A big data approach for investigating the performance of road infrastructure

Federico Perrotta[1], Tony Parry[1], Luis C. Neves[1], Mohammad Mesgarpour[2], Emma Benbow[3], Helen Viner[3]
[1]Faculty of Engineering, University of Nottingham, University Park, Nottingham, NG7 2RD, England
[2]Microlise Ltd, Farrington Way, Eastwood, Nottingham, NG16 3AG, England
[3]TRL Ltd, Crowthorne House, Nine Mile Ride, Wokingham, RG40 3GA, England
e-mail: federico.perrotta@nottingham.ac.uk, tony.parry@nottingham.ac.uk, luis.neves@nottingham.ac.uk

ABSTRACT: "Using truck sensors for road pavement performance investigation" is a research project within TRUSS, an innovative training network funded from the EU under the Horizon 2020 programme. The project aims at assessing the impact of the condition of the road pavement unevenness and macrotexture, on the fuel consumption of trucks to reduce uncertainty in the framework of life-cycle assessment of road pavements.
In the past, several studies claimed that a road pavement in poor condition can affect the fuel consumption of road vehicles. However, these conclusions are based just on tests performed on a selection of road segments using a few vehicles and this may not be representative of real conditions. That leaves uncertainty in the topic and it does not allow road mangers to review the current road maintenance strategies that could otherwise help in reducing costs and greenhouse gas emissions from the road transport industry.
The project investigated an alternative approach that considers large quantities of data from standard sensors installed on trucks combined with information in the database of road agencies that includes measurements of the conditions of the road network. In particular, using advanced regression techniques, a fuel consumption model that can take into consideration these effects has been developed.
The paper presents a summary of the findings of the project, it highlights implications for road asset management and the road maintenance strategies and discusses advantages and limitations of the approach used, pointing out possible improvements and future work.

KEY WORDS: Fuel Consumption; Road Performance Evaluation; Big Data Analysis; TRUSS ITN.

1 INTRODUCTION

In the past several studies focused on assessing the impact of road surface unevenness and macrotexture on vehicle fuel economy (Beuving et al., 2004). This could be of particular interest for governmental authorities and road managers since, if a certain amount of fuel is consumed due to the poor condition of the road surface, maintenance would represent an opportunity for road agencies to reduce costs and the emissions of pollutants from road vehicles significantly.

Recent studies, for example, stated that road unevenness and macrotexture can affect up to 5% of fuel consumption (Beuving *et al.*, 2004; Chatti and Zaabar, 2012). That could mean that road maintenance may allow the United States to reduce costs by $400 billion and the United Kingdom by £1 billion (with current fuel prices). This, at global level, implies huge possible reduction of costs and emissions from the road transport industry. An opportunity that cannot be neglected.

However, currently road maintenance policies do not account for the extra costs and environmental impact that the poor state of the road infrastructure can generate (Beuving *et al.*, 2004; Chatti and Zaabar, 2012; EP, 2014). This is mainly due to the fact that different studies show different results and because study data are not considered reliable. In the past, researchers collected data by testing a few vehicles driven at a constant speed or performing coast-down measurements on selected and short road segments. Therefore, what still remains unclear is: are those experimental data really representative of real driving conditions at network level?

Because of this uncertainty road maintenance strategies cannot be currently justified and road managers do not account for the direct and indirect costs that the condition of the road surface can generate for society when making decisions towards road maintenance.

"Using truck sensors for road pavement performance investigation" is a project funded by the European Union under the Horizon 2020 programme within the TRUSS ITN framework (visit www.trussitn.eu for more information). The main aim of the project is to assess the impact of road surface characteristics, such as unevenness and macrotexture, on vehicle fuel economy.

Large quantities of data from trucks driving all across the UK are analyzed in the project and used to model the excess fuel consumption affected by the condition of the road surface. A 'Big Data' approach is undertaken combining advanced statistics and machine learning techniques to point out complex correlations across the data and in particular between fuel consumption of the considered vehicles and road surface properties (Perrotta, Parry and Neves, 2017c, 2017b). This represents the main difference with previous studies and testing the feasibility of this innovative approach represents one of the main contributions of the study.

A new fuel consumption model has been developed based on data that comes directly from vehicles driving across the UK. This allows the model to be continuously updated and representative of real driving conditions. This gives to road managers a new method to test the performance of the road

infrastructure in terms of vehicle fuel economy in almost real time. That will help engineers in justifying a review of the current road maintenance policies and drivers in selecting the most eco-friendly route between two locations making the road transport industry more efficient.

The paper summarizes results of the project, highlighting implications for road asset managers and discussing advantages and disadvantages of the developed method and possible future developments.

2 DATA

Data about the performance of trucks come anonymized from the database of Microlise Ltd. Microlise is a company that offers fleet management services to its clients by collecting, storing and analyzing data from any vehicle type but mostly from trucks. In fact, in accordance to SAE J1939 (SAE International, 2016) modern trucks are equipped with many sensors that constantly monitor the performance of this vehicle type.

The data include characteristics of the vehicle itself (e.g. type of truck, number of wheels, manufacturer, etc.), vehicle speed, engine parameters (e.g. used torque and revolutions), GPS location and fuel consumption to the nearest 0.001 liter among other parameters and is collected every minute or mile (that is ~1609 m). By analyzing these data, Microlise helps its clients (e.g. big chains of supermarkets, multinational delivery companies, truck manufacturers, etc.) in optimizing the operational costs of their vehicle fleets. That can be in regards to fuel consumption, maintenance of vehicles, training of drivers, optimization of delivery time, etc.

On the other hand, road agencies periodically collect data to analyze the performance of their infrastructure and to inform their decisions towards maintenance of road pavements. These data include road geometry (e.g. gradient, crossfall, radius of curvature, etc.), materials used to build (or rehabilitate) road pavements, measurements of road unevenness and macrotexture, skid-resistance, etc.

These data combined represent an opportunity to model fuel consumption with the possibility of modelling the excess of fuel spent because of the conditions of the road surface. This in the past has been done using experimental data and fuel consumption has been modelled using a physical/mechanistic approach (e.g. Emma Benbow, Brittain, & Viner, 2013; Chatti & Zaabar, 2012). A new challenge for future work would be to model fuel consumption and the excess of fuel spent based on data from vehicles driving under real driving conditions and in situations representative of the whole network.

Previous studies on the topic typically used measurements of IRI (International Roughness Index) for road unevenness and MPD (Mean Profile Depth) for macrotexture. However, in UK, road unevenness is measured as Longitudinal Profile Variance (LPV) at 3, 10 and 30 meters wavelength and macrotexture is measured as Sensor Measured Texture Depth (SMTD). Benbow, Nesnas, & Wright (2006) and Viner et al. (2006) established that these unevenness and texture parameters are closely related.

Measurements are taken at high frequency by a diagnostic vehicle thanks to the lasers and standard sensors installed on it (Highways Agency, 2008). The data is stored in the Highways Agency Pavement Management System (HAPMS) and can be accessed through authorization of Highways England, the strategic road agency in England.

3 METHODOLOGY

This section describes the techniques and methods used in the project to analyse data and construct the fuel consumption model. First this section introduces the methodology applied to clean the data and enrich information from the truck fleet management system with data from the HAPMS.

Then, it focuses on variable selection and on the techniques used for assessing the significance of correlation between road surface characteristics and truck fleet fuel consumption. Finally, techniques used to build the developed model and make estimates of the effect of the road surface conditions on the amount of fuel spent by the considered fleet of trucks are described.

R ver. 3.4.1 (CRAN, 2017) is the main software used for analyzing the data, with 'glmnet' (Friedman et al., 2017), 'caret' (Kuhn, 2017), 'e1071' (Meyer et al., 2017), 'randomForest' (Liaw et al., 2017), and 'neuralnet' (Fritsch, et al., 2016) as main packages used within the project.

3.1 Data mining

As this project represents an initial study which aims at testing the feasibility of a 'Big Data' approach for assessing the impact of road surface characteristics on vehicle fuel consumption, a few assumptions have been made in order to simplify the data analysis and reduce the effect of nonlinearity on the fuel consumption of the considered fleet of trucks. For this reason, only data from heavy trucks driving at constant speed (+/- 2.5 km/h) and collected on motorways (at relatively high speed, ~70-100 km/h) have been analysed. These filters reduce the amount of data available from more than a million rows to a few thousands, which still represent a significant amount of data rarely examined in the literature before.

3.2 Variable selection

In regression analysis one of the most delicate and controversial phases for modelling is the selection of predictors. In fact, identifying causation and distinguishing that from correlation is not an easy task and may require time and multiple tests. For this reason and because linear regression was initially used to model fuel consumption in a first phase of the project, statistics like the Pearson's correlation coefficient, the Akaike Information Criterion (AIC) (Akaike, 1974), and analysis of p-values and the adjusted-R^2 were used to select the best predictors that increase accuracy and reliability of the developed model.

Although these statistics can be used when dealing with linear regression, recent studies showed that these do not always work well, especially when the data comes in large quantities, from different sources and may hide highly non-linear correlations (Lew, 2013; Baker, 2016).

For this reason, alternative statistics and mathematical methods have been used and tested in combination. in order to obtain more accurate and reliable results.

In particular, methods such as:

- the Lasso (least absolute shrinkage selection operator) (Tibshirani, 1996), a method similar to linear regression that performs variable selection through regularization

with the aim of enhancing interpretability and accuracy of the developed models (James, et al. 2013);
- Principal Components Analysis (Pearson, 1901; Hotelling, 1933), a statistical method based on orthogonal transformation commonly used in regression analysis for dimensionality reduction and feature selection (Song, et al. 2010),
- Random Forests (Breiman, 2001), a machine learning method based on the theory of decisions trees (Breiman, et al. 1984) commonly used to solve complex classification or regression problems and that is able to perform variable selection;
- and, Boruta Algorithm (Kursa and Rudnicki, 2010), an evolution of the random forests method specifically designed for variable selection.

have been used to identify the predictors that give the highest correlation with fuel consumption, that increase accuracy and reliability of the developed model, avoiding overfitting.

3.3 Modelling

In the first phase of the study various attempts at fitting a linear regression to the data have been performed (Perrotta et al., 2017; Perrotta, Parry and Neves, 2017c, 2017a) with variables that have been selected by analyzing the Pearson's correlation coefficients, p-values, adjusted-R^2, AIC and Lasso. Selection of the best predictors to use in the developed models results from an analysis and comparison of all of these statistics listed and causation identified based on the findings of experimental studies (e.g. Emma Benbow et al., 2013; Chatti & Zaabar, 2012).

In the second phase of the project, due to the high quantity and variety of data available application of machine learning has been tested for modelling the fuel consumption of the considered fleet of trucks and performance compared to the linear regression model developed in the first phase. In fact, using machine learning algorithms allows non-linear correlations, avoiding overfitting and issues related to the multicollinearity of predictors. That allows for example, to consider the effect of different wavelengths of unevenness on fuel consumption that may vary in different situations (e.g. different vehicle speed).

For instance, in this second phase of the project, three different models have been developed using different machine learning techniques (Perrotta, Parry and Neves, 2017b, 2018). A Support Vector Regression (SVR) (Gunn, 1998), a Random Forest (RF) (Breiman, 2001) and a back-propagation Artificial Neural Network (ANN) (Goh, 1995) models have been developed. The methods have been chosen as it was found in the literature that these have already been used for modelling the fuel consumption of road vehicles successfully (e.g. Laxhammer & Gasc´on-Vallbona, 2015; Xu & Zhao, 2010; Zeng et al., 2015). However for different tasks it is not possible to know a priori which method works best and performance may vary based on the accuracy and quantity of the data available. For this reason, all three methods have been tested and performance compared in terms of root-mean squared error (RMSE), mean absolute error (MAE), R^2 and calculation time.

Due to the way machine learning algorithms work they need to be trained and fed with data before they can give accurate and reliable estimates (James *et al.*, 2013; Ng, 2018). Usually the higher the amount of data fed for training and the higher the performance of the developed machine learning model (Ng, 2018). However, sometimes performance of the model can depend on how the data is split and generate overfitting (James *et al.*, 2013; Ng, 2018). Crossvalidation (James *et al.*, 2013) is used in order to avoid that.

In the project ten-fold crossvalidation was used in order to develop the SVR, RF and ANN models. This consists in repeating the training process ten times with different subsets of data in order to test the variability of results for different training sets. Constant performance of the model among the ten trials prove that results do not depend on how the data is split and that bias is spread homogenously across the data.

Initially, a 4 ± 2.5 % set is extracted from the data. That is used to test the performance of the developed models. This is chosen as containing records from trucks that are not included in the rest of the data (96 ± 2.5 %). Similarly, 25% of the remaining data is used for validation and 75% for training. The fact that training, validation and test sets do not contain data from the same trucks allows to test the ability of the developed models to generalize patterns and correlations through the data, avoiding overfitting.

One of the major criticisms in Engineering in regards to the use of machine learning models such as SVR, RF and ANN is the fact that these work as black-boxes and have low interpretability in comparison to linear regression or physical based models. For this reason, a parametric analysis has been conducted in order to 'open' the black box and see how SVR, RF, and ANN approximate the correlation between predictors and fuel consumption. This consists of using SVR, RF and ANN to predict the fuel consumption of trucks for fifty different values of each of the considered variables. The values are chosen to be evenly distributed within the 5^{th} and 95^{th} percentiles of the distribution of the considered predictor. While the value of one predictor changes all others are set to their average.

4 RESULTS

In the first phase of the study, a multiple linear regression model has been developed using data from different truck fleet systems. This was to try to understand the type of information available and explore opportunities within the data.

For example, in one of these initial studies 1420 records from 260 trucks, driving at constant speed on the M18, a motorway in UK, have been analyzed. The data were used to develop a linear model of fuel consumption with measurements of unevenness and macrotexture among the predictors.

In particular, the model presented in (Perrotta, Parry and Neves, 2017c) has equation:

$$FC = 62.42 + 0.00024\,GVW + 14.84\,g\% - 0.57\,s + 0.26LPV10 + 0.87SMTD \quad (1)$$

where FC is fuel consumption (in l/100km), GVW the gross vehicle weight (in kg), $g\%$ the road gradient (in %), s the vehicle speed (in km/h), $LPV10$ unevenness as longitudinal profile variance at 10 m wavelength (in mm^2), and $SMTD$

macrotexture sensor measured texture depth (in mm), see Figure 1.

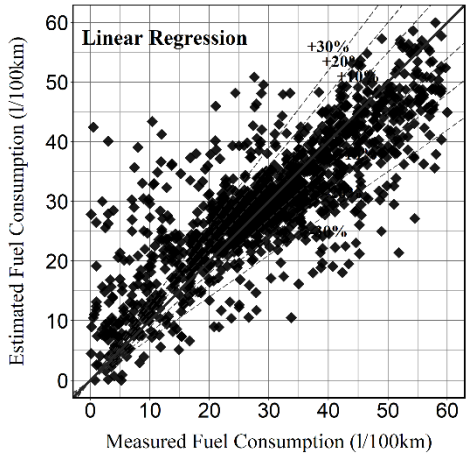

Figure 1. Fit of the linear model. Adapted from (Perrotta, Parry and Neves, 2017c).

Performance of the model in terms of RMSE, MAE and R^2 are summarized in Table 1:

Table 1. Summary of performance of the model presented in (Perrotta, Parry and Neves, 2017c).

RMSE	MAE	R^2
7.80	5.55	0.68

For the considered fleet of trucks the model estimates that 3% of fuel consumption is affected by LPV10 (unevenness) and 5% by the SMTD (macrotexture) of the road surface. As predictors, the model contains only 5 of the 45 variables available. These have been chosen based on the AIC and adjusted-R^2 statistics. Similar results have been obtained previously with other samples of data (Perrotta *et al.*, 2017; Perrotta, Parry and Neves, 2017a).

In the second phase of the study, application of machine learning to fuel consumption modelling of the considered fleet of trucks have been performed.

In this study (Perrotta, Parry and Neves, 2017b), 14,281 records from 1110 articulated trucks driving along the whole M18 and part of the M1 (probably the most important motorway in UK and part of the strategic road network in England) for a total length of ~300 km of road were investigated.

The study investigated the capabilities of three machine learning techniques and their performance compared to those of the linear regression performed on the same data. In particular, the developed machine learning models are a SVM, a RF and an ANN (Perrotta, Parry and Neves, 2017b). The models contain 14 out of 56 variables initially available that include the gross vehicle weight, the road gradient, radius of curvature of the road, the vehicle speed, average acceleration, engine parameters such as revolutions and torque (at start and end of the record), gear used, activation of cruise control (0/1), measurements of road unevenness at three, ten and thirty metres wavelength (i.e. LPV03, LPV10 and LPV30) and measurement of the macrotexture of the road pavement (SMTD). These have been identified to significantly impact the fuel consumption of the considered fleet of vehicles (Perrotta, Parry and Neves, 2017b) by using the random forest algorithm for variable selection (Breiman, 2001).

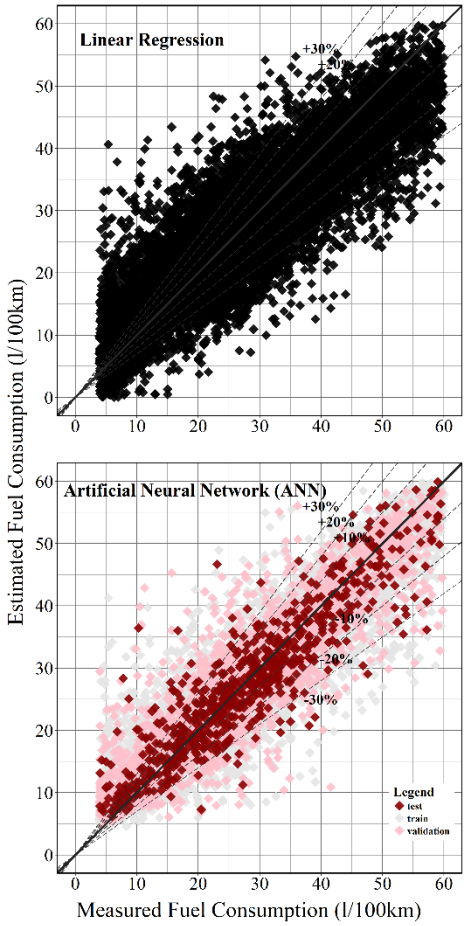

Figure 2. Fit of the ANN model published in (Perrotta, Parry and Neves, 2017b) and comparison with a linear regression model developed using the same data.

Figure 2 reports the fit of the developed ANN model and compares it to the fit of linear regression for the same data. Ten-fold crossvalidation is performed on the machine learning models for training but only one of the models is shown in the Figure. Due to the fact that machine learning algorithms need training before they can be used for making any estimate on new situations three colors are used in order to distinguish estimations made by the developed model on the training, validation and test sets.

Performance of the models in terms of RMSE, MAE and R^2 (for test sets and averaged on the ten crossvalidation processes for the ANN) are summarized in Table 2.

The authors show also that the RF model gives the highest performance in terms of RMSE, MAE and R^2, slightly better than the ANN (i.e. $R^2 = 0.87$). However, this is also the model that requires the highest crossvalidation time for training (double that required by the ANN).

Table 2. Summary of performance of the ANN model presented in (Perrotta, Parry and Neves, 2017b) and comparison with performance of a linear regression model built on the same data.

Model	RMSE	MAE	R^2
Linear reg.	6.02	4.42	0.76
ANN	4.88	3.46	0.85

Finally, estimation of the effect of road surface conditions on fuel consumption can be computed by performing a parametric analysis. Figure 3 reports some examples of parametric analysis performed on the linear regression and ANN model developed.

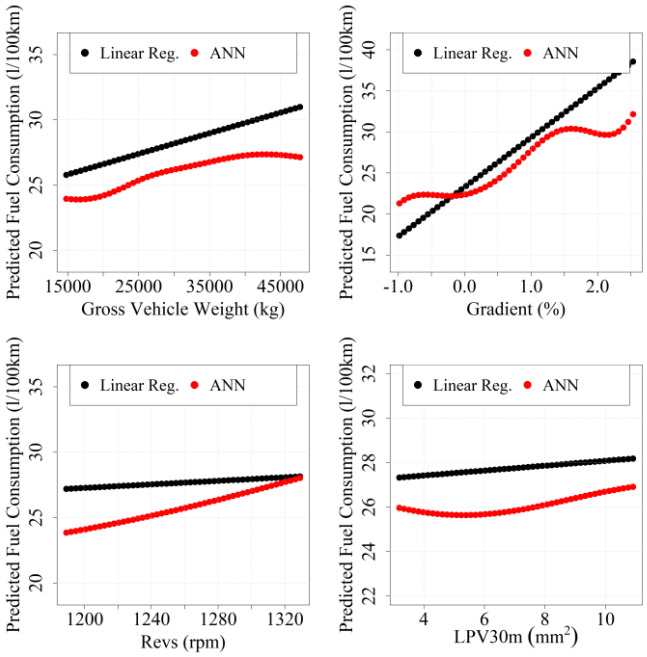

Figure 3. Examples of a parametric analysis and comparison of estimates between the linear regression and machine learning models developed in (Perrotta, Parry and Neves, 2017b).

This allows to compute the effect of each of the considered variables on vehicle fuel consumption for the fleet of trucks analyzed and also makes the developed SVR, RF and ANN partially interpretable.

5 CONCLUSIONS

Results of the project show great potential for the 'Big Data' approach to be used for modelling the fuel consumption of road vehicles including the possibility of estimating the influence that road surface characteristics have on this.

Initial results showed that a linear regression of the variables is able to predict the fuel consumption of the considered fleet of trucks and that this is able to estimate the impact of each of the single variables used as predictors on fuel consumption. This allows to estimate the impact of road surface properties on vehicle fuel economy that has been assessed at approximately 3% - 4% for road unevenness (LPV10) and at 1.25% - 5% for macrotexture (SMTD) (Perrotta et al., 2017; Perrotta, Parry and Neves, 2017a). This substantially confirms results reported by experimental studies (Sandberg, 1990; Beuving et al., 2004; Chatti and Zaabar, 2012) and gives more confidence in the used approach.

However, the fact that linear regression is not reliable for very large quantities of data, it does not allow to identify and estimate non-linear effects and it cannot consider non-numerical or non-continuous variables, represent limitations that can be overcome by using machine learning. This is why in the second part of the project performance of SVR, RF and ANN have been investigated.

Results show that machine learning is able to outperform linear regression, in terms of RMSE, MAE and R^2 and that inclusion of more data and types of measurements could allow the developed models to further improve in precision, accuracy and reliability of estimates.

One issue presented by machine learning algorithms is the time required for training that increase with 1) the number of repetitions used for crossvalidating the model, 2) quantity of data analyzed and 3) complexity of the structure of the model used. However, this remains lower than the time required to organize, perform and analyze data from experiments that was the approach used in the past (e.g. Chatti & Zaabar, 2012).

Exploration of a wider range of vehicle types and investigation of a more extensive road network could help in obtaining results representative of real driving conditions in UK. That will improve applicability of the study and may help engineers in justifying a review of the current road design and maintenance strategies that may help highway authorities in reducing the emissions of pollutants from the road transport industry and save significant costs.

Also, performing a sensitivity analysis (Cortez and Embrechts, 2013) could allow to improve interpretability of the results helping in further testing the reliability of the obtained estimates.

In future, when data for different vehicle types will be available and accessible, this approach could help engineers in estimating the excess costs and environmental impact that the conditions of the road surface have on vehicle fuel economy and inform their decisions towards road design and maintenance of the road surface. That would be important to consider also for electric vehicles since batteries could last longer on smoother road pavements and possibly increase the distance that a vehicle can travel with a single charge.

ACKNOWLEDGEMENTS

This project has received funding from the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 642453 and it is part of the Training in Reducing Uncertainty in Structural Safety project (TRUSS ITN, www.trussitn.eu).

REFERENCES

Akaike, H. (1974) 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control*, 19(6), pp. 716–723. doi: 10.1109/TAC.1974.1100705.

Baker, M. (2016) *Statisticians issue warning over misuse of P values*, *Nature*. Macmillan Ltd. doi: 10.1038/nature.2016.19503.

Benbow, E., Brittain, S. and Viner, H. (2013) *Potential for NRAs to provide energy reducing road infrastructure*. MIRAVEC - Deliverable D3.1.
Benbow, E., Nesnas, K. and Wright, A. (2006) *PPR131 - Shape (surface form) of Local Roads*. Crowthorne (London).
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*. doi: 10.1371/journal.pone.0015807.
Breiman, L. (2001) 'Random forests', *Machine Learning*, 45(1), pp. 5–32. doi: 10.1023/A:1010933404324.
Chatti, K. and Zaabar, I. (2012) 'Estimating the Effects of Pavement Condition on Vehicle Operating Costs', *NCHRP Report 720*. Washington, D.C.: Transport Research Board.
Cortez, P. and Embrechts, M. J. (2013) 'Using Sensitivity Analysis and Visualization Techniques to Open Black Box Data Mining Models', *Information Sciences, Elsevier*, (225), pp. 1–17. doi: http://dx.doi.org/10.1016/j.ins.2012.10.039.
CRAN (2017) *R ver. 3.4.1 'Single Candle', the Comprhensive R Archive Network*. Available at: https://cran.r-project.org/bin/windows/base/old/3.4.1/ (Accessed: 24 April 2017).
EP (2014) 'EU road surfaces: Economic and safety impact of the lack of regular road maintenance', *Policy Department for Structural and Cohesion Policies, European Parliament*.
Friedman, J. *et al.* (2017) 'R Package: glmnet, Ver. 2.0-13'.
Fritsch, S. *et al.* (2016) 'R Package: neuralnet, Ver. 1.33'.
Goh, A. T. C. (1995) 'Back-propagation neural networks for modeling complex systems', *Artificial Intelligence in Engineering*, 9(3), pp. 143–151. doi: 10.1016/0954-1810(94)00011-S.
Gunn, S. R. (1998) *Support Vector Machines for Classification and Regression, Faculty of Engineering, Science and Mathematics, School of Electronics and Computer Science*. Southampton. doi: 10.1039/B918972F.
Hotelling, H. (1933) 'Analysis of a complex of statistical variables into principal components', *Journal of Educational Psychology*, 24(6), pp. 417–441. doi: 10.1037/h0071325.
James, G. *et al.* (2013) *An Introduction to Statistical Learning*, *Springer Texts in Statistics*. New York, NY: Springer Science+Business Media New York (Springer Texts in Statistics). doi: 10.1007/978-1-4614-7138-7.
Kuhn, M. (2017) *R Package: caret - Classification and Regression Training, Ver. 6.0-76, https://Cran.R-Project.Org/Package=Caret*. Available at: https://github.com/topepo/caret/.
Kursa, M. B. and Rudnicki, W. R. (2010) 'Feature Selection with the Boruta Package', *Journal Of Statistical Software*, 36(11), pp. 1–13.
Laxhammer, R. and Gasc´on-Vallbona, A. (2015) 'D4.3. Vehicle models for fuel consumption', *Seventh Framework Programme - COMPANION*, (610990), pp. 1–13. doi: 610990.
Lew, M. J. (2013) 'To P or not to P: on the evidential nature of P-values and their place in scientific inference', (December 2012). Available at: http://arxiv.org/abs/1311.0081.
Liaw, A. *et al.* (2017) 'R Package: randomForest, Ver. 4.6-12'. doi: 10.5244/C.22.54.
Meyer, D. *et al.* (2017) *R Package: e1071, Ver. 1.6-8*. Available at: https://cran.r-project.org/web/packages/e1071/e1071.pdf.
Ng, A. (2018) *Machine Learning Yearning - Technical Strategy for AI Engineers in the Era of Deep Learning*. Stamford, CA, USA: Draft.
Pearson, K. (1901) 'On Lines and Planes of Closest Fit to Systems of Points in Space', *Philosophical Magazine*, 2(11), pp. 559–572. doi: 10.1080/14786440109462720.
Perrotta, F. *et al.* (2017) 'Route level analysis of road pavement surface condition and truck fleet fuel consumption', in Al-Qadi L., I., Ozer, H., and Harvey, J. (eds) *Pavement Life-Cycle Assessment*. Champaign, Illinois: CRC Press, pp. 51–57. doi: doi:10.1201/9781315159324-7.
Perrotta, F., Parry, T. and Neves, L. C. (2017a) 'A big data approach to assess the influence of road pavement condition on truck fleet fuel consumption', in Dell'Acqua, G. and Wegman, F. (eds) *Transport, Infrastructure and Systems: Proceedings of the AIIT International Congress on Transport, Infrastructure and Systems*. Rome, Italy: CRC Press, pp. 33–38. doi: https://doi.org/10.1201/9781315281896-7.
Perrotta, F., Parry, T. and Neves, L. C. (2017b) 'Application of Machine Learning for Fuel Consumption Modelling of Trucks', *2017 IEEE International Conference on Big Data*.
Perrotta, F., Parry, T. and Neves, L. C. (2017c) 'Using truck sensors for road pavement performance investigation', in Čepin, M. and Briš, R. (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. doi: 10.1201/9781315210469-343.
Perrotta, F., Parry, T. and Neves, L. C. (2018) 'Evaluation of road pavements fuel efficiency using truck sensors data', in *TRAVISIONS Young Researchers Competition 2018, Transportation Research Arena (TRA) 2018*. Vienna (16-19 April, 2018).
SAE International (2016) 'SAE J1939-71, Vehicle Application Layer - Surface Vehicle Recommended Practice', *SAE International Standards*. SAE International. Available at: http://standards.sae.org/j1939/71_201610/.
Sandberg, U. (1990) 'Road macro-and megatexture influence on fuel consumption', in Meyer, W. E. and Reichert, J. (eds) *Surface Characteristics of Roadways: International Research and Technologies*. Philadelphia, pp. 460–479. doi: 10.1520/STP23382S.
Song, F., Guo, Z. and Mei, D. (2010) 'Feature Selection Using Principal Component Analysis', *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization*, pp. 27–30. doi: 10.1109/ICSEM.2010.14.
The Highways Agency (2008) 'HD 30/08 Maintenance Assessment Procedure', *Design Manual for Roads and Bridges, Volume 7, Section 3*.
Tibshirani, R. (1996) 'Regression Shrinkage and Selection via the Lasso', *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), pp. 267–288. doi: http://dx.doi.org/10.1111/j.1467-9868.2011.00771.x.
Viner, H. *et al.* (2006) *PPR148 - Surface Texture Measurement on Local Roads*. Crowthorne (London).
Xu, X. and Zhao, Y. (2010) 'Prediction of fuel consumption per 100km for automobile engine based on Gaussian processes machine learning', in *International Conference on Mechanical Engineering and Green Manufacturing 2010, MEGM 2010*, pp. 1951–1955. doi: 10.4028/www.scientific.net/AMM.34-35.1951.
Zeng, W., Miwa, T. and 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.