large quantities of data for fleet managers and manufacturers aiming to optimize the operational costs of their fleets.

By default records are triggered every minute or mile (that is approximately 1.6km). The data considered in this study include records from 473 trucks equipped with ~11 litres euro 6 engines, driving at a constant speed on two segments of motorway in England; part of the M1 and the entire M18. This corresponds to approximately 300km of motorway. 5,423 records are available in total. These include the geographical position (GPS) of the vehicles, the vehicle speed, the torque and revolutions from the engine, an approximation of the vehicle weight, and the fuel consumption measured to the closest 0.001 litre. It is important to highlight the fact that the vehicle weight is an estimate given by an algorithm based on engine parameters. The data from Microlise are enriched by extracting information from the HAPMS (the Highways Agency Pavement Management System) the management system of Highways England, the strategic road agency in UK. In particular, the average gradient, radius of curvature, and measurements of roughness and macrotexture are attached to the data of Microlise based on GPS position. This corresponds to the procedure presented in Perrotta et al. (2017b).

The data contains 56 different measurements characterizing the performance of each single vehicle in the database, the road geometry, and the condition of the road surface. Roughness and macrotexture are measured as Longitudinal Profile Variance (LPV) at three different wavelengths (3, 10 and 30 meters), and Sensor Measured Texture Depth (SMTD), respectively, which are the standard in UK.

3 METHODOLOGY

BA is applied first to select the most significant variables aiming to avoid overfitting. Then, a NN that includes all the variables selected by the BA has been developed. This has been trained using approximately 75% of the data, validated with 20% and tested using the remaining 5%. Ten-fold cross-validation has been performed. Thus, the training, validation and test process has been repeated 10 times.

Similarly to work described in Perrotta et al. (2017a), the test sets have been extracted first by selecting data from trucks that are not to be included in the training or validation sets. This filter has been applied by randomly selecting all the records that come from the same trucks (using an identification number that refers to the single truck) until approximately 5% of data is obtained. This avoids the possibility that the model can learn the behaviour of the single truck and reproduce the performance of the vehicles in the test set.

Therefore, in each iteration of cross-validation about 3,904 records are used for training, 1,301 for validation, and 218 for testing the developed model. However, this can vary slightly due to the initial extraction of the test set.

The structure of the developed NN has been chosen for optimization of the R^2, root mean squared error (RMSE), mean absolute error (MAE), and the calculation time required to perform the ten iterations of cross validation. As the randomized processes of cross-validation comes up with 10 different measures of R^2, RMSE and MAE, the average value of the ten iterations for the test set is used as a measure of performance for the developed model.

As a last test for the model, the predicted fuel consumption for a truck driving at a constant speed of 87km/h (that is the average across the considered fleet of trucks) on a flat and straight road (that is considered to be representative of a motorway) is compared to the average of all measured fuel consumption of the considered fleet of trucks. This is meant to give an idea of the error made by the developed model from a quantitative perspective. The error in this case is calculated as the difference between the estimated value and the real average fuel consumption as a percentage of the latter.

Finally, the estimated volume of fuel spent by the vehicles can be converted to GHG emissions using the emission factors published in EPA (2016).

3.1 *Machine Learning*

3.1.1 *Boruta Algorithm*

In Perrotta et al. (2017a) the authors used random forests (Breiman 2001), a machine learning algorithm based on the theory of decision trees (Breiman et al. 1984), for variable selection. In this study the Boruta Algorithm (BA) (Kursa & Rudnicki 2010) is used instead. BA is actually based on the random forests algorithm (Breiman 2001) but it is optimized for variable selection specifically. It actually represents and improvement to the random forest algorithm as, although based on a similar approach, it adds randomness to the processes by creating shadow features. These are generated by shuffling the available data. Then the BA compares the accuracy of the model that uses the original feature with the model that uses the shadow feature. The BA considers the variable to be significant if the accuracy of the model developed by using the original feature is higher than the model that uses the shadow version of the feature. More detail can be found in Kursa & Rudnicki (2010).

In this study, the BA has been applied to determine the most important variables among the 56 initially available in the considered dataset. The BA returns only the list of variables that help the model in explaining a larger portion of variance, improving its accuracy.

3.1.2 *Artificial Neural Networks*

Neural networks (NNs) (McCulloch & Pitts 1943) is a machine learning algorithm that simulates the way human brains work and that is commonly used to solve complex regression problems. Recently NNs have been applied to a number of fields that include the modelling of fuel consumption of road vehicles (e.g. Ahn 1998; Zeng et al. 2015; Perrotta et al. 2017a).

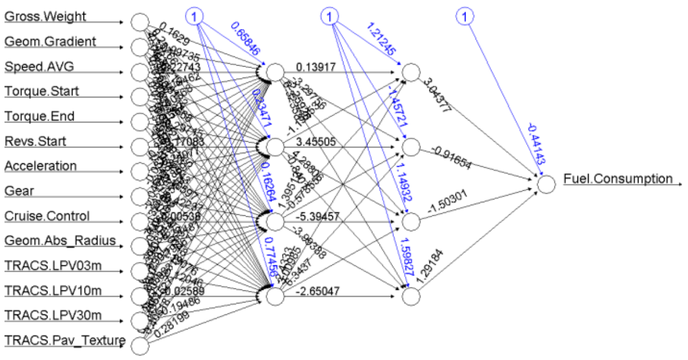

Figure 1. Structure of the developed NN.

Several different types of NNs are available in the literature. In this study, as we analyse data similar to those in Perrotta et al. (2017a), we use the resilient propagation algorithm with backtracking (rprop+) with logistic activation function. The structure of the developed NN being a network with 2 hidden layers with 4 neurons each (Figure 1).

4 RESULTS

Figure 2 shows the results of the BA. The histogram shows how each of the considered variables impact the accuracy of the developed model. Variables that do not show any correlation with the fuel consumption of the considered fleet of vehicles have not been included in the graph.

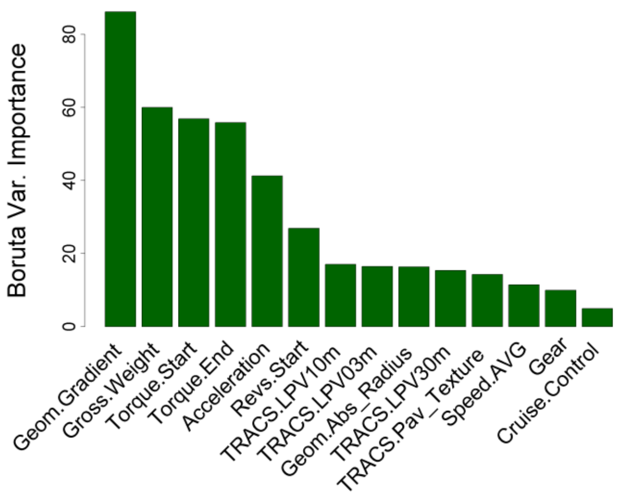

Figure 2. Outcome of the BA. Selected features.

From the figure it is possible to see that the BA considers significant many of the parameters that, from a physical point of view, can be considered important for modelling fuel consumption (Haider & Kriegisch 2013) avoiding overfitting due to incorrect variable selection. These are 14 of the 56 initially available and correspond to those that have been found significant for a different type of truck in Perrotta et al. (2017a). These include, for instance, the road gradient, the vehicle weight, the vehicle speed and performance of the engine (e.g. torque and revolutions).Among the significant variables, the model identifies the LPV03m, LPV10m, LPV30m, and the SMTD.

Table 1 reports the average performance of the model for the test set over the ten iterations of the cross-validation process used to train the NN.

Table 1. Summary of the performance of the developed NN for the test set.

	Average over 10-fold cv
$\mathbf{R^2}$	0.88
RMSE	4.02 l/100km
MAE	2.59 l/100km

Figure 3 shows the fit of the developed model for the training, validation and test sets of one of the ten randomized processes of cross-validation performed on the NN.

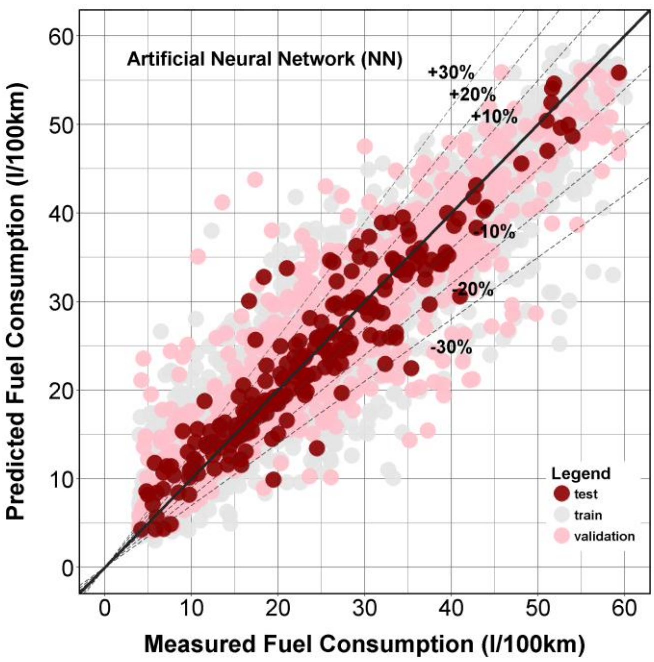

Figure 3. Fit of the developed NN.

Finally, Table 2 shows the results of the test performed on the model where the fuel consumption of a truck driving at 87km/h on a motorway has been estimated and compared to the average measured consumption of the fleet of trucks considered.

Table 2. Outcome of the test performed on the developed NN.

Avg Measured FC	Avg Estimated FC*	Error
24.73 l/100km	24.35 l/100km	-1.5%

* for a truck driving on a flat and straight road at 87km/h

5 CONCLUSIONS

The study shows how the BA (Kursa & Rudnicki 2010) is quickly and effectively able to identify the most significant variables for the considered problem. The algorithm performs well, identifying the impact of variables similar to those identified by previous studies (Perrotta et al. 2017b) that used similar data and a combination of multiple statistics variable selection.

The BA also confirms findings of Perrotta et al. (2017a) that, using similar data but different algorithms, found that different wavelengths of roughness and macrotexture can impact significantly the fuel consumption of articulated trucks.

The study shows how using the methodology set out in Perrotta et al. (2017a), NN can be used for modelling the fuel consumption of the considered fleet of trucks effectively.

Looking at the structure of the developed NN (Figure 1), it is possible to see that it is different from the structure of the NN presented in Perrotta et al. (2017a), for the analysis of a truck fleet with larger engines and higher average fuel consumption. One question still open is regarding if and how the structure and performance of the NN change when data from similar vehicle models from different producers or driving on different roads are considered.

Although results of this study confirm the findings of previous studies in regard to the effect of road roughness and macrotexture on the fuel consumption of the considered fleet of trucks (Sandberg 1990; Chatti & Zaabar 2012; Haider & Kriegisch 2013), further research is still needed.

Exploration of a wider range of road conditions, road materials, effect of weather, effect of acceleration and wider range of vehicle speeds, tire pressure, and driver behaviour may help in obtaining more generally applicable results reducing uncertainties towards the estimation of costs and environmental impact of the use phase of road pavements. In fact, investigating more general conditions may help the model in explaining a wider variance across the data and improve confidence in the obtained results.

Performing a parametric and sensitivity analysis of the models may help in understanding how the NN model addresses the impact of each of the considered variables on fuel consumption and how robust the NN is to errors of measurement or faulty sensors.

Also, at the moment, the current frequency of acquisition of the data, that is every one minute or one mile, may represent a limitation of the study. For example, increasing the frequency of acquisition may open possibilities to extend the analysis to urban areas where the change of route of single vehicles over a minute of travel can be various and introduce bias in the developed model due to incorrect attribution of road characteristics. Further investigation may help in verifying this and address the issue.

All of this will make machine learning, and NN models in particular, a better and reliable decision support tool for engineers and road managers reducing uncertainties in the field of pavement LCA.

AKNOWLEDGEMENTS

This project has received funding from the EU Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 642453.

The authors would like to thank Ian Dickinson from Microlise Ltd for allowing the use of an anonymized sample of their database and Emma Benbow and Helen Viner from TRL for the help in the interpretation of some of the outcomes of the research.

REFERENCES

Ahn, K., 1998. *Microscopic fuel consumption and emission modeling*. Virginia Politechnic Institute and State University.

Beuving, E. et al., 2004. Environmental Impacts and Fuel Efficiency of Road Pavements. , (March).

Breiman, L. et al., 1984. Classification and Regression Trees. *The Wadsworth Statistics/Probability Series*.

Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp.5–32.

Chatti, K. & Zaabar, I., 2012. Estimating the Effects of Pavement Condition on Vehicle Operating Costs. *NCHRP Report 720*.

EPA, 2016. Greenhouse Gas Inventory Guidance - Direct Emissions from Stationary Combustion Sources. *Report of the United States Environmental Protection Agency*.

Haider, M. & Kriegisch, M., 2013. *Modeling Infrastructure Influence on Road Vehicle Energy Consumption*, Deliverable D5.3, Final report.

Karlsson, R. et al., 2011. *Road surface influence on rolling resistance Coastdown measurements for a car and an HGV*,

Kerali, H.R., Odoki, J.B. & Stannard, E.E., 2006. Overview of HDM-4, Volume 1. *The Highway Development and Management Series, International Study of Highway Development and Management (ISOHDM)*.

Kursa, M.B. & Rudnicki, W.R., 2010. Feature Selection with the Boruta Package. *Journal Of Statistical Software*, 36(11), pp.1–13. Available at: http://www.jstatsoft.org/v36/i11/paper.

McCulloch, W.S. & Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), pp.115–133.

Perrotta, F. et al., 2018. Comparison of HDM-4 fuel consumption estimates with real measurements from trucks on motorways: a UK case study. *Transportation Research Board 97th Annual Meeting. January 7-11, Washington, D.C.*

Perrotta, F., Parry, T. & Neves, L.C., 2017a. Application of Machine Learning for Fuel Consumption Modelling of Trucks. *2017 IEEE International Conference on Big Data*.

Perrotta, F., Parry, T. & Neves, L.C., 2017b. Using truck sensors for road pavement performance investigation. In M. Čepin & R. Briš, eds. *Safety and Reliability – Theory and Applications, Proceedings of 27th annual European Safety and Reliability Conference, ESREL 2017*. Portoroz, Slovenia: CRC Press Taylor & Francis Group, pp. 392–396.

SAE International, 2016. SAE J1939-71, Vehicle Application Layer - Surface Vehicle Recommended Practice. *SAE International Standards*. Available at: http://standards.sae.org/j1939/71_201610/.

Sandberg, U., 1990. Road macro-and megatexture influence on fuel consumption. In W. E. Meyer & J. Reichert, eds. *Surface Characteristics of Roadways: International Research and Technologies*. Philadelphia, pp. 460–479.

Sandberg, U. et al., 2011. *Road surface influence on tyre/road rolling resistance*, Available at: http://miriam-co2.net/Publications/MIRIAM_SP1_Road-Surf-Infl_Report 111231.pdf.

Trupia, L. et al., 2016. Rolling resistance contribution to a road pavement life cycle carbon footprint analysis. *International Journal of Life Cycle Assessment*, pp.972–985.

Zeng, W., Miwa, T. & Morikawa, T., 2015. Exploring trip fuel consumption by machine learning from GPS and CAN bus data. *Journal of the Eastern Asia Society for Transportation Studies*, 11(June 2016), pp.906–921.