

University of Nottingham
UK | CHINA | MALAYSIA

Exploring the impact of road surface conditions on truck fleet fuel consumption through Big Data

Federico Perrotta

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

June 2019

Table of contents

Table of contents ..... **i**

Abstract ..... **vii**

Acknowledgements ..... **x**

List of abbreviations ..... **xi**

List of publications ..... **xv**

List of figures ..... **xvii**

List of tables ..... **xxx**

1 Introduction **1**

1.1 Background *4*

1.2 Aims and objectives *6*

1.3 Scope of the study *7*

1.4 Structure of the thesis *8*

2 Literature Review **10**

2.1 Road surface characteristics *11*

2.1.1 IRI vs LPV 12

2.1.2 MPD vs SMTD 13

2.2 Pavement LCA 15

2.3 Vehicle fuel consumption 20

2.4 Principles of Vehicle Dynamics 23

2.5 Factors influencing truck fuel consumption 26

2.5.1 Vehicle type 26

2.5.2 Vehicle weight and payload 27

2.5.3 Engine 31

2.5.4 Fuel type 31

2.5.5 Aerodynamics 33

2.5.6 Driving context 33

2.5.7 Vehicle speed 34

2.5.8 Driver behaviour 35

2.5.9 Road geometry 36

2.5.10 Weather conditions 38

2.5.11 Other factors 39

2.6 The tire-pavement interaction 40

2.7 RR, pavement characteristics and fuel consumption 42

2.7.1 Pavement stiffness 44

2.7.2 Pavement type 46

2.7.3 Surface conditions 47

2.8 Big Data 69

2.9 Summary and Work Plan 73

3 Data 77

3.1 Truck fleet management system 77

3.1.1 SAE J1939 78

3.1.2 Telematic data 79

3.1.3 DVLA data 82

3.2 Road condition data 83

3.2.1 The UK road infrastructure 83

3.2.2 HAPMS and TRACS data 85

3.3 Meteorological data 88

3.4 Data collected 90

3.4.1 Initial dataset (Dataset 1) 92

3.4.2 Final dataset (Dataset 2) 96

4 Methodology 105

4.1 Description of case study 105

4.2 Data cleaning 106

4.2.1 Geographic data cleaning 106

4.2.2 Completion of a dataset 107

4.3 Variable selection 114

4.3.1 Pearson's correlation coefficient 115

4.3.2 Regression analysis 116

4.3.3 Lasso 120

4.3.4 Random Forests: for Variable Selection 121

4.4 Linear Regression (LR) 122

4.5 Machine Learning (ML) 124

4.5.1 Support Vector Regression (SVR) 128

4.5.2 Random Forests (RF): for Regression 130

4.5.3 Artificial Neural Networks (ANN) 132

4.6 Measurement of performance 136

4.7 Learning curves 136

4.8 Grid-search 139

4.9 Cross-validation 143

4.10 Validation *144*

4.11 Parametric analysis *144*

5 Results and Discussion **147**

5.1 Variable selection *147*

5.2 Learning curves *154*

5.3 Modelling *156*

5.4 Paramteric analysis *162*

5.5 Summary *174*

6 Conclusions and Recommendations **185**

6.1 Conclusions *185*

6.2 Limitations *191*

6.3 Recommendations *193*

References **196**

Appendix A Hardware and software **1**

Appendix A.1 Computer used *1*

Appendix A.2 R & RStudio *2*

Appendix A.2.1 R 2

Appendix A.2.2 RStudio 3

Appendix A.2.3 R functions & packages used 4

Appendix A.3 Microsoft SQL Server *6*

Appendix A.4 ArcGIS *7*

Appendix A.5 Tableau *7*

Appendix B Selection of case study **1**

Appendix C Data distributions **1**

Appendix D R code **1**

Appendix D.1 Data procurement *1*

Appendix D.1.1 Truck telematic data 1

Appendix D.1.2 Road conditions data 5

Appendix D.1.3 Creation of Data.T 8

Appendix D.1.4 Append road conditions to Data.T 14

Appendix D.1.5 Append weather conditions to Data.T 26

Appendix D.1.6 Complete Data.T 28

Appendix D.2 Variable selection *30*

Appendix D.2.1 p-values 30

Appendix D.2.2 Adjusted-R^2 30

Appendix D.2.3 Lasso 31

Appendix D.2.4 Random Forest 32

Appendix D.3 Cross-validation *33*

Appendix D.3.1 Data split 33

Appendix D.3.2 10-fold cross-validation 34

Appendix D.4 Modelling *37*

Appendix D.4.1 LR 37

Appendix D.4.2 SVR 37

Appendix D.4.3 RF 38

Appendix D.4.4 ANN 39

Appendix D.5 Parametric analysis *41*

Appendix E Results **1**

Appendix E.1 Analysis of the Dataset 1 ..... *1*

Appendix E.1.1 Variable Selection 1 2

Appendix E.1.2 Modelling 1 8

Appendix E.1.3 Parametric Analysis 1 21

Appendix E.2 Analysis of the Dataset 2 ..... *34*

Appendix E.2.1 Variable Selection 2.....35

Appendix E.2.2 Modelling 2 69

Appendix E.2.3 Parametric Analysis 2.....123

Abstract

The thesis presents a novel approach for estimating the impact of road roughness and macro-texture on truck fleet fuel consumption based on Big Data.

An extensive literature review was carried out to provide a comprehensive background for the study. Objectives of the study are defined as, primarily, to reduce the uncertainties left by previous studies, which based on experimental data (likely not be representative of real driving conditions) claimed that road roughness and macro-texture could affect the fuel consumption of road vehicles. In fact, this may represent an opportunity for road agencies that, in the light of these results, may wish to review the current maintenance strategies of road pavements to save significant direct and indirect costs for the society. Therefore, instead of carrying out new experiments, by exploring opportunities in the large amount of real-world data available in England, from truck fleet managers, road agencies and weather institutions, the method introduced in the present study promises to be able to provide estimates representative of real driving conditions. The method also differs from those presented in the past for its characteristic repeatability and adaptability to new situations. That includes adaptation of the model to different vehicles and to countries other than England, all without performing any additional field tests. This makes the introduced methodology

unique. Advanced statistics and machine learning (ML) techniques have been used to analyse the data. In particular, a multiple linear regression (LR), a support vector regression (SVR), a random forest (RF) and an artificial neural network (ANN) model, including the effect of road roughness and macro-texture as longitudinal profile variance (LPV) and sensor-measured texture depth (SMTD) respectively, have been developed for light, medium and heavy trucks driving along motorways and A roads in England. A parametric analysis was then used to interpret the obtained results and quantify the impact that each of the considered variables, including LPV and SMTD, have on the fuel consumption of the considered truck types.

Results show that, although the present study confirms that road surface characteristics, such as roughness and macrotexture, can affect the fuel consumption of trucks, due to the low quality of the data available, that is currently difficult to quantify. A comparison of the results obtained with the findings of studies conducted in the past, shows that there is some match in the order of magnitude of the estimates made, but this is not always the case. For instance, the impact of road roughness, measured as LPV at 3, 10 and 30 metres wavelength, was estimated to be between -2.91% and +6.27% for light trucks, between -8.56% and +3.97% for medium trucks and between -6.55% and +6.28% for heavy trucks, while the effect of macrotexture, measured as SMTD, was estimated to be between +1.04% and +1.21% for light trucks, between -1.21% and +5.98% for medium trucks and between -2.33% and +7.84% for heavy trucks. Variance among the obtained estimates is high and therefore, although results of the present study seem to be promising, it is fair to say that, at the current state of technology, while this approach is feasible, it is difficult to estimate the effect of road surface conditions on truck fleet fuel consumption using the Big Data approach, and further investigation is required to optimise the methodology.

Future work will have to consider a wider range of road conditions, increased vehicle reporting frequencies, vehicle speed and vehicle types, including cars. This will extend the applicability of the study and is necessary before estimation of the impact of road surface characteristics on vehicle fuel consumption can be considered reliable and ready to be implemented in the LCA analysis and maintenance programming of road pavements, to support decision making at strategic level.

Acknowledgements

First, I would like to thank both my supervisors Dr Tony Parry and Dr Luis Neves for their guidance and tutoring throughout the whole course of my studies.

I would like to thank Microlise Ltd and, in particular, Dr Mohammad Mesgarpour and Ian Dickinson for their help and for authorizing the use of an anonymized portion of their databases in my research.

I would like to thank also Dr Emma Benbow and Dr Helen Viner from TRL for their advices and suggestions towards the interpretation of the results of my research.

Also, I am grateful to Kamila, my girlfriend, for her great support and patience in these almost three years together. I love her so much.

I would like also to thank my family, and in particular my parents and my brother, for their continuous encouragement and endless support.

Finally, but not less importantly, I would like to thank all my friends, close and far, that helped, supported and continuously encouraged me all throughout my PhD studies.

List of abbreviations

AASHTO, American Association of State Highway and Transportation Officials

AFDC, Alternative Fuels Data Centre - U.S. Department of Energy

AI, Artificial Intelligence

ANN, Artificial Neural Network

ARTEMIS, Assessment and Reliability of Transport Emission Models and Inventory Systems

AVG, average

BNG, British National Grid

BSI, British Standards Institution

CAN, Control Area Network

CV, cross-validation

DEFRA, Department for Environment, Food & Rural Affairs

DfBEIS, Department for Business and Energy of the United Kingdom

DfT, Department for Transport of the United Kingdom

DVLA, Driver and Vehicle Licensing Agency

EB, East Bound

EC, European Commission

ECRPD, Energy Conservation in Road Pavement Design

ECU, Electronic Control Unit

EEA, European Environmental Agency

EERE, Office of Energy Efficiency and Renewable Energy - U.S. Department of Energy

EIA, Energy Information Administration

eLPV, Enhanced Longitudinal Profile Variance

ENERGY.GOV, U.S. Department of Energy

EP, European Parliament

EPA, Environmental Protection Agency

ERF, European Road Federation

EU, European Union

FC, Fuel Consumption

FHWA, Federal Highway Administration

FTP, Federal Test Procedure

GHG, Greenhouse Gas

GIS, Geographical Information Systems

GOV.UK, Government of the United Kingdom

GPRS, General Packet Radio Services

GPS, Global Positioning System

GSM, Global System for Mobile communications

GVW, Gross Vehicle Weight

GVWR, Gross Vehicle Weight Rating

HA, Highways Agency

HAPMS, Highways England Pavement Management System

HDM-4, Highway Development & Management 4

HDV, Heavy Duty Vehicle

HE, Highways England

HGV, Heavy Good Vehicle

IBM, International Business Machines Corporation

ID, Identification

IRI, International Roughness Index

ISO, International Organisation for Standardisation

Lasso, Least absolute shrinkage and selection operator

LCA, Life-Cycle Assessment

LDV, Light Duty Vehicle

LPG, Liquefied Petroleum Gas

LPV, Longitudinal Profile Variance

LR, Linear Regression

LSM, Least-Squared Method

MAE, Mean Absolute Error

MDV, Medium Duty Vehicle

Met Office, Meteorological Office of the United Kingdom

MIRAVEC, Models for rolling resistance In Road Infrastructure Asset Management systems

MIRIAM, Models for rolling resistance In Road Infrastructure Asset Management

ML, Machine Learning

MPD, Mean Profile Depth

MS, Microsoft

MSE, Mean Squared Error

NAPA, National Asphalt Pavement Association

NAS, National Academy of Sciences

NB, North Bound

NCHRP, National Cooperative Highway Research Program

NRC, National Research Council

NVDB, National Road Data Base of the Swedish Road Administration

OBD, On-Board Diagnostic

PA, Parametric Analysis

PCR, Product Category Rules

PIARC, World Road Association

RBF, Radial Basis Function

RF, Random Forest

RMSE, Root-Mean Squared Error

RR, rolling resistance

RSS, Residual Sum of Squares

SAE, Society of Automotive Engineers

SB, South Bound

SCRIM, Sideway-force Coefficient Routine Investigation Machine

SMTD, Sensor Measured Texture Depth

SQL, Structured Query Language

SRN, Strategic Road Network

SVR, Support Vector Regression

TR, training

TRACS, Traffic Speed Condition Survey

TRL, Transport Research Laboratory Ltd

TT, Test

TUG, Technical University of Gdansk

UCPRC, University of California Pavement Research Center

UK, United Kingdom

UN, United Nations

US, United States of America

Val, Validation

VETO, VEjstandard og Transport-Omkostninger

VTI, Swedish National Road and Transport Research Institute

WB, West Bound

WGS, World Geodetic System

List of publications

Perrotta, F., Parry, T., Neves, L. C., Buckland, T. and Benbow, E. (2019) 'Verification of the HDM-4 fuel consumption model using a Big Data approach : a UK case study', Transportation Research Part D: Transport and Environment. Vol. 67, February 2019, pp. 109 - 118 doi: 10.1016/j.trd.2018.11.001.

Perrotta, F., Parry, T., Neves, L. C., Buckland, T., Benbow, E. and Viner, H. (2018) 'Comparison of truck fuel consumption measurements with results of existing models and implications for road pavement LCA', in Caspeele, R., Taerwe, L., and Frangopol, D. (eds) The 6th International Symposium on Life-Cycle Civil Engineering (IALCCE) 2018. October 28 - 31, 2018. Ghent, Belgium.

Perrotta, F., Parry, T., Neves, L. C. and Mesgarpour, M. (2018) 'A machine learning approach for the estimation of fuel consumption related to road pavement rolling resistance for large fleets of trucks', in Caspeele, R., Taerwe, L., and Frangopol, D. (eds) The 6th International Symposium on Life-Cycle Civil Engineering (IALCCE) 2018. October 28 - 31, 2018. Ghent, Belgium.

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. April 16-20, 2018. Vienna, Austria. (Report)

Perrotta, F., Parry, T., Neves, L. C., Buckland, T. and Mesgarpour, M. (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, 2018. Washington, D.C. (Poster)

Perrotta, F., Parry, T. and Neves, L. C. (2017) 'Application of Machine Learning for Fuel Consumption Modelling of Trucks', in 2017 IEEE International Conference on Big Data. December 11-14, 2017. Boston, Massachussets.

Perrotta, F., Parry, T. and Neves, L. C. (2017) '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: 10.1201/9781315281896-7.

Perrotta, F., Trupia, L., Parry, T. and Neves, L. C. (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.

List of figures

Figure 1. Summary of a) energy and b) GHG shares by transport sector in the EU. Adapted from EEA (2016c) and EP (2017). 3

Figure 2. Summary of the four different types of wavelengths that characterise the shape of road surfaces. Adapted from Sandberg *et al.* (2011a). 11

Figure 3. Comparison of MPD and SMTD measurements for two different type of surface. Figure taken from Viner *et al.* (2006). 14

Figure 4. Representation of a) the deformation at the tire-pavement interaction (not to scale), and b) of the hysteretic loading-unloading cycle. 43

Figure 5. Example of the drum test machine for RR measurements. Different configurations exist. 48

Figure 6. Example of a trailer for RR measurements. The TUG from Technical University of Gdańsk is represented in the figure. Different configurations exist. 49

Figure 7. Schema of how the SAE J1939 system works. 79

Figure 8. Map representing all motorways (in blue) and primary roads (in red) part of the SRN of England (Spencer-Bickle *et al.*, 2017). 84

Figure 9. Map of the weather stations considered in the study. 89

Figure 10. Example of light truck analysed in the study. 91

Figure 11. Example of medium truck analysed in the study. 91

Figure 12. Example of heavy truck analysed in the study. 92

Figure 13. Map representing the location of data points considered in the initial phase of the study. These are 14,291 records from 1110 heavy trucks. 93

Figure 14. Distributions of data in the Dataset 1. 95

Figure 15. Map representing the location of data points considered in the second phase of the study. These are 119,667 records from 580 light trucks. 98

Figure 16. Example of data collected from the A41. For instance, these are from medium trucks. 99

Figure 17. Distributions of data in the Dataset 2 from light trucks from A roads only. 100

Figure 18. Distributions of data in the Dataset 2 from light trucks from all roads considered (M and A roads). 101

Figure 19. Distributions of data in the Dataset 2 from medium trucks from A roads only. ... 102

Figure 20. Distributions of data in the Dataset 2 from medium trucks from all roads considered (M and A roads). 103

Figure 21. Distributions of the data in Dataset 2 from heavy trucks from A roads only. 103

Figure 22. Distributions of the data in Dataset 2 from heavy trucks from all roads considered (M and A roads). 104

Figure 23. Representation of how SVRs work. 129

Figure 24. Schematic representation of how RFs work. 131

Figure 25. Scheme of a a) human and b) ANN's neurons. 132

Figure 26. Schema of a generic ANN model. 134

Figure 27. Example of convergence of learning curves. 137

Figure 28. Example of learning curves which do not converge at the expected/desired level of performance. 138

Figure 29. Example of learning curves which do not converge at the expected number of training samples. 139

Figure 30. Example of the results of the variable selection operated by the RF algorithm (in terms of %IncMSE) for (from left to right) light trucks App 1 and heavy trucks App 1. 150

Figure 31. Example of how each of the developed models LR, SVR, RF and ANN fit the available data. This represents the case of the heavy trucks App 2. 157

Figure 32. Example of the PA performed for (left to right) the medium trucks App 2 and heavy trucks App 1 for variation of the *Gross.Weight_kg* (kg). 165

Figure 33. Example of the PA performed for (left to right) the medium trucks App 1 and heavy trucks App 1 for variation of the *Geom.Gradient* (%)..... 166

Figure 34. Example of the PA performed for (left to right) the medium trucks App 1 and heavy trucks App 2 for variation of the *TRACS.LPV03m* (mm^2)..... 171

Figure 35. Example of the PA performed for (left to right) the light trucks App 1 and heavy trucks App 1 for variation of the *TRACS.LPV10m* (mm^2)..... 172

Figure 36. Example of the PA performed for (left to right) the Dataset 1 (heavy trucks) and light trucks App 2 for variation of the *TRACS.LPV30m* (mm^2)..... 173

Figure 37. Example of the PA performed for (left to right) the light trucks App 2 and medium trucks App 2 for variation of the *TRACS.Pav_Texture* (SMTD, mm). 174

Figure 38. Example of overfitting by the ML models on the *Geom.Abs_Radius* and *TRACS.Pav_Texture* for the case of heavy trucks App 2. 177

Figure 39. Correlation matrix for data in the Dataset 1. 3

Figure 40. Analysis of the adjusted-R^2 for the Dataset 1..... 6

Figure 41. Rank of variable importance as increase in MSE on the Dataset 1, as computed by the RF algorithm. 8

Figure 42. Fit of the developed LR to the Dataset 1..... 9

Figure 43. Ten-fold cross-validation of the SVR model for the Dataset 1. 11

Figure 44. Fit of the developed SVR model to the Dataset 1. 12

Figure 45. Ten-fold cross-validation of the RF model for the Dataset 1. 13

Figure 46. Fit of the developed RF model to the Dataset 1. 14

Figure 47. Ten-fold cross-validation of the ANN model for the Dataset 1. 16

Figure 48. Fit of the developed ANN model to the Dataset 1. 17

Figure 49 – RMSE learning curve for the ANN fitted on the Dataset 1. 19

Figure 50. MAE learning curves for the ANN fitted to the Dataset 1. 19

Figure 51. Parametric analysis of *Gross.Weight_kg* (kg) for the Dataset 1. 21

Figure 52. Parametric analysis of *Geom.Gradient* (%) for the Dataset 1. 22

Figure 53. Parametric analysis of *Torque.Start* (%) for the Dataset 1. 23

Figure 54. Parametric analysis of *Torque.End* (%) for the Dataset 1. 24

Figure 55. Parametric analysis of *Revs.Start* (rpm) for the Dataset 1. 25

Figure 56. Parametric analysis of *Speed.AVG* (km/h) for the Dataset 1. 25

Figure 57. Parametric analysis of *Acceleration* (m/s^2) for the Dataset 1. 27

Figure 58. Parametric analysis of *Geom.Abs_Radius* (m) for the Dataset 1. 28

Figure 59. Parametric analysis of *TRACS.LPV03m* (mm^2) for the Dataset 1. 29

Figure 60. Parametric analysis of *TRACS.LPV10m* (mm^2) for the Dataset 1. 30

Figure 61. Parametric analysis of *TRACS.LPV30m* (mm^2) for the Dataset 1. 31

Figure 62. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the Dataset 1. 32

Figure 63. Correlation matrix for the data from light trucks App 1, from the Dataset 2. 37

Figure 64. Analysis of the adjusted-R^2 for the light trucks App 1, from the Dataset 2. 39

Figure 65. Rank of variable importance as increase in MSE for the light trucks App1, from the Dataset 2, as computed by the RF algorithm. 41

Figure 66. Correlation matrix for the data from light trucks App 2, from the Dataset 2. 42

Figure 67. Analysis of the adjusted-R^2 for light trucks App 2, from the Dataset 2. 45

Figure 68. Rank of variable importance as increase in MSE for light trucks App 2, from the Dataset 2, as computed by the RF algorithm. 47

Figure 69. Correlation matrix for the data for the medium trucks App 1, from the Dataset 2. 48

Figure 70. Analysis of the adjusted-R^2 the medium trucks App 1, from the Dataset 2. 50

Figure 71. Rank of variable importance as increase in MSE for the medium trucks App 1, from the Dataset 2, as computed by the RF algorithm. 52

Figure 72. Correlation matrix for the data from the medium trucks App 2, from the Dataset 2. 53

Figure 73. Analysis of the adjusted-R^2 for the medium trucks App 2, from the Dataset 2.....55

Figure 74. Rank of variable importance as increase in MSE for the medium trucks App 2, from the Dataset 2, as computed by the RF algorithm. 57

Figure 75. Correlation matrix for the data from the heavy trucks App 1, from the Dataset 2. .58

Figure 76. Analysis of the adjusted-R^2 for the heavy trucks App 1, from the Dataset 2. 60

Figure 77. Rank of variable importance as increase in MSE the heavy trucks App 1, from the Dataset 2, as computed by the RF algorithm. 62

Figure 78. Correlation matrix for the data from heavy trucks App 2, from the Dataset 2.....63

Figure 79. Analysis of the adjusted-R^2 for the heavy trucks App 2, from the Dataset 2. 65

Figure 80. Rank of variable importance as increase in MSE for the heavy trucks App 2, from the Dataset 2, as computed by the RF algorithm. 67

Figure 81. Fit of the developed LR for the light trucks App 1, from the Dataset 2. 69

Figure 82. Ten-fold cross-validation of the SVR model for the light trucks App 1, from the Dataset 2. 71

Figure 83. Fit of the developed SVR model for the light trucks App 1, from the Dataset 2. ...72

Figure 84. Ten-fold cross-validation of the RF model for light trucks App 1, from the Dataset 2. 73

Figure 85. Fit of the developed RF model for the light trucks App 1, from the Dataset 2. 74

Figure 86. Ten-fold cross-validation of the ANN model for the light trucks App 1, from the Dataset 2. 75

Figure 87. Fit of the developed ANN model to the data for the light trucks App 1, from the Dataset 2. 76

Figure 88. Fit of the developed LR for the light trucks App 2, from the Dataset 2. 78

Figure 89. Ten-fold cross-validation of the SVR model for the light trucks App 2, from the Dataset 2. 79

Figure 90. Fit of the developed SVR for the light trucks App 2, from the Dataset 2. 80

Figure 91. Fit of the developed RF to light trucks for the light trucks App 2, from the Dataset 2. 81

Figure 92. Ten-fold cross-validation of the RF model for the light trucks App 2, from the Dataset 2. 82

Figure 93. Fit of the developed ANN for the light trucks App 2, from the Dataset 2. 83

Figure 94. Ten-fold cross-validation of the ANN model for the light trucks App 2, from the Dataset 2. 84

Figure 95. Fit of the developed LR for the medium trucks App 1, from the Dataset 2. 86

Figure 96. Ten-fold cross-validation of the SVR model for the medium trucks App 1, from the Dataset 2. 88

Figure 97. Fit of the developed SVR for the medium trucks App 1, from the Dataset 2. 89

Figure 98. Ten-fold cross-validation of the RF model for the medium trucks App 1, from the Dataset 2. 90

Figure 99. Fit of the developed RF for the medium trucks App 1, from the Dataset 2. 91

Figure 100. Ten-fold cross-validation of the ANN model for the medium trucks App 1, from the Dataset 2. 92

Figure 101. Fit of the developed ANN for the medium trucks App 1, from the Dataset 2. 93

Figure 102. Fit of the developed LR for the medium trucks App 2, from the Dataset. 95

Figure 103. Ten-fold cross-validation of the SVR model for the medium trucks App 2, from the Dataset 2. 96

Figure 104. Fit of the developed SVR for the medium trucks App 2, from the Dataset 2. 97

Figure 105. Ten-fold cross-validation of the RF model for the medium trucks App 2, from the Dataset 2. 98

Figure 106. Fit of the developed RF for the medium trucks App 2, from the Dataset 2. 99

Figure 107. Ten-fold cross-validation of the ANN model for the medium trucks App 2, from the Dataset 2. 100

Figure 108. Fit of the developed ANN for the medium trucks App 2, from the Dataset 2. ...101

Figure 109. Fit of the developed LR for the heavy trucks App 1, from the Dataset 2. 103

Figure 110. Ten-fold cross-validation of the SVR model for the heavy trucks App 1, from the Dataset 2. 104

Figure 111. Fit of the developed SVR for the heavy trucks App 1, from the Dataset 2. 105

Figure 112. Ten-fold cross-validation of the RF model for the heavy trucks App 1, from the Dataset 2. 106

Figure 113. Fit of the developed RF for the heavy trucks App 1, from the Dataset 2. 107

Figure 114. Ten-fold cross-validation of the ANN model for the heavy trucks App 1, from the Dataset 2. 108

Figure 115. Fit of the developed ANN for the heavy trucks App 1, from the Dataset 2. 109

Figure 116. Fit of the developed LR for the heavy trucks App 2, from the Dataset 2. 111

Figure 117. Ten-fold cross-validation of the SVR model for the heavy trucks App 2, from the Dataset 2. 112

Figure 118. Fit of the developed SVR for the heavy trucks App 2, from the Dataset 2. 113

Figure 119. Ten-fold cross-validation of the RF model for the heavy trucks App 2, from the Dataset 2. 114

Figure 120. Fit of the developed RF for the heavy trucks App 2, from the Dataset 2. 115

Figure 121. Ten-fold cross-validation of the ANN model for the heavy trucks App 2, from the Dataset 2. 116

Figure 122. Fit of the developed ANN for the heavy trucks App 2, from the Dataset 2. 117

Figure 123. RMSE learning curve for the ANN fitted for the heavy trucks App 2, from the Dataset 2. 118

Figure 124. MAE learning curve for the ANN fitted for the heavy trucks App 2, from the Dataset 2. 119

Figure 125. Ten-fold cross-validation of the ANN1 model for the heavy trucks App 2, from the Dataset 2. 120

Figure 126. Fit of the developed ANN1 for the heavy trucks App 2, from the Dataset 2. 121

Figure 127. Parametric analysis of *Gross.Weight_kg* (kg) for the light trucks App 1. 123

Figure 128. Parametric analysis of *Geom.Gradient* (%) for the light trucks App 1. 124

Figure 129. Parametric analysis of *Torque.AVG* (%) for the light trucks App 1. 125

Figure 130. Parametric analysis of *Revs.AVG* (rpm) for the light trucks App 1. 125

Figure 131. Parametric analysis of *Speed.AVG* (km/h) for the light trucks App 1. 126

Figure 132. Parametric analysis of *Acceleration* (m/s^2) for the light trucks App 1. 127

Figure 133. Parametric analysis of *Traveled.Time* (s) for the light trucks App 1. 128

Figure 134. Parametric analysis of *Geom.Abs_Radius* (m) for the light trucks App 1. 129

Figure 135. Parametric analysis of *Geom.Crossfall* (%) for the light trucks App 1. 129

Figure 136. Parametric analysis of *TRACS.LPV03m* (mm^2) for the light trucks App 1..... 130

Figure 137. Parametric analysis of *TRACS.LPV10m* (mm^2) for the light trucks App 1..... 131

Figure 138. Parametric analysis of *TRACS.LPV30m* (mm^2) for the light trucks App 1..... 132

Figure 139. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the light trucks App 1. 133

Figure 140. Parametric analysis of *Altitude.AVG* (mamls) for the light trucks App 1..... 134

Figure 141. Parametric analysis of *Atm_Pressure_hPa* (hPa) for the light trucks App 1..... 134

Figure 142. Parametric analysis of *Temperature_deg_C* (°C) for the light trucks App 1..... 135

Figure 143. Parametric analysis of *Humidity_*. (%) for the light trucks App 1. 136

Figure 144. Parametric analysis of *WindSpeedDir_kn* (kn) for the light trucks App 1. 137

Figure 145. Parametric analysis of *Gross.Weight_kg* (kg) for the light trucks App 2. 140

Figure 146. Parametric analysis of *Geom.Gradient* (%) for the light trucks App 2. 141

Figure 147. Parametric analysis of *Torque.AVG* (%) for the light trucks App 2. 141

Figure 148. Parametric analysis of *Revs.AVG* (rpm) for the light trucks App 2. 142

Figure 149. Parametric analysis of *Speed.AVG* (km/h) for the light trucks App 2. 143

Figure 150. Parametric analysis of *Acceleration* (m/s^2) for the light trucks App 2. 143

Figure 151. Parametric analysis of *Traveled.Time* (s) for the light trucks App 2. 145

Figure 152. Parametric analysis of *Geom.Abs_Radius* (m) for the light trucks App 2. 145

Figure 153. Parametric analysis of *Geom.Crossfall* (%) for the light trucks App 2. 146

Figure 154. Parametric analysis of *TRACS.LPV03m* (mm^2) for the light trucks App 2. 147

Figure 155. Parametric analysis of *TRACS.LPV10m* (mm^2) for the light trucks App 2. 148

Figure 156. Parametric analysis of *TRACS.LPV30m* (mm^2) for the light trucks App 2. 149

Figure 157. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the light trucks App 2. 150

Figure 158. Parametric analysis of *Altitude.AVG* (mamls) for the light trucks App 2. 150

Figure 159. Parametric analysis of *Atm_Pressure_hPa* (hPa) for the light trucks App 2. 151

Figure 160. Parametric analysis of *Temperature_deg_C* (℃) for the light trucks App 2. 152

Figure 161. Parametric analysis of *Humidity_*. (%) for the light trucks App 2. 152

Figure 162. Parametric analysis of *WindSpeedDir_kn* (kn) for the light trucks App 2. 153

Figure 163. Parametric analysis of *Gross.Weight_kg* (kg) for the medium trucks App 1. 156

Figure 164. Parametric analysis of *Geom.Gradient* (%) for the medium trucks App 1. 157

Figure 165. Parametric analysis of *Torque.AVG* (%) for the medium trucks App 1. 158

Figure 166. Parametric analysis of *Revs.AVG* (rpm) for the medium trucks App 1. 159

Figure 167. Parametric analysis of *Speed.AVG* (km/h) for the medium trucks App 1. 160

Figure 168. Parametric analysis of *Acceleration* (m/s^2) for the medium trucks App 1..... 161

Figure 169. Parametric analysis of *Traveled.Time* (s) for the medium trucks App 1. 161

Figure 170. Parametric analysis of *Geom.Abs_Radius* (m) for the medium trucks App 1. ... 162

Figure 171. Parametric analysis of *Geom.Crossfall* (%) for the medium trucks App 1. 163

Figure 172. Parametric analysis of *TRACS.LPV03m* (mm^2) for the medium trucks App 1... 164

Figure 173. Parametric analysis of *TRACS.LPV10m* (mm^2) for the medium trucks App 1... 165

Figure 174. Parametric analysis of *TRACS.LPV30m* (mm^2) for the medium trucks App 1... 165

Figure 175. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the medium trucks App 1. 166

Figure 176. Parametric analysis of *Altitude.AVG* (mamls) for the medium trucks App 1. 167

Figure 177. Parametric analysis of *Atm_Pressure_hPa* (hPa) for medium trucks App 1. 168

Figure 178. Parametric analysis of *Temperature_deg_C* (°C) for the medium trucks App 1. 169

Figure 179. Parametric analysis of *Humidity_*. (%) for the medium trucks App 1. 169

Figure 180. Parametric analysis of *WindSpeedDir_kn* (kn) for the medium trucks App 1.... 170

Figure 181. Parametric analysis of *Gross.Weight_kg* (kg) for the medium trucks App 2. 173

Figure 182. Parametric analysis of *Geom.Gradient* (%) for the medium trucks App 2. 174

Figure 183. Parametric analysis of *Torque.AVG* (%) for the medium trucks App 2. 175

Figure 184. Parametric analysis of *Revs.AVG* (rpm) for the medium trucks App 2. 175

Figure 185. Parametric analysis of *Speed.AVG* (km/h) for the medium trucks App 2. 176

Figure 186. Parametric analysis of *Acceleration* (m/s^2) for the medium trucks App 2..... 177

Figure 187. Parametric analysis of *Traveled.Time* (s) for the medium trucks App 2. 178

Figure 188. Parametric analysis of *Geom.Abs_Radius* (m) for the medium trucks App 2. ... 179

Figure 189. Parametric analysis of *Geom.Crossfall* (%) for the medium trucks App 2. 179

Figure 190. Parametric analysis of *TRACS.LPV03m* (mm^2) for the medium trucks App 2. .. 180

Figure 191. Parametric analysis of *TRACS.LPV10m* (mm^2) for the medium trucks App 2. .. 181

Figure 192. Parametric analysis of *TRACS.LPV30m* (mm^2) for the medium trucks App 2. .. 182

Figure 193. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the medium trucks App 2. 183

Figure 194. Parametric analysis of *Altitude.AVG* (mamls) for the medium trucks App 2..... 183

Figure 195. Parametric analysis of *Atm_Pressure_hPa* (hPa) for medium trucks App 2. 184

Figure 196. Parametric analysis of *Temperature_deg_C* (°C) for the medium trucks App 2. 185

Figure 197. Parametric analysis of *Humidity_*. (%) for the medium trucks App 2. 186

Figure 198. Parametric analysis of *WindSpeedDir_kn* (kn) for the medium trucks App 2. ... 186

Figure 199. Parametric analysis of *Gross.Weight_kg* (kg) for the heavy trucks App 1..... 189

Figure 200. Parametric analysis of *Geom.Gradient* (%) for the heavy trucks App 1. 190

Figure 201. Parametric analysis of *Torque.AVG* (%) for the heavy trucks App 1. 191

Figure 202. Parametric analysis of *Revs.AVG* (rpm) for the heavy trucks App 1. 192

Figure 203. Parametric analysis of *Speed.AVG* (km/h) for the heavy trucks App 1..... 192

Figure 204. Parametric analysis of *Acceleration* (m/s^2) for the heavy trucks App 1..... 193

Figure 205. Parametric analysis of *Traveled.Time* (s) for the heavy trucks App 1. 194

Figure 206. Parametric analysis of *Geom.Abs_Radius* (m) for the heavy trucks App 1. 195

Figure 207. Parametric analysis of *Geom.Crossfall* (%) for the heavy trucks App 1..... 195

Figure 208. Parametric analysis of *TRACS.LPV03m* (mm^2) for the heavy trucks App 1. 196

Figure 209. Parametric analysis of *TRACS.LPV10m* (mm^2) for the heavy trucks App 1. 197

Figure 210. Parametric analysis of *TRACS.LPV30m* (mm^2) for the heavy trucks App 1. 197

Figure 211. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the heavy trucks App 1..... 198

Figure 212. Parametric analysis of *Altitude.AVG* (mamls) for the heavy trucks App1. 199

Figure 213. Parametric analysis of *Atm_Pressure_hPa* (hPa) for the heavy trucks App 1.... 200

Figure 214. Parametric analysis of *Temperature_deg_C* (°C) for the heavy trucks App 1. .. 201

Figure 215. Parametric analysis of *Humidity_*. (%) for the heavy trucks App 1. 202

Figure 216. Parametric analysis of *WindSpeedDir_kn* (kn) for the heavy trucks App 1. 202

Figure 217. Parametric analysis of *Gross.Weight_kg* (kg) for the heavy trucks App 2. 205

Figure 218. Parametric analysis of *Geom.Gradient* (%) for the heavy trucks App 2. 207

Figure 219. Parametric analysis of *Torque.AVG* (%) for the heavy trucks App 2. 208

Figure 220. Parametric analysis of *Revs.AVG* (rpm) for the heavy trucks App 2. 208

Figure 221. Parametric analysis of *Speed.AVG* (km/h) for the heavy trucks App 2. 209

Figure 222. Parametric analysis of *Acceleration* (m/s^2) for the heavy trucks App 2. 210

Figure 223. Parametric analysis of *Traveled.Time* (s) for the heavy trucks App 2. 210

Figure 224. Parametric analysis of *Geom.Abs_Radius* (m) for the heavy trucks App 2. 211

Figure 225. Parametric analysis of *Geom.Crossfall* (%) for the heavy trucks App 2. 212

Figure 226. Parametric analysis of *TRACS.LPV03m* (mm^2) for the heavy trucks App 2. 213

Figure 227. Parametric analysis of *TRACS.LPV10m* (mm^2) for the heavy trucks App 2. 213

Figure 228. Parametric analysis of *TRACS.LPV30m* (mm^2) for the heavy trucks App 2. 214

Figure 229. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the heavy trucks App 2. 215

Figure 230. Parametric analysis of *Altitude.AVG* (mamls) for the heavy trucks App 2. 216

Figure 231. Parametric analysis of *Atm_Pressure_hPa* (hPa) for the heavy trucks App 2.... 217

Figure 232. Parametric analysis of *Temperature_deg_C* (°C) for the heavy trucks App 2. .. 217

Figure 233. Parametric analysis of *Humidity_*. (%) for the heavy trucks App 2. 218

Figure 234. Parametric analysis of *WindSpeedDir_kn* (kn) for the heavy trucks App 2. 219

List of tables

Table 1 - Summary of energy consumption (in %) by sector for the EU, the UK and the US. Adapted from DfBEIS (2017), EEA (2017a) and EIA (2018a, 2018b). 2

Table 2 - Summary of GHG emissions (in %) by sector for the EU, the UK and the US. Adapted from EEA (2017b), EPA (2017a) and DfBEIS (2018). 2

Table 3 - Classification of trucks as specified by the EU (EP, 2006). 28

Table 4 - Classification of trucks in the US as specified by the FHWA (FHWA, 2014). 29

Table 5 - Road vehicle classification by vehicle weight as suggested by EPA (adapted from (EPA, 2017b)). 29

Table 6 - Classification of trucks in the UK (adapted from (DfT, 2003)). 30

Table 7 - Summary of the existing methods currently used to evaluate RR. 52

Table 8 - Summary of the major findings from studies conducted in the past regarding the impact of RR and road surface characteristics on vehicle fuel consumption. 68

Table 9 – Summary of the data provided by Microlise. 81

Table 10 - Summary of the DVLA data included by Microlise in the Vehicle data table. 83

Table 11 – Summary of the TRACS data (from the HAPMS of HE) used in the study. 87

Table 12 - Summary of weather condition measurements considered in the study. 90

Table 13 – Summary of the final dataset. 97

Table 14 - Summary of data in Data.T data table.....110

Table 15 - Summary of road condition measurements considered in the study. 111

Table 16 - Summary of weather condition measurements considered in the study.....113

Table 17 – Summary of the results from the performed variable selection.....153

Table 18 - Learning curves: approximated amount of data needed for each ML algorithm to converge. 155

Table 19 - Comparison of the performance of fuel consumption models for heavy trucks. ..158

Table 20 - Comparison between studies conducted in the past and the present study of the approaches adopted to model the fuel consumption of road vehicles. 160

Table 21 - Summary of the results of the PA for the *Gross.Weight_kg* (kg).....165

Table 22 - Summary of the results of the PA for the *Geom.Gradient* (%).....166

Table 23 - Summary of the results of the PA for the *TRACS.LPV03m* (mm^2).....171

Table 24 - Summary of the results of the PA for the *TRACS.LPV10m* (mm^2).....172

Table 25 - Summary of the results of the PA for the *TRACS.LPV30m* (mm^2).....173

Table 26 - Summary of the results of the PA for the *TRACS.Pav_Texture* (mm).....174

Table 27 - Summary of the results obtained for light trucks (at 87.5 km/h). 179

Table 28 - Summary of the results obtained for medium trucks (at 87.5 km/h).....180

Table 29 - Summary of the results obtained for heavy trucks (at 87.5 km/h). 182

Table 30 - Summary of the computed p-values for the highest correlated variables for the Dataset 1.....5

Table 31 - Summary of the Lasso coefficients computed for the selected variables, for the Dataset 1.....7

Table 32 - Summary of the performance of the SVR model over the ten-fold cross-validation process on the Dataset 1. 10

Table 33 - Summary of the performance of the RF model over the ten-fold cross-validation

process on the Dataset 1. 12

Table 34 - Summary of the performance of the ANN model over the ten-fold cross-validation process on the Dataset 1. 15

Table 35 - Summary of performance of the regression models developed using the Dataset 1.17

Table 36 - Comparison of performance between the small (ANN) and larger (ANN1) neural networks fitted to the Dataset 1. 20

Table 37 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the Dataset 1. 33

Table 38 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the Dataset 1. 34

Table 39 - Summary of the computed p-values of the correlated variables for light trucks App 1, from the Dataset 2. 38

Table 40 - Summary of the Lasso coefficients for the light trucks App1, from the Dataset 2. 40

Table 41 - Summary of the computed p-values of the correlated variables for light trucks App 2, from the Dataset 2. 44

Table 42 - Summary of the Lasso coefficients for light trucks App2, from the Dataset 2..... 46

Table 43 - Summary of the computed p-values of the correlated variables for the medium trucks App 1, from the Dataset 2. 49

Table 44 - Summary of the Lasso coefficients the medium trucks App 1, from the Dataset 2.51

Table 45 - Summary of the computed p-values of the correlated variables for the medium trucks App 2, from the Dataset 2. 54

Table 46 - Summary of the Lasso coefficients for the medium trucks App2, from the Dataset 2. 56

Table 47 - Summary of the computed p-values of the correlated variables for the heavy trucks App 1, from the Dataset 2. 59

Table 48 - Summary of the Lasso coefficients for the heavy trucks App 1, from the Dataset 2.

.....61

Table 49 - Summary of the computed *p-values* of the correlated variables for the heavy trucks App 2, from the Dataset 2. 64

Table 50 - Summary of the Lasso coefficients for the heavy trucks App 2, from the Dataset 2. 66

Table 51 - Summary of performance of the regression models developed for the light trucks App 1.....77

Table 52 - Summary of performance of the regression models developed for the light trucks App 2.....85

Table 53 - Summary of performance of the regression models developed for the medium trucks App 1.....94

Table 54 - Summary of performance of the regression models developed for the medium trucks App 2.....102

Table 55 - Summary of performance of the regression models developed for the heavy trucks App 1.....110

Table 56 - Summary of performance of the regression models developed for the heavy trucks App 2.....122

Table 57 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the light trucks App 1.....138

Table 58 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the light trucks App 1. 139

Table 59 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the light trucks App 2.....154

Table 60 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the

5th and 95th percentiles of each of the considered variable, for the light trucks App 2. 155

Table 61 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the medium trucks App 1. 171

Table 62 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the medium trucks App 1. 172

Table 63 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the medium trucks App 2. 187

Table 64 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the medium trucks App 2. 188

Table 65 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the heavy trucks App 1. 203

Table 66 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the heavy trucks App 1. 204

Table 67 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the heavy trucks App 2. 220

Table 68 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the heavy trucks App 2. 221

1 Introduction

Road vehicles are major contributors to air pollution and GHG gas emissions, with the first that significantly impacts on human health, increasing the risk of respiratory diseases and allergic reactions, and the second, which increase the progression of climate changes, causing extreme temperatures, more frequent and severe droughts, flooding, wildfires, etc. (EPA, 2018a). This can harm human health and cause significant costs for the society. For these reasons, it is paramount to invest in reducing the impact of road vehicles on the environment, and particular effort is required from those countries that contribute the most to the production of air pollutants.

In 2011, the European Commission released the White Paper on Transport (EC, 2011) which states that, by 2050, the EU should meet a target of, at least, 60% reduction of greenhouse gas (GHG) emissions from the transport sector in comparison to 1990 levels. This ambitious goal requires the effort of all countries in the EU and all sectors. Action to reduce GHG emissions in the EU have been underway for decades. However, accordingly to the most recent projections (EP, 2017), if no further action is undertaken, by 2050 the GHG emissions from the transport sector in the EU will be 15% higher than in 1990. This is due to the increased travelled distances and

goods transported within Europe, which are a reflection of the economic growth of the countries members of the EU (EP, 2017).

(DfBEIS, 2017; EEA, 2017a; EIA, 2018a, 2018b) show that about 30% of the whole energy demand in Europe is used by transport (see Table 1). Of this, due reliance of road vehicles on oil derived fuels, road transport is the sector that contributes the most to the production of GHG emissions, with cars making the highest contribution followed by trucks and buses (EP, 2017) (see Table 2).

Table 1 - Summary of energy consumption (in %) by sector for the EU, the UK and the US. Adapted from DfBEIS (2017), EEA (2017a) and EIA (2018a, 2018b).

	Commercial & Residential	**Industry**	**Transportation**
EU	36 %	31 %	33 %
UK	29 %	31 %	40 %
US	39 %	32 %	29 %

Table 2 - Summary of GHG emissions (in %) by sector for the EU, the UK and the US. Adapted from EEA (2017b), EPA (2017a) and DfBEIS (2018).

	Commercial & Residential	**Industry**	**Transportation**
EU	11 %	65 %	24 %
UK	17 %	47 %	34 %
US	11 %	59 %	28 %

Specifically, the road transport sector accounts for about the 74% of the energy demand from transport in Europe, which makes approximately 95% of the GHG emissions (EP, 2017) (see Figure 1). However, road transport is vital for the economy and society. Therefore, curbing transportation is not an option, but the need to meet the requirements set in the White Paper 2011 (EC, 2011) and agreed in the

Paris Agreement (UN, 2015) for limiting the effects of global warming requires immediate action at a strategic level.

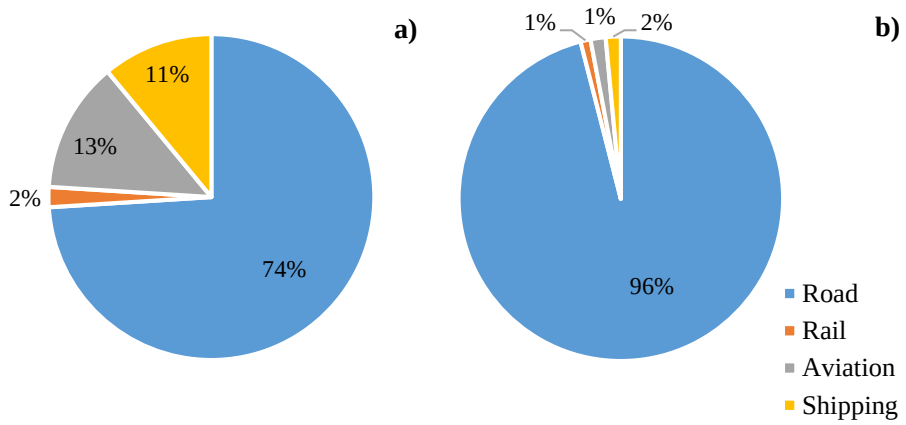


Figure 1. Summary of a) energy and b) GHG shares by transport sector in the EU. Adapted from EEA (2016c) and EP (2017).

With about 5.5 million km of roads and an estimated value of € 8,000 billion, the road infrastructure represents the most extended and valuable asset in the EU (ERF, 2014). It makes possible the exchange of goods and travel of citizens without obstacles. For this reason, maintaining a good level of service on the network by improving the reliability of the infrastructure is essential to ensure efficiency, competitiveness and safety.

In the last few decades, several studies claimed that smoother roads can help in reducing the fuel consumption of road vehicles (e.g. Sandberg, 1990; Beuving *et al.*, 2004; Chatti *et al.*, 2012). This should require attention from engineers and raise awareness at political and strategic levels for the need to invest in road maintenance. This may represent an opportunity for road agencies to review the current maintenance strategies in consideration of the extra costs and environmental impact that derive from poorly maintained roads. This could help the EU in meeting the targets fixed in the White Paper on Transport in 2011 and recently renewed in the Paris Agreement in

2015 in terms of emissions. However, the conclusions of existing studies are based on results from experiments where a few vehicles have been tested at constant speed on a few selected road segments under controlled conditions (e.g. no wind, dry roads, etc.). This may not be representative of what happens in real driving conditions, limiting the usefulness of these results in supporting changes to maintenance policies.

This study presents an innovative methodology for the estimation of the fuel consumption of road vehicles and the impact of road surface conditions. In particular, large quantities of data from standard sensors installed on trucks (SAE, 2016) are analysed in combination with data from Highways England (HE), the road agency in England, who routinely monitor the performance of the road infrastructure, and and data from the UK's Meteorological Office (Met Office), who continuously monitor the weather conditions in England. Application of multiple machine learning (ML) techniques supports the development of models that, thanks to their ability to identify complex correlations across large datasets, can estimate the influence of the road surface characteristics on truck fuel consumption. This represents a significant change in paradigm in the development of fuel consumption models, from an experimental to a Big Data approach. The proposed method, can take into account variables that were not considered or considered in simplistic ways in previous studies, including air temperature, wind speed, acceleration, deceleration, traffic and, being based on measurement taken from vehicles in normal operational conditions, the model can provide approximation of the fuel consumption in real conditions.

1.1 Background

Recently, the EP (2014) highlighted the opportunity for road agencies to contribute to the reduction of GHG emissions from the road transport industry by performing road maintenance. Extensive literature has pointed out the correlation

between the poor conditions of the road surface with discomfort in driving (e.g. Wang *et al.*, 2014; Hu *et al.*, 2017), noise (Birkner *et al.*, 2016) and accidents (Kane *et al.*, 2009). However, others claimed that this could also impact fuel economy (e.g. Laganier *et al.*, 1990; Sandberg, 1990; Chatti *et al.*, 2012; Haider *et al.*, 2012; Benbow *et al.*, 2013), and thus GHG emissions from road vehicles (Araújo *et al.*, 2014). This is due to the impact of road surface characteristics, i.e. roughness and macro-texture, on rolling resistance (RR), which consumes part of the power produced by vehicles' engine (e.g. Haider *et al.*, 2011; Sandberg *et al.*, 2011a; Hooghwerff *et al.*, 2013; Sohaney *et al.*, 2013; Ejsmont *et al.*, 2017). Generally speaking, RR is the force that resists the rolling motion of any physical body on a given surface. In the case of road vehicles, this refers to the resistance generated by the tire/pavement interaction and the energy lost due to the non-elastic deformation of the tire. This is hystereticly dissipated, resulting in an increased fuel consumption. The hysteretic behaviour depends mostly on the physical properties of the material composing the tire, but it is also affected by the characteristics of material properties and geometry of the pavement (Beuving *et al.*, 2004; Lee *et al.*, 2014). A further cause of RR is the slippage between the wheel and the road surface. This dissipates energy because not all the power transmitted from the engine to the tires is effectively used for producing motion of the vehicle.

It has been estimated that a reduction of RR by 10% is technically feasible (e.g. due to new materials for road pavements and tires, by performing road maintenance, etc.) and can improve the fuel economy of road vehicles by 1% – 2% (NRC, 2006; Chatti *et al.*, 2012). Considering that there are approximately 300 million vehicles in Europe (ACEA, 2018), which yearly consume about 275 Btoe (billion tonnes of oil equivalent), reducing the total amount of fuel consumed by 2% would

represent a meaningful accomplishment, which could help to save enormous costs to the society and reduce GHG emissions from the road transport industry. However, existing studies (e.g. Descornet, 1990; du Plessis *et al.*, 1990; Laganier *et al.*, 1990; Sandberg, 1990; Sandberg *et al.*, 2011a; Chatti *et al.*, 2012; Hammarström *et al.*, 2012; Benbow *et al.*, 2013; Haider *et al.*, 2013; Birkner *et al.*, 2016; Ejsmont *et al.*, 2017) based their conclusions on experimental data only. These were measured by using specifically designed devices for measuring RR or performing tests on a limited number of vehicles under carefully controlled conditions (e.g. steady speed, on flat road segments, etc.), and just for a few short sections of road. Therefore, this may not be representative of what happens at network level under real driving conditions, resulting in significant uncertainty on the impact of roughness and macro-texture on RR (Willis *et al.*, 2015). For this reason, the impact of road conditions on fuel consumption is rarely considered in the life-cycle studies of road pavements (Santero *et al.*, 2011b), causing significant direct and indirect costs for the society (Araújo *et al.*, 2014; Trupia, 2017).

1.2 Aims and objectives

The main aim of this research is to assess the impact of road surface characteristics, such as road roughness and macro-texture, on truck fleet fuel consumpion, for conditions representative of the Strategic Road Network (SRN) of England.

With a focus on truck fleets, using information from existing databases and data from standard sensors installed on trucks (SAE, 2016), the study answers to both questions, supporting road managers decision making in regard to the maintenance and the design of road pavements considering user's and environmental impacts. Thus, the objectives of this study are to:

i) test the usability of truck sensors' data for capturing the impact of road roughness and macro-texture on truck fleet fuel consumption, for conditions typical of the SRN of England;

ii) develop a new methodology for statistically assessing the impact of road surface on vehicle fuel consumption under real driving conditions at network level;

iii) validate the proposed methodologies on a large dataset and make results representative for the whole SRN of England;

iv) provide more realistic fuel consumption models to be used in LCA studies of road pavements, and decision making in maintenance and repair of road pavements, in England.

It was not within the aims of the study to investigate the impact of the pavement stiffness, pavement deflection and pavement type on truck fleet fuel consumption. Therefore, although for completeness these topics are considered in the literature review (see sect. 2.7.1 and 2.7.2), it is out of the scope of the present research to analyse that.

1.3 Scope of the study

Truck fleet managers and road agencies collect large quantities of data regarding their assets. This is to support decision making in regard to operation of large vehicle fleets and road maintenance. The data is collected as a requirement of existing standards (Highways Agency, 2008a; SAE, 2016) that specify modality, frequency, accuracy, precision, etc. by which the data is collected. Among others, data from the truck fleet management systems include fuel consumption and data from the databases of road agencies include measurements of the conditions of the road infrastructure such as measurements of the road surface roughness and macro-texture.

The fundamental concept of the current study is that it is possible to assess the influence of the conditions of the road surface on truck fleet fuel consumption by making secondary use of data currently available. In fact, although the considered data may not be the most precise and do not include information about all the variables that affect fuel consumption, it is expected that, thanks to the large amount of information available, it is possible to neglect the bias in the data and still model the fuel consumption of the considered vehicles and the impact of road surface conditions with a good level of accuracy. Combination of advanced statistics and automated pattern recognition techniques have been used to explore hidden patterns. Based on the results presented by Almér (2015), application of ML was explored. Using these techniques, optimised fuel consumption models able to take into account the effect of various wavelengths of road roughness and macro-texture on truck fleet fuel economy were developed and validated. Finally, a parametric analysis was utilised to quantify the impact that each of the considered variables have on the fuel consumption of the considered fleets of truck, which helped interpreting the results of the study. This represents an innovative approach, previously untested in the field, which is shown to be promising in terms of results.

1.4 Structure of the thesis

Chapter 1 corresponds to the present introduction.

Chapter 2 provides background information of road surface characteristics, pavement LCA, vehicle fuel consumption, factors affecting the energy consumption of trucks, RR, theoretical approaches, field measurements and laboratory tests to assess the influence of road surface characteristics on vehicle fuel consumption. Also, in this chapter the concept of ‘Big Data’ is introduced.

Chapter 3 describes in detail the data used in the study. Firstly, an overview of the truck fleet management system, the SAE J1939 standard and the origin and type of data from trucks used in the study is presented. Then, the data collected by road managers and Highways England (HE) in the UK is described, including an overview of the current state of the road infrastructure in the UK. Finally, the chapter describes the data obtained from the Met Office.

Chapter 4 describes the methodology applied in the study; it explains the type of pre-processing performed on the data collected, provides details about the methods used for variable selection, introduces the theory of the regression analysis and ML techniques applied to model the fuel consumption of the considered fleets of truck, and describes the methods used in the study to investigate the impact of road surface characteristics.

Chapter 5 summarises and discusses the results of the study. First, result from an initial case study, on a small dataset, from heavy trucks driven at relatively constant speed on motorways are analysed. This helped at testing feasibility of the study, defining a methodology, and allowing to make a first comparison with conclusions from studies conducted in the past. Then, six additional cases study are analysed. Larger datasets from different truck types driving at relatively constant speed along motorways and A roads are considered. That allows to test the developed methodology and is to improve the applicability of the study. Results are then discussed in a critical way and compared to those of studies conducted in the past, to help identifying strengths and limitations of the presented approach.

Chapter 6 gives conclusions and recommendations.

2 Literature Review

This chapter presents a review of existing knowledge on the impact that road infrastructure conditions have on vehicle fuel economy and the environment. Firstly, methods for the evaluation of road surface characteristics, in the UK and at international level, are presented and briefly compared (sect. 2.1). Then, methods for pavement Life-Cycle Assessment (LCA) and vehicle fuel economy modelling are discussed, providing additional insight into the motivation of the study (sect. 2.2). This is followed by two sections dedicated to vehicle fuel economy (sect. 2.3), and principles of vehicle dynamics (sect. 2.4). Following this, the major factors that affect fuel consumption and estimation of their impacts are considered in the literature (sect. 2.5). A section (sect. 2.6) is then dedicated to the the tire-pavement interaction and its effects on vehicle fuel consumption. After that, the concept of RR is introduced, with focus on parameters that influence this type of pavement vehicle interaction (sect. 2.7). The section (sect. 2.7) provides also an extensive and comprehensive critical review of existing studies on the methods used to quantify the impact of pavement type, RR and other pavement related parameters, including rod surface characteristics, such as road unevenness and macotexture, on vehicle fuel consumption. Finally, sect. 2.8

introduces the concept of 'Big Data' and discusses its application in road asset management.

2.1 Road surface characteristics

A pavement surface is usually characterised by four different wavelengths of road surface profile (Figure 2).

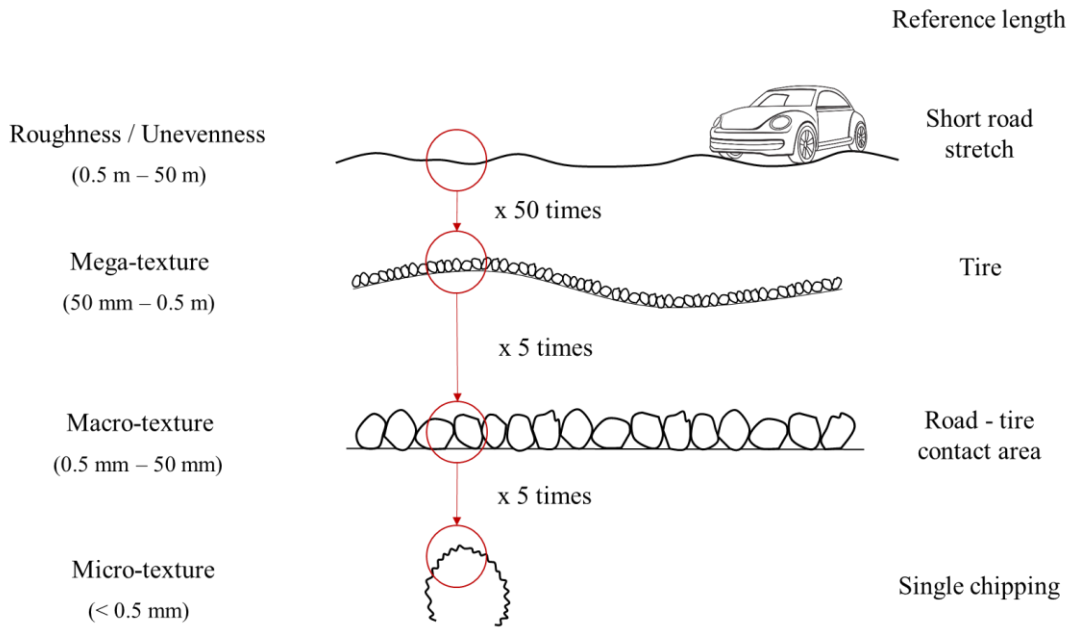


Figure 2. Summary of the four different types of wavelengths that characterise the shape of road surfaces. Adapted from Sandberg *et al.* (2011a).

- Roughness (alternatively referred to as unevenness) (wavelength between 0.5 m to 50 m) describes amplitudes at longer wavelengths. That affects vehicle suspensions (Sandberg, 1990; Sandberg *et al.*, 2011a; Haider *et al.*, 2012).
- Mega-texture (wavelength between 50 mm to 500 mm) represents the degree of smoothness of the surface.
- Macro-texture (wavelength between 0.5 mm to 50 mm) represents the height above a road surface of the aggregate chipping, or the depth of texture below

the road surface. That affects deformation of the tire (Klein *et al.*, 2004; Sandberg *et al.*, 2011a; Ejsmont *et al.*, 2017).

- Micro-texture (wavelength less than 0.5 mm) describes the roughness of the surface aggregate, which is associated with the crystalline structure of the coarse aggregate and the sand particles in the surface laitance of a brushed concrete surface.

2.1.1 IRI vs LPV

Worldwide, the International Roughness Index (IRI) is used to characterise longitudinal road roughness and to evaluate ride quality. This is computed from measurements of slope of the road surface taken by a quarter car model driven at 80 km/h (Sayers, 1989). Basically, IRI is a mathematical representation of the suspension motion that a vehicle accumulates while driving on a certain road surface over a certain distance (Benbow *et al.*, 2006). For this reason, it has the same units of a slope, and usually it is represented in mm/km (or inches/mile). However, in the UK, eLPV (enhanced Longitudinal Profile Variance) is used to assess road roughness and ride quality (Benbow *et al.*, 2006; Highways Agency, 2008a). This usually is measured as a modified moving average of the profile amplitude variance squared, at two points, measured every 1 metre (Benbow *et al.*, 2006; Highways Agency, 2008a). Units of eLPV are mm^2. Measurements are averaged over three different wavelengths, 3, 10 and 30 metres, to capture different levels of deformation of the longitudinal profile of the road. For instance, whereas the 3 m variance reflects small undulations of the road surface, which may be more significant at lower speeds, 30 m variance reflects long wavelengths, which would be most likely to affect ride quality at higher speeds (Benbow *et al.*, 2006). The measurement is then transformed to eLPV in order to take in consideration of the effects of road geometry (such as crossfall and road gradient).

eLPV is the only data stored in the Highways England Pavement Management System (HAPMS) and raw data of LPV are not publicly accessible (Highways Agency, 2008a). Therefore, although LPV and eLPV are well correlated (Highways Agency, 2008a), these are not the same, but in the present study eLPV is always referred as LPV for simplicity. For additional information about how eLPV is calculated from the LPV raw data, please, refer to the TRL Published Report PR/CSN/08/02 (McRobbie *et al.*, 2002) or to the Appendix E of the TRL Report 131 (Benbow *et al.*, 2006).

Clearly, IRI and LPV measure different things and comparing the two is not simple. For this reason, recently, in order to make comparison possible to a certain extent, independent research was carried out. The study showed that, in the UK, LPV is closely related to IRI (Benbow *et al.*, 2011). In particular, the authors showed that IRI can be derived from LPV at 3 and 10 m wavelengths using Equation (1):

$$IRI = \max(\sqrt{10 \cdot LPV03m/3} + \sqrt{LPV10m} - 0.1, 0) \quad (1)$$

Because in the UK LPV is the standard to assess ride quality of roads, this is the type of information stored in the databases of HE and therefore used in this study. Comparison with IRI can be made using Equation (1), which does not consider variation of LPV at 30 m wavelength.

2.1.2 MPD vs SMTD

Both Mean Profile Depth (MPD) and Sensor Measured Texture Depth (SMTD) are measurements used by road managers to assess macro-texture and evaluate road safety. The MPD is currently the parameter most used internationally for characterising macro-texture at international level. To measure that, the road is divided into segments of 10 cm each and the average of all peaks of the road surface within these segments calculated. Then two portions of 5 cm each are considered and, for each of them, the highest peak of the road surface measured. From that the average

elevation of all peaks within 10 cm is subtracted and, finally, the MPD is given by the average of the two measurements. For SMTD, 10 m sections are considered. Of these, the average peak height is calculated and the root mean square of peaks above and below the average computed. That represents the standard deviation of texture profile measurements over 10 m. In order to make differences between MPD and SMTD clearer, Figure 3 exemplifies evaluation methods from a graphical perspective. Please, note that, although Figure 3 does not show it, MPD is evaluated on 10 cm sections while SMTD is evaluated on 10 m sections.

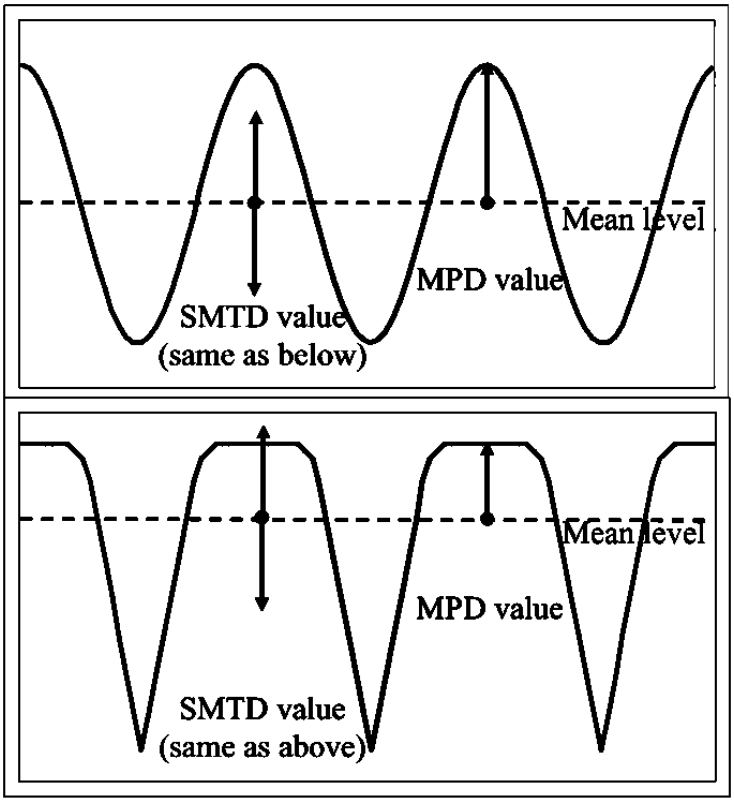


Figure 3. Comparison of MPD and SMTD measurements for two different type of surface. Figure taken from Viner *et al.* (2006).

In 2006, Viner *et al.* analysed the relationship between MPD and SMTD and developed a model (Equation (2)) to transform SMTD measurement into MPD with good level of accuracy allowing comparison between the two, for UK conditions.

$$MPD = 1.42 \cdot SMTD^{0.840} \quad (2)$$

Because, in the UK, SMTD is the standard indicator, this is the information stored in the databases of HE and therefore used in this study.

2.2 Pavement LCA

Based on the definition given by the International Organisation for Standardisation (ISO), Life-Cycle Assessment (LCA) is the "compilation and evaluation of the inputs, outputs and the potential environmental impacts of a product or system throughout its life cycle" (ISO, 2006). In practice, LCA consists of a framework of methods, tools and techniques to analyse, understand and quantify the environmental impacts that any product, process or service generates during every phase of its life, from the extraction of raw materials to disposal.

Energy use and GHG emissions are actually significant for each phase of the life-cycle of road pavements (Santero *et al.*, 2009; Harvey *et al.*, 2010; Wang *et al.*, 2014b; Trupia *et al.*, 2016). That includes the extraction and production of materials, construction, use, maintenance and end-of-life. For this reason, re-evaluation of the current management procedures of the road infrastructure might represent a huge opportunity to reduce costs and emissions from the road transport industry (Santero *et al.*, 2011a, 2011b, 2011c; Wang *et al.*, 2014b; Santos *et al.*, 2015a, 2015b, 2017). Because of this, some governmental institutions, road authorities, companies and organisations have started using LCA to inform their decisions in consideration of sustainability (Trupia *et al.*, 2016; Trupia, 2017). However, due to complexity of road pavements, this does not represent a simple task.

In the last years, several studies (Carlson, 2011; Santero *et al.*, 2011b, 2011c; Azari Jafari *et al.*, 2016) focused on reviewing existing pavement LCA studies trying to identify issues and research gaps within the framework. In particular, Santero *et al.*

(2011b, 2011c) reviewed 15 studies from all around the world from 1996 to 2010. In their critical review, the authors identified that:

- 9 of the considered studies included transportation as part of the material production phase;
- 11 of these studies estimated the impact of on-site equipment, but only 3 of them considered the consequences of this on traffic delay;
- only 2 studies considered the use-phase as part of the system boundaries, and none of them included the effect of RR and albedo;
- 10 studies considered maintenance, sometimes partially, and just as a simplified series of events;
- only one study considered the end-of-life phase.

Similar conclusions can be found in the study conducted by Carlson, (2011) who reviewed 11 studies conducted just in Europe from 1996 to 2011.

More recently, Azari Jafari *et al.* (2016) reviewed 12 studies completed between 2010 and 2015. In this more recent review of the pavement LCA framework the authors identified that:

- since 2011, consideration of the effects of the work-zone on traffic delay and of RR on excessive vehicle fuel consumption has increased significantly;
- regarding the effects of RR, as the excess of fuel consumption and emissions due to road surface roughness, different models are currently used that can lead to different conclusions. For this reason, further research is required in this area;
- also, regarding the effects of road roughness, future challenges and research opportunities are represented by an exploration of the influence of roughness on vehicle maintenance and repair;

- other challenges are represented by the implementation of models to estimate the effects of lighting, albedo, carbonation and earthworks. These were not considered by any of the studies analysed.

The major limitation identified by all the authors, from Carlson (2011) to Santero *et al.* (2011b, 2011c), and Azari Jafari *et al.* (2016), is the fact that none of the pavement LCA studies performed up to 2016 is really complete. Due to the diversity of the considered pavement projects, different authors apply different methodologies and consider different phases of pavement LCA in their studies and, for this reason, no detailed comparison is possible. This is a consequence of the lack of specific procedures or guidelines for applying LCA to road pavements, the lack of suitable and customisable tools and the many uncertainties that the methodology still presents (Santero *et al.*, 2011b, 2011c; Bryce, 2014; Trupia, 2017). This, together with the general opinion that sees environmental-friendly solutions to be initially expensive and not particularly cost-effective, hinders the implementation of LCA in the framework of road asset management, thus not allows engineers and road agencies to account for the costs and emissions caused by their assets and users of the infrastructure (Trupia, 2017).

The two most recent and complete guidelines specifically developed for performing LCA studies of road pavements are the UCPRC Pavement Life-Cycle Assessment Guidelines (Harvey *et al.*, 2010), and the FHWA pavement LCA framework (FHWA, 2016).

In the UCPRC Guidelines, since the system boundaries of the maintenance and construction phases are consistent, similar equipment and methods are used, with similar consequences on traffic, both phases have been included in the framework and considered in the same process box. The guidelines consider the impact of construction

regarding work-zone traffic and effects on traffic at the network level. The guidelines exclude from the system boundaries events of preservation, maintenance or rehabilitation of road pavements that address only surface treatments and do not affect the base course, the subgrade and drainage systems, considering these minor works that are negligible for the computation of costs and emissions from road pavements. The guidelines suggest considering this type of works only for comparison purposes.

On the other side, what characterises the framework proposed by the FHWA (2016) is that it includes the design process as a phase of pavement LCA. In fact, Rebitzer *et al.* (2004) argued that, if the LCA aims at reducing the impact of a product or service on costs and the environment, then the study should be carried out early enough to be able to influence the choices made during the design phase of the product. For instance, in regard to road pavements, choosing to use a certain mix design of asphalt instead of another can affect the construction phase related costs, while choosing between building a road at 2% gradient or 3% can affect both the costs related to the construction and use phase of the road.

The pavement LCA framework is dynamic and in constant evolution with continuous updates, new tools and methods that become available year by year. For example, in January 2017, the National Asphalt Pavement Association (NAPA) released the latest version of the Product Category Rules (PCR) for asphalt mixtures. Other recent research works in the area of pavement LCA focused on the impact of the use-phase of road pavements on the environment (Wang *et al.*, 2012c, 2014b; Araújo *et al.*, 2014; Bryce *et al.*, 2014a; Xu *et al.*, 2015). This includes the influence of the road surface characteristics and RR related parameters on vehicle fuel economy (Araújo *et al.*, 2014; Bryce, 2014; Bryce *et al.*, 2014b; Santos *et al.*, 2015a; Trupia *et al.*, 2016). Indeed, for road pavements, the use-phase represents the longest phase

(FHWA, 2016), lasting up to 40 years and thus it has significant environmental impact (up to 97% for high-trafficked roads (Santos *et al.*, 2015a)). For this reason, estimating the impact of RR (and in particular that of road surface characteristics) in the use phase is critical in the definition of accurate LCA studies. Trupia *et al.*, (2016) pointed out that application of existing models (e.g. Chatti *et al.*, 2012; Hammarström *et al.*, 2012) can lead to significantly diverse results in terms of use phase impacts. Trupia (2017) explored the system boundaries expansions in the pavement LCA framework with a focus to traffic delay and RR contribution to the use-phase, showing that:

- traffic delay and RR must be included in road pavement LCA studies;
- due to the lack of standardisation of the LCA framework for road pavements and limitations of the currently adopted models, existing studies draw very different conclusions.
- the author recommends that sensitivity analysis should be implemented and performed in pavement LCA studies in order to estimate the variability of results and conclusions.

This review shows that the level of uncertainty regarding the system boundaries and methodological assumptions in the framework of pavement LCA are still high. In particular, the factors that need most of the attention are: estimation and prediction of road vehicle fuel efficiency, quantification of the fuel consumption due to road surface conditions, uncertainty in the estimation of traffic growth factors, effects of pavement albedo (Noshadravan *et al.*, 2012; Bryce, 2014; Trupia, 2017), and introduction of the effects of the design phase (FHWA, 2016).

The analysis of the existing literature shows that a widely accepted methodology does not exist within the framework of road pavement LCA.

Regarding the impact of road surface characteristics on vehicle fuel consumption, a critical review of existing pavement LCA studies highlighted that further research is needed in the area, since existing models result in large dispersion, being based on experimental data, they are site-specific and cannot always be reliably extrapolated. The lack of direct measurements of the impact of road surface characteristics on fuel consumption and GHG emissions, together with the fact that existing methods and models find different answers, represent a primary reason why, in road asset management, LCA is held back. This represents a motivation for the present research. Development of a standard methodology to assess these impacts using similar assumptions and data, representative of real driving conditions, will help in reducing uncertainties in the framework, allowing comparison between alternative design and maintenance policies applied. The new methodology should be reliable, repeatable, cost-effective and able to give results representative of reality and not only of controlled conditions (e.g. steady speed, no gradient, no wind, etc.). This will support including the impact of road surface properties in the studies of pavement LCA, informing road managers about the effects that different strategies can have on direct and indirect costs for the society.

2.3 Vehicle fuel consumption

Vehicle fuel consumption, sometimes also referred to as vehicle fuel economy, represents the amount of fuel required by a vehicle to travel a unit distance or, vice versa, the distance travelled per unit of fuel consumed, and is used to evaluate vehicle performance (EPA, 2018c). This depends firstly from fuel efficiency, which represents the capability of an engine to transform the chemical energy from the fuel to torque at the wheels. In fact, although the two quantities are somewhat related one to the other, fuel efficiency differs from fuel consumption as the latter is determined

by the interaction of the vehicle with external factors such as environmental conditions, road properties driver behaviour, etc.

Over 90% of vehicles in Europe are equipped with internal combustion engines (EC, 2017b). Although advances have been made recently (EERE, 2015), these are historically quite inefficient motors that, in the best of cases, have fuel efficiency of about 30% up to 45% (Stone, 2012). Due to the inefficiency of internal combustion engines and the dependence of these on oil derived fuels the road transport sector currently represents the greatest share of GHG emissions in the EU (EEA, 2016b; EP, 2017). Recent statistics show that road transport is the only sector in the EU that did not record any significant decrease in GHG emissions since 1990 (EEA, 2016b; EP, 2017). These also show that, in regard to road transport, it is expected that by 2050 passenger and freight transport will grow by about 42% and 60% respectively (EC, 2017a). Therefore, it is expected that, without further action, the amount of GHG emissions produced by the sector will not decrease in the foreseeable future (EP, 2017). For decades, governmental authorities in Europe are trying to reduce the production of GHG emissions from road vehicles by focusing on fuel efficiency, introducing more and more stringent standards and directives for vehicle emissions and promoting research activity from manufacturers and academia for developing alternative eco-friendly power-trains (e.g. EP, 2007, 2009a, 2009b, 2018 and amends). For example, in order to reduce emissions of pollutants from the road transport industry, countries such as the UK (DEFRA *et al.*, 2017) together with France, Germany and Norway are planning to ban the sale of new fossil-fuelled cars by 2030. This will force vehicle manufacturers to invest in the production of new engine technologies among which electric engines seem to be the most promising option. However, since there are still a few barriers to overcome before electrification of the

road transport can take place (EC, 2016) lots of research worldwide is still investigating the feasibility of alternatives. Barriers to the electrification of road transport can be identified in the lack of investments from governmental authorities, lack of recharging points in private buildings, lack of a distributed grid of re-charging stations all around Europe, lack of interoperability rules across the EU, etc. These together with the facts that currently electric vehicles do not have long operation range (max 200-300 km), that price of this vehicle type is on average slightly higher than others, and that recharging a vehicle battery takes a few hours, makes switching from internal combustion engines to electric ones difficult. Because of this, vehicle fleet managers cannot justify a renovation of all their vehicles with electric models, which would represent a very important investment at high risks. Therefore, although this solution might be very effective in the long term, this may be ineffective in the short term. In fact, recent statistics show that by 2030 90% of vehicles in circulation in Europe will be still equipped with petrol or Diesel fuelled engines with projections of 86% in 2050 (EC, 2016). Although this represents a reduction from the current figures, this might not represent a significant reduction, and shows how the strategy of governmental authorities in the EU may take time before it can produce actual changes. This should raise awareness upon the need of finding alternative solutions to reduce emissions that can be sustainable and require less time for implementation. Significant effort focused on reducing fuel consumption by improving efficiency of internal combustion engines (EERE, 2015), training drivers behaviour (e.g. Figueredo et al., 2017; Figueredo, Quinlan, Mesgarpour, Garibaldi, & John, 2015; Walnum & Simonsen, 2015; Werner et al., 2013), reducing travel time through vehicles re-routing, (Fawcett *et al.*, 2000; Feillet *et al.*, 2004; Zhou *et al.*, 2016; Chen *et al.*, 2017b), reducing vehicle weight through application of new materials (Hall, 2011;

Horvath *et al.*, 2012), and improving vehicle aerodynamics (Watkins, 2005; Sudin *et al.*, 2014). However, all of these factors depend substantially on vehicle characteristics and, although this may be of quicker implementation than electrification, changes in terms of emissions of pollutants may require time and would not happen before the whole fleet of existing vehicles changes.

2.4 Principles of Vehicle Dynamics

Vehicle dynamics refers to the study of the physical phenomena and processes that take place when a vehicle is in motion. In its broadest sense, it encompasses all forms of vehicles from cars to trucks, trains, aeroplanes and ships. However, in this thesis, the focus is just on trucks and the movement of this vehicle type along the road. Trucks are characterised by four or more rubber-tired wheels and by the forces that the interaction between the tires and the road pavement produces. This contact between tires and road surface produces one of the dominant forces in the system and controls the performance of the vehicle (Gillespie, 1992). In fact, in order to understand vehicle dynamics, it is important to identify all forces that act on vehicles and define why and how these are generated and act against motion. This can be accomplished at two levels: i) empirical and ii) analytical (or mechanistic). The first requires extensive tests to understand which factors influence vehicle performance, how and in which conditions. This is extremely difficult due to the variety of vehicles, drivers, roads and situations that in reality can take place. For this reason, without a mechanistic understanding of how changes in vehicle design or properties affect performance, an experimental approach will likely fail (Gillespie, 1992). A model based on physical laws and mechanical models allows the evaluation of the role played by the vehicle properties. This provides means to identify important factors that

operate under the considered conditions providing the capability to make future estimates.

Most of existing mechanical models are usually composed by a system of algebraic or differential equations that describe the interaction of vehicles with the surrounding environment and the road. In spite the advent of computers, from the mathematical point of view this type of problems remains of difficult solution and requires time and considerable computational power for each scenario investigated. This is essentially due to the wide range of components, systems and sub-systems (and their interaction) that make a vehicle and that make comprehensive modelling virtually impossible. Nowadays, thanks to the increasing power of available desktops, this has been overcome. It is possible for engineers to build component-specific models that, integrated, allow simulation and evaluation of the system before the vehicle goes into production. This provides engineers with powerful tools that make testing and investigating the performance of vehicles through less expensive than time-consuming experimental campaigns.

According to Guzzella *et al.* (2015), the most relevant forces acting on a moving vehicle are:

a) Aerodynamic resistance (F_a): caused by the friction between the air surrounding the vehicle and by the pressure difference generated between the front and the rear of the vehicle. It is usually modelled as:

$$F_a(v,t) = \frac{1}{2} \cdot \rho_a \cdot A_f(t) \cdot c_d(v(t),...) \cdot v(t)^2 \tag{3}$$

that depends on air density, ρ_a, frontal area $A_f(t)$, and vehicle speed, $v(t)$. The drag coefficient, c_d, takes into account for the non-linearity of the problem and depends on vehicle speed, shape of the vehicle, etc.

b) Resistance of gravity (F_g), due to the mixed effect of gravity and road gradient. It resists the motion of vehicles for positive gradients (uphill) and helps the motion for negative slopes (downhill). This is usually modelled as:

$$F_g(t) = m \cdot g \sin(\theta(t)) \tag{4}$$

that depends on the mass of the vehicle, m, the gradient of the road $\theta(t)$, and the gravity, g.

c) RR (F_r), caused by the interaction of the tires with the road pavement, which can be modelled as:

$$F_r(v, p, t, \ldots) = c_r(v, p, \ldots) \cdot m \cdot g \cos(\theta(t)) \tag{5}$$

where c_r is the RR coefficient, dependent on the tire properties (e.g. texture of the tire, material, etc.), vehicle speed (v), tire pressure (p), contact area between tire and pavement and characteristics of the road surface (i.e. roughness and macro-texture).

In order that the vehicle can move it is required that the engine transmits to the tires a traction torque, which is transformed into a traction force (F_t) due to the tire-pavement interaction, equal or higher than the sum of all mentioned forces above. In consideration of this, the Equation (6) can be used to describe the longitudinal dynamics of road vehicles:

$$m\frac{d}{dt}v(t) = F_t(t) - (F_a(t) + F_g(t) + F_r(t) + F_d(t)) \tag{6}$$

where $\frac{d}{dt}v(t)$ is the derivative of velocity, v, in respect to time, t, that is the instantaneous acceleration of the vehicle and F_d is the 'disturbance force' that models all the other forces that act on the vehicle in motion that have not been taken into account explicitly in Equation (6).

2.5 Factors influencing truck fuel consumption

This study focuses on assessing the impact of road surface characteristics on truck fleet fuel consumption. For this reason, the present section focuses on identifying the key factors that influence truck and heavy-goods vehicles (HGVs) fuel consumption, analysing first the effects that vehicle related properties have on fuel consumption, and then those of road surface characteristics and other external factors. Estimates are commented based on findings of previous studies.

2.5.1 Vehicle type

Vehicle type is one of the major factors that influence vehicle fuel consumption. Each vehicle to carry out a specific task or match the requirements of a certain segment of customers within the market. For this reason, a sport-car is equipped with very powerful engine and engineers test its aerodynamics in wind tunnels while a passenger car is equipped with a small or medium size engine. That determines firstly the price of the vehicle but also its energy requirements and thus its typical fuel consumption.

Specifically, trucks are designed to carry goods for long distances. That requires this type of vehicle to be equipped with large and powerful engines, large tanks of fuel, comfortable cabins for the driver, wide and long trailers for transporting goods, high number of tall and wide tires for better transmission of the load to pavement, etc. That makes trucks complex systems that cannot be treated, defined or modelled just as “big cars”.

For example, nowadays typical average fuel consumption of a medium car on highway is ~ 6 l/100 km (adapted from EPA, 2018c), approximately 17 l/100 km for medium trucks (adapted from GOV.UK, 2017), and higher than 25 l/100 km for heavy trucks (adapted from GOV.UK, 2017) depending on the considered drive-cycle

and payload. Small differences can be seen also between similar truck models from different manufacturers due to differences in the design or engine technologies adopted. These values result from decades of continuous improvement dictated by advances in technology and improvements in the design of modern road vehicles.

2.5.2 Vehicle weight and payload

Vehicle weight is one of the most impactful factors on vehicle fuel consumption. This is because the weight of a vehicle influences the resistance produced by gravity and the RR. ENERGY.GOV (2012) shows that reducing a vehicle's weight by 10% can improve fuel consumption by 6% to 8%. This can be achieved by reducing at a minimum the use of steel and substitute that with light composite materials (more expensive) or plastic parts whether possible (if that does not compromise the safety of the vehicle).

As the focus of this study is the analysis of the fuel consumption of trucks, it is important to highlight that the weight of this type of vehicle can vary from empty to full capacity and that this is restricted by road authorities' regulations to 44 tonnes in the UK (DfT, 2018c; GOV.UK, 2018). The weight of goods transported by trucks (or good vehicles in general) is usually referred as 'payload'. That recalls the fact that companies make customers pay for transporting their goods based on value, volume and weight of the goods transported. Due to the high amount of goods transported by trucks, it becomes immediately clear that variations of the payload can affect fuel consumption significantly. However, curbing on the transportation of goods cannot be considered as an option and alternatives reduce to the design of lighter vehicles, thanks to the use of innovative materials or optimisation of the distribution of goods for each single truck. In particular, the latter is a technique currently used by fleet managers,

who try to maximise profits through maximization of the payload transported by each single vehicle in consideration and respect to the law.

Because of the different weight and duty of each vehicle, many countries use different classification methods.

For example, in Europe, trucks are classified based on vehicle weight (Gross Vehicle Weight, GVW) and licence required to drive them (EP, 2006). That makes for trucks four different classes (see Table 3).

Table 3 - Classification of trucks as specified by the EU (EP, 2006).

EU class (and license)	Weight limit (GVW, kg)
C1	3,500-7,500
C1E	3,500-12,000
C	> 3,500 (trailer ≤ 750 kg)
CE	> 3,500 (trailer > 750 kg)

In US, the Federal Highway Administration (FHWA, 2014; AFDC, 2018) classifies trucks based on their duty (i.e. light, medium and heavy), vehicle weight and axles configuration. Considering the different combination of these, eight different classes of truck are defined in the US (see Table 4). However, a different classification system is adopted by the U.S. Environmental Protection Agency based on vehicle weight and expected levels of emissions (EPA, 2017b) (see Table 5).

In the UK, trucks are classified based on their duty, gross vehicle weight, axle configuration and shape (DfT, 2003). Although the UK classification can sound similar to that adopted by the FHWA in US, classes of vehicle weight are different (see Table 4 and 6 for comparison). Although the UK classification system is quite detailed, the lack of information in the Microlise data would not allow a complete

identification of the considered trucks. For this reason, the present study makes use of the EPA classification system which, although, less representative of the considered case study, allows identification of the considered vehicles based on GVW only. For instance, while using the UK classification system all the considered trucks would result 'heavy good vehicles', using the US system (see Table 5) these can be distinguished in 'light', 'medium' and 'heavy'.

Table 4 - Classification of trucks in the US as specified by the FHWA (FHWA, 2014).

US truck class	Duty classification	Weight limit (GVWR, kg)
Class 1	light truck	0–2,722
Class 2	light truck	2,722–4,536
Class 3	medium truck	4,536–6,350
Class 4	medium truck	6,351–7,257
Class 5	medium truck	7,258–8,845
Class 6	medium truck	8,846–11,793
Class 7	heavy truck	11,794–14,969
Class 8	heavy truck	> 14,969

Table 5 - Road vehicle classification by vehicle weight as suggested by EPA (adapted from (EPA, 2017b)).

<table>
<tr><th colspan="11">Gross Vehicle Weight Rating (kg)</th></tr>
<tr><td></td><td colspan="10">2,722 3,856 4,536 6,351 7,257 8,845 11,793 14,969 27,216</td></tr>
<tr><td rowspan="4">Federal</td><td colspan="2">LDV</td><td>MDPV</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="2">LDT</td><td colspan="8">HDV/HDE</td></tr>
<tr><td>LLDT</td><td>HLDT</td><td colspan="4">LHDDE</td><td colspan="2">MHDDE</td><td colspan="2">HHDDE/Urban Buses</td></tr>
<tr><td>LDT 1 & 2</td><td>LDT 3 & 4</td><td>HDV 2b</td><td>HDV3</td><td>HDV4</td><td>HDV5</td><td>HDV6</td><td>HDV7</td><td>HDV8a</td><td>HDV8b</td></tr>
</table>

Table 6 - Classification of trucks in the UK (adapted from (DfT, 2003)).

Recommended description		**Identifier**	**UK Max GVW allowed (t)**
Light good vehicles		2 axles	3.5
Heavy good vehicles	Smaller 2-axle trucks	2 axles	> 3.5 up to 7.5
	Bigger 2-axle trucks	2 axles	> 7.5 up to 18
	Multi-axle trucks	3 axles rigid	25, 26*
		3 axles artic.	26
		4 axles rigid	30, 32*
		4 axles artic.	36, 38*
		Vehicle and draw-bar trailer 4-axles	30, 36*
		5 axles artic.	40
		Vehicle and draw-bar trailer 5-axles	40**
		6 axles artic. 1	41*
		6 axles draw-bar 1	41* and **
		5 or 6 axles artic.	44* and ***
		6 axles draw-bar 2	44*, ** and ***
		6 axles artic. 2	44*
		6 axles draw-bar 3	44 and ***

* If the driving axle, if it is not a steering axle, has twin tires and road friendly suspension, or each driving axle is fitted with twin tires and the maximum weight for each axle does not exceed 8.55 tonnes;
** Distance between the rear axle of the motor vehicle and the front axle of the trailer is not less than 3 metres;
*** If the vehicle is being used for combined transport.

Considering the above classification systems, in the past, different fuel consumption models have been developed for different vehicle classes. That simplifies the modelling process but, in general, can bring also significant discrepancies between

actual measurements (from a specific vehicle) and estimates from models (calibrated for a whole class of vehicles).

2.5.3 Engine

The engine is the system that allows transformation of the chemical energy stored in fuel into mechanical energy. That allows transmission of a torque to the wheels of road vehicles and thus allows road vehicles to move.

Currently, various types of engine are available, but, in regard to road vehicles, internal combustion engines are the most popular being installed on over 90% of vehicles in circulation in Europe (EC, 2017b). For a typical internal combustion engine over 60% of fuel is transformed into waste heat, used to pump air in and out from the engine and to overcome internal frictions (Stone, 2012). Also, it is estimated that about 18% of the energy generated by fuel combustion is actually transferred to the driveline. That means that only 12.6% of fuel is finally used to generate the force that help a vehicle to overcome the resistance to motion produced by vehicle weight, aerodynamic drag and RR together. That makes this type of engine quite inefficient but, at the same time, they represent (today) the only reliable alternative for vehicles that requires travelling long distances as in the case of buses or trucks.

2.5.4 Fuel type

The vast majority of trucks and HGVs in Europe are equipped with Diesel fuelled internal combustion engines, which tend to be more efficient than the Otto cycle engines (typical of petrol engines). Reason for this is that Diesel has slower combustion properties as compared to petrol which allows for higher torque at low

speeds (Stone, 2012). Alternative fuels such as LPG, Methane, Hydrogen and electricity recently found more and more space in the automotive industry.

Use of different fuel types typically lead to different fuel efficiencies (e.g. Bielaczyc *et al.*, 2015). That is because different fuels require different engine's technologies and because of the characteristic physical and chemical limitations of different fuel types (Stone, 2012).

Alternatives to petrol and Diesel have been studied and yet currently subject of research. Examples are electric vehicles which have been proven as cleaner solutions compared to petrol and Diesel vehicles, sometimes even more cost-effective solutions than traditional fuels (e.g. Bielaczyc *et al.*, 2015), which could help reducing the emission of pollutants from the road transport industry, significantly (e.g. Moro *et al.*, 2017). However, for the time being their technologies are not ready to be fully implemented and most of newly installed engines adopt hybrid solutions that see traditional internal combustion engines at support of alternatively fuelled ones. However, hybridization recently demonstrated to be able reduce also vehicle operating costs as it can help in improving fuel efficiency by up to 35% (EPA, 2018b). For this reason, the future of road vehicle fuels seems to be electricity, which could help reducing the emissions of pollutants to nearly-zero (EC, 2011; United States Enviromental Protection Agency, 2011; EEA, 2016a). However, especially for trucks and HGV this technology is not ready for deployment yet. In fact, one of the major issues with electric vehicles is their autonomy, often below 200 km. That may not be enough even for using a vehicle for just one single day, and the charging process is slow (up to a full day for a truck). Although in the medium-long terms electricity seems to be set for becoming the next fuel for cars (as it is already for medium buses), that does not seem to be the case for trucks and HGVs which may require a bit longer.

2.5.5 Aerodynamics

Aerodynamics play a very important role in vehicle dynamics and thus affects fuel consumption significantly.

Excluding thermal wastes, it is estimated that, for a 40 t truck driving at a constant speed of 80 km/h, about 40% of the energy from the engine is used to overpass the air drag (Sandberg, 2001a). For this reason, a lot of effort has been put in decreasing the effect of aerodynamic forces on the fuel consumption of trucks. However, it is not possible to reduce their size, and different approaches have been undertaken. For instance, the most critical factors identified are the tractor streamlining and management of the airflows around the tractor-to-trailer gap, under the trailer and at the rear of the trailer (NAS, 2010). That lead to the redesign of the shape of the top of vehicles, introduction of lateral bands that partially cover the wheels and other mechanical parts that help in reducing the differences in pressure between the front and rear of trucks (NAS, 2010). Implementation of these can lead to reductions in fuel consumption of up to 15% for a single truck (NAS, 2010).

More recently, platooning has been investigated as a method to reduce air drag and cut costs from the transport of freights. Platooning corresponds to driving multiple trucks very close to each other and showed to be able to reduce the fuel consumption of these from 6 up to 23 % (Zabat *et al.*, 1995), due to reductions in the aerodynamic resistance, depending on the number of vehicles forming the platoon and the distance between one truck and the other (Zabat *et al.*, 1995; Dávila *et al.*, 2010).

2.5.6 Driving context

Understanding the driving context is key for estimating road vehicles fuel consumption. There are mainly three types of driving context that can be considered: 1) urban driving, 2) extra-urban driving and 3) mixed-mode. Each of them has

characteristic driving patterns with significant influence on fuel consumption. For example, in urban context driving is characterised by low speed, frequent start-and-stop, acceleration-deceleration and roads in poor conditions, which can affect the fuel consumption of road vehicles negatively (Ferreira, 1982). On the other hand, on motorways these phenomena rarely happen. In this case, drivers tend to drive at relatively high and constant speed and accelerate just from time to time for overpassing a vehicle. In addition, on motorways usually the engine is already warmed-up while in urban contexts it is common to start a journey with cold-engine, which does not allow the motor to perform at its maximum capabilities (Stone, 2012). For these reasons, fuel consumption tends to be lower on highways than in the urban context (Ferreira, 1982). For example, Greenwood *et al.* (2007) show that for cars there is a difference of 13% fuel consumption between congested (typical of urban contexts) and uncongested (motorway) conditions. More recently, Chen *et al.* (2017a) showed that for buses the effect of congestion (more common in urban context) can increase the typical fuel consumption of buses (24 l/100 km) by 5.56%. No study specifically dedicated to the effect of congestion on truck fuel consumption has been found in the literature. This is because, trucks usually travel on motorways or major roads and not in urban areas where traffic jams are more frequent. However, due to similarities of vehicle shape and engines, it is possible to assume that the impact of traffic congestion on the fuel consumption of trucks is similar to that of buses, and that variations could be due to the effect of payload, which for trucks is often higher than for buses.

2.5.7 Vehicle speed

At high speeds a vehicle consumes more fuel, as there is a higher difference in relative speed between the air and the vehicle itself that makes the aerodynamic drag increase (Michelin, 2003; Goodyear, 2008). That produces an increase in the

power demand from the vehicle and thus an increase in fuel usage from the engine (Guzzella *et al.*, 2015).

In the past, a few studies (e.g. Zaniewski *et al.*, 1982; Chatti *et al.*, 2012) considered the effect of vehicle speed on vehicle fuel consumption. For that, researchers tested four/five vehicle types at constant speeds and analysed the difference in fuel consumption between them. In particular, Chatti *et al.* (2012) found that in the US passenger cars fuel consumption reaches a minimum at around 50 km/h on average, while this increases for lower or higher speeds. Zaniewski *et al.* (1982) published similar results. While the fact that at high speeds a vehicle consumes more fuel can be intuitive, this is not of easy interpretation for low speeds. However, that can be explained by the fact that vehicle engines are designed, and their performance optimised to operate in specific conditions. Moreover, a lower speed requires a lower gear, which reduces revolutions of the engine to transmit a higher torque to the wheels, demands a higher power from the engine, and produces higher fuel consumption (Stone, 2012).

2.5.8 Driver behaviour

The driver behaviour plays a key role when this is asked to make decisions (Evans, 1979). The more decisions the driver has to make and the more that can affect the fuel consumption of a vehicle. That happens more often in urban contexts where the surrounding environment can change sometimes suddenly (e.g. when a traffic light switches from green to red or when a child crosses the road incautiously) and the driver is asked to make decisions frequently affecting vehicles performance and fuel economy. Several studies carried out work specifically focused at estimating the effect of driver behaviour on car fuel consumption. Berry (2010) used simulation and data from vehicles tested under standard drive-cycles (FTP-72 and FTP-75, Federal Test

Procedures) to assess the aggressiveness of drivers in US. Results in this case show that use of lower accelerations and speed from drivers can help in reducing the fuel consumption of road vehicles significantly. Carrese *et al.* (2013) analysed the impact of driving behaviours on bus fuel consumption for a case study in the city of Rome. Ma *et al.* (2015) also analysed the effect of driving style on the fuel consumption of city buses finding that the acceleration process can significantly affect fuel consumption (56.5%), while they classified deceleration to have a lower impact (5.7%). Moreover, there is a significant effect of the average depth of the accelerator pedal, gear position and gear shift timing. Results showed also particular sensitivity to driving context.

More recent work analysed the effect of driver behaviour on the performance of large fleets of HGVs (Figueredo *et al.*, 2015, 2017), presenting a new framework for ranking driver behaviour based on vehicle and engine parameters from the telematics database of truck fleet managers (SAE, 2016). Results indicate that similar methods could be used to link the identified driver classes to fuel consumption allowing to continuously monitor the performance of single drivers in terms of fuel efficiency. However, Walnum *et al.* (2015) showed that, although driver behaviour has certainly a significant impact on the fuel consumption of heavy-duty trucks, sometimes the effect of the characteristics of the road infrastructure (e.g. road geometry) can be between 10 and 12 times higher than is the effect of driver behaviour.

2.5.9 Road geometry

It is possible to characterise a road from the geometrical point of view essentially by specifying three components: 1) the gradient of the road, which describes vertical variations of the road profile, 2) the crossfall, which indicates a

transversal variation of the road profile and 3) the radius of curvature, which indicates changes in the horizontal profile of the road.

Although in the past some models have been developed to implement the effect of crossfall on vehicle fuel consumption (e.g. Benbow *et al.*, 2013), these have never been validated using actual data. Therefore, due to the lack of literature in this regard, future research should focus on estimating this impact more precisely.

Other studies, specifically focused their analysis on the effect of road gradient and radius of curvature on vehicle fuel consumption (e.g. Webb, 1952; Boriboonsomsin *et al.*, 2009; Ko, 2011a; Svenson *et al.*, 2016). Road gradient seems to be, together with the vehicle weight, the most important factor that determines the fuel consumption of road vehicles (Svenson *et al.*, 2012, 2016). Using analytical procedures, Webb (1952) found that the fuel consumption of a private car driving between two points on roads at a gradient of 25% can be three times higher than that of the car driving on nearly flat roads. The author also specifies that theoretically the impact should be even higher for heavier vehicles such as buses or trucks. Boriboonsomsin and Barth (2009) instead tested a medium car on two different routes, a flat one and a hilly one (± 6%) and found out that the fuel economy of the hilly route exceeded that of the flat route by up to 20%. More recently, Svenson and Fjeld (2016) found that combination of gradient and vehicle weight can lead from 86% to 104% increase in fuel consumption between empty and fully loaded trucks (60 tonnes) driving between high- and low-class roads in Sweden.

Regarding the radius of curvature, results of the study conducted by Ko (2011, 2014, 2015) show that this can affect the fuel consumption of road vehicles significantly. In particular, Ko (2011, 2014) showed that, for a passenger car, reducing the radius of curvature of a road at a half of a minimum standard radius (AASHTO,

2011), causes a vehicle driving at 70 km/h fuel consumption to increase by approximately 10%. That increase is approximately 35% in the case of heavy trucks (Ko, 2011a). Also, results show that increasing the radius of curvature from the minimum allowed slightly decreases fuel consumption because of shorter trajectories (Ko, 2014). Because crossfall varies with changes of the radius of curvature (Highways Agency, 2002), results by Ko (2014) could be also affected by this variable, which was not considered in their study. Ko (2014, 2015) concludes also that, development of a tool able to predict the fuel consumption and emissions of road vehicles in consideration of a comprehensive pool of variables, including geometrical characteristics of the road infrastructure, avoiding complex calculation and iterative processes, would be beneficial, and help highway designers in estimating the environmental impact of geometric and non-geometric factors. However, similar models could be used also for eco-routing (e.g. Svenson and Fjeld, 2016) or for intelligent systems that control the vehicle speed of road vehicles (i.e. cruise control). In this direction, Ding and Jin (2018) recently showed that using a dynamic programming algorithm it is possible to estimate and optimise the speed profile of road vehicles and improve their fuel economy by up to 17% on curved roads.

2.5.10 Weather conditions

Another factor that can affect the fuel consumption of road vehicles and that of trucks is the weather. In particular, wind can affect the aerodynamics of road vehicles and make these consume more fuel (Haider *et al.*, 2012). Also, Stone (2012) showed that air temperature and air pressure can affect the efficiency of the vehicle engine, as these affect the efficiency of the chemical reactions that allow the engine to transform the fuel in mechanical energy. Additionally, air temperature can affect air density and air pressure. That can influence the aerodynamics of trucks and road

vehicles and thus affect fuel consumption (Sandberg, 2001a; Sudin *et al.*, 2014). Another factor that depends on weather and that recently has been found to play a significant role in the process that brings road vehicles consuming fuel is rain. Ejsmont *et al.* (2015) found that the water film generated on the road pavement by rain can affect tire RR on a trailer measurement by 30 to 50 % depending on the thickness of the water film, air temperature and vehicle speed. Moreover, a reduction in tire temperature of 6 to 7 ºC due to road wetness directly increases RR by about 10% (Ejsmont *et al.*, 2015). Ejsmont *et al.* (2018) found that temperature can also affect tire RR. In particular, three temperatures influence the behaviour of road vehicle tires and their RR: i) air temperature, ii) tire tread temperature and iii) temperature of the road pavement (Ejsmont *et al.*, 2018). The authors also found that different tires react to the effect of temperature differently (Ejsmont *et al.*, 2018). That highlights how weather conditions can significantly affect the energy demand of road vehicles and why it is important considering the effect of this on vehicle fuel consumption. In fact, one should consider that, especially at network level, weather conditions can change between hours, and different locations may be characterised by different weather conditions over the year. That may lead to differences in fuel consumption between a vehicle driven along the same road segments, at two different times over the day, or for the same vehicle driven at the same time, on similar roads, in two different locations.

2.5.11 Other factors

As seen in the previous section, fuel consumption is affected by vehicle characteristics (e.g. vehicle type, engine type and size, aerodynamic characteristics and fuel type) and external factors (e.g. driver behaviour, meteorological effects and characteristics of the road infrastructure) (e.g. Haider *et al.*, 2012). Some of these have

been listed in the previous sections, the effects that each of them has on vehicle fuel consumption was described, with particular focus on trucks. However, this list is not and cannot be exhaustive. For example, other factors that have been demonstrated to have an effect on vehicle fuel consumption are road surface characteristics (e.g. du Plessis *et al.*, 1990; Laganier *et al.*, 1990; Sandberg, 1990; Sandberg *et al.*, 2011a; Chatti *et al.*, 2012), the tire inflation pressure, wheel alignment and the type of wheels used by the vehicle (e.g. single, super-single or double) (Goodyear, 2008), which generally influence the pavement-vehicle interaction (e.g. Cebon, 1988; Cole *et al.*, 1989; Delanne, 1994; Louhghalam *et al.*, 2015; Sakhaeifar *et al.*, 2017) and RR (e.g. Descornet, 1990; Cenek, 1994; Hammarström *et al.*, 2008; Karlsson *et al.*, 2011; Sandberg, Bergiers, *et al.*, 2011; Ejsmont *et al.*, 2016, 2017). When modelling vehicle fuel consumption, controlling too many factors might become expensive and, for this reason, all of these effects are simplified by grouping them under a few coefficients, which allow to achieve the required level of accuracy for specific the specific application considered (e.g. Carlsson, 1986; Hammarström *et al.*, 1987; Kerali *et al.*, 2006; Guzzella *et al.*, 2015). For this reason, it is necessary to establish a priori the level of precision/accuracy that can be considered 'good enough' for the considered case study, bearing in mind that although the developed model performs well for the current application, this might require re-calibration or adaptation for others (e.g. Bennett and Paterson, 2000).

2.6 The tire-pavement interaction

In vehicle dynamics, the tire-pavement interaction represents the mean through which motion take place. Indeed, thanks to the friction generated between the tires and the pavement, a vehicle is able to transform the power generated by its engine into traction at the tires and finally motion. However, although the friction at the tire-

pavement interaction represents the only way a vehicle can move, it is also one of the main causes of resistance to the motion itself and is what allows brakes to stop a vehicle.

Several studies focused on the tire-pavement interaction and specifically on how that affects safety (e.g. Sandberg, 1997; Kane *et al.*, 2009; Nitsche *et al.*, 2009), noise (e.g. Sandberg, 2001b; Sakhaeifar *et al.*, 2012, 2017) and fuel consumption (e.g. Jiao, 2015; Jiao *et al.*, 2016; Hernandez *et al.*, 2017). These investigated how different tire materials and tread patterns interact with roughness and macro-texture of the road surface and materials that compound the pavement. This is important to ensure the minimum friction between the tire and the pavement that allows a vehicle to move but also stop as quickly as possible when in danger. Traditionally, this is measured using the British Pendulum test (BSI, 2006) or sideways force skid resistance survey devices (e.g. SCRIM) (Highways Agency, 2006b). More sophisticated approaches consist in the use of complex finite element models that can simulate different situation and allow optimisation of material properties, tire tread patterns and road surface characteristics (e.g. Wang *et al.*, 2010, 2012a, 2014a).

Experiments also highlighted the strong correlation between the tire-pavement interaction with noise (Sandberg, 2001b; Sandberg *et al.*, 2002). In particular, Sandberg (2001b) states that part of the vehicle noise is produced by the tire-pavement interaction and that this depends on the type of vehicle, vehicle speed, road surface characteristics and properties of the tire, concluding that further research was needed before we could have silent and safe roads. In this sense, further investigation saw a consortium of various institutions in Europe grouping for harmonization of the characterisation of skid resistance (Greene, 2014) and noise (Birkner, 2016). Birkner *et al.* (2016) showed that a correlation between the noise

produced by the tire/pavement interaction and RR (and thus fuel consumption) exists. This was in consideration of the reciprocal correlation of noise and fuel consumption with road surface properties but represents jut a hypothesis, which has not been validated yet (Birkner *et al.*, 2016).

All mentioned studies agree upon the importance of characterising correctly materials of the tire, as much as, those of the pavement materials, rolling conditions of the tire, speed, temperature, pavement stiffness, tire tread patterns, unevenness and macro-texture of the road pavement, etc. This explains how complex the tire-pavement interaction is and how many variables are involved that can modify the output of tests and experiments.

2.7 RR, pavement characteristics and fuel consumption

Based on the definition given by Michelin (2003) and ISO (2009), RR (RR) is the "energy consumed by a tire per unit of distance covered".

Generally, materials deform when subjected to the action of external forces. For instance, road vehicle tires deform and similarly so does the road pavement, due to the tire-pavement mutual interaction and vehicle weight (see Figure 4). Due to the rotation of the tires and translation of the vehicle, loading and unloading cycles are applied to both to the tire and the pavement. This generates stresses beyond the elastic range, and, consequently, some energy is dissipated in both the tire and the pavement, in the hysteretic cycle (see Figure 4). Deformations in both the tire and the road pavement, make some energy dissipated, which represent the main physical cause of RR. As an energy (expressed in Joules or Newton-metres) divided by a distance in metres (see Equation (7)), the RR is equivalent to a force that depends on factors such as the tire characteristics, stiffness of the stress absorbers of the vehicle, pavement

materials, stiffness of the pavement, air temperature, etc. and the weight of the considered vehicle:

$$RR = c_r \cdot Z \tag{7}$$

where, c_r is the RR coefficient, which summarises the effect of all the factors mentioned above, and Z is the sum of vertical forces acting on the road pavement through the contact area with the tires, which is equal to:

$$Z = m \cdot g \cdot \cos\theta \tag{8}$$

where m is the mass of the vehicle, g the acceleration of gravity and θ the inclination of the road pavement as slope.

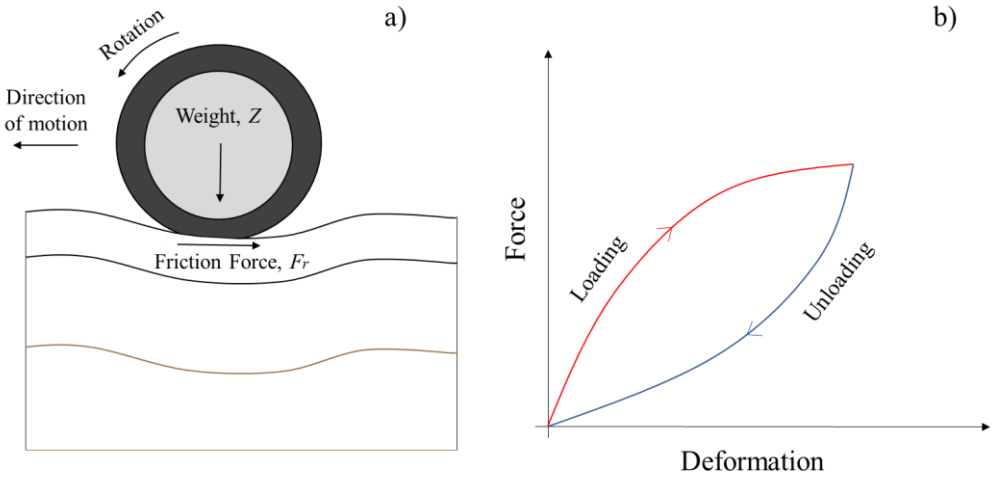


Figure 4. Representation of a) the deformation at the tire-pavement interaction (not to scale), and b) of the hysteretic loading-unloading cycle.

However, recent studies showed that RR is not just affected by the characteristics of the tire, but also by the properties of the road surface (Sandberg *et al.*, 2011a; Ejsmont *et al.*, 2015, 2017), weather conditions (Ejsmont *et al.*, 2018), and other factors such as tire pressure (Ejsmont *et al.*, 2016). Moreover, at high speed, such as on motorways, the effect of RR may be overcome by other factors such as the

aerodynamic drag (Michelin, 2003; Chatti *et al.*, 2012). For this reason, quantifying RR from experimental tests is not an easy task and estimating the effect of RR from one single fuel consumption model that accounts for all the factors involved and that affect RR may require high computational power and time. In fact, mechanical properties of materials composing vehicle suspensions, tires and road surface can change from one manufacturer to the other and vary due to the effect of temperature, depending on the load applied, and because of ageing of materials. For this reason, although in the past several methods have been proposed for measuring RR only laboratory tests have been standardised while no official guideline is available about on-road tests.

2.7.1 Pavement stiffness

Usually road pavements are classified as rigid or flexible based on their stiffness. In particular, flexible pavements usually refer to pavements consisting of one or multiple layers of asphaltic or bituminous mixtures, while rigid pavements refer to those consisting of one or multiple slabs of concrete. The main differences between rigid and flexible pavements is the flexural response of these to the vehicular load and how this is transferred to the subgrade. Because of the difference in structural response between rigid and flexible pavements, there is an open debate about the influence of pavement type on vehicle fuel consumption. In particular, it is expected that lower deflections can decrease hysteretic energy losses, making rigid pavements preferable.

Zaniewski (1989) found that, there is no difference in fuel consumption between asphalt and concrete roads for light vehicles. However, Zaniewski (1989) showed also that for a 25 t truck, there is a 20% difference in fuel consumption between concrete and asphalt pavements.

Benbow *et al.* (2007) compared a rigid pavement to a flexible pavement in laboratory tests resulting in rigid pavements reducing RR by 5.7%, which correspond to 1.14% fuel saving.

Tao (2010) used finite element modelling to analyse the impact of pavement stiffness on RR and vehicle energy consumption, concluding, together with Thom *et al.* (2010), that the most significant factor that determines variations of RR is the stiffness of the surface layer. In particular, the authors found that flexible (asphalt) pavements can absorb more energy than rigid with a factor of around 1.7 to 4.5 (Lu, 2010; Thom *et al.*, 2010). This is essentially due to changes in the mechanical properties of asphalt with ambient temperature but that can be affected also by vehicle speed.

More recently, Booshehriam *et al.* (2016) showed that by adapting the Westergaard model using historical Falling Weight Deflectometer data, it is possible to capture the behavior of both rigid and flexible pavements under dynamic loading with good accuracy. This allowed to identify that, although temperature variations may cause higher deflections on flexible pavements, which can cause higher fuel consumption, a significant portion of energy dissipation is due to the structural response of the foundation of the pavement itself (Booshehriam *et al.*, 2016). Ignoring this effect may lead to underestimation of the energy dissipation, especially for rigid pavements (Booshehriam *et al.*, 2016).

Coleri *et al.* (2017) analysed variation of the excess of fuel consumption induced by the viscous behavior of asphalt pavement materials by using a 3D viscoelastic finite element model. Based on the simulation of the response of 234 different structures, different vehicle speeds and temperatures, the authors found that for an average truck at highway speed (97 km/h) the excess of fuel consumptions

ranged between 0.1 % to 0.35 % (Coleri *et al.*, 2017). Interestingly, results were not significantly influenced from the presence of a concrete slab under the asphalt overlay and subgrade properties (Coleri *et al.*, 2017).

Hernandez *et al.* (2017) investigated the interaction between a deformable tire and elastic pavement using a 3D finite element model. In this study, the authors found that a reduction of pavement stiffness can increase the energy dissipation caused by frictional forces by 9.3% (Hernandez *et al.*, 2017). Although this might change when variables such as the viscoelasticity of the pavement are considered, the RR force was found to decrease of about 5.2% from the stiffest to the most flexible pavement (Hernandez *et al.*, 2017).

As it is possible to see, regarding the effect of pavement stiffness on vehicle fuel consumption there is not large consensus overall. Although all the studies considered find that stiffer pavements cause lesser energy dissipation and excess of fuel consumption, due to the lower deflection induced in the pavement, it is difficult to defined this effect from a quantitive perspective. Different authors use different techniques and different sets of assumptions are adopted to make conclusions, which lead to slightly different results. Due to the uncertainty regarding the quantification of this effect, although scientists agree that the structural properties of the pavement can impact the fuel consumption of road vehicle, this is not always considered in the framework of pavement LCA and more research work needs to be done before this can be considered as part of the use phase of the road pavements LCA with confidence.

2.7.2 Pavement type

Because of the difference in structural response that rigid pavements to guarantee in comparison to flexible ones, a few studies claimed that concrete could be a more sustainable material than asphalt (e.g. Zaniewski, 1989; Akbarian *et al.*, 2012;

Balzarini *et al.*, 2017). However, the existing literature does not allow a clear conclusion. In fact, the Scottish Enterprise (2003), Benbow *et al.* (2007) and Tao (2010) agree on the fact that paved roads can help in saving fuel as there is a clear difference between vehicles driving on gravel in comparison with vehicles driving on concrete or asphalt roads. However, evidences of the impact of the pavement type on fuel consumption are currently not strong enough to allow conclusions on this matter.

2.7.3 Surface conditions

Sandberg (1990), Cenek (1994), Hammarström *et al.* (2009), Kane *et al.* (2009), Haider *et al.* (2011), Sandberg *et al.* (2011a) Ejsmont *et al.* (2017) and several other in the past claimed that one factor that can affect RR, and thus the fuel consumption of road vehicles, is road surface properties, in particular unevenness and macro-texture. The first influences comfort of ride, by inducing vibrations in the wheel suspensions and the driver, that affects RR (Haider *et al.*, 2012). The second instead is important as it affects deformation of the tires, skid-resistance and noise produced by the passing of vehicles on the road surface. That induces higher RR and thus affects the fuel consumption of road vehicles (Klein *et al.*, 2004; Sandberg *et al.*, 2011a; Zöller, 2014).

In 2011, Sandberg et al. summarised and classified the techniques commonly used for measuring RR as:

a) Drum tests;
b) Trailer methods;
c) Coast-down tests;
d) Fuel consumption methods.

a) Drum tests are performed to measure RR of tires and consist of a tire held against the internal surface of a steel drum. Then a standard load is applied to the tire

and make the drum rotate. RR is deduced as the force that resist rotation of the drum due to the contact with the tire (see Figure 5). Advantage of this setup is the exclusion of various factors that influence RR. Also, this method has the advantage of being able to measure the tire related RR accurately with precise and repeatable results that allow manufacturers to compare the performance of different materials and textures of tires. However, exclusion of various factors represents also a limitation of the test, as the estimated quantity is characteristic of the specific tested tire and does not consider the interaction of this with a real road surface, the effect of aerodynamics, etc. In fact, usually the drum is made of steel that as a material cannot be compared for characteristics to asphalt mixtures. That make results good for comparison between tires as tested in similar conditions, but it may not be representative of the performance of these on roads. Today this procedure is standard [SAE J1269 (SAE, 2006), ISO 18164 and ISO 28580 (ISO, 2005, 2009)] and used by tire manufacturing companies to test and compare the performance of their products in terms of RR.



Figure 5. Example of the drum test machine for RR measurements. Different configurations exist.

b) The trailer method is characterised by one or more tires towed by a trailer (see Figure 6). A load, which can be characteristic of a car or a truck, is applied to the tested tires. In this case the torque needed to make the tires roll on the road surface at a certain speed is measured and related to RR. The method has been under development since the 1980s (Descornet, 1990) and further improved over the years (Sandberg *et al.*, 2011b; Andersen *et al.*, 2014). Currently, trailer methods are considered key for developing models of RR in dependence with road surface characteristics (Sandberg *et al.*, 2011b; Andersen *et al.*, 2014).



Figure 6. Example of a trailer for RR measurements. The TUG from Technical University of Gdańsk is represented in the figure. Different configurations exist.

However, one of the main criticisms is the poor day-to-day repeatability of results (Bergiers *et al.*, 2011). For this reason, several patented methods are currently available, and no standard is released yet. The main advantage of this method is certainly the fact that tires can be tested at any desired speed on real roads. However, in this case, the current poor repeatability of results makes this procedure unreliable for the time being. Another critic in regard to this method is that results may be typical just of vehicles driving at a constant speed and may not be representative of real

driving conditions. Indeed, the method is not designed to account for the impact of driver behaviour, traffic, and a series of other factors, which might be difficult to reproduce in an experimental set-up, and that can affect fuel consumption significantly, minimising the effect of other factors such as that of RR and related parameters.

c) Coast-down measurements consist in accelerating a vehicle until a certain speed and then let it roll freely along the road in neutral gear (Sandberg *et al.*, 2011b). This method does not allow direct measurement of RR that must be estimated as a parameter (using e.g. the least square method) from mathematical RR macro-models (Andersen *et al.*, 2014). Indeed, this method was initially deigned for assessing vehicle aerodynamic drag. However, as the method is essentially based on an application of the 2nd Newton's law, that allows to include all possible forces that act on a vehicle (see sect. 2.4) and for this reason coast-down measurements have been recently used to quantify the RR impacts on vehicle fuel economy (e.g. Karlsson et al. 2011; Hammarström et al. 2009). In fact, coast-down measurements are just a particular case of fuel consumption modelling where the developed model is used to describe the particular situation of a vehicle that from travelling at a certain speed it constantly decelerates until stops or reaches a certain other speed. The major advantage of coast-down measurements is that a real vehicle is used to run the test so estimates are characteristic of that vehicle and can be considered representative of vehicles that have similar characteristics. However, the same can be seen also as a disadvantage as results may not be representative of a wide range of vehicles, or different engines, fuels, etc. Furthermore, results may be typical of the tested vehicle only when coasting-down and not in other situations that happens under real driving conditions where different vehicles, drive-styles and drive-cycles occur.

d) Fuel consumption models are the most general way for assessing RR. Indeed, as a direct measurement of energy consumption, fuel consumption is directly affected by all factors that resist the motion of road vehicles including RR. Majority of modern fuel consumption models are based on vehicle dynamics (see sect. 2.4), can be complex and include various sub-models that describe particular situations or systems that affect vehicle fuel economy (Bennett *et al.*, 2000a; Sandberg, 2001a; Guzzella *et al.*, 2015). This allows to address the causes of excess of fuel consumption and optimisation of the design of vehicle components. In regard to RR, estimation of the excess of fuel consumption due to tire and road surface characteristics could allow engineers in optimising design, materials and maintenance strategies for these two components. The main advantage of this method is the possibility of accounting for any variable that can affect fuel consumption. However, in order to consider all possible situations that occur at network level under real driving conditions this method may require collection of a huge amount of data from tests performed running vehicles through standard drive-cycles, in coast-down mode or at constant speed. Additionally, re-calibration of the model should be performed each time new assumptions or situations are required for consideration. Because of this, the method can be quite expensive and time consuming.

For completeness, Table 7 (in the following page) summarises what each of the described methods measures, how that is done, advantages and disadvantages of the listed methods and interested parties.

Table 7 - Summary of the existing methods currently used to evaluate RR.

Method	What?	How?	Advantages	Disadvantages	Interested parties
Drum-test	RR	RR is measured in the lab by a tire rolling internally or externally to a drum made of steel.	• Stanadardised procedure; • Repeatibility; • Accuracy.	• Cannot capture the effect of the road surface.	Tires manufacturers
Trailer	RR	RR is measured as the torque (or angle from the rest position) that resists the movement of a rolling tire on a specifically designed trailer.	• Repeatibility; • Accuracy; • Measures the RR as influenced by the road pavement.	• Non-standard yet; • Not sure whether results are characteristic of real vehiclesin real driving conditions.	Road managers & tire manufacturer
Coast-down	*a* or FC	The deceleration (*a*) or extra FC of a vehicle is measured from a certain speed A to B along a road surface of known characteristics. This is then correlated to the RR, using the 2nd Newton law.	• Stanadrdised procedure; • Can be used to calculate the portion of FC influenced by the road pavement.	• Poor repeatability; • Requires controlled conditions (e.g. no gradient, no wind, etc.); • Not sure whether results are characteristic of real vehiclesin real driving conditions.	Road managers
Fuel consumption	FC	FC is measured for specific conditions along road surfaces of known characteristics, usually at constant speed, and regression analysis used to estimated the impact of the variables of interest on fuel consumption.	• Potentially able to consider any number of variables and their effect on FC; • Repeatibility; • Accuracy.	• Non-standard; • Potentially expensive; • Computationally demanding.	Road managers

The following sections summarise results of existing studies. First, an overview of results from experimental studies is provided. Then, results from studies that tackled the problem using an analytical approach are also reviewed.

2.7.3.1 Experimental approach

By measuring the fuel consumption of a car driven at constant speed along a few road segments, Sandberg (1990) found that the fuel consumption of this type of vehicle can be affected by road roughness and macro-texture by approximately 10% and 5%, respectively.

Using coast-down tests, du Plessis *et al.* (1990) found that road unevenness can affect the fuel consumption of passenger cars by up to 7% (at 100 km/h), and that of trucks or buses by up to 20% (at 80 km/h). Regarding the macro-texture, du Plessis *et al.* (1990) found that this can affect the fuel consumption of cars by approximately 7%, while it was found to be non-significant for trucks and buses.

Similarly, based on the outcome of both laboratory and in-situ tests, Laganier and Lucas (1990) stated that fuel consumption is affected by road unevenness and macro-texture by up to 7%.

Recently, the two methods that saw the largest utilisation and advances in technology to measure RR and analyse the effect of road surface characteristics on vehicle fuel consumption are the trailer method and coast-down measurements. The following two sections are dedicated to the results from studies that in the past made use of these methods to investigate the RR associated to road surface properties.

Trailer method

Recently, in order to enhance the reliability of the trailer method, Ejsmont *et al.* (2015, 2016, 2017, 2018) focused their research on assessing the impact of a few specific factors affecting the measures of the TUG trailer. In particular, the authors

worked on separating the effect of different variables trying to understand how each of them affect the measurement of the trailer. For example, Ejsmont *et al.* (2015) found that increase of RR due to heavy rain can be as high as 50% for certain tires, which is mainly due to the mixed effects of cooling of the tire and hydrodynamic turbulences generated at the front edge of the tire/road contact area. Ejsmont *et al.* (2016) found instead that tire load generally increases RR, but this depends on the tire and pavement characteristics, while increasing the inflation pressure of the tires of about of 10 kPa leads to a reduction of RR of 3.5% but, again, this depends on the tested tire. Ejsmont *et al.* (2017) analysed the effect of road texture on tire RR. For this, measurements from the TUG showed that differences in tire RR due to surface texture can reach up to 100%, which translates in differences of fuel consumption of 20% - 30%. The authors suggest that further work should work on reviewing the current definition of MPD is necessary as it does not account for the enveloping phenomenon that can influence RR (Ejsmont *et al.*, 2017). That may help in obtaining stronger correlation between macro-texture measurements and tire RR, which can help in understanding better the phenomenon and its correlation with the fuel consumption of road vehicles. Finally, Ejsmont *et al.* (2018) studied the influence of temperature on the tire RR, using data from the TUG trailer and different tires. Based on these tests, the authors concluded that temperature has a very important influence on tire RR and, for comparison, all measurements performed at temperature different from 25 ºC should be transformed to the standard one. The authors also noticed differences in RR measurements due to differences in temperature between the road pavement and the tire tread. The author suggest that in order to perform TUG measurements a warm-up period of 20 - 30 minutes should be performed in order to bring the tires to adequate temperature and improve the repeatability of the measurements. Eventually, that

should be included also in a standardized methodology of RR measurements through the trailer method that, currently, does not exist. Additionally, the trailer method does not allow direct calculation of the effect of RR and, in particular, of that caused by road surface roughness and macro-texture on fuel consumption and GHG emissions of road vehicles.

Although the trailer method could be used for evaluating the performance of the pavement in terms of RR it may not be used to estimate the consequences of a deterioration of the pavement on direct and indirect costs for the society. That could be done using coast-down measurements or modelling fuel consumption of road vehicles including the effect of road surface characteristics.

Coast-down method

After du Plessis *et al.* (1990), Cenek (1994) used coast-down measurements for investigating RR and the tire-pavement interaction. By testing a passenger car results by Cenek (1994) evidenced that there can be up to 55% increase in RR for 2-mm higher texture depth, depending on speed. Using the same method, Cenek *et al.* (1996) and Jamieson *et al.* (1997) found that pavement deflection (rigidity) seems to be, among the considered variables, the parameter with greater influence on the fuel consumption of trucks. The authors also found that, among the investigated range of wavelengths, short (0.5-5m) wavelengths seems to have higher influence on the fuel consumption of the considered trucks than medium (5 – 15 m) or long (15 – 50 m) wavelengths. However, for the case study, effect of short wavelength of surface roughness seem to be three times lower than the effect of pavement deflection. That was even more pronounced when unbound materials or unsealed roads were considered (Jamieson *et al.*, 1997).

Hammarström *et al.* (2008) investigated the effect of road surface characteristics on RR by validating VETO, a fuel consumption model developed years before by the same authors (Carlsson, 1986; Hammarström *et al.*, 1987), using coast-down measurements. A linear relationship between IRI and RR was used as a polynomial update of the model was giving just marginal, sometimes negligible, improvements. Results showed that changes in crossfall can influence fuel consumption as that affects the presence of water on the pavement and the side force. The authors also observed that coast-down measurements are probably not the best test to investigate the effects of speed and tire temperature on the driving resistance (Hammarström *et al.*, 2008).

Recently, MIRIAM - "Models for RR In Road Infrastructure Asset Management systems" (Carlson, 2011; Haider *et al.*, 2011; Sandberg *et al.*, 2011b, 2011a; Hammarström *et al.*, 2012) - a joint project between Europe and North America focused on the energy consumption associated with the tire-pavement interaction. In a first phase, MIRIAM focused on investigating road surface characteristics and their influence on RR and fuel economy (Sandberg *et al.*, 2011a). In particular, first, definitions of tires RR in regard to their interaction with road pavements, together with parameters influencing the phenomena and existing methods for measuring RR have been reviewed, evaluated and compared (Bergiers *et al.*, 2011; Sandberg *et al.*, 2011b, 2012). Results of this first phase of MIRIAM brought to the development of a model that linearly relates RR with road surface characteristics (Sandberg *et al.*, 2011a). The model was developed for three different vehicle categories: car, heavy truck and heavy articulated truck, based on coast-down measurements (Hammarström *et al.*, 2012). In a second phase, the MIRIAM model has been further refined and the correlation between RR, road surface characteristics and emission of pollutants from road vehicles

investigated (Carlson, 2011; Haider *et al.*, 2011; Sandberg *et al.*, 2011a; Hammarström *et al.*, 2012; Karlsson *et al.*, 2012). In particular, based on coast-down measurements a novel RR model has been developed for cars, rigid trucks and articulated trucks (Hammarström *et al.*, 2012). This model updated the previous used in VETO (Hammarström *et al.*, 2012). Results of the analysis performed show that contribution of roughness and macro-texture is statistically significant, particularly for heavier vehicles. However, due to the instability of results obtained for trucks, it was not possible to make precise estimates of how much that could be (Karlsson *et al.*, 2011; Hammarström *et al.*, 2012). Also, the more breaking points on a road the less the effect of road conditions, probably due to higher fuel consumption during idle time (Hammarström *et al.*, 2012). The authors also conclude that maximum effects of optimisation of MPD may lead to a reduction of fuel consumption and GHG emission in the order of 1% - 2% for cars. Similar effects may be achieved adjusting road alignment but that may require more time and higher investments.

2.7.3.2 Analytical modelling

Physical-mechanistic approach

Dynamic forces at the tires are caused by the vehicle vibrations when this is excited by the roughness of the road surface and normally occur at frequencies below 20 Hz (which corresponds to different levels of IRI and LPV depending on vehicle speed). This produces energy dissipation in the suspension dumpers, hysteretic losses in the contact patch and microslips, which contribute to the RR (Miège *et al.*, 2005). Modelling the dynamic response of vehicle suspensions in a rigorous manner is quite complicated and estimating the energy losses caused by the tire-pavement interaction as those solely due to the conditions of the road surface is not an easy task (Miège *et al.*, 2005; Yang *et al.*, 2010; Smith *et al.*, 2011).

In the past several mathematical models have been developed to understand which factors affect RR and how. Among these we can mention the Klingbeil's model (Klingbeil, 1980), or the Pacejka's transient tire model (Pacejka, 2002), but many others are available in the literature, and allow to investigate the RR under specific sets of assumptions. For example, using a physical/mechanistic approach Cebon (1985a, 1985b) found that different frequencies of vibration can excite the vehicle suspensions differently, producing different energetic losses in the contact patch and the suspension dumper. Interestingly, this may highlight that use of IRI to evaluate the fuel efficiency performance of road pavements might represent a limitation of studies conducted in the past, and of LPV, which enables to explore different wavelengths of road unevenness, may be revealed to be a more appropriate measure for the scope. In particular, at conclusion of his studies, Cebon (1985a, 1985b) found that depending on vehicle speed the suspension springs are more likely to get excited, vibrate, and produce energy losses when the frequency of vibration is close to the frequency of resonance typical of the sprung mass of the vehicle at the considered speed. This may imply that at a certain speed longer wavelengths of road unevenness (e.g. LPV at 10 or 30 metres) may produce higher energetical losses for trucks than for cars which, conversely, might be affected more by shorter wavelengths of road unevenness (e.g. LPV at 3 metres wavelengths).

However, none of the existing models is currently able to model the phenomenon accurately for a wide range of vibration frequency, which limits their applicability (Miège *et al.*, 2005).

About a decade ago, Yang *et al.* (2010) developed a novel 3D model to investigate the response of vehicle-pavement-foundation systems, which found that the vehicle-pavement coupling action cannot be neglected when vehicles run on

smooth road pavements. However, the developed model did not allow to clarify the effect of the dynamic pavement-vehicle interaction on the displacement of the pavement, which resulted of complicated reproducibility (Yang *et al.*, 2010).

Smith *et al.* (2011) investigated the effect of the variation of suspensions and tire properties on the power dissipation in automotive suspensions using a quarter- and a half-car model. Interestingly, in their study Smith *et al.* (2011) found that power dissipation is insensitive to variations of in vehicle and suspensions parameters, while it seems to depend mostly on tire related parameters. However, this would contradict the findings of Sandberg (1990) who, testing a car to estimate the direct impact of road unevenness and macrotexture on fuel consumption, made the hypothesis that his results could be affected by the stiffness and size of vehicle suspensions.

In the attempt of tackling the problem in a rigorous manner, accounting for the viscoelastic properties of materials such as asphalt, and yet reducing calculation time, without compromising on the interpretability and accuracy of the developed models, also finite element modelling was used to investigate the effect of road roughness on the pavement-vehicle interaction. For example, using a 3-D model, Wang *et al.* (2009) analysed the perpetual response of flexible pavements subject to a moving wheel load and variation of the tire-pavement contact stresses, concluding that using a non-uniform distribution of vertical contact stresses at the tire-pavement interface can help at simulating more realistic response. Wang *et al.* (2012a) confirmed that the vertical contact stresses and localised tangential contact stresses at the tire-pavement interface are non-uniform, and found that this depends on the load, inflation pressure and rolling conditions.

Khavassefat (2014), used finite element modelling to investigate the acceleration of the pavement deterioration process due to enhanced vibrations induced

by changes of the road roughness profile. In particular, Khavassefat (2014) and Khavassefat *et al.* (2015) conclude that a rougher profile can produce a vehicle dynamic load to increase by approximately 40% in comparison with the smoother profile, producing an increase of the horizontal stresses of approximately 50%, independently from the frequency ranges explored. This contradicts the findings from Cebon (1985a) who assumed variation of the induced stresses due to different vibration frequencies. In addition, a limitation of the study by Khavassefat (2014) is that the authors consider a uniform distribution of tire-pressure, and arbitrary rolling conditions, which can affect the results of the study negatively (Wang *et al.*, 2009, 2012b; Khavassefat *et al.*, 2015).

More recently, Mansura *et al.* (2017) developed a novel multiscale numerical model for the prediction of texture-related impacts on fuel consumption. In their study, Mansura *et al.* (2017, 2018b, 2018a) found that the pavement surface texture seems to be more influential on energy losses than the vertical load, the vehicle speed and the pavement structural properties, which contradicts the findings of other studies.

Contradictions in the literature between physical/mechanistic based models and experimental findings are not uncommon and highlight the fact that high complex physical based models do not always allow to interpret the effect of road surface characteristics on the energy usage of road vehicles correctly. Therefore, it is difficult to make definitive conclusions purely on the basis of physical/mechanistic models, and further research work is required before the impact of road surface characteristics on vehicle fuel consumption can be included in pavement LCA studies with confidence.

Fuel consumption modelling

In the past, various attempts at assessing what is the influence of road unevenness and macro-texture on RR and vehicle fuel consumption have been performed. Different techniques have been used such as: drum tests, trailer methods, coast-down measurements and fuel consumption modelling (Sandberg *et al.*, 2011b). However, among these, those that seem to be closer at determining the effect of road surface characteristics on vehicle fuel consumption are coast-down measurements and fuel consumption modelling.

Fuel consumption models allow estimating the energy requirements of road vehicles in consideration of variables that affect the performance of a vehicle and its fuel usage. Thus, a model able to estimate the fuel consumption of road vehicle, including the impact of road surface characteristics, would allow road managers to estimate the extra costs and emissions resulting from roads in a poor state, supporting their decisions in road pavement LCA. Additionally, a major advantage of fuel consumption models is that they can be re-adapted to the current state of technology of road vehicles by collecting new data to validate and re-calibrate the model. However, while this represents an advantage of the technique, it might also be seen as a disadvantage. In fact, different vehicles typically produce slightly different fuel consumption due to technological differences, and this means that, to collect data representative of real driving conditions at network level, a huge quantity of sensors should be installed on large quantities of vehicles worldwide, which would require unsustainable investments. This has led to the development of models based on physical and mechanistic concepts, which require fewer experimental results for calibration. For example, VETO (Carlsson, 1986; Hammarström *et al.*, 1987), initially developed by the VTI (the Swedish National Road and Transport Research Institute),

estimates fuel consumption in consideration of four main effects: road properties, vehicle technology, driver behaviour and weather conditions (Karlsson *et al.*, 2012). In particular, in regard to road surface characteristics the model considers the effects of roughness (as IRI), macro-texture (as MPD) and age of the pavement surface course (Hammarström *et al.*, 1987; Karlsson *et al.*, 2012). Most recent calibration of the model has been performed in MIRIAM by Hammarström *et al.* (2012)) using measurements from coast-down tests (Hammarström *et al.*, 2008). Based on results of this study it was concluded (Hammarström *et al.*, 2012) that the effect of IRI is dependent on the vehicle speed and at 90 km/h can vary between 4.6% to 15.1% for cars, between 7.1% and 18.4% for heavy trucks and between 7.9% and 20.3% for heavy trucks with trailer. Hammarström *et al.* (2012) also concluded that the effect of MPD at 90 km/h is between 0.8% and 2.8% for cars, 1.3% and 3.4% for rigid heavy trucks and between 1.7% and 5.3% for articulated heavy trucks. Recently, VETO has been used also in the MIRAVEC project (Haider *et al.*, 2012). MIRAVEC aimed at building on existing knowledge and models to produce an asset management tool usable in practice (Haider *et al.*, 2012, 2013). For this reason, among other tools, VETO was selected in the project to support national road agencies in estimating potential reductions of costs and emissions from the road transport industry by improving the fuel efficiency of road infrastructures (Benbow *et al.*, 2013). The model has been used for evaluating the effect of characteristics of the road infrastructure for different vehicle types in different scenarios in the UK and in Austria (Benbow *et al.*, 2013). Results of the sensitivity analysis conducted show that, in certain circumstances, higher levels of IRI and MPD can affect the fuel consumption (and thus emissions of GHG emissions) of road vehicles significantly. However, that depends on the traffic composition, traffic volume and vehicular speed, idling time,

etc. For this reason, there are cases where, due to a reduction of speed, an increase of IRI and MPD leads to an actual reduction of fuel consumption and GHG emissions so that, in these situations, the effect of road conditions is considered finally negligible (Benbow *et al.*, 2013). The model has been implemented in an Excel spreadsheet and is available for road agencies to use (Benbow *et al.*, 2013; Haider *et al.*, 2013). However, the model has never been validated using real data and presents some clear limitations (Haider *et al.*, 2013). Therefore, in practice, road managers do not use the MIRAVEC spreadsheet for estimating the vehicle operating costs associated with the condition of the road infrastructure and prefer other commercially available solutions.

The Highway Development and Management software (HDM-4), (Kerali *et al.*, 2006) is a powerful decision support tool developed by the World Bank and used by road agencies and managers of road infrastructure in various applications worldwide (e.g. Jianhua *et al.*, 2004; Thube, 2011; Valdés *et al.*, 2011; Odoki, Anyala and Bunting, 2013; Prasad, Swamy and Tiwari, 2013; Odoki, 2016). It includes dedicated tools that cover all the management processes of highway infrastructures from planning, to the evaluation of investments, work programming, etc. that helps in taking decisions at a strategic level. One of the models implemented in HDM-4 aims at estimating the fuel consumption of road vehicles. That was derived from the mechanistic based model implemented in HDM-III as described by Bennett and Greenwood (2000). However, before the model can be used with confidence it must be calibrated to local conditions (Bennett *et al.*, 2000b). Zaabar *et al.* (2010) and Chatti *et al.* (2012) performed re-calibration of the HDM-4 to estimate the effects of pavement condition on vehicle operating costs that reflect current vehicle technology in the US. Chatti *et al.* (2012) describe in details the experimental set up and procedure used by Zaabar *et al.* (2010) for calibrating the HDM-4 for the US. For instance, five

different locations were selected for the field tests. Six different vehicles, including a medium car, a SUV, a Van, a light truck and an articulated truck were chosen in order to cover all 'typical vehicles' in US. Each vehicle was run along the selected road segments at 56, 72, 88, 96 and 112 km/h. Trucks were tested fully loaded and completely unloaded. Tests were repeated in winter and summer. Wind speed was monitored, and tests performed only when this was lower than 3 m/s. Similarly, to data collected using SAE International (2016) sensors, in this case vehicle data were collected using an on-board diagnostic system (OBD) and used to calibrate the model. Calibration of the HDM-4 was performed using the procedure described in Bennett and Paterson (2000). Based on the results of an ANCOVA (analysis of covariance), Chatti *et al.* (2012) and Zaabar *et al.* (2014) concluded that, independently of the vehicle speed, an increase in IRI of 1 m/km could increase the fuel consumption of a car from 2% to 3%. For heavy trucks, instead, the fuel consumption increases by 1% to 2% at 112 km/h (highway speed in the US) and between 2% and 3% at 56 km/h (typical of urban contexts in the US). The authors also conclude that MPD does not have significant effect on the fuel consumption of road vehicles except for heavy trucks. For the latter, the authors show that an increase in MPD of 1 mm could increase fuel consumption by up to 1.5% at 88 km/h and 2% at 56 km/h. Results show also significant increase of vehicle repair and maintenance costs when IRI goes over 4 m/km. In the study, also the effect of road type was investigated. However, in this regard the authors concluded that there is a difference only for heavy trucks in Summer time at low speed. No statistically significant difference was noticed for other vehicle types or different conditions. However, in this regard, recent studies conducted in Europe showed a significant effect of both roughness and macro-texture affecting RR and thus the fuel consumption of road vehicles (e.g. Hammarström *et al.*, 2009;

Karlsson *et al.*, 2011; Sandberg, Bergiers, *et al.*, 2011; Haider *et al.*, 2013; Ejsmont *et al.*, 2017). In particular, these studies showed that, on European roads lower texture reduces pavement surface friction, which positively affects RR (reducing at the meantime fuel consumption). However, that can represent also a risk for safety and therefore low RR materials and their optimal texture depth might have to be designed also in consideration of skid-resistance.

Other existing fuel consumption models do not include possibility of estimating the effect of road surface unevenness and macro-texture (or RR) on vehicle fuel consumption (e.g. Rakha *et al.*, 2011; Chen *et al.*, 2017b; Wang *et al.*, 2017; Zhou *et al.*, 2017), and therefore cannot be used for applications in road asset management.

A step in direction to the inclusion of the effect of road characteristics on vehicle fuel consumption in road pavement LCA and road asset management decision making is presented by Svenson *et al.* (2012, 2016), who assessed the influence of road geometry and surface roughness on the fuel consumption of logging trucks in Sweden. The study considered a single 60 t truck driven in Summer time, by a seven years experienced driver, along 300 km road. That included different road classes, from motorway to forest roads and was to include a range of geometry (i.e. slope and curvature) and road conditions (i.e. roughness (IRI)) representative of Swedish conditions. Measurements included fuel consumption, vehicle speed, road gradient, curvature and IRI. Fuel consumption and vehicle speed have been recorded using the system of sensors that, as required by standards (SAE, 2016), is installed on the utilized truck. Road geometry and IRI were taken from the NVDB, the National Road Data Base from the Swedish Road Administration. Although measurements of road surface macro-texture is available in the NVDB, the study did not consider its effect on vehicle fuel consumption and focused just on IRI to characterise the road surface.

Two linear regression (LR) models, a complete and a simplified one, have been developed and performance compared to those of the ARTEMIS tool (André *et al.*, 2009; Sjödin *et al.*, 2009). Analysis of Pearson's correlation coefficients and p-values was used to identify correlations and test the significance of each of the considered variables. That highlighted the significance of correlation between IRI and fuel consumption of the considered truck, which finds confirmation from results of previous studies. Although the approach of Svenson and Fjeld (2012, 2016) seems to be interesting, its results may be representative just of the tested trucks and not of all trucks driving in Sweden. Also, future work should focus on extending the approach to different vehicle types such as articulated trucks and rigid trucks normally used for transporting other type of goods. That will help in extending the applicability of results of the study.

A summary of the results of the main studies which in the past focused on estimating the impact of RR and road surface characteristics on fuel consumption (including findings from both experimental and analytical approaches) is provided in Table 8. From the table, it is possible to see that there is significant difference between the results of one study and another. In fact, although each of the studies conducted in the past seems to have identified the possibility for road surface characteristics to affect the vehicle fuel consumption, quantification of this impact seems to be dependant on a variety of factors (sometimes typical of certain geographical areas) which also means it is difficult to make robust estimates. Additionally, results from studies conducted in the past may be representative of experimental conditions only and not of any situation typical of real driving. Furthermore, due to updates in vehicle technology and new road pavement materials, results from studies conducted in the past may no longer be representative of the effect of road surface conditions on the

fuel consumption of modern vehicles and new tests (followed by a re-calibration of existing models) should be performed to re-assess that. Therefore, further research is required in the area before the impact of road surface characteristics can be quantified and included in pavement LCA studies with confidence, and alternative approaches should be explored to try investigating how these changes under real driving conditions and at network level.

Table 8 - Summary of the major findings from studies conducted in the past regarding the impact of RR and road surface characteristics on vehicle fuel consumption.

Study	Method	Vehicle types	Range of speed tested (km/h)	Explored range of macro-texture (IRI, m/km)	Estimated effect of roughness (IRI) on RR	Estimated effect of roughness (IRI) on fuel consumption	Explored range of macro-texture (MPD, mm)	Estimated effect of macro-texture (MPD) on RR	Estimated effect of macro-texture (MPD) on fuel consumption
Sandberg (1990)	fuel consumption measurements	Car	50 60 70	0.5 to 10** (52 - 66 dB L_{SU})	-	10%	0.2 to 3** (44 - 66 dB L_{MA})	-	5%
Descornet (1990)	rolling res. measure using the trailer method	Car	from 10 to 80	0.5 to 10** (10 - 120 CP [2.5 m])	47%	9%	-	-	-
du Plessis *et al.* (1990)	fuel cons. measure using the coast-down method	Car	100 (car) 80 (truck and bus)	1.0 to 16** (15 - 200 co/km QI)	-	7%	0.2 to 3** (0.5 - 4.0 mm TD)	-	7%
		Bus			-	20%		-	0%
		Medium truck			-	20%		-	0%
Laganier *et al.* (1990)	lab tests and fuel consumption measurements	Car	40, 80, 120 lt 60, 100 fc	0.5 to 10** (1 - 10 UL)	-	5.7%	0.15 to 2.5** (0.3 - 1.3 mm SMTD)	-	5.7%
Hammarström *et al.* (2008)	rolling res. measure using the coast-down method	Car	60, 65, 70 cd 70, 80, 90 cd	0.4 to 9.3	1.8% to 6%	0.36% to 1.2%***	0.3 - 2.3	17% to 30%	3.4% to 6%***
		Van			not enough of data	not enough of data		not enough of data	not enough of data
		Heavy truck			unreliable estimates	unreliable estimates		unreliable estimates	unreliable estimates
Chatti *et al.* (2012)*	fuel consumption measurements and modelling	Car	56 72 88 96 112	0.8 to 8.5	-	2.6% to 4.8%	0.2 - 2.7	-	0%
		SUV			-	0.8% to 4.1%		-	0%
		Van			-	0.8% to 1.8%		-	0%
		Light truck			-	0.5% to 1.6%		-	0%
		Heavy truck			-	0.9% to 2.9%		-	5% to 7%
Ejsmont *et al.* (2017)	rolling res. measures using lab tests and the trailer method	Car	50 80 100	-	-	-	0.4 - 1.3	100%	20% to 30%
		Heavy truck			-	-		25% to 37.5%	5% to 7.5%***

* study conducted in the United States;
** this range is approximated as measurements were done in different units. Equation (7) and Equation (8) were used to make the transformation when possible;
*** calculated assuming that 10% increase in RR produces 2% increase in FC (Chatti *et al.*, 2012) and that this is equally distributed between IRI and MPD.

2.8 Big Data

As defined by Gartner Inc. (2018), 'Big Data' is "high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enables enhanced insight, decision making, and process automation". However, as highlighted by Ciklum (2017), often the term "Big Data" does not refer "only to large datasets, but also to the frameworks, techniques, and tools used to analyse it."

Nowadays data can be collected through any data-generating process from proper sensors to social media, public utility infrastructure, and search engines. The data is collected at specific intervals or continuously depending on the purpose of the business and analysed to offer instant insights to users and to enhance customer experience. Recent statistics (IBM, 2016) show that over 90% of existing data has been collected over the last three years. Anything can bring data and unlocking the hidden power of that is key for the business in any sector. Data are collected today in large quantities because the existing hardware allows storage and communication of this amount of information that, together with the application of advanced algorithms, produces valuable information. Analysing such an amount of data requires extensive computational power, and that would not be possible using old computers or traditional approaches. However, exponential growth of computational power has led to the ability of analysing ever growing datasets with increasingly complex algorithms. Many of the algorithms used today to analyse "Big Data" were invented decades ago. However, the capability of modern machines of applying these algorithms repeatedly and faster than ever before has significantly increased their applicability.

Based on the definition given by Samuel (1959), ML is "the field of study that focus on giving computers the ability to learn without being explicitly

programmed". Or, as defined instead by Mitchell (1997), ML is when a computer program is able "to learn from experience E with respect to some task T and some performance measure P, so that its performance on T, as measured by P, improves with experience E". These algorithms were initially born to be used in pattern recognition and with a theory that computers could learn without being programmed for solving a specific task (Samuel, 1959). However, recently these algorithms are applied to "Big Data" to evaluate if computers could highlight patterns and identify correlations through large datasets producing reliable and repeatable results. Today, effectiveness of this type of algorithms has been proven in a number of applications in a variety of different fields Examples are the self-driving car by Google (www.waymo.com), the customized online offer recommendations (Tucker, 2014), the fraud detection systems that modern banks use to protect current accounts (Maes *et al.*, 1993). These are just a few of the great number of incredible applications of ML in Science. The main reason for the application of ML to "Big Data" is that traditional statistics and analytical tools are not well suited for capturing the full value buried within large datasets. The volume of data to consider is too large for a comprehensive analysis, and the range of potential correlations and relationships between very diverse data sources is too wide for analysts to test all hypotheses and maximise information that can be extracted from the data. Instead, ML has been shown to be ideal for exploiting these opportunities with performance that thrive on growing datasets. The more information fed into ML the higher the quality of insights and results generated. This is why recently, with the growing amount of data available in any field, these techniques have been applied for solving the most disparate problems with success. For this reason, governmental authorities worldwide recently decided to make large

investment into this technology and promote the development of new ML techniques and application of these to new fields.

Due to the increasing interest in this technology and availability of specialists that are able to implement, apply and interpret ML techniques and their results, today there could be the possibility of applying these methods for assessing the effect of road surface characteristics on RR and vehicle fuel economy and emissions. In practice, this is not an easy task. Indeed, the main difficulty in this regard is the lack of large datasets. Existing data from experimental tests is limited in size, and not always representative of real driving conditions, which may not allow ML algorithms to produce robust models of the considered phenomenon.

Modern trucks are fitted with standard sensors (SAE, 2016) to continuously collect data that monitor their performance. This includes, for example, GPS location, engine revolutions and fuel consumption. The data is usually analysed by fleet managers or manufacturers aiming at detecting irregularities or uncommon patterns that indicate the need for maintenance of vehicles or training for drivers. This is to optimise the operational costs of large vehicle fleets (e.g. Roeth *et al.*, 2018).

Similarly, road managers collect road condition data to monitor the performance of roads and prioritise maintenance actions to ensure safety across the network (Highways Agency, 2008b). This include measurements of road roughness and macro-texture that are parameters that in the past have been found to be correlated with RR and fuel consumption (Beuving *et al.*, 2004). The data in this case is collected on an annual basis and stored in the databases of the road agency (Highways Agency, 2008a).

Together these data could be used for analysing the correlation between road surface characteristics and RR giving clearer insights towards the significance of these

effects on vehicle fuel consumption. In this regard, due to the amount of data available ML algorithms could represent a solution to identify complex correlations and relationships that would be difficult to observe otherwise. For instance, although at network level the amount of variables influencing fuel consumption can be large (Haider *et al.*, 2012), it is expected that, for the central limit theorem (Jaynes, 2014), on average, the difference between the fuel consumption of vehicles driving on roads in poor condition and that of vehicles driving on roads in good conditions can be assessed to be statistically significant. Testing this assumption using traditional methods would be impossible or leading to poorly repeatable and reliable results, while it could be a relatively easy task for ML algorithms.

In the past, a few studies focused on modelling fuel consumption using data from in-vehicle sensors, similar to those regulated by SAE (SAE, 2016) for trucks (Xu *et al.*, 2010; Almér, 2015; Laxhammer *et al.*, 2015; Zeng *et al.*, 2015). However, none of these considered the effect of road conditions and only one considered to apply ML to modelling the fuel consumption of large vehicle fleets (Almér, 2015; Laxhammer *et al.*, 2015), with no further application of the developed model. In their study Almér (2015) and Laxhammer *et al.* (2015) analysed 33,677 records, collected at 10-minute frequency, from 62 Scania trucks driving across Sweden between June 2013 and October 2014. The data included parameters characteristic of truck configuration, driver behaviour and road geometry, excluding characteristics of the road surface. The study also considered the effect of weather conditions on the fuel consumption of the considered fleet of vehicles with data collected from all weather stations across the territory of Sweden. Aims of the study were to tests applicability of various regression techniques, including LR and a few ML methods (i.e. decision trees, support vector regression (SVRs), random forests (RFs) and artificial neural networks (ANNs)), for

modelling fuel consumption using the large quantity of data available. Results showed that, for the considered case study, ML techniques were able to outperform LR, for modelling the fuel consumption of the considered fleet of trucks. That points out the possibility of using 'Big Data' for modelling the fuel consumption of truck fleets and evidences possible non-linearities in the correlation between the considered variables and the fuel consumption of the considered fleet of trucks and needs further investigation. That will be useful also for updating and improving the performance of existing tools that use LR for modelling the fuel consumption of road vehicles. Results of the study do not show whether the regression model used has a statistically significant impact on the predictive performance. Also, the authors conclude that is not clear whether the performance of the developed model is good enough for practical applications (Almér, 2015; Laxhammer *et al.*, 2015).

Because of the many similarities between the method applied in the present research and that of Almér (2015), differences, in terms of algorithms applied and performance, have been summarized in Table 19 sect. 5.3.

2.9 Summary and Work Plan

Analysis of the literature highlighted that there is large consensus regarding the fact that under certain conditions road roughness and macrotexture, characteristics of the road surface, can affect RR and thus the fuel consumption of road vehicles. In particular, the research conducted in the past highlighted that these can affect the fuel consumption of road vehicles by up to 5% (Beuving *et al.*, 2004), with common surface treatments able to improve on average the fuel consumption of road vehicles by up to 2% (Chatti *et al.*, 2012). This, together with a re-evaluation of the current road design and maintenance strategies could represent an opportunity for road authorities to reduce both the direct and indirect costs associated with the use phase of

the infrastructure, which could produce significant benefits for the society (Beuving *et al.*, 2004). However, the fact that these results have been obtained using experimental approaches, of doubtful representativeness of real driving and network level conditions, mines their credibility with doubts regarding the fact that under more general conditions the actual possible savings might be negligible (Willis *et al.*, 2015).

In particular, experimental set-ups considered:

a) a limited number of vehicles or specifically designed equipments;
b) tests were performed at constant speed or coasting-down;
c) tests were performed along limited segments of road of known characteristics;
d) no effect of traffic, weather, etc. was considered.

This highlighted how strict the considered set of assumptions is and how far this can be from real driving at network level conditions, which reduces the boundaries of validity of the obtained results.

Similarly, there are motivations regarding the reason why existing models are currently considered unreliable and held back from being applied in the field of pavement LCA with confidence. In fact, existing models are:

a) calibrated based on results from experimental studies and therefore cannot be applied to model conditions different from those tested;
b) need adaptation to local conditions;
c) need periodical re-calibration (and new data to be collected);
d) different models sometimes lead to completely different results.

Recent research showed that there is great potential for applying the Big Data approach to model the fuel consumption of trucks in Scandinavia (Svenson *et al.*, 2012, 2016; Almér, 2015; Laxhammer *et al.*, 2015). This would allow to obtain

estimates representative of real driving, fleet and network level conditions. However, existing studies considered a limited amount of data, from a specific truck model from a known manufacturer and this limits the applicability of the presented method significantly. Additionally, the existing work on the use of ML for fuel consumption modelling does not include the effect of road surface conditions and does not clarify whether the obtained level of performance is good enough for actual application of the developed method.

Based on the above, in order to meet the objectives of the study and fill some of the existent research gaps in the area, the work plan for the present research was defined as follows:

1. First, a small dataset from heavy trucks driving along selected roads part of the SRN of England will be considered;
2. Applicability of multiple algorithm is tested to identify which variables are to be considered by the developed models. This is to reduce variance and improve interpretability of the developed models;
3. Then, different algorithms are tested to model the fuel consumption of heavy trucks driving along the SRN of England, including comparison of the effect of the road surface conditions and their performance;
4. Finally, a parametric analysis allows identification of how each of the considered variables affects the fuel consumption of the considered fleet of trucks.

These first steps are to set up a methodology and identify the key challenges to be addressed in the present research work.

5. After this, a larger dataset is considered to include multiple truck types, a wider range of vehicle speeds, and road conditions typical of the SRN of England.
6. This is used to validate the developed methodology and improve the applicability of the developed models. This also establishes which algorithms perform best to model the fuel consumption of the considered fleets of truck and the impacts of road roughness and macrotexture;
7. On the results of the parametric analysis performed on the models developed using this second dataset, conclusions are finally drawn.

It is expected that, thanks to the nature of the data used in the present research, adaptability of the presented innovative method and repeatability of the presented results, further development of the presented method will be of interest for the scientific community, and possibly enable quantification of the effect of road roughness and macrotexture under real driving and network level conditions, more reliably, avoiding that in the future the impact of road surface properties on pavement LCA is again neglected or wrongly estimated.

3 Data

This chapter provides references, background and technical details about formats, origin, collection methods, level of accuracy and distributions of the data analysed. Sect. 3.1 and sect. 3.2 describe collection methods and data from the truck fleet management system of Microlise and road conditions from the databases of Highways England, respectively. Sect. 3.3 provides description of the weather condition data from the Met Office, and sect. 3.4 includes information about the data used to perform the presented analysis, including post-processed data.

3.1 Truck fleet management system

Modern trucks are fitted with standard sensors which help fleet operators to continuously monitor the performance of their vehicles and drivers (SAE J1939) (SAE, 2016).

The International Society of Automotive Engineers (SAE International) is a US based professional association globally active, which provides standards and professional development for engineers in various sectors of industry. The SAE J1939 (SAE, 2016) standard provides the recommended practices for communication and diagnostics among vehicle components (see sect. 3.1.1 for further details). In vehicle diagnostic systems data flows from sensors through a Controller Area Network (CAN)

bus system, and then to an Electronic Control Unit (ECU), which collects the data and transmit this to satellites and then to the database of fleet managers telematically (SAE, 2016). This is why, in the following, this type of data will be referred as 'truck telematic data' or simply 'telematic data'. The data is usually analysed by fleet managers to help their vehicle operators in reducing the costs and environmental impact from fleet operations. That influences routing and training of drivers and can help to increase efficiency, improving driver performance and safety (Roeth *et al.*, 2018).

For this study, telematic data were provided anonymised by Microlise Ltd (Microlise). Microlise is a fleet management specialised company based in Eastwood, close by Nottingham, which manages data from fleet managers and manufacturers, from over 60,000 trucks in the UK.

3.1.1 SAE J1939

The standard SAE J1939 (SAE, 2016) was originally introduced in the car and heavy-duty truck manufacturing sector, but it is now used in many other vehicle types and other sectors of industry.

SAE J1939 standard system has three main components:

i) Sensors, which are devices capable to detect and react to inputs from the environment. Specifically, truck sensors monitor physical changes (e.g. deformation, temperature, etc.) or performance (e.g. speed, torque, engine's revolutions, etc.) of vehicle components.

ii) CAN-bus, which is the system of cables or wireless connections which enable the ECU to communicate with sensors installed on the vehicle (or machine).

iii) ECU, which is a micro-computer, commonly referred also as 'black-box', which is installed in vehicles to collect data. In the past, data had to be collected manually, however, modern ECUs can pre-process, format and transmit data directly from sensors to the databases of fleet managers telematically (e.g. GSM, GPRS, etc.).

Figure 7 shows how the data is collected and transmitted to fleet managers (e.g. Microlise) for analysis.

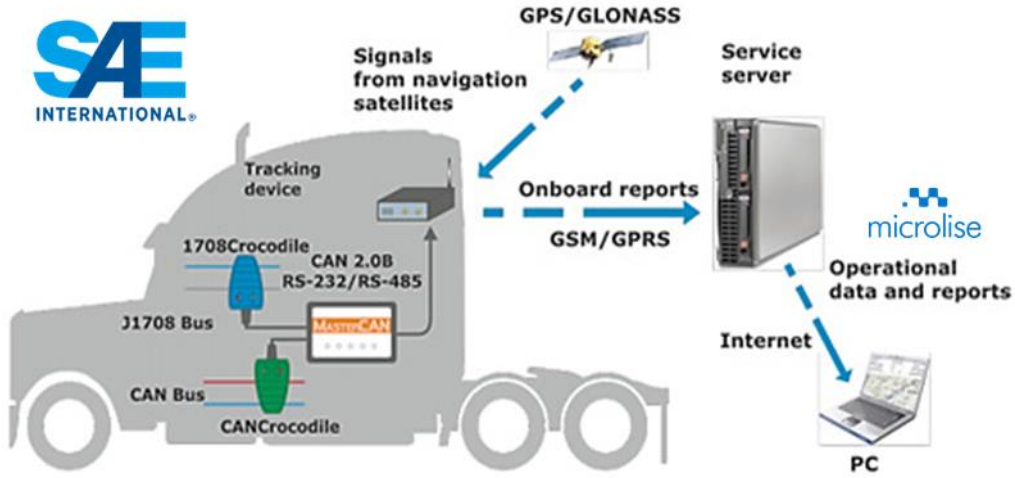


Figure 7. Schema of how the SAE J1939 system works.

SAE J1939 specifies the technical characteristics of components for vehicle diagnostics, how they interconnect, the type and frequency of data to be collected, and accuracy of the sensors to be used (SAE, 2016).

3.1.2 Telematic data

For this study, anonymised telematic data have been provided by Microlise. No information about the driver, maker, or company owning each vehicle is available. The data, stored by Microlise in Microsoft (MS) SQL databases, includes information

such as date and time of measurement, geographical position (in WGS84 reference system) and vehicle speed. The data include also vehicle performance parameters such as percent torque, revolutions per minute (rpm), use of cruise control (1 is active and 0 inactive), use of brakes and acceleration pedal (% press on pedal), cumulative travelled distance, and total fuel used among many others. Also an estimate of gross vehicle weight (± 400 kg), as calculated by a Microlise proprietary algorithm based on vehicle speed and engine parameters is part of the data.

From 2016, cumulative fuel used is measured to the nearest 0.001 litres, but previous versions of SAE J1939 required lower precisions. The data is georeferenced with the geographical position of all vehicles constantly monitored through GPS. All information is collected by default every minute or mile (approximately 1.6 km) travelled, or when triggered by other events, such as overspeed, over revs, change in gear, etc. This frequency of data collection is the most common within Microlise's customers. However, different timing or mileage can be set and adjusted to control costs or issues with storage size.

An ECU number, vehicle ID, wheel configuration, and type of engine allow identification of each vehicle. Microlise also gives an unambiguous ID number to each record; that makes each measurement unique and allows duplicates to be identified and removed from the analysis. All this information allows the whole journey of each vehicle to be traced. Although sensors and data collected are standard, Microlise method of collecting and storing information may differ from that of other fleet managers, which may lead to small discrepancies in terms of results. However, this should not affect the methodology of the present research. A full list with description of the data collected by Microlise using SAE J1939 sensors, which is stored by Microlise, is reported in Table 9.

Table 9 – Summary of the data provided by Microlise.

Parameter	Description	Resolution/Units
ID	ECU identification	-
VehicleID	Vehicle identification	-
eventID	Event identification (progressive)	-
eventDateTime	Date and time of the event	"DD/MM/YYYY HH:mm:ss"
latitude	Latitude in WGS84	0.00001 deg (~5m)
longitude	Longitude in WGS84	0.00001 deg (~5m)
courseInDegrees	Direction of truck to the geographic North	1 deg
speedInMetresPerSecond	Speed of the vehicle in m/s	0.1 m/s
eventCode	Code specifying the type of event triggered	-
canDistanceInMetres	Travelled distance as measured by the CAN sys.	0.1 m
cruiseControl	Activation of cruise control, 1 is active	Boolean: 0/1
canStatus	Checks if CAN bus is working correctly (1)	Boolean: 0/1
CANSpeedInMetresPerSecond	Speed of the vehicle in m/s as measured by the CAN sensor system	0.1 m/s
altitude	Altitude of the place where data is recorded in m	1 m
CalculatedFuelUsedInPoint00001-OfALitre	Fuel used in 0.001 l	0.001 l
AcceleratorPedalPosition-Percentage	Percent of pressure put on the accelerator pedal	0.1 %
CurrentGear	Current gear used	-
EngineLoadPercentage	Percent of load carried by the	1 %
EnginePercentageTorque	Percent of torque used by the driver	1 %
FuelSource	Fuel type (usually Diesel for trucks)	-
EngineSpeedInRPM	Revolutions per minute of the engine in rpm	1 rpm
GrossCombinationVehicleWeight-InKg	GCVW in kg (estimated)	400 kg

Not all collectable SAE J1939 data is used in this study; for additional details in regard to these and other types of data considered by SAE standards refer to SAE International (2016). For example, one parameter recently introduced by SAE in its J1939 standard collection system is the tire pressure. That has been shown to significantly affect RR and its impact on vehicle energy requirements (e.g. Ejsmont *et al.*, 2016) and, for this reason, it would be of particular interest for this study to consider this measurement as it might affect also the direct impact of road surface

characteristics on vehicle fuel consumption. However, since at the time this study was performed tire pressure was not collected by Microlise, this has not been considered in the present study.

3.1.3 DVLA data

In this study, information regarding the vehicle type, engine type, size of the engine, axle configuration have been identified to be necessary for modelling fuel consumption for each vehicle type, which is characterised by specific mean values and distribution.

DVLA (Driver and Vehicle Licensing Agency) is the agency of the UK government, which is responsible for issuing driving licenses, for collecting road taxes and for maintaining a database of drivers and vehicles in the UK.

Because vehicle characteristics cannot be measured by SAE J1939 sensors, Microlise usually buys this data from the DVLA in order to better correlate characteristics and vehicle (or driver) performance of the tested fleet. DVLA data is usually stored by Microlise in the 'Vehicle' data table and is also anonymised.

This is done so that the same vehicle ID is considered, allowing vehicle characteristics to each record to be assigned (see sect. 4.1). For this study, Microlise provided DVLA information for academic purposes within the scope of the present research. A full list with description of this information as provided by Microlise is reported in Table 10.

Table 10 - Summary of the DVLA data included by Microlise in the Vehicle data table.

Parameter	Description
registration_new_id	Anonymised registration number of the vehicle
VehicleID	Vehicle identification
AllocatedTrackingUnitID	ECU identification
Model	String containing axle configuration and code specific to vehicle and engine type
Range	String specifying the type of truck
Series	String specifying the engine type

3.2 Road condition data

3.2.1 The UK road infrastructure

With an estimated value of approximately £344 billion, the Strategic Road Network (SRN) (see Figure 8) is the most valuable infrastructure asset in England (House of Commons, 2014). The SRN includes all of England's motorways and major A-roads, with a total length of over 7000 km and is directly managed by Highways England (HE). The remaining 295,000 km of road is managed and maintained by local highway authorities.

Although the SRN represents just the 2.4% of the total road network, it carries approximately 33% of the whole motor traffic. A roads managed by local authorities, which represent 9% of the whole road network in length, carry another 32% of motor traffic. The remaining 88% of road network, made of non-principal roads, carries the remaining 35% (DfT, 2018b).

Recent studies showed how repairing a pothole can cost about £70 in London and approximately £50 elsewhere in England (House of Commons, 2014). Yet the cost of compensations requested by road users for damages arising to their vehicles due to poor road conditions is much higher and reached over £30 million in total in 2013/14

(House of Commons, 2014). Improving road conditions is critical for ensuring road safety (Viner, *et al.*, 2006) and, in relation to the extra costs that poor road conditions produce from the damage caused to vehicles and accidents, repairing roads may be a much cheaper solution to not doing so. Accordingly, the UK Government spends about £4 billion a year on roads in England (House of Commons, 2014) with most of it spent on maintaining and repairing the SRN.

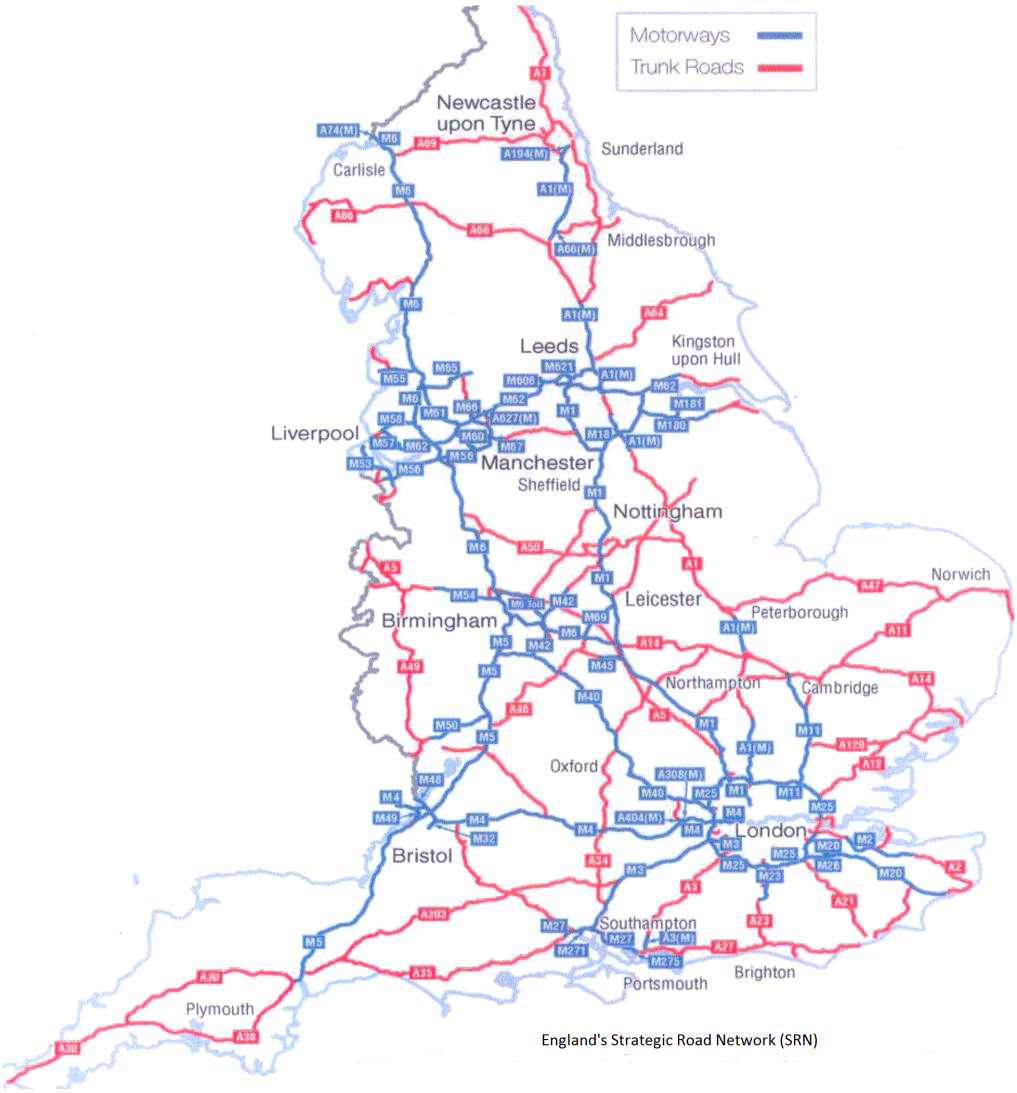


Figure 8. Map representing all motorways (in blue) and primary roads (in red) part of the SRN of England (Spencer-Bickle *et al.*, 2017).

In order to inform decision makers at strategic level upon budget allocation for road maintenance and repair, condition surveys to monitor the performance of the road infrastructure are run periodically by road managers.

3.2.2 HAPMS and TRACS data

The Highways England Pavement Management System (HAPMS) stores data from all roads that HE is responsible for. The system is maintained by HE and service providers such as TRL (the Transportation Research Laboratory). For this study, some of this data has been provided by TRL for selected roads under authorization from HE. The remaining road condition data from non-primary roads have been acquired from managers of those portions of the network. Data in the HAPMS is stored for 10 and 100 meters sections and can be extracted in various formats. For this study, data have been provided as shapefiles (.shp) for 10 metres sections. Shapefile is a popular geospatial vector data format which can spatially describe vector features (e.g. points, lines or polygons) and assign them attributes (i.e. characteristics, information, data). The shapefile data format was developed and nowadays is regulated by Esri to ensure the interoperability between different Geographical Information System (GIS) products. Thanks to its open specifications the format is also readable using R, the software used in this study for carrying out the data analysis (see Annex A). The data analysed in this study includes road geometry and road conditions. This is measured by a monitoring vehicle, called TRACS (TRAffic-speed Condition Survey), which operates at traffic speed, and is used by the managers of the infrastructure to collect data periodically across the network (Highways Agency, 2008a). This is also the reason why road geometry and road conditions data are usually referred as TRACS data.

3.2.2.1 TRACS data

In order for road managers to assess road pavement conditions and plan maintenance, an overview of the conditions and trends in condition of road pavements is required. For this reason, HE collects data about the conditions of their infrastructure which inform their decision on when is optimal to intervene.

In the UK, road condition data are collected usually on an annual basis using TRACS vehicles (Highways Agency, 2008a). The TRACS is equipped with lasers, video image collection and inertia measurement apparatus to enable surveys of road conditions to be carried out at up to 100 km/h (Highways Agency, 2008a). The fact that this vehicle can operate at traffic speed allows HE to monitor the conditions of the road infrastructure and cause minor disruption to road users. Surveys usually include lanes 1 and 2 of the main carriageway on dual carriageway roads (such as motorways) and both lanes of single carriageways (such as primary roads) (Highways Agency, 2008a). Since in the UK, the speed limit for trucks is lower than that for cars (DfT, 2018c), drivers of these vehicles must drive on lane 1 and 2 of motorways and they can use the latter only to overpass another slower vehicle (DfT, 2018c). For this reason, in order to simplify the data analysis, the assumption that trucks drive only in lane 1 was made in this study. This also makes easier to assign road conditions to each journey performed by the considered fleet of trucks. In particular, in order to investigate the correlation between road characteristics and truck fleet fuel consumption, the average road geometry and conditions from lane 1 of the travelled road segment is assigned to each record of truck data. This allows variations in fuel consumption and road properties to be compared directly and, on large datasets, it may permit patterns and trends in the data to be identified.

Table 11 – Summary of the TRACS data (from the HAPMS of HE) used in the study.

Parameter	Description	Resolution/Units
Sect_label	Code assigned to identify the considered road section	-
Survey.ID	ID of the survey	-
Survey_date	Date of the survey	'YYYY-MM-DD'
Sect_length	Length of the whole road sect.	1 m
Direction	Direction of traffic	NB/SB/WB/EB
Northing	Northing coordinate referring to British National Grid (BNG) reference system	0.01 N
Easting	Easting coordinate referring to BNG reference system	0.01 E
Xsp_code	Code assigned to identify lanes	-
S_chainage	Start measuring point	0.001 m
E_chainage	End measuring point	0.001 m
Radius	Radius of curvature	0.01 m
Abs_radius	Absolute value of Radius	0.01 m
Gradient	Road gradient value	0.1 %
Crossfall	Road crossfall value	0.1 %
Left_rut	Rutting measured on the left side of the TRACS vehicle	0.1 mm
Right_rut	Rutting measured on the right side of the TRACS vehicle	0.1 mm
Greate_rut	Greatest value of rutting measured	0.1 mm
Lef_lpv_3m	LPV at 3 m wavelength measured on the left side of the TRACS vehicle	0.01 mm^2
Rig_lpv_3m	LPV at 3 m wavelength measured on the right side of the TRACS vehicle	0.01 mm^2
Grea_lpv_3m	Greatest value of LPV at 3 m measured	0.01 mm^2
Le_lpv_10m	LPV at 10 m wavelength measured on the left side of the TRACS vehicle	0.01 mm^2
Ri_lpv_10m	LPV at 10 m wavelength measured on the right side of the TRACS vehicle	0.01 mm^2
Gre_lpv_10m	Greatest value of LPV at 10 m measured	0.01 mm^2
Le_lpv_30m	LPV at 30 m wavelength measured on the left side of the TRACS vehicle	0.01 mm^2
Ri_lpv_30m	LPV at 30 m wavelength measured on the right side of the TRACS vehicle	0.01 mm^2
Gre_lpv_30m	Greatest value of LPV at 30 m measured	0.01 mm^2
Texture	SMTD measured value	0.01 mm

Table 11 presents a description of the TRACS data retrieved from the HAPMS for this study. For the present study, also local authorities provided the same list of data normally available in the HAPMS and reported in Table 11. However,

generally, in the UK, type of data and quality of these from local authorities differs from those in the HAPMS (Ekins *et al.*, 2003).

3.3 Meteorological data

Sect. 2.4 and sect. 2.5 describe how external factors such as meteorological conditions, including wind speed and direction, air pressure, humidity and air temperature, can affect vehicle fuel consumption, affecting their aerodynamics and performance of their engines. Hence, in this study, meteorological data were considered to evaluate the impact of these factors on truck fleet fuel consumption.

The Meteorological Office (Met Office) is the part of the Department for Business, Energy and Industrial Strategy (DfBEIS), which is responsible for providing weather services to the UK. The Met Office works actively with governmental authorities in the UK (and overseas) to help them improve the resilience and effectiveness of public services. They collaborate with third parties to enable reuse of their information in commercial applications, broadcast media and academia, and they develop and provide their customers with data and weather forecasting models to support their decisions in their day to day activities. For example, one of the services provided by the Met Office to the public is the 'Data Point'. This allows anyone interested to access historical weather conditions data from the Met Office and, for academic purposes, the service is completely free.

The Met Office owns over 200 meteorological stations installed all around the UK. Stations are usually around 40 km away from each other and autonomously collect data to help experts in monitoring weather conditions and make forecasts. Since the Met Office stations can be distant from the analysed road segment, assumptions were made in order to assign meteorological data to each location. In particular, assuming that data collected by a weather station can be considered representative of

a large territory including the road on which trucks are travelling, information was taken from the nearest weather station to the truck. Full details on the proposed methodology are described in sect. 4.2.2.



Figure 9. Map of the weather stations considered in the study.

Figure 9 maps the considered weather stations across the UK territory. The data, retrieved from these and described in Table 12, is available for each of the Met Office's weather stations considered in the study. In alphabetic order, weather stations considered in the study are: Bedford, Herstmonceux West, High Bradfield, Leek Thorncliffe, High Bradfield, London St. James, Mahram, Nottingham Watnall, Ryhill, Scampton, Shawbury, Shobdon Airfield, and Wittering.

Table 12 - Summary of weather condition measurements considered in the study.

Variable name	Description	Resolution
Date	Date of collection	'DD-MM-YYYY'
Hour	Hour of collection	'HH'
Hourly Temperature	Hourly air temperature	0.1 ℃
Hourly Relative Humidity	Hourly relative humidity	0.1 %
Hourly Pressure at msl	Hourly atmospheric pressure at mean sea level (msl)	1 hPa
Hourly Rainfall Total	Total hourly rainfall	0.1 mm
Hourly Mean Windspeed	Hourly mean wind speed	1 kn
Hourly Mean Wind Direction	Hourly mean wind direction	10 deg

3.4 Data collected

In this research project two main different cases study were considered. In the first case study, impact of road surface conditions on the fuel consumption of trucks driving on motorways were evaluated. This was to evaluate feasibility of the study, with the current state of sensor, accuracy of data and computational power available. In the second phase of the study, a new case study was considered to include conditions representative of the whole SRN of England. For this reason, trucks driving on both motorways and primary roads were investigated. This allows to consider conditions typical about 65% of the road traffic in the UK (DfT, 2018b). Additionally, although out of scope, in this second case study, the impact of weather conditions measured by weather stations nearby the considered roads was analysed. In fact, although weather stations might be miles away from the exact location where trucks are driving, it is expected that data about meteorological phenomena such as atmospheric pressure, humidity, etc. can be considered relevant for large areas surrounding the considered weather stations. Therefore, that might help in explaining a larger variance, which should allow to improve accuracy of the developed models.

The truck data analysed includes records from three different type of vehicle. Light heavy-duty trucks (simply ‘light trucks’ in the following), medium heavy-duty trucks (simply ‘medium trucks’ in the following) and heavy heavy-duty trucks (simply ‘heavy trucks’ in the following).



Figure 10. Example of light truck analysed in the study.



Figure 11. Example of medium truck analysed in the study.



Figure 12. Example of heavy truck analysed in the study.

Classes are defined as suggested by the United States Environmental Protection Agency (EPA, 2017b) (see Table 5). In particular, in lieu of Table 5 it is possible to associate HDV4 and HDV5 to light trucks, HDV6 and HDV7 to medium trucks and HDV8a and HDV8b to heavy trucks. Weights refer to empty trucks (EPA, 2017b). Figure 10, 11 and 12 show examples of trucks considered in the study as classified by EPA (EPA, 2017b). However, due to the fact that classification by EPA (EPA, 2017b) is based on vehicle weight only, data used in the study may come from trucks of different brand, colour, and number of axles from those represented in the figures.

3.4.1 Initial dataset (Dataset 1)

The initial dataset (also referred as Dataset 1 in the following) was from heavy trucks and collected during a week in October 2016. This was provided

anonymised by Microlise. The data come only from heavy trucks driving along the M18 and part of the M1. Heavy trucks were selected as these are equipped with the SAE J1939 (SAE, 2016) sensors and represent the vast majority of the fleet managed by Microlise.



Figure 13. Map representing the location of data points considered in the initial phase of the study. These are 14,291 records from 1110 heavy trucks.

This dataset had a total of 1,658,532 records from 2,546 heavy trucks. However, this was reduced to just 14,291 records from 1,110 trucks when only journeys performed at relatively constant speed were considered. That is to simplify the data analysis while representing a significant amount of data in comparison to any of previous studies, which also considered fairly controlled conditions to test their

methodologies (e.g. Hammarström *et al.*, 2008, 2012; Zaabar *et al.*, 2010; Karlsson *et al.*, 2011; Chatti *et al.*, 2012). Further details about how this is defined are provided in sect. 4.2.2. For the sampled dataset, Microlise records data every minute (60 s) or mile (~ 1609 m) and measures fuel usage with 0.001 litres resolution. However, for others of Microlise's customer, this can vary.

M1 and M18 were selected as they are considered representative for conditions of the road surface of the SRN of England (see Annex B). Although M18 includes some segments of road surfaced with brushed concrete, it was not within the objectives of this study to investigate the difference in fuel consumption between roads built with concrete and those built using asphalt. Segments from the M1 were selected in order to include 'smart' sections of motorway. These are sections of road where, during specific times of the day or under specific conditions, vehicles can use hard shoulders or reduced speed limits are applied interactively in order to control the traffic flow. That is meant to reduce congestion by temporarily increasing capacity of the road. Because of that, it was assumed that considering these road segments would allow to include in the data analysis records from trucks driving on motorways at relatively low speed (lower than speed limit fixed by law (DfT, 2018c)). For this initial case study the whole length of the considered road segments, given by the sum of M18 and part of M1, is of approximately 300 km. Road condition data were provided as shape files in reference to the British National Grid (BNG) coordinates system. Because Microlise collects their data in the World Geodetic System 1984 (WGS84) coordinate system this requires transformation of one of the systems to the other. That is to allow comparison of data and investigation of correlations. For this reason, the WGS84 reference system was selected to be the unique reference system of the study and data provided by TRL transformed into this.

Figure 14 reports the data distributions from the Dataset 1. The figure shows the distributions of variables included in the developed models only. For the remaining distributions of data, please, refer to the Annex C.

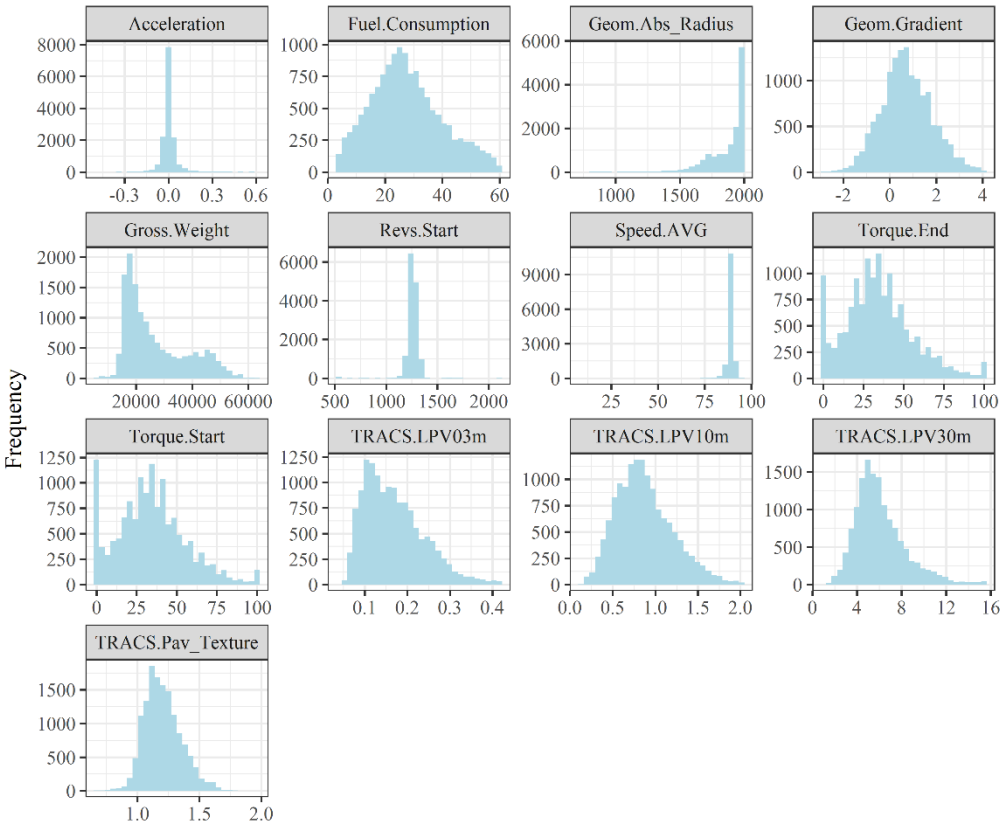


Figure 14. Distributions of data in the Dataset 1.

This data was to initiate the project, understand the type of data available, set up a methodology and test accuracy of the data in consideration to the purpose of the study. For this reason, no weather conditions were considered in this initial phase of the study and more focus was put on setting up a methodology to assess and estimate the impact of road surface conditions on truck fleet fuel consumption.

3.4.2 Final dataset (Dataset 2)

After a methodology was set up and a model to estimate the impact of road surface characteristics on truck fleet fuel consumption developed, a new dataset was requested from Microlise. That was to expand the analysis performed, extending it from just motorways to some primary roads from the SRN of England and some A roads managed by local authorities. That was to have a dataset that include road conditions representative of the whole United Kingdom. In particular, the new dataset had to satisfy the following criteria:

- Include a wider range of vehicle types. So, in this second phase of the study not only heavy trucks were considered, but also medium and light trucks (defined as specified in sect. 2.5.2 and sect. 3.4).
- Include road surface conditions characteristic of the whole SRN of England. In particular, in consideration of the damage level thresholds for road maintenance and repair defined in the DMRB (Highways Agency, 2008a), part of the M1, A47 and A49, and the whole M18, A50 and A259 have been included in the data analysis. Additionally, part of the A41 managed by the Council of Oxfordshire was also included. That corresponds to analyse approximately 900 km of road which represents about 7.5% of the whole SRN of England in length (DfT, 2018a).
- Also in this second phase of the study only journeys of trucks performed at constant speed are considered. However, in this second phase of the study roads have selected to include a range of vehicle speeds characteristic of trucks driving all across the SRN of England, which is expected to be wider than that considered in the initial phase when only motorways were analysed.

- Include data from six months to study the effect of seasonality in the data and explore the effect of air temperature on truck fleet fuel consumption. For this reason, data from January to March 2017 and from July to September 2017 have been considered.
- Include meteorological information from the Met Office (other than just air temperature) to explore possibility of using 'Big Data' to investigate the effect of weather (e.g. atmospheric pressure, rainfall, wind speed, etc.) on truck fleet fuel consumption.

All of this allowed to consider the new dataset (also referred as Dataset 2 in the following) and results of the data analysis representative of conditions typical of trucks driving across the SRN of England and local roads, which extends the applicability of the study. Therefore, the Dataset 2 includes data from three types of trucks: light, medium and heavy defined as specified in sect. 2.5.2 and sect. 3.4 based on vehicle weight (EPA, 2017b). In particular, from 11,896,669 records available in total, by applying the filters specified in section 4.2, six different subsets of data were obtained (see Table 13). For this dataset, the default frequency of data collection is set to every 1 mile (~ 1,609 m) and no restriction was set on the time spent on travelling. In the data analysis, the case of analysis of A roads only is referred also as Application 1 (or simply App 1), while the case of M and A roads together as Application 2 (or simply App 2).

Table 13 – Summary of the final dataset.

App	Roads	Truck type		
		light	**medium**	**heavy**
1	**A**	12,844	167,111	278,607
2	**M + A**	119,661	197,753	291,073

Figure 15 shows the location where data of light trucks was collected. The figure reports only data from light trucks, but identical plots can be obtained for medium and heavy trucks, which have though a much larger amount of data points recorded over the six months considered.



Figure 15. Map representing the location of data points considered in the second phase of the study. These are 119,667 records from 580 light trucks.

In comparison to the present study, experimental studies (e.g. Sandberg, 1990; Hammarström *et al.*, 2008; Chatti *et al.*, 2012) collected data at much higher frequency, which is to validate the assumption of constant speed or constant acceleration of their models. However, those data have been collected along much

shorter road segments than those analysed in the present study. In fact, as it was seen above, while in this study a total of 900 km from the SRN and local roads in England are analysed, none of previous studies took in consideration similar road stretches or amount of data (e.g. Sandberg, 1990; Hammarström *et al.*, 2008; Chatti *et al.*, 2012).



Figure 16. Example of data collected from the A41. For instance, these are from medium trucks.

Figure 16 reports instead a representation of part of the data collected along the segment of A41 managed by the Council of Oxfordshire. As specified in the caption, as an example, the figure reports data from medium trucks only. From the figure it is possible to see that the highest concentration of data from trucks driving at relatively constant speed comes from straight road segments. This highlights the fact that, in an urban environment, where turns are more frequent than on motorways or primary roads, data from vehicles driving at relatively constant speed might be very rare.

Figure 17 shows the distributions of data from light trucks from A roads only, while Figure 18 shows the distributions of data from light trucks from all roads considered in the study (M and A roads).

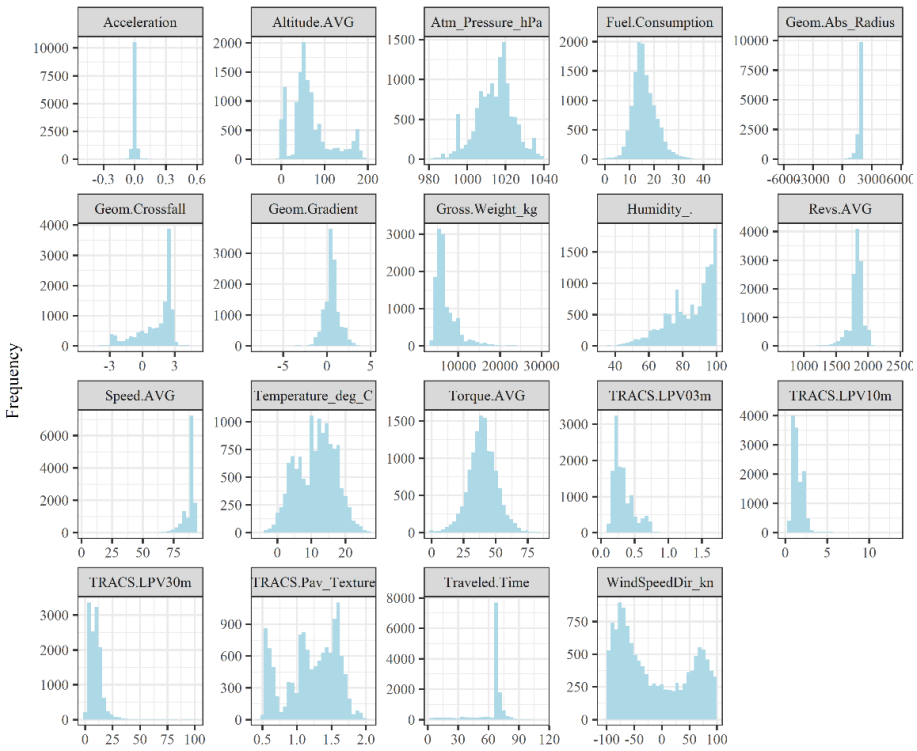


Figure 17. Distributions of data in the Dataset 2 from light trucks from A roads only.

From Figure 17 and 18, it is possible to see that the amount of data from A roads is limited compared to that from M type roads. That makes the available dataset more representative of motorways than primary roads in England. For this reason, two separate models have been developed. One considers data from trucks driving on A roads only and the other from M and A roads together. This is to evaluate if any difference can be identified and if road conditions can affect fuel consumption differently on these two types of road. The same split of data and separate analysis was performed also for medium and heavy trucks.

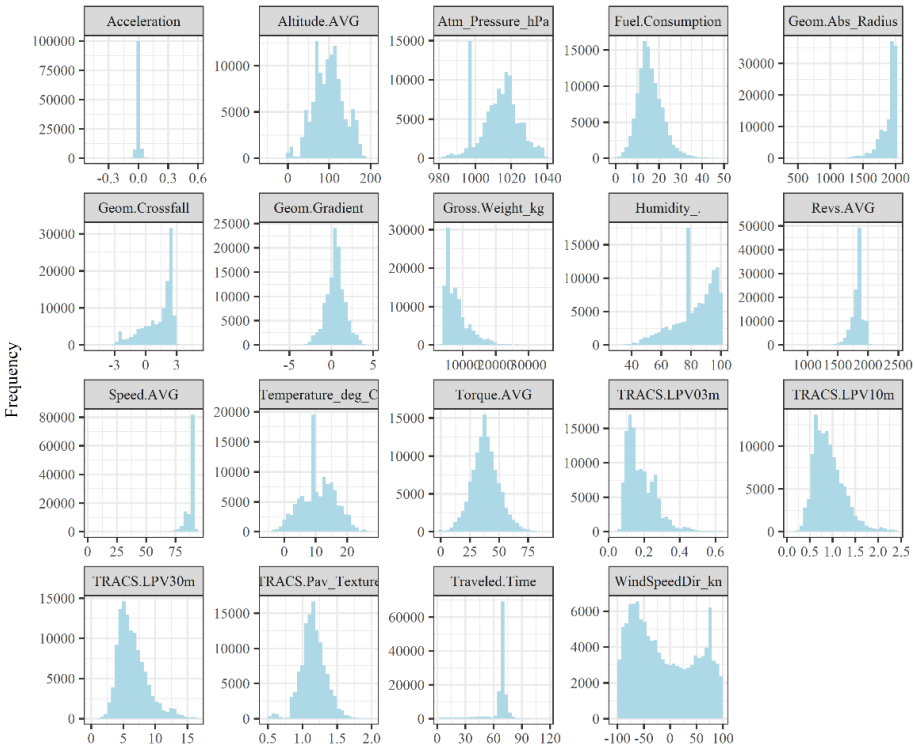


Figure 18. Distributions of data in the Dataset 2 from light trucks from all roads considered (M and A roads).

Figure 19 and 20 respectively show the distributions of data from medium trucks from A roads only, and from all roads considered in the study (M and A roads). From an analysis of Figure 19 and 20, it is possible to see that for this case study there is not much difference between the amount of data considered for A roads only and the case that considers A roads together with M type roads. This does not mean that medium trucks are more frequently used on A roads but it is a result of a data split operated on the second set of data.

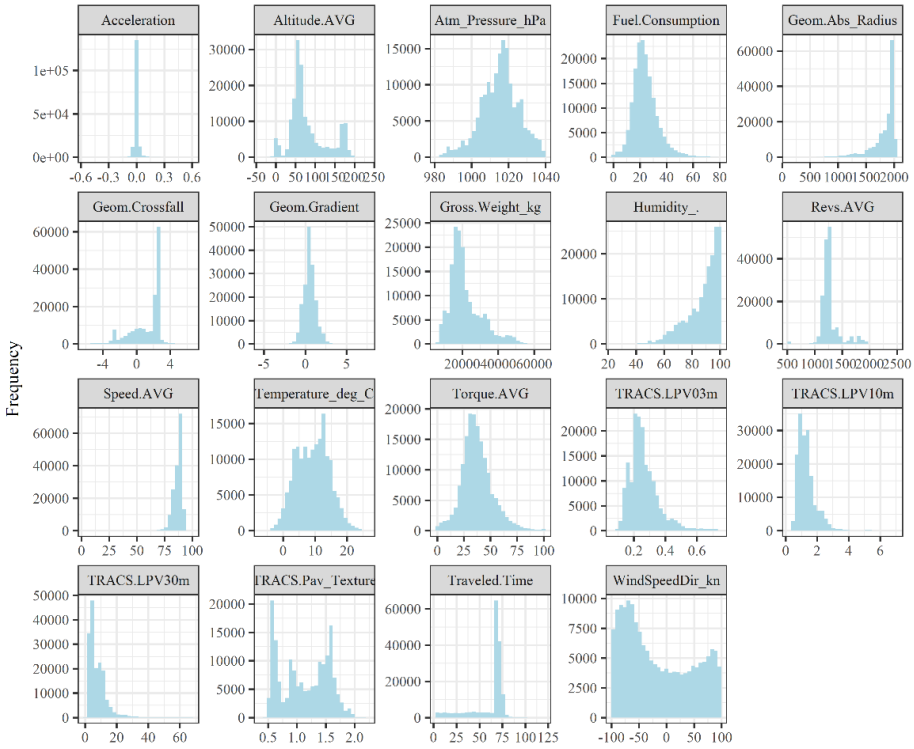


Figure 19. Distributions of data in the Dataset 2 from medium trucks from A roads only.

In fact, due to the large amount of information available for M and A roads together it has been decided to consider in the study all data from road segments in poor conditions and in particular from those road segments showing LPV 30 lower than the 5th percentile of its data distribution and higher than the 95th percentile. Then, the remaining data (between the 5th and 95th percentiles of the LPV 30 data distribution) have been split in ten bins and from each of these 15,000 records have been randomly extracted. This is to not allow the ML algorithms to put too much focus on variations of fuel consumption due to the average road conditions, but emphasize extreme values, which represent the main target of the present study. Similarly, also the data from heavy trucks where selected for the case study which considers M and A roads together.

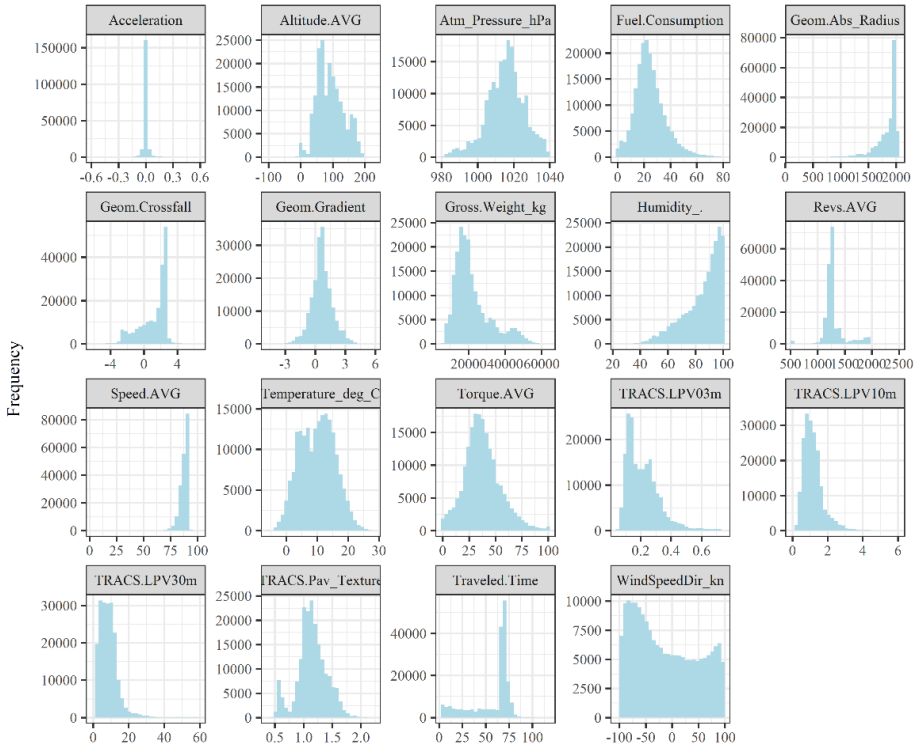


Figure 20. Distributions of data in the Dataset 2 from medium trucks from all roads considered (M and A roads).

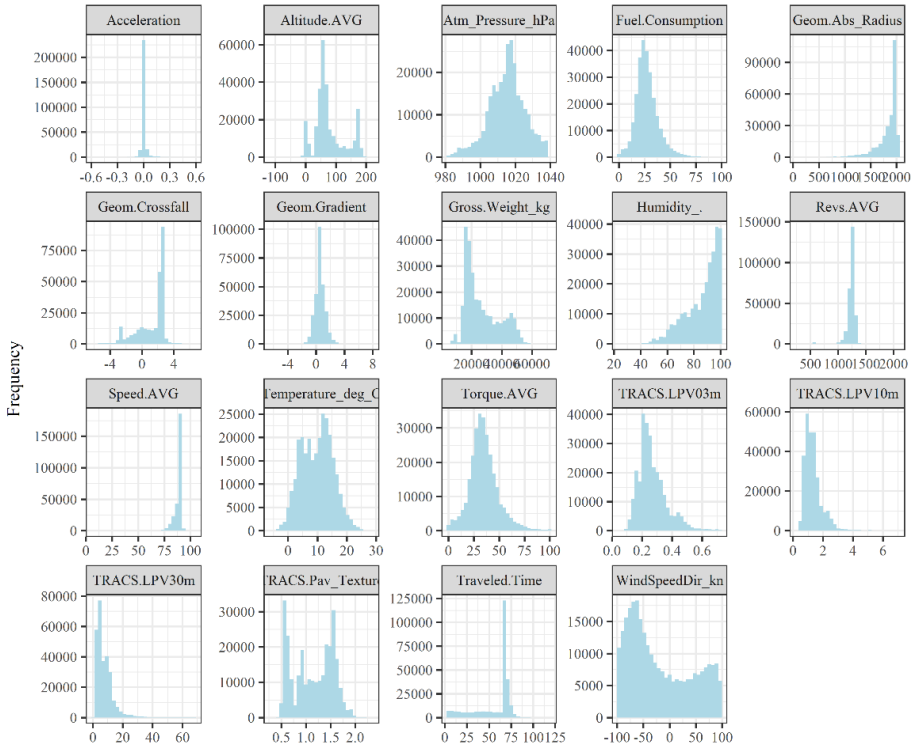


Figure 21. Distributions of the data in Dataset 2 from heavy trucks from A roads only.

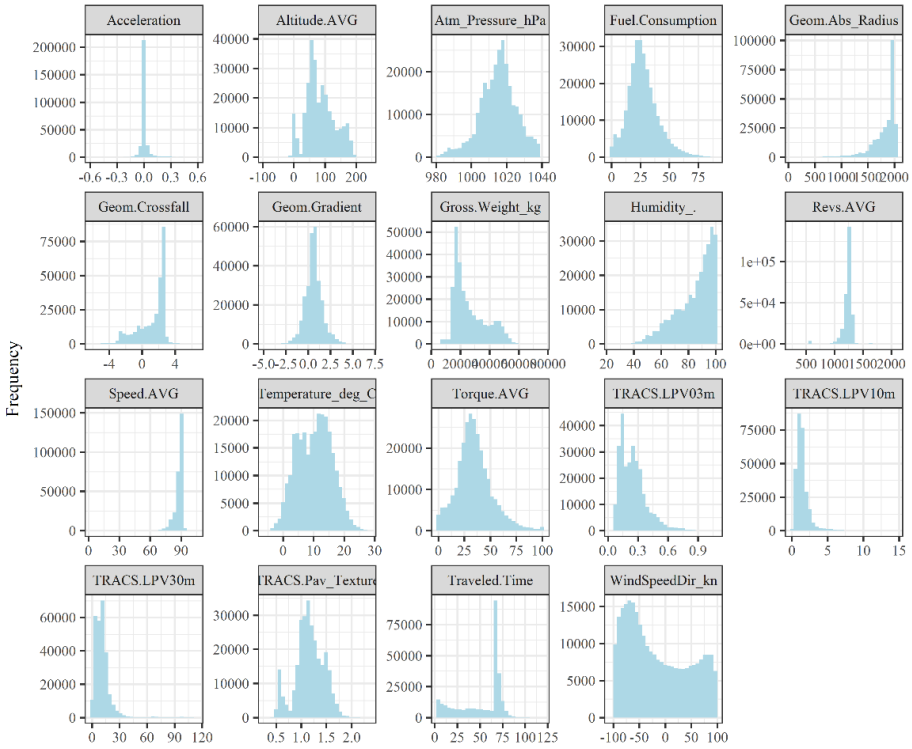


Figure 22. Distributions of the data in Dataset 2 from heavy trucks from all roads considered (M and A roads).

Figure 21 reports the distributions of data from heavy trucks from A roads only, while Figure 22 reports the distributions of data from heavy trucks from all roads considered in the study (M and A roads). Again, as explained before, similarly to what happens for medium trucks, for heavy truck only part of the data available from M and A roads is considered in the study. This allows the ML algorithm to emphasize the variation of fuel consumption due to poor road conditions.

To highlight is the fact that, staying to what reported by Benbow *et al.* (2011), over 95% of roads in the UK are characterised by road conditions within a range similar to that investigated in this second case study. This makes the analysed conditions typical of the vast majority of the SRN of England.

4 Methodology

This chapter describes in detail the process followed to clean, mine and analyse the data provided by Microlise, TRL and the Met Office. Information about the R code used to analyse the data can be found in Annex D.

4.1 Description of case study

As described in sect. 32.8 two datasets, Dataset 1 and Dataset 2, have been analysed. While Dataset 1 contains data from heavy trucks only, the Dataset 2 contains data from light, medium and heavy trucks. These were analysed separately, and for each of them different models have been developed. Additionally, in order to show how different road conditions affect fuel consumption on different vehicle types, and in order to investigate different range of vehicle speed for each set part of the Dataset 2, two different applications of the developed methodology have been performed. In particular, for each vehicle type two different models were developed by considering first data from A roads (A roads) only, and then including records from motorways (M roads). Additionally, while for light trucks all data available were considered in both applications, for medium and heavy trucks, when data from M and A roads were analysed together, only part of the data was considered. Section 3.4.2, which is dedicated to the description of the Dataset 2, provides details about how this was done.

4.2 Data cleaning

Any data mining problems require first to understand the data available. This is even more important when data is not collected directly by the analyst, or if the analyst does not have much experience with the type of data available.

The data needs to be cleaned of inaccurate or corrupted records, which could be due to non-working sensors, wrong readings, wrong transmission, wrong processing of the data that can sometimes depend also on external factors influencing the measurements. That is to reduce the amount of noise within the data and, consequently, increases confidence in the results of the analysis.

Data cleaning is the process of identifying/detecting and correcting (or removing) corrupted or inaccurate records from a certain data set. That can be done interactively using data wrangling tools, such as Tableau (see Annex A), or as batch processing through coding (see Annex D). In this study, depending on the situation, both techniques have been used.

4.2.1 Geographic data cleaning

As mentioned in sect. 3.1.2, Microlise provided two datasets, a week of data from heavy trucks collected in October 2016 and six months from light, medium and heavy trucks collected in between January - March and July - September, in 2017.

Based on a comparison of LPV and SMTD distributions, some M and A roads (see sect. 3.4.2) were selected and considered representative of the whole SRN of England (see Annex B). Therefore, only data from trucks driving along the selected road segments were initially selected from the database of Microlise. That is done in MS SQL by selecting multiple rectangles (delimited by latitude and longitude) along the considered roads. However, the method lead to the selection of datapoints in the vicinity of the road as the rectangle method does not allow to completely eliminate

data from adjacent roads. That is due to the fact that the rectangle method does not allow to completely eliminate data from adjacent roads. Additionally, the rectangle method does not allow to understand whether a truck is driving on lane one, lane two or others. However, for this study we make the assumption that when a truck drives at constant speed it does that just on lane one of motorways and primary roads. This is a reasonable assumption which allows to consider all data in the study from truck driving on lane one and allows to assign road conditions to each record from trucks.

Finally, in order to remove records from trucks driving on adjacent roads from the geographical areas of interest, data is imported in Tableau (see Annex A). Within Tableau it is possible to visualise the data on a map and clean them from all data collected outside the roads to consider, interactively. This procedure is applied to data from Microlise only. Another possibility to clean up these data would be to run a query to a map service provider (such as Google Maps) and ask for each record in the database to retrieve the address based on GPS location, and then filter out data from roads which are not of interest for the study. However, in a few tests performed this method demonstrated to be costly (due to charges that the map service provider applies after reaching a certain amount of queries) and time consuming, and therefore abandoned.

4.2.2 Completion of a dataset

In order to analyse fuel consumption a single record from the database of Microlise is not enough. In fact, SAE J1939 data give just a cumulative fuel used and not exactly the amount of fuel effectively used by a vehicle in the journey between two consecutive records. Therefore, that must be calculated.

For this reason, a code which iterates across all data in the database of Microlise and identifies journeys of trucks performed at constant speed has been

developed. For each journey, the code calculates also fuel consumption between consecutive records and extracts truck engine performance.

To do that, for each record from Microlise, the code assigns first an identifier of the road along which the truck is travelling using the GPS coordinates of each record.

After that, the code lists all records in order of Vehicle.ID and date/time and then iterates across the data. At each iteration the script checks that consecutive records come from the same truck and same ECU. For these records the code extracts information from the database of Microlise such as characteristics of the vehicle, details of the journey and calculates some parameters such as the distance travelled, the time spent travelling and fuel consumption. In particular, the latter is calculated as the ratio between the fuel used during the journey and the distance travelled. That normalises the fuel used in consideration of the length of the journey. That is then transformed to litres per 100 kilometres. Equation (9) summarises the fuel consumption calculation process:

$$FC = \frac{(FU_2 - FU_1)}{(s_2 - s_1)} \cdot 100{,}000 \qquad \text{[l/100 km]} \qquad (9)$$

where, FU_2 and FU_1 are the cumulative fuel usage at the final and initial record of the considered journey, and s_2 and s_1 are the cumulative distance travelled by the truck when the final and initial records have been triggered during the journey. Result from Equation (9) is then multiplied by 100,000 to transform micro-litres in litres and metres in 100 kilometres. That gives l/100 km which is a standard way of measuring fuel consumption at International level.

Based on GPS position and road identifier the code also assigns to each record a direction of travel (NB, SB, EB and WB for North, South, East and West bound

respectively). That is useful information for assigning road conditions to the telematic data.

Finally, the code saves all processed data in the Data.T data table which include all post-processed telematic data from trucks (from Microlise). Full details of the post-processing data is given in Table 14.

Once the Data.T table is ready, the data is then filtered in order to include just data from trucks driving at constant speed and that take no longer than the default triggering events (60 s or 1 mile for most customers of Microlise).

Then, the code iterates through the Data.T, and it considers that a truck travels at constant speed when the initial and final speed do not differ by more than ± 2.5 km/h, and if the average of speeds (initial and final) does not differ by more than ± 2.5 km/h from the mean speed (calculated as the ratio between distance and travel time). This filter reduces the number of records in Data.T to about a 10%. That should raise concerns about validity of the present study and experimental studies conducted in the past (e.g. Hammarström *et al.*, 2008; Karlsson *et al.*, 2011; Sandberg *et al.*, 2011a; Chatti *et al.*, 2012), which considered only vehicles driving at constant speed or used coast-down measurements to measure the impact of road conditions on vehicle fuel consumption because those conditions represent just a small percent of all journeys performed at network level. However, as previously mentioned, primary objective of this study is to test and introduce a new method for estimation of the impact of road surface characteristics on vehicle fuel consumption. For this reason, in the present study, focus is put on journeys performed at constant speed to evaluate whether the developed method works well by comparing results to those of previous studies.

Table 14 - Summary of data in Data.T data table.

Variable name	Description	Resolution
Vehicle.ID	Vehicle identification code	-
EventID.Start	ID code for the initial event	-
EventID.End	ID code for the final event	-
Wheel.Plan	Configuration of wheels	-
Euro.Type	Type of engine	-
Date.Start	Date/time of initial location	'dd/MM/yyyy hh:mm:ss'
Date.End	Date/time of final location	'dd/MM/yyyy hh:mm:ss'
Lat.Start	Latitude of initial event	0.0001º
Long.Start	Longitude of initial event	0.0001º
Lat.End	Latitude of final event	0.0001º
Long.End	Longitude of final event	0.0001º
Direction.Start	Heading at the initial event	1.00º
Direction.End	Heading at the final event	1.00º
Altitude.Start	Altitude at initial event	1.00 m
Altitude.End	Altitude at final event	1.00 m
Travelled.Dist	Travelled distance	0.01 m
Travelled.Time	Travelled time	0.1 s
Road.Type	Type of road	1 = A road 2 = motorway
Road.ID	ID of the considered road	-
Gross.Weight_kg	Estimated vehicle weight	400 kg
Speed.Start	Instantaneous initial speed	0.01 km/h
Speed.End	Instantaneous final speed	0.01 km/h
Acceleration	Average acceleration	0.01 m/s^2
Speed.AVG	Average speed	0.01 km/h
Travel.Direction	Road direction (North bound, etc.)	NB, SB, EB, WB
Gear.Start	Initial gear used	-
Gear.End	Final gear used	-
CruiseControl.Start	Activation of cc (1/0) init.	-
CruiseControl.End	Activation of cc (1/0) final	-
Torque.Start	Torque %, initial event	1.00 %
Torque.End	Torque %, final event	1.00 %
Torque.AVG	Mean value of torque	1.00 %
Revs.Start	Instantaneous initial rpm	1.00 rpm
Revs.End	Instantaneous final rpm	1.00 rpm
Revs.AVG	Mean value of rpm	1.00 rpm
Used.Fuel	Quantity of fuel used	0.001 l
Fuel.Consumption	Used.Fuel div. by Travelled.Distance	0.01 l/100 km
Const.Speed	1/0 message if speed is constant	-

This study assumes that, if the developed methodology and models perform well under controlled conditions, than future work could focus on generalising the obtained results for more diverse situations characteristic of network level conditions.

After Data.T has been subset in order to obtain just data from trucks driving at constant speed, road geometry and surface characteristics are appended to the data.

Table 15 - Summary of road condition measurements considered in the study.

Variable name	Description	Resolution
Geom. Radius	Mean value of geom. radius	0.01 m
Geom. Radius_sd	Std deviation of radius	-
Geom.Gradient	Mean of road gradient	0.1 %
Geom.Gradient_sd	Std deviation of road grad.	-
Geom.Crossfall	Mean value of crossfall	0.1 %
Geom.Crossfall_sd	Std deviation of crossfall	-
TRACS.Rutting	Mean of rutting depth along the considered segment of road	0.1 mm
TRACS.Rutting_sd	Std deviation of rutting depth along the considered segment of road	0.1 mm
TRACS.LPV03m	Mean of LPV at 3 m wavelength along the considered segment of road	0.01 mm^2
TRACS.LPV03m_sd	Std deviation of LPV at 3 m wavelength along the considered segment of road	0.01 mm^2
TRACS.LPV10m	Mean of LPV at 10 m wavelength along the considered segment of road	0.01 mm^2
TRACS.LPV10m_sd	Std deviation LPV at 10 m wavelength along the considered segment of road	0.01 mm^2
TRACS.LPV30m	Mean of LPV at 30 m wavelength along the considered segment of road	0.01 mm^2
TRACS.LPV30m_sd	Std deviation of LPV at 30 m wavelength along the considered segment of road	0.01 mm^2
TRACS.Pav_Texture	Mean of SMTD along the considered segment of road	0.01 mm
TRACS.Pav_Texture_sd	Std deviation of SMTD along the considered segment of road	0.01 mm

This is done by assigning to each record in Data.T the average of measured characteristic of the travelled road segment. Also, standard deviation of all parameters is calculated as a measurement of variation of the parameter along the selected road

segment. Full details regarding the road characteristic parameters considered in the study are given in Table 15.

Finally, meteorological conditions are assigned to Data.T. This is done by assigning, to each record, the reading from the nearest weather station measured as close as possible to the time when the data was collected by Microlise. Additionally, some meteorological parameters are calculated. For instance, wind speed (*WindSpeed_kn*) and its direction (*WindDirection_deg*) are retrieved from weather stations as measured at their location. However, these should be transformed in reference to the location and direction of travel of the considered truck. Therefore, first an approximated direction of travel of the considered truck (*Direction_rad* or *Direction_deg*, if in radiants or degrees respectively) is calculated. This is done by mean of Equation (10) and (11):

$$Direction_rad = atan2\begin{pmatrix} sin\,(Long.End - Long.Start) \cdot cos\,(Lat.End), \\ cos\,(Lat.Start) \cdot sin\,(Lat.End) - sin\,(Lat.Start) \cdot \\ cos\,(Lat.End) \cdot cos\,(Long.End - Long.Start) \end{pmatrix} \quad (10)$$

$$Direction_deg = \frac{180 \cdot Delta.Angle_rad}{\pi} \quad (11)$$

In case the result of Equation 11 is negative, this is transformed to an angle between zero and 360º by simply adding 360º. Then, a *Delta.Angle_deg*, angle between the direction of travel of the considered truck and the wind as measured by the weather station, in degrees, is calculated using Equation (12):

$$Delta.Angle_deg = WindDirection_deg - Direction_deg \quad (12)$$

Again, if the result of Equation (12) is negative this is transformed to an angle between zero and 360º by adding 360º. Then, once *Delta.Angle_deg* is known, the *WindSpeed_kn* is split in two components; one in opposite direction to the travel of the considered truck, and one in transversal direction.

Table 16 - Summary of weather condition measurements considered in the study.

Variable name	Description	Resolution
Weather_Station	Name of weather station where data come from (the nearest)	-
Temperature_deg_C	Temperature from the nearest weather station in ℃	0.1 ℃
Humidity_%	Humidity from the nearest weather station in %	0.1 %
Atm_Pressure_hPa	Atmospheric pressure from the nearest weather station in hPa	1 hPa = 0.001 bar
Rainfall_mm	Layer of rain felt on the ground during the last hour in mm	0.1 mm
WindSpeed_kn	Wind speed from the nearest weather station in knots	1 deg
WindDirection_deg	Wind direction from the North from the nearest weather station in degrees	1 kn = 1.85 km/h
Delta.Angle_deg	Wind direction calculated in reference to the direction of travel of the truck in degrees	1 deg
WindSpeed.Dir_kn	Wind speed component in opposite direction to that of travel of the truck in knots	1 kn = 1.85 km/h
WindSpeed.Cro_kn	Transversal wind speed component as absolute value in knots	1 kn = 1.85 km/h

As *Delta.Angle_deg* is calculated in reference to the direction of the wind, positive values of this indicate a truck going in opposite direction to the component of the wind parallel to the direction of travel (*WindSpeedDir_kn*, see Equation (13)). Instead, since it is reasonable to assume that no difference exist between a cross wind (*WindSpeed.Cro_kn*), which comes from one side or the other of the truck, an absolute value of this is calculated (see Equation (14)).

$$WindSpeed.Dir_kn = WindSpeed_kn \cdot cos(\frac{Delta.Angle_deg}{180} \cdot \pi) \quad (13)$$

$$WindSpeed.Cro_kn = WindSpeed_kn \cdot sin(\frac{Delta.Angle_deg}{180} \cdot \pi) \quad (14)$$

Table 16 provides full details of the type of weather condition parameters considered in the study. For full details about the code, which operates the processing, please, refer to Annex D.

4.3 Variable selection

Variable selection, also known as variable selection or attribute selection, is the process of selecting a subset of relevant variables (variables, attributes or predictors) to use in the development of a model. In fact, when dealing with a large number of measurements, it is likely that data contain redundant or irrelevant variables which can be removed without incurring a loss of information. Variable selection algorithms can be seen as a combination of trends and pattern recognition techniques, which identify produce a subset of variables that contribute the most to explaining variance across the multi-dimensional space of initial variables/measurements available (James *et al.*, 2013). Therefore, these techniques are used to simplify models, which enhances interpretability, reduces training time of ML algorithms, and improves generalization of the developed models by reducing variance and therefore overfitting (James *et al.*, 2013; Ng, 2018). For this, various techniques and methods are available nowadays, these can minimize the error rate, use proxy measure or regression coefficients, shrinkage techniques, etc. However, some methods may perform better than other depending on how strongly variables are correlated with each other and with the dependent variable, and if there is a linear or non-linear correlation between them. Since it is not known a priori which performs best in which case (Wolpert *et al.*, 1997), in this thesis a set of methods is investigated and performance compared. Finally, only variables selected by most of the methods applied are used for modelling.

The following sections introduce each of the used methods and provides details about how each of them is applied to select variables to model then the fuel consumption of the considered fleet of trucks and the effect of road surface conditions.

4.3.1 Pearson's correlation coefficient

The Pearson's correlation coefficient, usually referred as Pearson's *r*, is a measure of linear correlation between two variables. It can take values between -1 and 1, where 1 is total positive linear correlation, -1 is total negative linear correlation and 0 for no linear correlation.

For a sample the Pearson's correlation coefficient can be calculated as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})}} \quad (15)$$

where, n is the sample size, x_i and y_i are the individual sample points indexed with i, and $\bar{x}, \bar{y}$ are the sample mean e.g. for $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

In this thesis, the Pearson's correlation coefficient is used for a preliminary selection of variables. For instance, variables which result in an absolute value of the correlation (r) with truck fleet fuel consumption (*Fuel.Consumption*) lower than 0.05 and that, in consideration of the results of existing studies (see sect. 2.4), should not produce any significant effect on vehicle fuel consumption, are not included in the following data analysis. That is to simplify the investigation by focusing only on variables which numerically or physically show some correlation (or possible causation) with truck fleet fuel consumption. For example, having to investigate the correlation of road surface characteristics with truck fleet fuel consumption, use of a technique such as the Pearson's *r* may be restrictive, and may not allow to identify the trends existing through the data, due to the non-linearity of the problem. For this reason, no threshold of *r* is applied to the correlation coefficients calculated for LPV (*TRACS.LPV03m*, *TRACS.LPV10m*, and *TRACS.LPV30m*) and SMTD

(*TRACS.Pav_Texture*) with truck fleet fuel consumption, and other techniques are used to assess their importance in the study (see sect. 4.3.2, 4.3.3 and 4.3.4)

In regression analysis, it is not just important to identify variables correlated to the target, but also possible collinearities among the predictors. In fact, including in a regression model collinear variables, may result in feeding to the model redundant information which may produce overfitting, loss of accuracy and inability of the model to generalise the defined correlations to new sets of data from the same population (James *et al.*, 2013). For this reason, in this thesis correlation matrices are used to identify correlation among all the considered variables. That usually helps in identifying dependent and independent variables, collinearities across the data, and might help in improving the accuracy of traditional regression models.

However, modern regression techniques, such as ML, usually already implement methods to reduce the effect of collinearity and establish differences of impacts also from variables which are linearly correlated (Goodfellow *et al.*, 2016; Ng, 2018; Burkov, 2019). That cannot be done by traditional methods (James *et al.*, 2013) since the variables influencing the dependent variable (and therefore included in the developed model) must be defined by the expert a priori, and including collinear variables in the developed model may result in reducing its accuracy (James *et al.*, 2013).

4.3.2 Regression analysis

Regression analysis is a set of statistical processes used to estimate the impact that one or a set of variables (features or predictors) have on another (outcome, predicted or dependent variable) (James *et al.*, 2013). Specifically, regression analysis helps experts understand how the value of the dependent variable changes typically when the independent variables vary.

Due to the large amount of information available, it was decided to use regression analysis to investigate the relationship between truck fleet fuel consumption and road surface conditions and their impact.

The simplest type of regression is LR. By fitting an LR to all variables which show a certain correlation (r) with the dependent variable, it is possible to test the significance of this correlation. Different tests exist to test the significance of variables included in a LR but some of the most commonly used consist of i) p-values, and ii) adjusted-R^2.

i) By fitting an LR of all independent variables available to determine the dependent variable of interest, *p*-values can be computed. *p*-values are commonly used in the context of null hypothesis testing to quantify the statistical significance of evidence (James *et al.*, 2013). In regression analysis, p-values calculated for each independent variable test the null hypothesis that the coefficient assigned to the variable is equal to zero or, in other words, that the independent variables do not affect the dependent.

   A low *p*-value indicates that it is possible to reject the null hypothesis and, hence, that the regression coefficient assigned to the independent and variable to estimate typical values of the dependent one (approximation of the correlation between the two) is statistically significant.

   Conversely, a large *p*-value suggests that changes in the independent variable do not produce (or at least are not associated with) changes in dependent variable. Because of this, statisticians typically use *p*-values to determine which predictors to keep in regression models.

   However, low and high is relative. Therefore a threshold must be defined for the p-value to show significance. Usually, the limit is set to 0.05 meaning that

hypothesis is tested at 95% (0.95) confidence level. However, for different purposes, other confidence levels can be used for different hypothesis testing depending on the risk which is involved in the investigated process.

ii) In regression analysis, the coefficient of determination, R^2, represents the proportion of variance that the combination of the independent variables (predictors) included in a model explain.

R^2 takes values between 0 and 1, where 0 corresponds to the case of independent variables that cannot explain any of the variance in the dependent variable and 1 to the case of variables that completely explain the variance in the dependent variable.

The coefficient of determination can be computed using:

$$R^2 = (r)^2 = 1 - \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \tag{16}$$

where, $\hat{y}_i$ is the estimated value corresponding to the i^{th} measurement (y_i) and $\bar{y}$ is the mean of all measurements. Basically, R^2 provides with a measure of how well observations are replicated by the develop model. However, the coefficient of determination tends to increase with an increase in the number of variables included in regression models, which makes this measurement inappropriate to measure significance of variables. In fact, increasing the number of variables in a model can help in improving accuracy but can also introduce noise and produce overfitting, reducing the ability of the model to generalise its estimates.

For this reason, it is important to reduce the number of variables to an optimum which can give accurate estimates without incurring overfitting.

This is why, statisticians introduced a different measurement called adjusted-R^2. This measure, as for the cases of the traditional R^2, shows how well data fit a certain curve but it takes consideration of the number of variables included in the developed model. In particular, adjusted-R^2 takes the form:

$$\text{adjusted-}R^2 = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1} \right] \tag{17}$$

where, R^2 takes can be computed using Equation (16), n is the total amount of measurements and k is the number of independent predictors considered.

Notice that, if the variables used in the model help in achieving higher levels of accuracy then adjusted-R^2 and R^2 tend to be equal. However, if some of the variables used in the model are not that significant, then adjusted-R^2 will be significantly lower than R^2. This way the adjusted R^2 can help in understanding whether it is important to include a certain variable in a regression model or not.

In this thesis, analyses of *p-values* and adjusted-R^2 are compared with each other and the results of Lasso and RFs (presented in sect. 4.3.3 and sect. 4.3.4) used to determine the optimal selection of features (variables) for the developed models. Other methods to perform variable selection exist in the literature (James *et al.*, 2013), however the selection considered in this study covers some of the most widely used ones (James *et al.*, 2013) and comparison of these, together with information from the literature review (e.g. Haider, Conter, *et al.*, 2011; Haider and Conter, 2012; Benbow, Brittain, *et al.*, 2013; EPA, 2018e), should help in improving the reliability of the final choice and that of the developed models.

As for the whole data analysis, all regression variable selection methods are implemented using the statistical software R (see Annex A).

4.3.3 Lasso

As an alternative to traditional variable selection models, it is possible to fit a model containing all predictors and use regularisation to shrink the coefficient estimates towards zero and drop out some variables (James *et al.*, 2013; Goodfellow *et al.*, 2016; Ng, 2018). That is to reduce variance in the model and improve its fit to the data (Tibshirani, 1996; James *et al.*, 2013; Ng, 2018).

The Lasso, least absolute shrinkage and selection operator, is a shrinkage technique very similar to the traditional least squares method applied in regression except for the fact that a penalty is applied to reduce the regression coefficients to zero if possible (Tibshirani, 1996; James *et al.*, 2013).

For instance, the Lasso coefficients are calculated to minimise the quantity:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij} \right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right| = RSS + \lambda\sum_{j=1}^{p}\left|\beta_j\right| \qquad (18)$$

where, the first term represents the formulation of the residual sum of squares (*RSS*) minimised in the least squares method, and the second factor is the Lasso shrinking penalty applied.

From Equation (18) it is possible to notice that while the least squares method attempts to minimise just the *RSS*, the Lasso introduces the shrinking penalty $\lambda\sum_{j=1}^{p}\left|\beta_j\right|$. The advantage of introducing the shrinking penalty is that, as λ increases, variance tends to decrease and bias to increase meaning that the Lasso will set all coefficients close to zero but not exactly to zero unless $\lambda \to \infty$ or if the set of data available for a variable does not allow the model to explain a larger variance across the data. That helps in producing simpler models of high accuracy and better interpretability (James *et al.*, 2013).

Other advantages of the Lasso are that it does allow identification of non-linear correlations among variables (Yamada *et al.*, 2013; Zhang *et al.*, 2016) and that it does not increase calculation time in comparison to the classic least squares method (James *et al.*, 2013).

Since the value of λ is fundamental for the Lasso to perform best, it is important to tune this parameter carefully. To do that, there are different methods applicable, but in this thesis the grid-search is used. Multiple values of λ (a grid of values) are tested and that which gives the lowest error is selected. In this study the Lasso and grid-search method for optimisation of the parameter λ has been implemented using the `glmnet` package available in R (see Annex A and Annex D for details about the `glmnet` package and implementation of the Lasso in the study).

4.3.4 Random Forests: for Variable Selection

Random Forests (RF) (Breiman, 2001) is a ML algorithm based on the theory of decision trees (Breiman *et al.*, 1984). More specifically, a forest is a combination of tree predictors (Breiman *et al.*, 1984; Mitchell, 1997a) such that each tree in the forest depends on a vector of independently and randomly sampled variable in the data (Breiman, 2001). This way, as the number of trees in the forest increases, the error, in prediction or classification, tends to converge depending on the strength of correlation between the target and the variables considered in the individual trees, and level of similarity between the trees (Breiman, 2001). Although the RF still remains a 'black box' model, since its results are substantially obtained by averaging the results of multiple decision trees, these models can result to be more interpretable than other ML techniques. However, the larger the model, the lower might be its interpretability.

RF is generally used for classification (James *et al.*, 2013; Goodfellow *et al.*, 2016) however, the algorithm is easily adaptable to the case of regression analysis

(Breiman, 2001). In the past, several studies (e.g. Deloncle *et al.*, 2007; Khaidem, Saha and Dey, 2016) demonstrated its suitability for regression and, due to the randomness of processes involved, the algorithm is demonstrated to be also robust with respect to outliers (Breiman, 2001).

Another characteristic of the RF algorithm is that it can produce also internal estimates to measure variable importance (Breiman, 2001). In particular, for each tree in the forest, the algorithm compares the accuracy (or error) that each tree in the forest scores when it does consider or not a certain variable. This can be done by measuring for each tree within the forest the decrease in accuracy (for classification) or change in mean squared error (MSE, see Equation (19)) (for regression) that considering a certain variable produces, in comparison to trees that do not consider that variable (Breiman, 2001; Liaw *et al.*, 2017). That is than compared to the accuracy or MSE of a tree developed including all variables (Breiman, 2001; Liaw *et al.*, 2017).

$$MSE = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{n} \tag{19}$$

where, $\hat{y}_i$ is the i-th measurement of the dependant variable, y_i is the i-th prediction of the dependant variable, and n is the number of the available measurements.

In this study, the increase in MSE (%IncMSE) was used to rank variable importance and quantify how each of the considered measurements can help in explaining the variance in the fuel consumption of the considered fleets of truck.

4.4 Linear Regression (LR)

LR takes a linear approach to modelling the relationship between two or more quantities (or variables). The main assumptions of LR are: 1) linearity, meaning that response is assumed to be linearly correlated to combination of predictors; 2)

homoscedasticity, meaning that variance of errors is constant and does not depend on the value of predictors; 3) independence of errors, meaning that residuals are normally distributed; 4) no (or low) multicollinearity of predictors, meaning that predictors should not be correlated with each other (James *et al.*, 2013).

LR represents the simplest case of regression. This is the case of a response (*y)* being estimated on the basis of measurements from a single predictor (*x)*, which takes the mathematical form:

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{20}$$

where β_0 and β_1 are constants, representing respectively the intercept and slope terms in the linear model, and ε is the average error made by the model in predicting the target *y* given a measurement of the explanatory variable *x*. However, in practice, β_0, β_1 and ε are unknown and must be estimated so that the resulting line is as close as possible to the actual data. This can be done in several ways, but the most common technique is the least squares method (LSM). In particular, the method estimates β_0, β_1 and ε as those that minimise the sum of squared differences between measurements and actual value of the target variable (*y)*, alternatively referred as residual sum of squares (*RSS*). In other words, if the difference between the i^{th} measurement (y_i) and i^{th} estimate ($\hat{y}_i$) is called e_i :

$$e_i = y_i - \hat{y}_i \tag{21}$$

and, based on the results from Equation (21), the *RSS* is defined as:

$$RSS = e_1^2 + e_2^2 + ... + e_n^2 \tag{22}$$

the LSM approach chooses β_0, and β_1 to minimise the *RSS* (Equation (22)).

Simple LR is a useful approach for predicting a response based on a single predictor variable. However, in practice, since a combination of more predictors

usually allows to achieve better levels of accuracy, multiple measurements are used. Suppose then to have *m* distinct type of measurements to use as predictors. Then a multiple LR of them takes the form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m + \varepsilon \tag{23}$$

where, again, β_0 is the intercept of the developed LR model, β_i is the coefficient assigned to the i^{th} predictor considered and ε the average error made by the model at estimating the actual value of the target variable (*y)*. In this case, based on the formulation of Equation (23), the *RSS* can be defined as:

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{24}$$

Consequently, the values of β_i which minimise the *RSS* (Equation (24)) are the multiple LSM regression coefficient estimates.

In this study, a multiple LR model has been developed to estimate the fuel consumption of large truck fleets including the effect of road surface conditions, and the LSM applied to estimate its coefficients.

In the past, several other studies used LR to model the fuel consumption of road vehicles including trucks (e.g. Clark *et al.*, 2009; Delgado *et al.*, 2011; Bifulco *et al.*, 2015; Çapraz *et al.*, 2016). For this reason, in this study multiple LR has been used to develop truck fleet fuel consumption models based on the large amount of standard (SAE, 2016) truck sensor measurements (Perrotta *et al.*, 2017a, 2017c, 2017b). That is to draw a baseline and allows comparison of performance with more advanced regression techniques used in the following.

4.5 Machine Learning (ML)

ML is the branch of Computer Science and Artificial Intelligence (AI) which makes use of statistical techniques to give machines the ability of learning from data

directly, without being explicitly programmed to do so (Mitchell, 1997b; Goodfellow *et al.*, 2016; Ng, 2018; Burkov, 2019). That is done thanks to geometric transformation in multi-dimensional spaces and application of advanced statistics to subset data and recognise hidden patterns within the them (James *et al.*, 2013; Goodfellow *et al.*, 2016). In practice, ML algorithms define how trends vary across the data by identifying thresholds and use functions (i.e. activation functions in neural networks, kernel functions for support vector machines and data separation for decision trees and RFs) to recognise when these thresholds are overcome and in which situations (Goodfellow *et al.*, 2016; Ng, 2018; Burkov, 2019).

ML was first cited by Samuel in 1959, however these techniques have recently gained fresh momentum thanks to the higher computational power and larger quantity of data available at scale. The main idea behind ML technique is that although large amount of data may contain noise, or sometimes include incomplete information, their effect may be marginal on average and that could allow machines to highlight trends and patterns across the data (Mitchell, 1997b; Goodfellow *et al.*, 2016). This is why these techniques may not perform well on small datasets, but tend to outperform traditional methods on larger sets (Ng, 2018).

Essentially, there are two types of ML techniques: 1) supervised and 2) unsupervised (Stergiou *et al.*, 1995; James *et al.*, 2013; Goodfellow *et al.*, 2016). The majority of ML techniques uses supervised learning, meaning that, initially, measurements of both input and output variables are available (Stergiou *et al.*, 1995; James *et al.*, 2013; Goodfellow *et al.*, 2016). From these, the algorithm learns correlations and identifies patterns across the data (Stergiou *et al.*, 1995; James *et al.*, 2013; Goodfellow *et al.*, 2016). The algorithm then defines models to approximate the trends identified, which allows modelling of the analysed phenomenon. Once a model

has been defined, it is possible to make estimates of the target variable based on new measurements of the explanatory variables. Although some algorithms can be very complex, methods such as LR can also be considered as a supervised learning algorithm, where the target is estimated based on a linear combination of the explanatory variables (Burkov, 2019). Examples of supervised ML algorithms are: support vector machines, classification and regression decision trees, and neural networks (Stergiou *et al.*, 1995; James *et al.*, 2013; Goodfellow *et al.*, 2016; Burkov, 2019).

Other types of ML algorithms are unsupervised. That means that, initially, the machine, computer or bot does not have any example of the output. Basically, without receiving any 'help', the algorithm classifies, identifies trends or subsets the data based on some difference in the structure of the data, which is identified by comparing multiple statistics (Stergiou *et al.*, 1995; James *et al.*, 2013; Goodfellow *et al.*, 2016). Examples of these algorithms are: k-means clustering, hierarchical clustering, autoencoders, and principal components analysis (Stergiou *et al.*, 1995; James *et al.*, 2013; Goodfellow *et al.*, 2016).

Nowadays ML is used in a number of different areas of science and to solve problems in different sectors of industry. That includes product personalisation, security enhancement and fraud detection, insurance and share pricing, cost and process optimisation, fault diagnostics, healthcare, etc. Again, this is thanks to the higher computational power available nowadays, which allows use of these complex algorithms, which need large amounts of information to perform well.

For this specific case study, the main reason to apply ML is exactly the ability of these techniques to learn from data which are available in large quantities provided by Microlise and TRL. Because data are labelled, meaning that measurements of the

output variable (fuel consumption) are available, multiple supervised ML techniques have been tested in the study, and performance compared to a multiple LR of the explanatory variables. In particular, in the past, SVR, RF and ANN have been used to model the fuel consumption of road vehicles for small fleets, along specific road segments and without considering the effect of road surface conditions (e.g. Xu *et al.*, 2010; Almér, 2015; Laxhammer *et al.*, 2015; Zeng *et al.*, 2015; Çapraz *et al.*, 2016; Wang and Rakha, 2017; Huang *et al.*, 2018). Due to the relatively high level of accuracy achieved by these studies, these techniques have been identified to be the most promising methods to perform regression analysis on large quantities of data such as in the present study. However, for the "no free lunch" theorem, for ML techniques it is not possible to say a priori which technique performs best in which conditions (Wolpert, 1996; Wolpert *et al.*, 1997). Therefore, before making decisions, different combination of learning and optimisation algorithms and sets of hyperparameters should be tested for the developed models and their performance compared. Due to the enormous amount of different learning and optimisation algorithms in the literature, and hyperparameters characterising the models to develop, in the present study only three different learning algorithms were explored, and grid-search (Goodfellow *et al.*, 2016; Burkov, 2019) used to determine the best combination of hyperparameters to use in the developed models.

As mentioned, in order to allow supervised ML techniques to identify trends, one has first to train the model on existing 'labelled' data and then test it in new 'unlabelled' situations. In practice, the available datasets are usually split in two, with 75% of data used for training and the remaining 25% used to test the developed model. To reduce bias in the final estimates the split of data must be random. That should allow to include all possible cases in the training set and similar scenarios in the test

set. However, this does not always represent the best solution for developing ML based models.

In the following sections, full details in regard to the methods applied, techniques and ways to tune their hyperparameters are provided.

4.5.1 Support Vector Regression (SVR)

Support vector machines are ML algorithms whose main characteristic is to perform separation across the data through definition of hyperplanes across multi-dimensional spaces (Cortes *et al.*, 1995; Gunn, 1998; Cristianini *et al.*, 2000). Due to their geometrical approach, support vector machines are commonly used to solve classification problems, but, as demonstrated by Gunn (1998), they can be easily adapted for regression analysis, in which case they are referred as Support Vector Regression (SVR). For this reason, capability of SVRs to estimate the fuel consumption of the considered fleets of truck and impact of road surface conditions was investigated in this study.

Figure 23 shows a representation of how SVRs work. In the figure, SVR identified the blue dots to follow a trend approximated by the regression line (the thick blue line), grey dots to be outliers, and red dots, to define limits (the two blue dashed lines) which allow the regression line to minimise the error ($\pm \varepsilon$) (and are called 'support vectors').

While support vector machines identify classes of data by separating them using hyperplanes, SVR defines the combination of hyperplanes that, within the multidimensional space of the considered variables, minimise the estimation error (Gunn, 1998; Cristianini *et al.*, 2000, 2002). Because performing this calculation in non-linear high dimensional spaces can be computationally expensive, similarly to support vector machines, SVR make use of kernel functions to transform the originally

linearly inseparable observations into higher-dimensional spaces to make data linearly separable and reduce calculation (Cristianini *et al.*, 2000, 2002). This is a similar concept to the geometric transformation that one can apply to describe the equation of a circle from cartesian to polar coordinates.

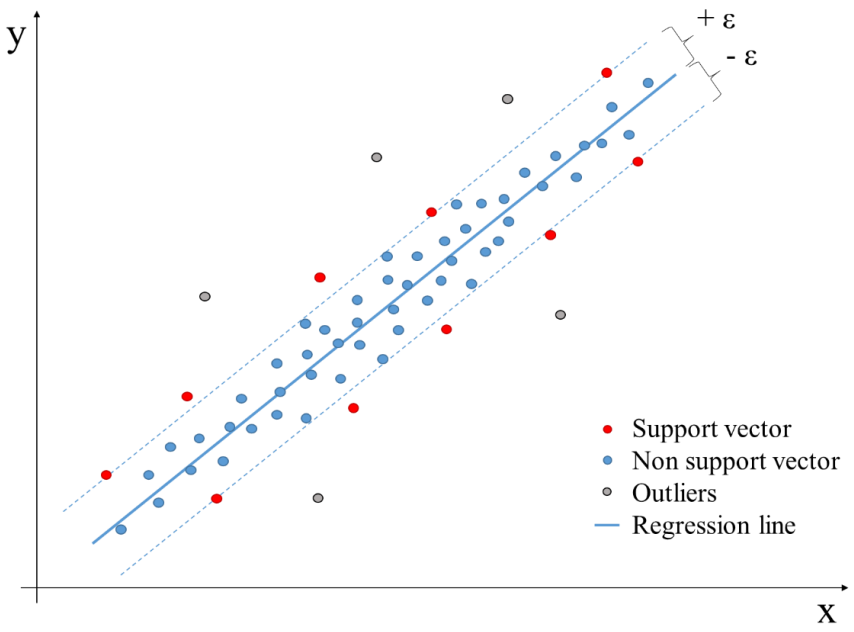


Figure 23. Representation of how SVRs work.

In the past, other studies concerning the fuel consumption of road vehicles made use of the Gaussian radial basis function (RBF) as kernel function (Zeng *et al.*, 2015). For this reason, the RBF (also called 'Gaussian function') has been chosen as kernel function of the developed SVR models.

The grid-search method (Goodfellow *et al.*, 2016; Burkov, 2019) was used to determine the optimal parameters to calibrate the model. In this method works that, from a selection of possible values for the parameters that characterise some SVR models are generated, and the combination that minimizes the error and makes the computational time reasonable is selected.

For this study, the *e1071* (Meyer *et al.*, 2017) R package was used to develop SVR models. This represents the most recently updated library available to develop SVR models using R.

4.5.2 Random Forests (RF): for Regression

As introduced in sect. 4.3.4, RF is a ML algorithm able to perform both classification and regression analysis which in the past has been demonstrated to be suitable for modelling the fuel consumption of trucks (Almér, 2015; Laxhammer *et al.*, 2015). For this reason, in this study, capability of RFs to model the fuel consumption of the considered fleets of truck and impact of road surface conditions was investigated.

The RF algorithm is based on the theory of decision trees. To perform regression analysis, the RF uses multiple trees developed using randomly selected variable and makes estimates by averaging results from all trees generated.

Figure 24 shows how RFs work. The figure shows how multiple (*n*) decision trees form the forest (RF). Result of the RF model are therefore given by an average of results of trees in the forest and depend on the number of trees and correlation of variables considered with the target. For this reason, RF are considered "ensemble averaging" models (Goodfellow *et al.*, 2016; Ng, 2018; Burkov, 2019), and this is why in order to characterise an RF model, the most important parameters to tune are the number of trees in the forest and the number of variables randomly sampled and considered in each tree. In particular, variable sampling is performed differently for each tree so that each tree differs from the others in the forest. This way, as previously mentioned (see sect. 4.3.4), as the number of trees in the forest increases, the error tends to converge (Breiman, 2001). That improves reliability of the estimates made by the forest and enhances resilience of the model to the presence of outliers in the data

(Breiman, 2001). A larger forest, with more trees, usually implies better precision and stability of the results, but at a higher computational cost and lower iterpretability.

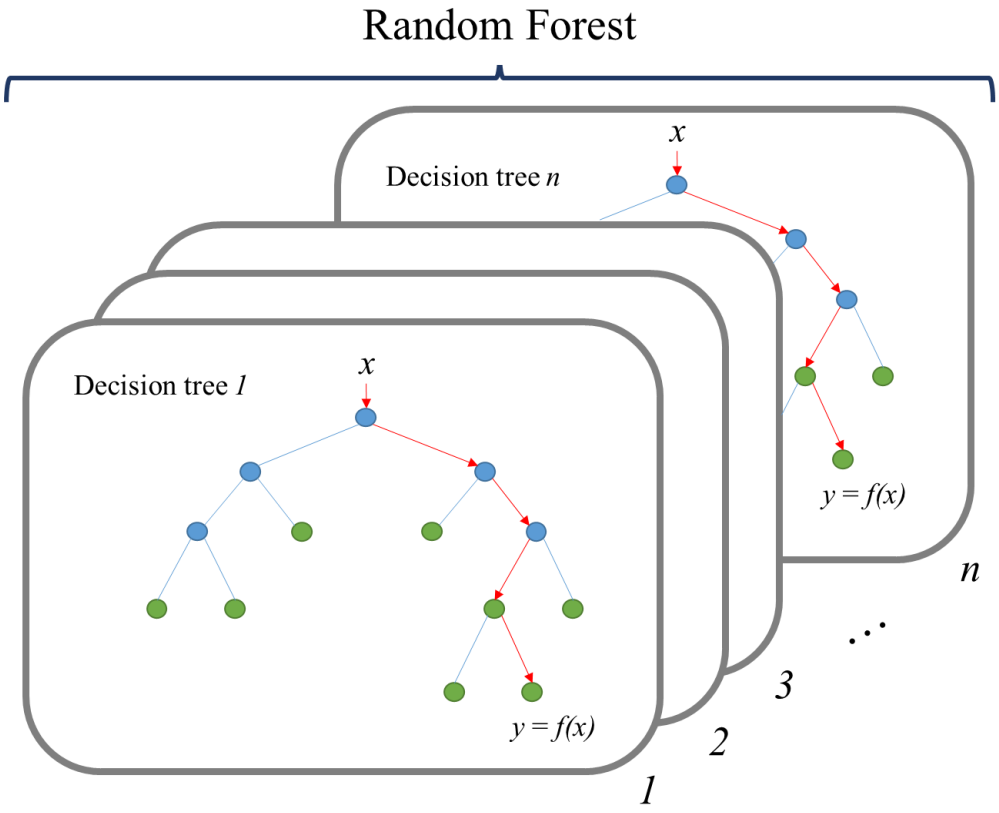


Figure 24. Schematic representation of how RFs work.

Additional details regarding the RF method have been provided in sect. 4.3.4. In order to determine the best number of trees and number of randomly sampled variables in each tree, which optimise error and calculation time, also in this case the grid-search method was used (Goodfellow *et al.*, 2016; Burkov, 2019).

For this study, RF models were developed using the *randomForest* (Liaw *et al.*, 2017) R package. This represents the most recently updated and widely used pakage available to develop RF models using R.

4.5.3 Artificial Neural Networks (ANN)

ANN (McCulloch *et al.*, 1943; Werbos, 1974) is an ML algorithm inspired by how the human brain processes information (Stergiou *et al.*, 1995). This is the reason why the simplest structure of an ANN is called a 'neuron' and, usually, multiple neurons, structured and interconnected in multiple layers, constitute an ANN.

Figure 25 compares a human neuron to the mathematical representation of a neuron in an ANN.

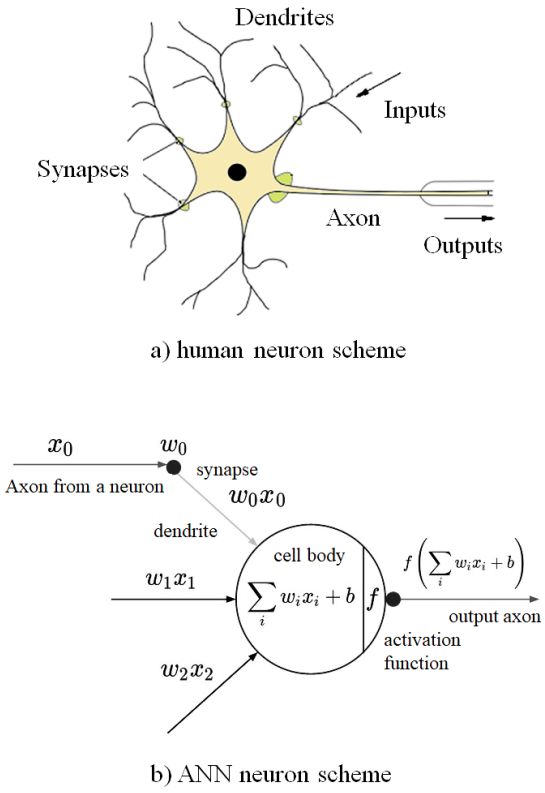


Figure 25. Scheme of a a) human and b) ANN's neurons.

A neuron is generally a device with several inputs and one output which, by communicating and passing information with other neurons in the network, makes estimates of the target from the inputs provided (Hecht-Nielsen, 1989; Stergiou *et al.*, 1995; Mitchell, 1997b). In practice, a neuron can work in two ways: 1) in 'training

mode' or 2) in 'use mode' (Stergiou *et al.*, 1995). In the training mode, the algorithm identifies patterns through the data and an 'activation function' is calibrated about when a particular pattern is statistically significant or not. In the use mode, instead, when a pattern is recognised, the neuron associates to this the learned output.

The case in which a single neuron constitutes the ANN is very similar to the traditional LR and corresponds to the logistic regression, where a combination of the input variable is used to estimate the target directly, as:

$$y = \sum_{i=1}^{n} \omega_i x_i + b \tag{25}$$

where y is the target (which can be also be an intermediate estimate), n the number of inputs considered, and ω_i the weights associated with the variable x_i and b the estimated error (see Figure 25).

A combination of multiple neurons interconnected through one or multiple layers makes an ANN (see Figure 26), which can therefore approximate complex functions, by combining and transforming the inputs through higher dimensional spaces (Hecht-Nielsen, 1989; Stergiou *et al.*, 1995; Mitchell, 1997b; Goodfellow *et al.*, 2016). That allows ANNs to identify trends and patterns which could not be detected otherwise.

Figure 26 shows the structure of a generic ANN model. As the figure shows, an ANN is composed by an input layer which contains measurements of all predictors considered, multiple hidden layers, and an output layer, which for this study is composed by only one output node, but can be multiple. Structure of ANNs can vary depending on the amount of data considered and complexity of the phenomenon to model. Again, for the "no free lunch" theorem (Wolpert, 1996; Wolpert *et al.*, 1997), the structure of the ANN is not known a priori and each hidden layer can be composed

of any number of neurons which may differ from layer to layer. Additionally, several types of ANN algorithms are available in the literature that use more or less complex neurons, different ways to learn trends and patterns in the data and different activation functions.

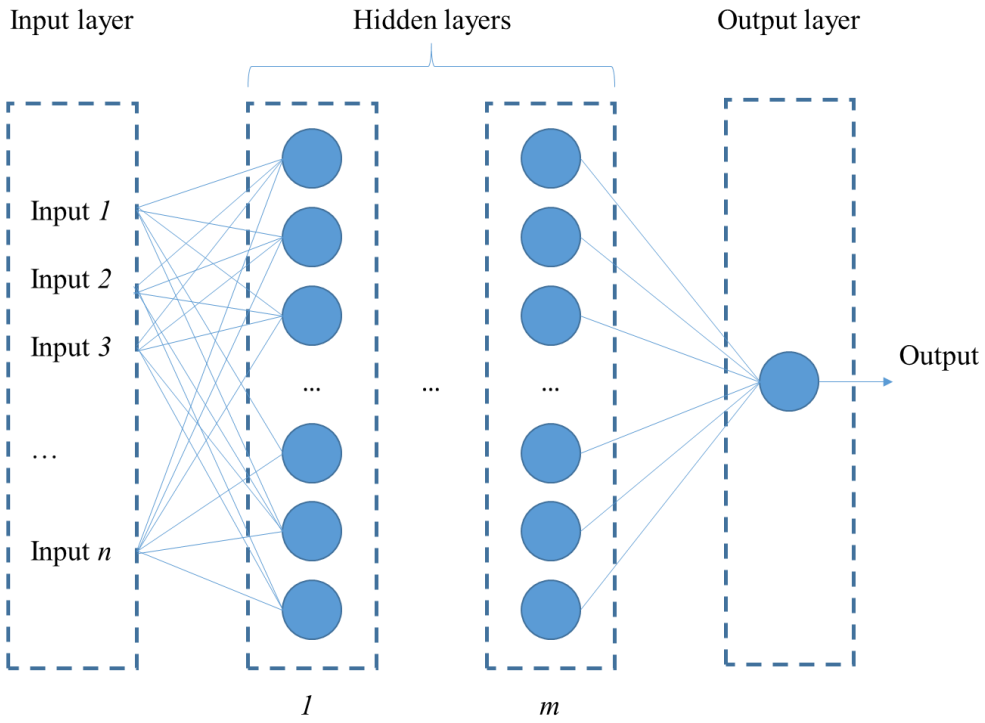


Figure 26. Schema of a generic ANN model.

Therefore, optimisation of the structure of ANNs can represent a complex problem, which include selection of the best algorithm to apply to specific datasets or problem and can be solved using various methods such as the grid-search, random-search or cross-validation methods (Goodfellow *et al.*, 2016; Ng, 2018; Burkov, 2019).

In the past, ANN has been applied for the estimation of the fuel consumption of aircraft (Schilling, 1997; Trani *et al.*, 2004), and more recently, of road vehicles (Almér, 2015; Zeng *et al.*, 2015). However, these studies considered smaller sets under

more specific condition than those analysed in this study, including only data from cars tested along a pre-defined road path (Zeng *et al.*, 2015) or considering just a specific heavy vehicle model from a specific manufacturer (Almér, 2015) which may not give opportunities to engineers to understand variations of fuel consumption at fleet and network level. Additionally, none of the existing studies considered the possibility of making secondary use of existing data, adding to these additional information from different databases. This is why none of the studies considered the effect of road geometry, limiting their ability to model fuel consumption under real conditions.

Sometimes in the literature deep learning or deep neural networks are mentioned (e.g. Goodfellow *et al.*, 2016; Ng, 2018; Burkov, 2019). These are ANNs characterised by a large number of layers (usually more than 2) and neurons per layer (usually more than the number of variables considered) (Goodfellow *et al.*, 2016; Ng, 2018; Burkov, 2019).

Also in this case, the grid-search method (Goodfellow *et al.*, 2016; Burkov, 2019) was used to identify the optimal structure of the ANN and values of the hyper-parameters selected to minimise the error and make the calculation time reasonable. For the same reason, the resilient backpropagation algorithm without backtracking (rprop-) algorithm (Riedmiller *et al.*, 1993) was chosen among others and the logistic activation function used to developed the ANN models, including the examples of deep neural network (ANN1).

The *neuralnet* (Fritsch *et al.*, 2016) R package was used to develop the ANN models produced. This represents the most recently updated package available to develop ANN models using R.

4.6 Measurement of performance

The coefficient of determination R^2, the root mean square error (RMSE) and the mean absolute error (MAE) are used to measure accuracy and compare the performance of the developed models.

$$RMSE = \sqrt{MSE} \tag{26}$$

$$MAE = \frac{\sum_{i=1}^{n} \left| \hat{y}_i - y_i \right|}{n} \tag{27}$$

In the equations above, $\hat{y}_i$ is the i-th measurement of the dependant variable, y_i is the i-th prediction of the dependant variable, and n is the number of available records.

Additionally, crossvalidation time (time required to train and test the model n-times) is used for comparison. This is because each algorithm learns in different ways from the data and their efficiency varies with the nature of the data available and their quantity.

4.7 Learning curves

Characteristic of ML algorithms is that they generally improve performance with the quantity of data which is fed to them (James *et al.*, 2013; Goodfellow *et al.*, 2016; Ng, 2018; Burkov, 2019). Similarly to what happens with human beings, the more examples the 'machine' has and the more precise its predictions will be. However, consideration of larger quantities of data can also increase bias due to the introduction of noise (Ng, 2018). In order to set up the optimal amount of data to feed to ML algorithms Data Scientists commonly use learning curves (Ng, 2018).

Learning curves plot variation of the training and test sets errors (e.g. RMSE or MAE, etc.) against the number of training examples (Ng, 2018). In particular, these

should be very different for small training datasets but eventually they should also asymptotically converge for a certain size of the training set (Ng, 2018).

By collecting more and more data, it is possible to see (Figure 27) that the training error tends to increase and that of test to decrease (eventually converging at some point). This is because the developed model is now considering a data distribution representative of the whole observed phenomenon.

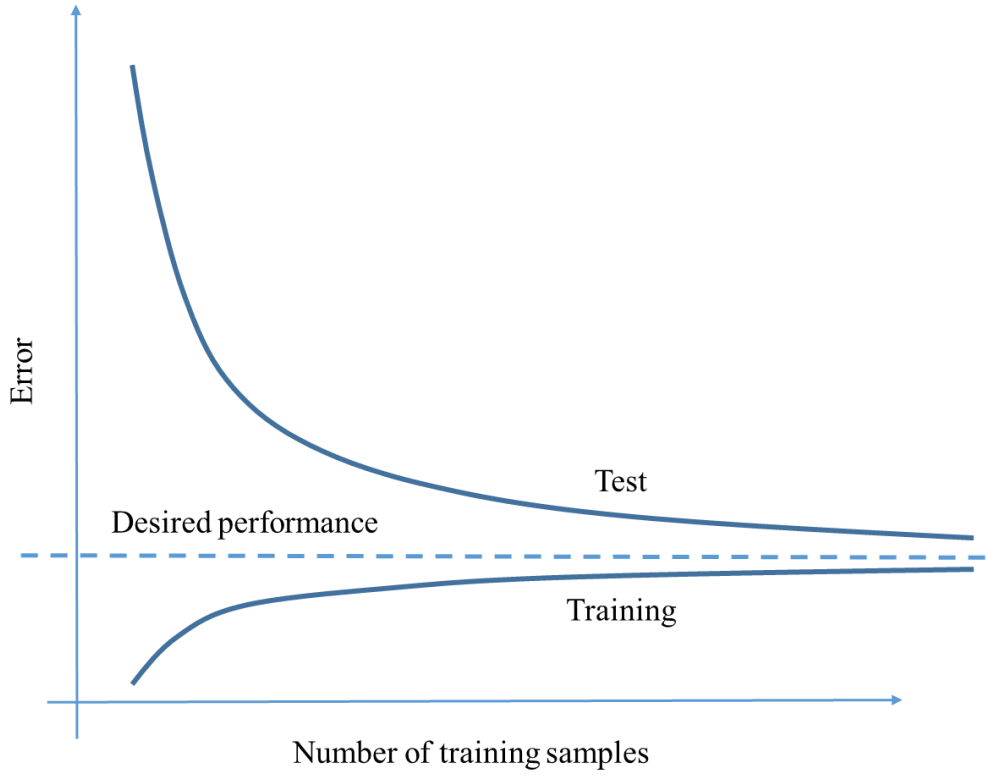


Figure 27. Example of convergence of learning curves.

By plotting variation of RMSE (or any other measure of error such as MAE, etc.) against the amount of data used to train a model it is possible to understand whether or not the amount of data used to train the developed model is enough to ensure coverage of the whole variance in the considered case study, and that is the aim of learning curves (Ng, 2018).

That can help also in assessing whether the model will be able to achieve the desired performance or if just feeding new data will not help the model in reaching the prefixed goals. In fact, if training and test sets RMSEs converge but not at the desired level of error, the problem may not be in the quantity of data used to feed the developed model, but in the variables used to train the model and its structure. It is said that, in this case, the model suffers high bias (see Figure 28), and that can be generally addressed by modifying the input variables, increasing the model size, or generally working the model architecture (Ng, 2018).

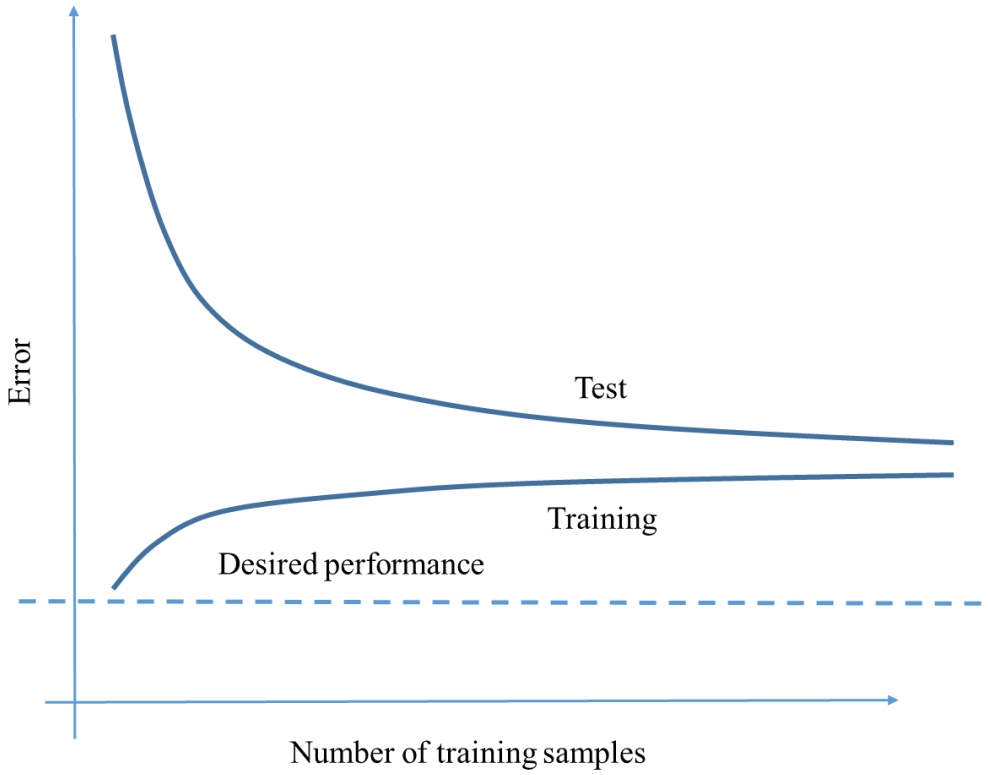


Figure 28. Example of learning curves which do not converge at the expected/desired level of performance.

Adding more data, variable selection or decreasing the model size can help only in the case of a model suffering a high variance which results in non-convergent learning curves (Ng, 2018) (see Figure 29).

In this study, learning curves have been used to assess whether the amount of data used in the study is enough to train the developed models. This has been done for the ANN model only and optimal amount of data determined is considered valid also for the SVR and RF models. This is possible thanks to the large quantity of data available but it is not always the case. In fact, although this is not the case of the present study, it is suggested that for small datasets learning curves are analysed separately for each of the developed models (Ng, 2018).

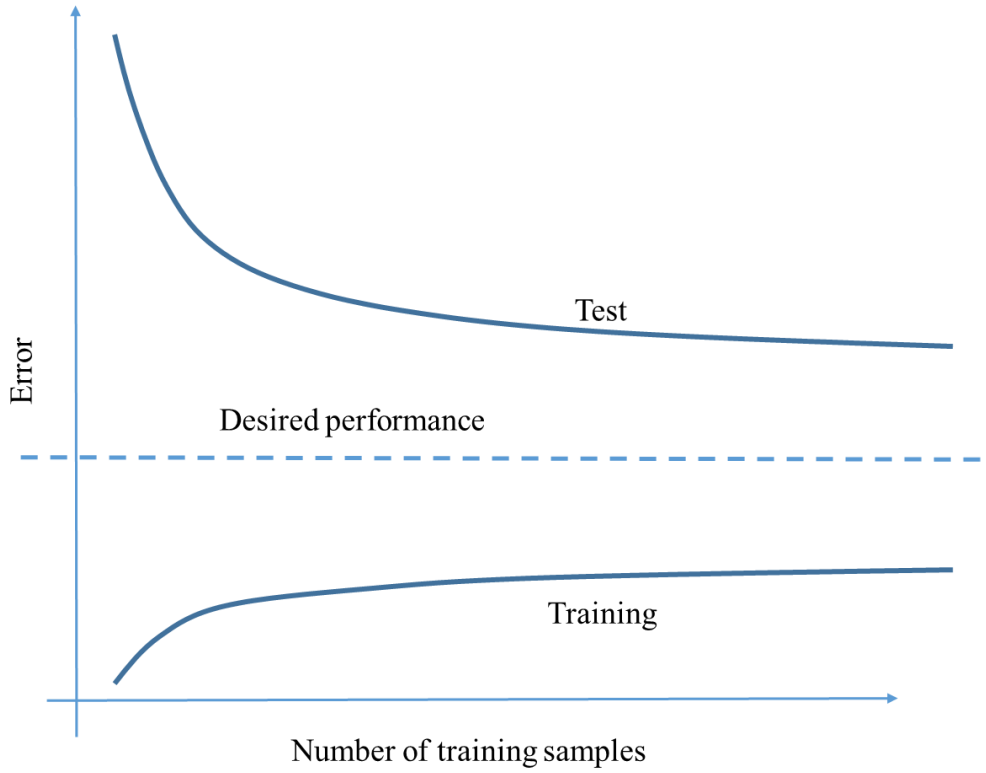


Figure 29. Example of learning curves which do not converge at the expected number of training samples.

4.8 Grid-search

Ideally ML algorithms would just take a dataset and output a function able to give robust and accurate estimates without requiring any supervision. This is in fact

what some unsupervised learning algorithms do but that is not the case for regression algorithms such as SVRs, RFs, and ANNs. Popularity of some algorithms such as logistic SVRs and RFs stems in part in their ability to perform well by hand tuning only one or two parameters. ANNs are other techniques which can sometimes perform quite well without tuning lots of parameters, but that essentially depends on the type of problem considered and adopted learning algorithm. However, most advanced recently developed ANN algorithms benefit significantly from tuning forty or more parameters in the model (Goodfellow *et al.*, 2016; Burkov, 2019). For this reason, when dealing with ML algorithms, hand tuning usually does not represent a feasible choice and exploring several combination of hyperparameters requires automation (Goodfellow *et al.*, 2016; Burkov, 2019).

When the amount of hyperparameters to tune is not very large, which is the case of this study, it is common practice to perform grid search (Goodfellow *et al.*, 2016; Burkov, 2019). Grid search is performed by selecting a small finite number of values for each of the hyperparameters in the model and exploring variation of performance of the developed model for each combination of selected values (Goodfellow *et al.*, 2016; Burkov, 2019). Usually, choice of hyperparameters is based on previous experience with similar experiments but picking values on a logarithmic scale should give an idea about the order of magnitude taken by each of the parameters that are to be estimated (Goodfellow *et al.*, 2016; Burkov, 2019). Also, repeating grid search on a smaller grid taken around the values from the first grid which gave the better performance can help in optimising results of the analysis (Goodfellow *et al.*, 2016; Burkov, 2019).

In this study, the grid search algorithm was used to calibrate the hyperparameters of SVR, RF and ANN models. In particular:

- for the SVRs, the only hyperparameters to calibrate consist of γ (*gamma*), ε (*insensitive-loss function*), and Δ (*tolerance of termination criterion*). γ is usually taken as the inverse of the number of variables considered (i.e. 1 / (data dimension)). ε and Δ were selected instead to be within the vector [0.0001, 0.001, 0.01, 0.1, 1, 10]. For SVRs, it is possible to apply grid search in R by using the *e1071* package (Meyer *et al.*, 2017). This allows the model to test all possible combinations of ε and Δ and output that which gives best performance in terms of accuracy (Meyer *et al.*, 2017) (see Appendix D for details about how this has been done). Another characteristic of the SVR which should be evaluated and carefully selected is the type of kernel used by the algorithm. However, based on the approach taken in previous studies (Zeng *et al.*, 2015), which modelled variation of fuel consumption of road vehicles using SVRs and used data similar to those of this study, the radial basis (RBF or Gaussian) function was adopted.
- for the RFs, only ntree (number of trees in the forest) and mtry (number of variables considered to build each tree in the forest) must be calibrated. In order to do so, a rule of thumb is to fix mtry to 1/3 of the number of variables considered in the study, while accuracy and robustness of the developed RF model should increase with the number of trees in the forest (Breiman, 2001; James *et al.*, 2013). In order to assess ntree the *randomForest* R package offers a plot of how accuracy of the model improves with ntree. However, that increases asymptotically and therefore a reasonable choice is to select ntree such as accuracy does not decrese significantly anymore.
- for the ANN, performance can vary due to the use of different learning algorithms, architecture (number of layers and neurons in each layer), activation functions and learning rates adopted.

In this study, the logistic activation function was used, while learning rate and architecture of the developed ANNs optimised using the grid search method. In particular, the *caret* package (Kuhn, 2018) was used to apply grid search with learning rate tested among the vector [0.00001, 0.0001, 0.001, 0.01] and architecture tested among all possible combinations of two hidden layers of maximum amount of neurons equal to the amount of input variables plus five.

In the initial phase of the study, which included use of ML, also very large architectures (deep neural networks) were tested as ANNs. That included three hidden layers and a large amount of neurons. However, for the type and quantity of data analysed and problem investigated that brought only to overfitting of the training set with performance that dropped below those of LR in the test and validation sets. However, that may require further investigation in the future, as adopting better learning algorithms, using more accurate data (which in the future might be available as standard) and optimising architecture of the ANN may allow to achieve better performance.

All combinations of the mentioned hyperparameters were explored and performance measured in terms of RMSE, MAE and training time for a single training set of data. The combination, providing the lower RMSE and MAE and a reasonable training time, was finally adopted in the study to model fuel consumption. In particular, with reference to the training time this was considered to be "reasonable" when an additional 30 minutes of calculation, for a single repetition of cross-validation (see following section), did not produce an improvement of RMSE and MAE of over 1%. This delivered a good level of accuracy of the developed models and yet contained the calculation time. Although it would have been interesting to show how changes to the hyperparameters of the developed models affect RMSE, MAE and calculation

time, for brevity, only the results from the optimal combination of hyperparameters are reported in the thesis.

4.9 Cross-validation

ML based models must be trained before being able to make estimates for new cases. However, training of ML models strongly depends on the quantity and quality of the data available, with the increase in data size and complexity of algorithms used, which should significantly reduce the error, eventually converging to a minimum. However, fortunate (or unfortunate) splits of data can affect reliability and robustness of the produced models (James *et al.*, 2013; Goodfellow *et al.*, 2016; Ng, 2018). Therefore, in order to reduce the impact of data splits, leading to predictions completely independent from the how the training and test sets have been obtained, ten-fold cross-validation (James *et al.*, 2013; Goodfellow *et al.*, 2016; Ng, 2018) was used. Basically, in cross-validation the randomized splitting process is repeated *k* times on the available data to generate ten different models for each method (SVR, RF, and ANN). Then for each technique used, performance indicators (e.g. measurements of the error in the generated models) are compared and the average of these used to characterise the developed models. Small variations in the performance of the model for different training and validation sets means that results are not affected by how the data is split, indicating a good level of reliability of the developed model (James *et al.*, 2013; Goodfellow *et al.*, 2016; Ng, 2018). In this study, only some 95 ± 2.5 % of data are used to apply ten-fold cross-validation. Of these, for each repetition, 75 % of the data are used for training the model and 25 % for testing.

Similarly to the process described in Perrotta, *et al.*, (2017a), the remaining 5 ± 2.5 % of the data is used to validate the developed models. Characteristic of this 5 ± 2.5 % of data is that it must not contain any record from trucks considered in the

remaining 95 ± 2.5 % of the data, which allows to validate robustness and reliability of the developed models on completely new records from untested vehicles. For light trucks driving on A roads only a few data are available these have been split in training and test set only and no validation was performed for this vehicle type in the study. However, constancy of performance of the developed fuel consumption models in the cross-validation, should allow anyway evaluation of robustness and reliability of the developed models for light trucks.

4.10 Validation

Following the cross-validation phase, the generated models require validation using new data. Therefore, fuel consumption is computed for the remaining 5 ± 2.5 % of data which is not considered in the cross-validation process. This is initially and randomly sampled by taking from the whole dataset available all records from a set of trucks to make 5 ± 2.5 % of the whole data and this is repeated for each of the ten cross-validation processes (Perrotta *et al.*, 2017a). This means that no data from these trucks is used to train or test the developed models and simulates the collection of new data. This procedure is used to test the reliability of the generated models and shows independence of the final estimates from the training dataset.

4.11 Parametric analysis

A parametric analysis has been performed to understand how each of the considered variables impacts the fuel consumption of the considered truck fleet. For continuous variables, the analysis consists of using the generated models to predict the fuel consumption of trucks for 50 different values of each of the variables considered in the model. The 50 values were chosen to be evenly distributed between the 5^{th} and 95^{th} percentile values of the distribution of the considered variable. In the case of non-

continuous variables, the models are used to make estimates of fuel consumption for the most common values of the considered variable. That is established by analysing the distribution of the data available, selecting the most frequent discrete values and performing the parametric analysis for those values only. While the value of one variable varies between the 5th and 95th percentiles all the others are set to their average. This shows how each of the developed models (LR, SVR, RF and ANN) approximates the relationships between fuel consumption and the explanatory variables, it tests the sensitivity of each of them, their consistency, and allows to investigate their interpretability in consideration of physical/mechanistic principles.

Although results of the parametric analysis can be considered descriptive of the identified trends and allow to visualise the detected patterns, two tables which summarise the results from a numerical perspective are produced. In particular, the tables report the impacts that each of the developed models estimate for each of the variables included. First, impacts are calculated as the difference between the 25th and 75th percentiles of results from the parametric analysis, and this is then compared to the average fuel consumption for the considered dataset (see Equation (28)).

$$Impact\ \% = \frac{PA_{75th} - PA_{25th}}{FC_{avg}} \cdot 100 \tag{28}$$

where, PA_{75th} and PA_{25th} respectively represent the 75th and 25th percentiles of the values of the parametric analysis, while the FC_{avg} is the average fuel consumption of the considered dataset.

Then, these differences are calculated for the range between the 5th and 95th percentiles of the results from the parametric analysis. Also in this case, comparison is made to the average fuel consumption characteristic of the considered dataset (see Equation (29)).

$$Impact\ \% = \frac{PA_{95th} - PA_{5th}}{FC_{avg}} \cdot 100 \tag{29}$$

where, PA_{95th} and PA_{5th} respectively represent the 95th and 5th percentiles of the values of the parametric analysis, while the FC_{avg} is the average fuel consumption of the considered dataset.

Finally, due to the fact that each of the developed models give different estimates of how each of the considered variables affect the fuel consumption of the considered fleets of truck, an average of the impacts calculated by SVR, RF and ANN is made. That corresponds to an "ensemble averaging" of the developed models and helps in reducing variance of the obtained estimates (Goodfellow *et al.*, 2016; Ng, 2018; Burkov, 2019). This also allows to reduce ranges to a single number, which facilitates the comparison of the obtained estimates with the results of studies conducted in the past.

5 Results and Discussion

In this chapter results of the study are presented. These include the results from the variable selection analysis, learning curves, modelling and PA, for each of the considered case study and datasets. For brevity, the chapter offers only a summary of the major findings and a brief discussion. This is to give only an overview of the obtained results and allows comparison with the findings of the studies conducted in the past. Further details regarding the application of the developed method and results of the study can be found in Appendix E.

5.1 Variable selection

Overall, from the analsyis of both the Dataset 1 and 2 it was observed that *Gross.Weight_kg*, *Geom.Gradient*, *Torque.AVG*, *Revs.AVG* and *Acceleration* are the variables that affect the fuel consumption of the considered fleets of truck the most, for any of the considered applications (App 1 and App 2). This is reasonable, as from a physical perspective these variables are expected to be those that affect the fuel consumption of road vehicles the most (see sect. 2.4 and sect. 2.5). Next, the other variable which seems to affect the fuel consumption of the considered fleets of truck significantly is the *Geom.Abs_Radius*. Again, this is in line with the physical/mechanistic principles of vehicle dynamics (e.g. Guzzella *et al.*, 2015), which

confirms the results of studies conducted in the past (e.g. Benbow *et al.*, 2013) and improves the confidence in the variable selection criteria applied.

Focusing on the analysis of the Dataset 1, which comprises data from heavy trucks driving along motorways only, this evidenced that the *Gross.Weight_kg, Geom.Gradient, Speed.AVG, Torque.Start, Torque.End, Revs.Start, Acceleration, Geom.Abs_Radius, TRACS.LPV03m, TRACS.LPV10m, TRACS.LPV30m*, and *TRACS.Pav_Texture* seem to significantly affect the fuel consumption of the considered fleet of trucks. These are twelve out of fifty-six measurements available within the Dataset 1. All these variables were considered in each of the developed models. Considering the same selection of variables in each of the developed models allows full comparability of their performance, which could be an argument of discussion otherwise. This selection was performed mainly because of the results of the Adjusted-R^2, Lasso and RF analyses, but also Pearson's correlation and *p*-values were considered. However, the fact that the correlation between *Fuel.Consumption* and the road surface characteristics is low and the fact that the *p-value* associated with *TRACS.LPV03m* and *TRACS.LPV30m* results are not statistically significant could be due mainly to non-linearity, which these two statistics cannot assess (James *et al.*, 2013).

The variables considered in the study seemed to show different levels of significance depending on the considered dataset (Dataset 1 or 2), truck type and situation (App 1 or App 2). For example, analysis of the Pearson's correlation coefficients and correlation matrices on the different case studies considered, showed that there is no (or very low) linear correlation between road surface characteristics (measured as LPV and SMTD) and the fuel consumption of the considered trucks in both the Dataset 1 and 2. Similarly, analysis of the *p-values* for the light trucks App 1 (driving along A

roads only) showed that *TRACS.LPV03m*, *TRACS.LPV10m*, and *TRACS.LPV30m*, do not seem to have statistically significant impact on the fuel consumption of the considered fleet of trucks. However, their *p-values* (and therefore their impacts on truck fleet fuel consumption) becomes statistically significant in all the other cases analysed (from Appendix E.2.1.2 to E.2.1.6), when the amount of data considered increases. Results from the Adjusted-R^2, Lasso and RF analyses also show that these seem to play a significant role. In particular, the RF algorithm ranks road surface characteristics such as *TRACS.LPV03m*, *TRACS.LPV10m*, *TRACS.LPV30m*, and *TRACS.Pav_Texture*, to be more important than other variables such as *Speed.AVG* (see Figure 30). Although it might not be immediately intuitive, this could be explained by the fact that in the present study only trucks driving at a relatively constant speed, within a narrow range of vehicle speeds (between 80 and 95 km/h) were considered. As an example of the major importance of road roughness and macrotexture on the fuel consumption of the considered fleet of trucks, Figure 30 reports an example of the results of the variable selection operated by the RF algorithm for the case of light and heavy trucks App 1. This is done based on the percentage increase in MSE (%IncMSE, see sect. 4.3.4) that including a certain variable in the developed model produces compared to a model that does not include that variable (Breiman, 2001). From the analysis of the results of the RF variable selection algorithm similar conclusions can be drawn for all the other considered case studies (see Appendix E.1.1 and E.2.1). Figure 30 shows that some of the considered variables are always classified as the highest influencers of the fuel consumption of the considered trucks. As a confirmation of what mentioned earlier, the top three are typically *Gross.Weight_kg*, *Geom.Gradient*, and *Torque.AVG*. This is also confirmed by the results of the performed PA (see Figure 32, 33, 34, 35, 36 and 37). Although

the road surface conditions are never amongst the first two or three positions of the rank, it is possible to see that, road roughness (as LPV) and the macro-texture of the road surface (as SMTD) are generally considered more important than other variables such as the *Speed_AVG* and *WindSpeedDir_kn* (see Figure 30), which other studies conducted in the past considered more important than the road surface characteristics for modelling the vehicle fuel consumption.

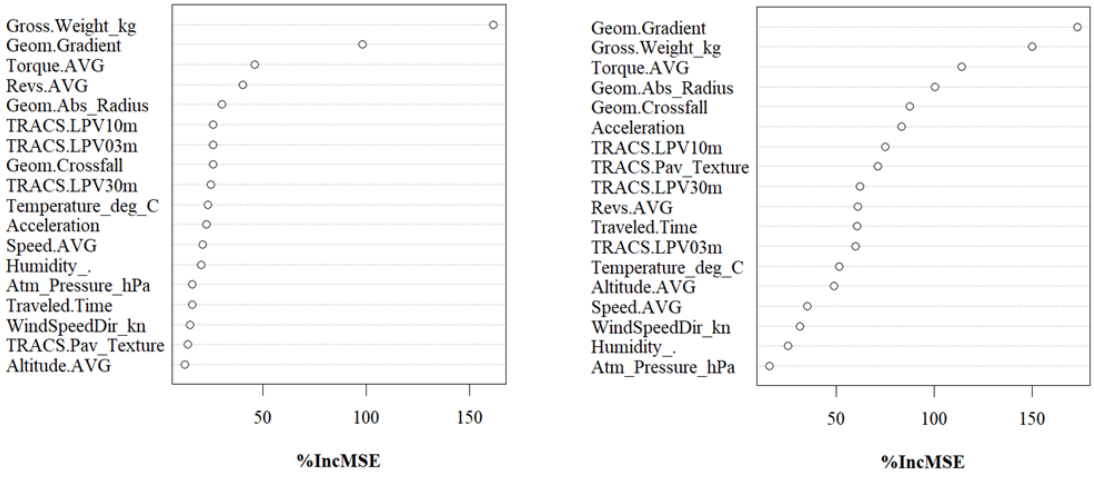


Figure 30. Example of the results of the variable selection operated by the RF algorithm (in terms of %IncMSE) for (from left to right) light trucks App 1 and heavy trucks App 1.

Regarding weather conditions, the variables, which all the applied techniques seem to identify to play a significant role in the estimation of the fuel consumption of the considered fleet of trucks, are *Atm_Pressure_hPa*, *Temperature_deg_C*, and *WindSpeedDir_kn*. Since these variables might affect vehicle aerodynamics and engine efficiency, this is reasonable and confirmed by results of studies that in the past applied a physical/mechanistic approach (see Figure 30).

Other variables, such as *Altitude.*AVG, *Traveled.Time*, *Geom.Crossfall*, and *Humidity_.*, seem not always correlated to the fuel consumption of the considered

fleets of truck, but they seem to be identified to significantly impact the fuel consumption of trucks by different statistics each time. For example, in regard to *Humidity_.*, this is identified to be significant by the Lasso and the RF techniques for the heavy trucks App 2 (see Appendix E.2.1.6), but the same variable is not considered significant by the same statistics for the medium trucks (App 1 and App 2) or the heavy trucks App 1 (see Figure 30). Similarly, for the *Geom.Crossfall*, the RF analysis identifies the variable to be significant for the medium trucks App 2 (see Appendix E.2.1.4) and the heavy trucks (App 1 and App 2) (see Figure 30 and Appendix E.2.1.6), but the same algorithm classifies this to be of low importance for the light (App 1 and App 2) (see Figure 30 and Appendix E.2.1.2) and the medium trucks App 1 (see Appendix E.2.1.3). This could mean that the *Geom.Crossfall* is more significant for heavier vehicles, but further investigation is required before definitive conclusions can be drawn. Additionally, for these variables, mixed results are obtained also when correlation matrices, *p-values* and adjusted-R^2 are considered, which makes hard the choice to include them in the modelling or not.

Generally, results of the variable selection analysis conducted show that different vehicle types could be affected by different variables in different ways and optimisation of the developed models could consider including different variables to model the fuel consumption of different vehicle types. However, in this study, in order to ensure comparability of the developed models and their performance, a final and homogenous decision about the variables to consider in the developed models was made. This was done based on the observations made in the study, results from the literature and to meet the objectives of the study. Because of this, variables such as the GPS location of the considered trucks (i.e. *Lat.Start*, *Lat.End*, *Long.Start*, and *Long.End*) was excluded, and variables such as *TRACS.LPV03m*, *TRACS.LPV10m*,

TRACS.LPV30m, and *TRACS.Pav_Texture* were always included in the modelling of fuel consumption. Additionally, because especially in regard to SVR, RF and ANN a reduced number of variables should enhance interpretability of the developed models, eighteen out of the seventy-seven variables available for the Dataset 2 were selected to be included in the developed fuel consumption models for light, medium and heavy trucks. These are the *Gross.Weight_kg*, *Geom.Gradient*, *Torque.AVG*, *Revs.AVG*, *Speed.AVG*, *Acceleration*, *Traveled.Time*, *Geom.Abs_Radius*, *Geom.Crossfall*, *TRACS.LPV03m*, *TRACS.LPV10m*, *TRACS.LPV30m*, *TRACS.Pav_Texture*, *Altitude.AVG*, *Atm_Pressure_hPa*, *Temperature_deg_C*, *Humidity_.*, and *WindSpeedDir_kn*. These represents some of the most common variables considered by other studies conducted in the past (e.g. Hammarström *et al.*, 1987, 2008; Chatti *et al.*, 2012; Benbow *et al.*, 2013; Almér, 2015) and their significance was assessed based on results from multiple statistics applied to a large quantity of data. This should improve confidence on the choice of variables made, which should allow the developed models to achieve a reasonable level of accuracy, avoiding any overfit, and ensuring, at the same time, a better interpretability of the estimates.

Finally, Table 17 summarises the results of the variable selection process, for each of the analysed datasets and applications of the developed method.

Table 17 – Summary of the results from the performed variable selection.

Variables	Light trucks		Medium trucks		Heavy trucks		
	App 1	App 2	App 1	App 2	Dataset 1	App 1	App 2
Gross.Weight_kg	•	•	•	•	•	•	•
Geom.Gradient	•	•	•	•	•	•	•
Torque.Start					•		
Torque.End					•		
Torque.AVG	•	•	•	•		•	•
Revs.Start					•		
Revs.AVG	•	•	•	•		•	•
Speed.AVG	•	•	•	•	•	•	•
Acceleration	•	•	•	•	•	•	•
Cruise.Control							
Traveled.Time	•	•	•	•		•	•
Geom.Abs_Radius	•	•	•	•	•	•	•
Geom.Crossfall	•	•	•	•		•	•
TRACS.LPV03m	•	•	•	•	•	•	•
TRACS.LPV10m	•	•	•	•	•	•	•
TRACS.LPV30m	•	•	•	•	•	•	•
TRACS.Pav_Texture	•	•	•	•	•	•	•
Altitude.AVG	•	•	•	•		•	•
Atm_Pressure_hPa	•	•	•	•		•	•
Temperature_deg_C	•	•	•	•		•	•
Humidity_.	•	•	•	•		•	•
WindSpeedDir_kn	•	•	•	•		•	•

• this indicates that the variable is considered in that case study.
Information regarding how each of these variables rank one compared to another for each of the considered case studies can be found in Figure 41, 65, 68, 71, 74, 77 and 80 in Appendix E.

5.2 Learning curves

Table 18 reports the results of the analysis of the learning curves calculated for each case study and algorithm tested. The table reports an approximation of the minimum amount of data required for the utilised algorithm to converge. In particular, an average of the data required to stabilise both the RMSE and MAE is reported. As a clarification, results from the learning curves are only indicative and different datasets may bring slightly different results. However, this technique seems to be a powerful and practical tool which can help one to determine an approximated amount of data required for the utilised algorithms to converge. From the table (see Table 18), it is possible to see that while some of the available datasets are not enough to allow the algorithms to stabilise their performance, analysis of larger quantities of data shows that some algorithms learn quicker than others requiring lower amounts of data to converge. Analysis of the data from heavy trucks App 1 and App 2 shows that the algorithms that achieves the more stable performance quicker is the ANN, followed by the SVR and then the RF. Due to time constraints, it was not possible within this study to determine the minimum amount of data to use to train the ANN1. In this regard, based on the results of the analysis of the data from heavy trucks App 1 and App 2, what it is possible to say is that the minimum amount of data required to train the ANN1 is higher than 291,073, possibly close to 300,000 records, for the considered selection of variables and driving conditions (e.g. constant speed).

In Appendix E.1.2 and E.2.2.6 results of the analysis of the learning curves for the ANN developed for the analysis of the Dataset 1 and for the ANN developed for the analysis of the data from heavy trucks App 2 from the Dataset 2 are reported as an example.

Table 18 - Learning curves: approximated amount of data needed for each ML algorithm to converge.

Algorithm	Light trucks		Medium trucks		Heavy trucks		
	App 1	App 2	App 1	App 2	Dataset 1	App 1	App 2
SVR	> 12,844	> 119,661	> 167,111	> 197,753	> 14,291	200,000	200,000
RF	> 12,844	> 119,661	> 167,111	> 197,753	> 14,291	250,000	250,000
ANN	> 12,844	> 119,661	> 167,111	180,000	> 14,291	180,000	180,000
ANN1	> 12,844	> 119,661	> 167,111	> 197,753	> 14,291	> 278,607	> 291,073

• this indicates that the variable is considered in that case study.

5.3 Modelling

Using the selection of variables outlined in sect. 5.1, LR, SVR, RF and ANN models have been developed for each of the considered datasets and case studies. In addition, application of deep learning (ANN1) was tested for the case of heavy trucks on the Dataset 1 and Dataset 2 App 2.

In comparison to the LRs, all the developed ML models are able to obtain better performance in terms of of R^2, RMSE and MAE. This was true for all the case studies examined and for any amount of data considered. This could be due to the effect of non-linearity, which may play an important role also in this case where relatively controlled conditions (e.g. relatively constant speed) have been investigated. Figure 31 shows how the LR, SVR, RF and ANN developed for the heavy trucks App 2 fit the available training, test and validation sets for one of the ten cross-validation repetitions performed. This represents only one of the seven case study considered in the present research work. As mentioned earlier, the full list of figures and description of the obtained results are reported in Appendix E.1.2 and E.2.2. From the figure (see Figure 31), it is possible to see that either SVR, RF or ANN reduce the variance of the obtained estimates significantly and appear more accurate than the LR. This is also confirmed by the analysis of the performance of the developed models performed in terms of R^2, RMSE and MAE (see Table 19).

It is important to highlight that while performance of the LR model tends to be stable independently of the amount of data considered, those of the ML models tend to increase with the amount of information available and utilised to train the developed models. In fact, it is with the ~ 200,000 records, used to train the developed ML models for the heavy trucks App 1 and App 2 (part of the Dataset 2) that SVR,

RF and ANN demonstrate the highest level of performance. This agrees with the result of the analysis of the learning curves (see sect. 5.2).

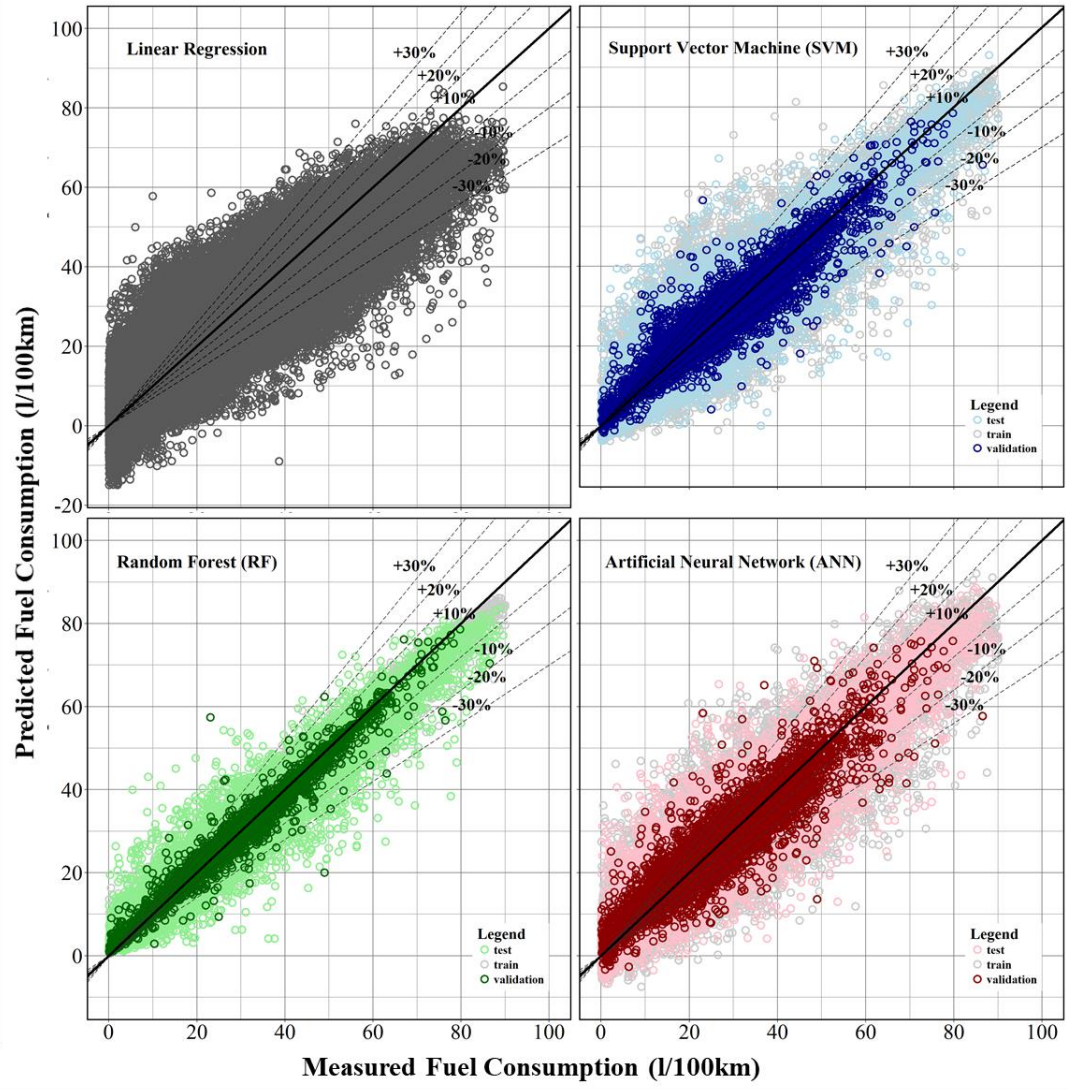


Figure 31. Example of how each of the developed models LR, SVR, RF and ANN fit the available data. This represents the case of the heavy trucks App 2.

In addition, based on the results of the analysis of the learning curves, it is possible to say that it is unlikely that increasing the amount of data available will cause a significant improvement of the performance of any of the developed models apart from the ANN1. In fact, due to the relatively small datasets analysed, it was not possible to determine the minimum amount of data necessary to stabilise the

Table 19 - Comparison of the performance of fuel consumption models for heavy trucks.

Study	Model	R^2	RMSE (l/100km)	MAE (l/100km)
Chatti *et al.* (2012)	HDM-4	0.88	5.29	-
Almér (2015)	LR	-	16.5	-
	SVR	-	14.1	-
	RF	-	12.9	-
	ANN	-	15.5	-
Perrotta (2019)* (*the present study*)	LR	0.76	6.50	4.67
	SVR	0.89	4.36	2.95
	RF	0.94	3.31	2.17
	ANN	0.88	4.54	3.18
	ANN1	0.92	3.76	2.64

* Results from the analysis of the Dataset 2 App 2 are reported.

performance of the ANN1, not even when considering the dataset from heavy trucks in the App 2 (which was the largest considered). Although it is recommended that a more precise estimate is made, it is believed that analysis of about 300,000 records should allow stabilised performance of the ANN1 for the considered vehicle type and selection of variables considered in the present study. On the other hand, in order to improve the performance of the other developed models, different solutions should be investigated and tested. This should include variation of the data collection frequency from trucks, analysis and identification of the outliers and hyperparameters optimisation (e.g. use of random search instead of grid search (Goodfellow *et al.*, 2016; Burkov, 2019)) to determine the structure and parameters of the developed models).

Due to use of the HDM-4 software in road asset management at global level and due to the similarities of the approach undertaken by Almér (2015), results from the studies by Chatti *et al.* (2012) and Almér (2015) have been used for performance evaluation and comparison. First, evaluation of the methods applied has been evaluated and results reported in Table 20. The table shows how, although the approaches used in this study and that by Almér (2015) are quite similar, main differences can be found in:

- the amount of data used by Almér (2015) does not allow one to consider the study as an application of Big Data,
- the reduced amount of statistics used by Almér (2015) to determine the selection of variables to develop his models, are demonstrated not to be optimal to ensure the repeatability of the obtained results (see sect. 5.1),
- no validation of the results has been performed by Almér (2015) and results might be affected by how the amount and type of data collected,

Table 20 - Comparison between studies conducted in the past and the present study of the approaches adopted to model the fuel consumption of road vehicles.

Study	Type of approach	Models	Variable selection	Validation	Impact assessment	Road conditions
Chatti *et al.* (2012)	physical/mechanistic	HDM-4	regression analysis (only *p-values*)	LR	ANCOVA	Yes
Almér (2015)	ML	LR SVR RF ANN	Pearson's correlation 2-way ANOVA Friedman Test	training and test	-	No
Perrotta (2019) (*the present study*)	Big Data / ML	LR SVR RF ANN ANN1	Pearson's correlation regression analysis (*p-values*, adjusted-R^2) Lasso RF	10-fold crossvalidation and validation	PA	Yes

- the study (Almér, 2015) does not consider the effect of road surface characteristics, which would not allow the application of the developed mdoels for applications in road asset management,
- no PA has been performed, which did not allow Almér (2015) to estimate the impact of each of the considered variables on the fuel consumption of the considered fleet of trucks.

All of these research gaps, which were partially considered and fulfilled in the study of Chatti *et al.* (2012), have been considered and addressed in the present research.

In terms of performance, Table 19 summarises the performance of the studies conducted by Chatti *et al.* (2012) and Almér (2015), together with those of the present study, for heavy trucks, in terms of R^2, RMSE and MAE. For brevity, results from the analysis of data from the heavy trucks App 2 are reported in the table. Results for the remaining case study can be found in Appendix E.1.2 and E.2.2. From Table 19, it is possible to see that previous studies did not always report a wide range of key performance indicators such as in the present case study. This makes the RMSE the only indicator which allows a comparison of performance between the three. In these terms, it is possible to see that all of the ML models developed in the present study outperform any of the models developed by both Chatti *et al.* (2012) and Almér (2015). Also in terms of R^2, the performance of each of the developed ML models are superior to those of the model developed by Chatti *et al.* (2012). This means that application of the developed methodology and used algorithms is able to produce models as accurate as, or sometimes superior, to commercially available solutions. Therefore, by making use of the large quantity of information collected in the HAPMS and by collaborating with truck fleet managers and weather institutions, application of

the developed method could represent an opportunity for road managers to obtain better estimates of the fuel consumption of road vehicles at network level.

Due to the lack of information, no comparison is possible between the present study and that of Almér (2015) in terms of R^2. However, due to the large difference noticed between the two studies in terms of RMSE, it is likely that performance of the models developed by Almér (2015) are in fact inferior, in terms of R^2, to the performance of the models developed in the present study and this is potentially due to Almér's (2015) reduced amount of data utilised, lack of optimisation of the variables selection process and lack of cross-validation, which have been performed in the present research.

5.4 Paramteric analysis

Results from the performed PAs (see Appendix E.1.3 and E.2.3) highlighted that while for some parameters the utilised algorithm seem to identify similar trends, this is not always the case for others.

Table 21 and 22 show how each of the utilised algorithms estimate variation of the fuel consumption of the considered fleet of trucks for variation of *Gross.Weight_kg* and *Geom.Gradient* for each of the considered case studies and datasets. For example, from these tables (Table 21 and 22) it is interesting to see that while from the analysis of the App 1 of light, medium and heavy trucks and heavy trucks Dataset 1 results are not quite consistent between algorithms, these tend to converge in the App 2 when higher quantities of data have been considered. This might be linked to the fact that ML algorithms are more stable for higher amounts of data fed to them during the training phase and could be, in fact, a confirmation of the findings from the analysis of the learning curves. In particular, this could highlight that for higher amounts of

data available, performance of SVR, RF and ANN tend to converge not only in terms of R^2, RMSE and MAE, but also in terms of PA.

As a confirmation of what has been said above, Figure 32 and 33 show a representation of the variation of the fuel consumption of the considered fleets of truck for variation of *Gross.Weight_kg* and *Geom.Gradient*, for all the developed models in two of the seven of the considered case study.

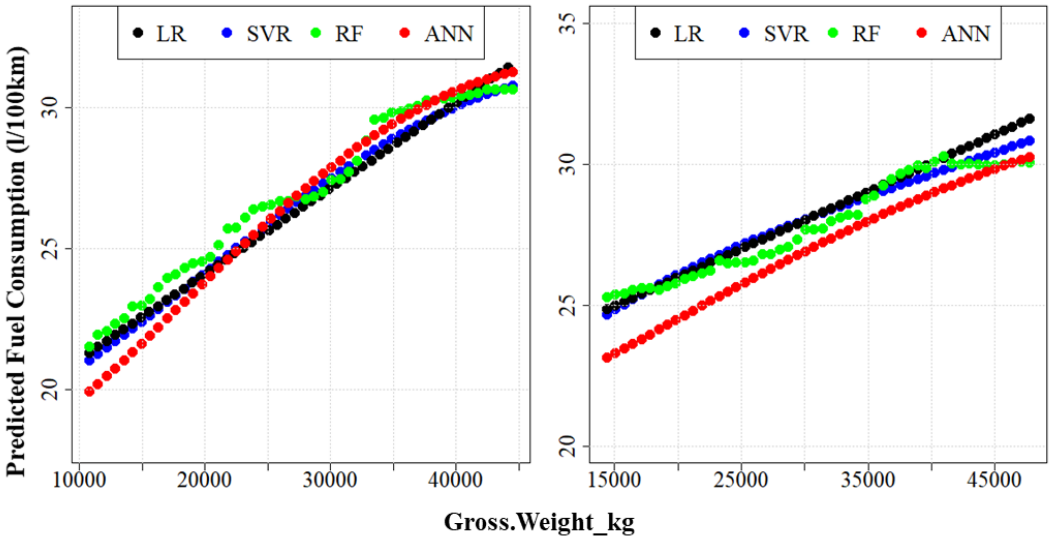


Figure 32. Example of the PA performed for (left to right) the medium trucks App 2 and heavy trucks App 1 for variation of the *Gross.Weight_kg* (kg).

Table 21 - Summary of the results of the PA for the *Gross.Weight_kg* (kg).

Truck	Data	LR	SVR	RF	ANN
Light	App 1	32.97	45.14	30.95	44.23
	App 2	43.45	45.24	44.90	48.57
Medium	App 1	37.65	39.76	25.27	35.13
	App 2	37.48	37.14	35.11	43.60
Heavy	Dataset 1	17.87	6.87	9.26	6.95
	App 1	22.65	20.72	16.77	24.29
	App 2	24.00	24.07	24.51	20.60

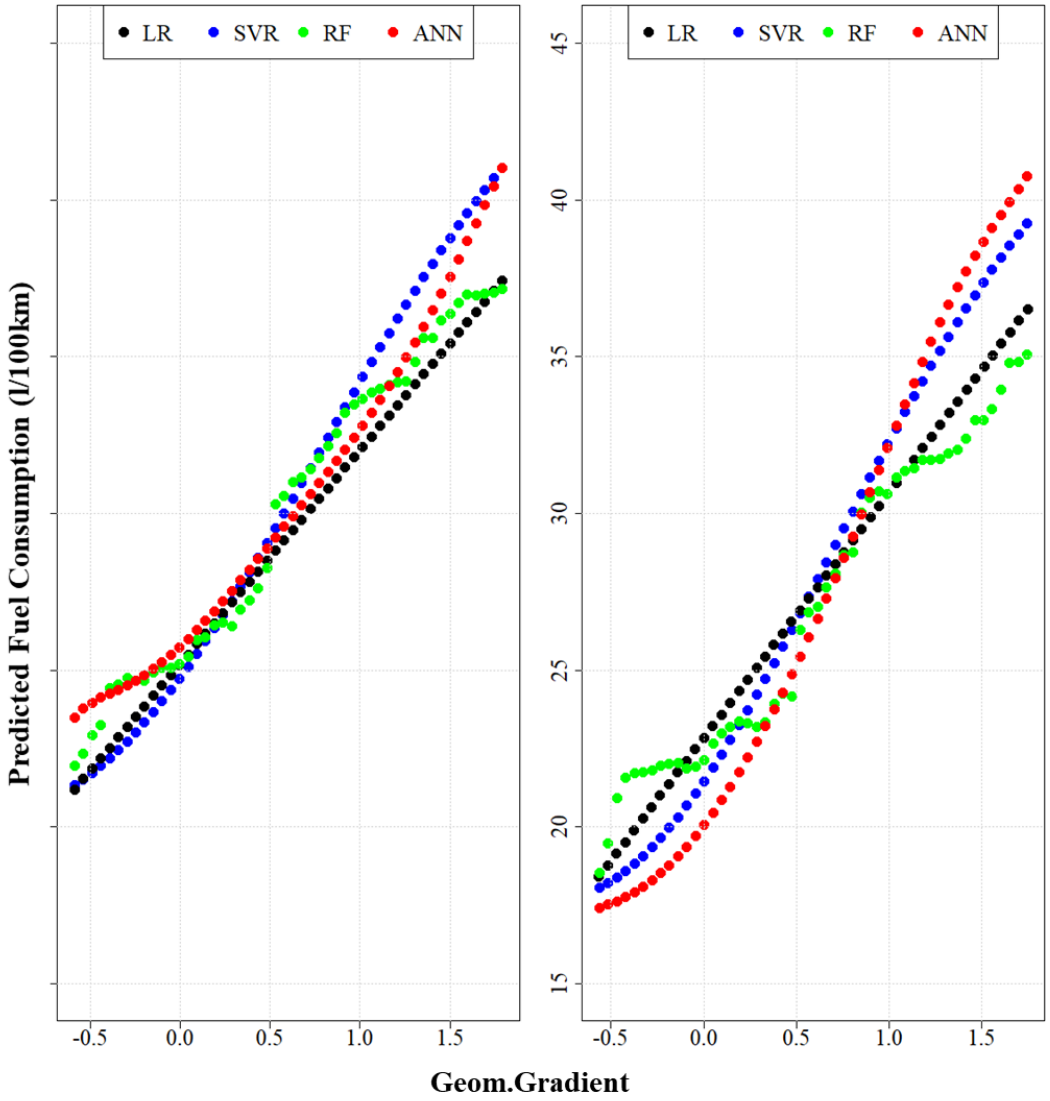


Figure 33. Example of the PA performed for (left to right) the medium trucks App 1 and heavy trucks App 1 for variation of the *Geom.Gradient* (%).

Table 22 - Summary of the results of the PA for the *Geom.Gradient* (%).

Truck	Data	LR	SVR	RF	ANN
Light	App 1	47.83	60.70	54.66	49.75
	App 2	66.76	70.10	71.86	76.24
Medium	App 1	61.72	76.41	60.60	64.21
	App 2	86.69	115.34	98.14	115.72
Heavy	Dataset 1	70.65	48.54	29.58	32.59
	App 1	60.79	72.88	52.88	80.42
	App 2	79.51	113.9	105.1	125.7

The two figures (Figure 32 and 33) show that all the developed models tend to identify similar trends for both the considered variables (*Gross.Weight_kg* and *Geom.Gradient*) in each of the considered case study. However, this is not the case for the road conditions for which, in fact, it is difficult to find consistencies between different algorithms across the different type of conditions investigated. For example, by looking at Table 23, it is possible to see that no couple of algorithms tend to agree on the impact of *TRACS.LPV03m*. Similarly, it is not possible to find any particular pattern among the other results from the analysis of variation of *TRACS.LPV10m, TRACS.LPV30m* and *TRACS.Pav_Texture* (see Table 24, 25 and 26). This could be due to the fact that although the currently analysed amount of data is good enough to stabilise the performance of the algorithms in terms of R^2, RMSE and MAE and for the PA estimation of the impact of the most significant variables such as *Gross.Weight_kg* and *Geom.Gradient* (see Table 21 and 22), this is not actually enough to stabilise the performance of the algorithms on variables that play a less evident role such as the road surface characteristics. This behaviour could be due to the presence of outliers which might have been determined by:

- the assumptions made in the study that trucks tend to drive mainly on lane 1, which might have resulted in assigning the wrong road surface characteristics to the truck performance;
- imprecisions from the GPS data from trucks and the TRACS, which did not allow the average road characteristics to be assigned to the actual truck location,
- actual observational errors from both the trucks and TRACS data, and
- the frequency of data collection from turcks. In fact, this is set by default to every minute or more by Microlise. However, averaging road conditions along the segment of road travelled by a truck at high speed, may minimise the effect of the

road defects and thus end up assigning a condition of the road surface which is non-representative of the actual conditions along the road.

For all these reasons, some bias might be present in the data which does not facilitate the ML algorithms to distinguish the actual trends from the outliers. This could be the reason also why correlation between the recorded road surface characteristics and the fuel consumption of the considered fleet of trucks is low and only a large amount of data allows the traditional statistics to identify certain variables as significantly impacting the target variable.

At confirmation of what has been mentioned above a graphical representation of the PA performed for some of the considered case study is reported. Figure 33, 34, 35 and 36 show the results of the PA performed for each of the developed models for variation of the *TRAC.LPV03m*, *TRACS.LPV10m*, *TRACS.LPV30m* and *TRACS.Pav_Texture* respectively.

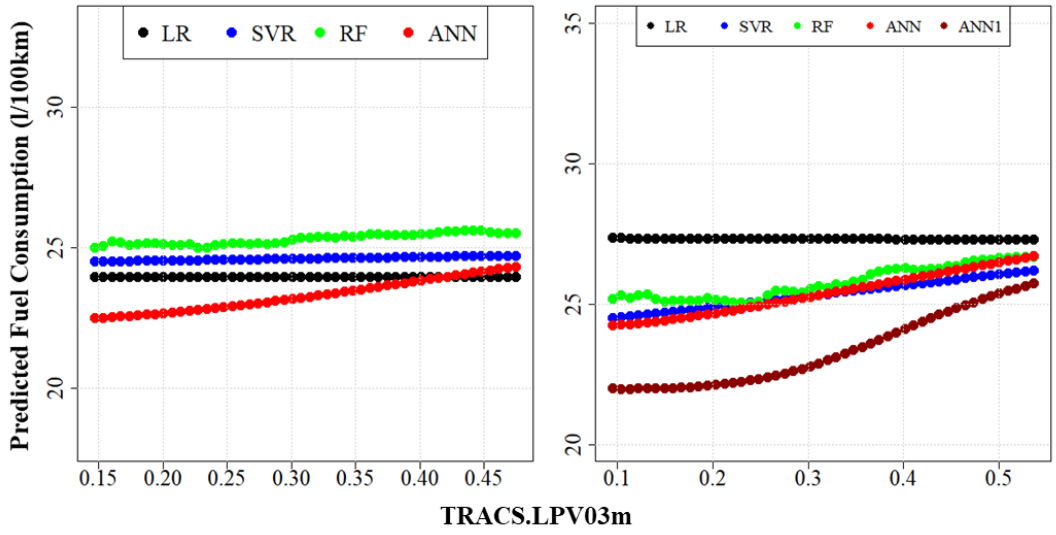


Figure 34. Example of the PA performed for (left to right) the medium trucks App 1 and heavy trucks App 2 for variation of the *TRACS.LPV03m* (mm^2).

Table 23 - Summary of the results of the PA for the *TRACS.LPV03m* (mm²).

Truck	Data	LR	SVR	RF	ANN
Light	App 1	0.04	0.72	4.72	13.38
	App 2	0.16	-0.18	-1.23	1.52
Medium	App 1	0.03	0.81	2.10	6.99
	App 2	-0.17	5.65	1.05	-3.27
Heavy	Dataset 1	-0.10	-2.89	0.80	4.56
	App 1	0.07	0.65	0.34	-2.53
	App 2	-0.17	5.72	4.80	8.33

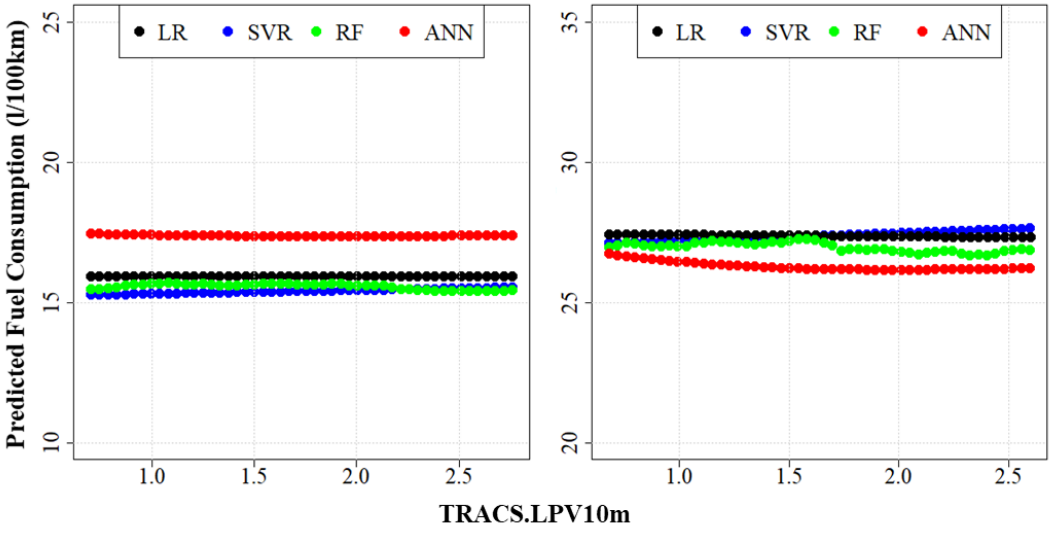


Figure 35. Example of the PA performed for (left to right) the light trucks App 1 and heavy trucks App 1 for variation of the *TRACS.LPV10m* (mm²).

Table 24 - Summary of the results of the PA for the *TRACS.LPV10m* (mm²).

Truck	Data	LR	SVR	RF	ANN
Light	App 1	0.09	1.50	-0.41	-0.34
	App 2	-0.76	-1.35	-0.33	0.83
Medium	App 1	-0.33	2.87	0.31	8.73
	App 2	-0.19	-3.67	0.52	0.14
Heavy	Dataset 1	-2.62	3.27	-0.78	-7.05
	App 1	-0.40	1.74	-0.73	-1.77
	App 2	0.13	-0.33	-0.57	4.48

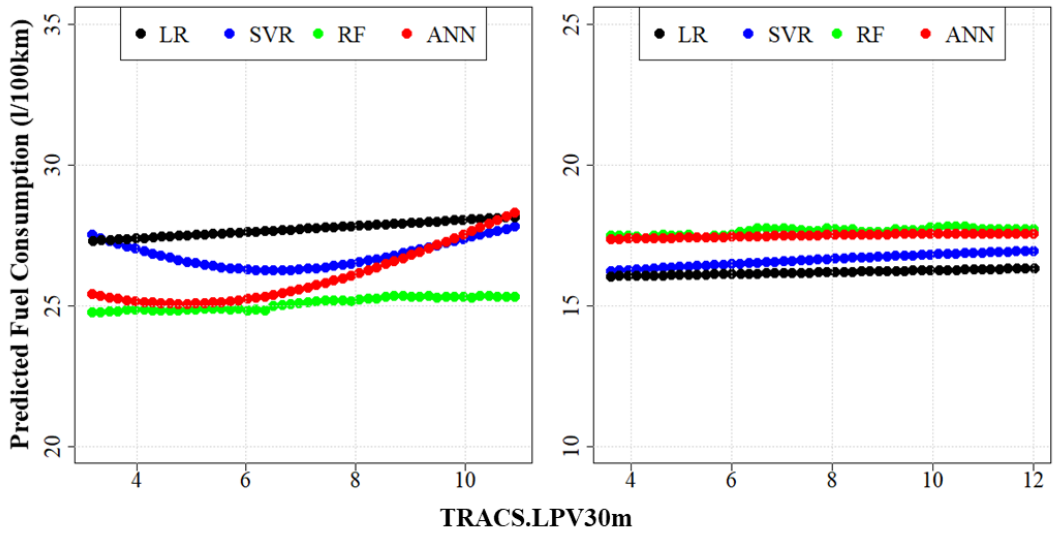


Figure 36. Example of the PA performed for (left to right) the Dataset 1 (heavy trucks) and light trucks App 2 for variation of the *TRACS.LPV30m* (mm^2).

Table 25 - Summary of the results of the PA for the *TRACS.LPV30m* (mm^2).

Truck	Data	LR	SVR	RF	ANN
Light	App 1	-0.08	0.14	-2.80	-6.07
	App 2	1.61	4.21	1.46	1.12
Medium	App 1	1.03	-3.75	-1.21	-20.71
	App 2	2.07	0.04	-0.58	-9.40
Heavy	Dataset 1	2.79	0.67	2.06	9.28
	App 1	0.61	-1.40	2.91	-8.06
	App 2	1.76	-10.58	-2.14	-0.23

The figures (see Figure 34, 35, 36 and 37) show that, regarding the effect of the road surface characteristics, although there are cases where (from a graphical perspective) some models agree regarding the impact of these variables on fuel consumption, generally this is not the case, with estimates that suffer of high variance.

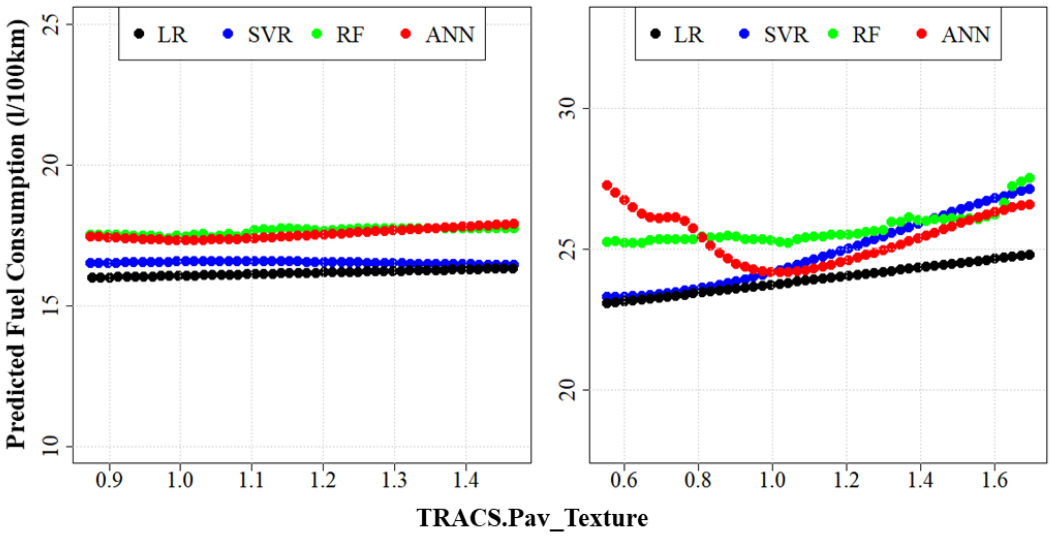


Figure 37. Example of the PA performed for (left to right) the light trucks App 2 and medium trucks App 2 for variation of the *TRACS.Pav_Texture* (SMTD, mm).

Table 26 - Summary of the results of the PA for the *TRACS.Pav_Texture* (mm).

Truck	Data	LR	SVR	RF	ANN
Light	App 1	3.61	3.41	-1.29	1.02
	App 2	1.91	-0.31	1.45	2.49
Medium	App 1	6.56	14.87	5.65	-2.57
	App 2	4.46	-3.91	2.38	-2.09
Heavy	Dataset 1	0.81	0.09	1.96	-3.59
	App 1	5.84	15.64	5.43	2.46
	App 2	5.42	-9.27	0.29	2.00

5.5 Summary

Analysis of the Dataset 1 showed that, even when a relatively small dataset is used to train SVR, RF and ANN, these are able to outperform the LR in terms of R^2, RMSE and MAE. The advantage of the LR model remains though the easier interpretability.

Analysis of the Dataset 1, inclusive of data from heavy trucks only, driving along part of the M1 and the whole M18, allowed investigation of only a limited range

of road surface conditions (between 1.58 m/km and 2.15 m/km IRI, and between 0.95 mm and 1.46 mm MPD). This, together with the relatively low amount of data available, resulted in the impossibility for the developed models to detect clearly defined trends. Results of the parametric analysis performed on the models developed using the Dataset 1 seem to indicate that small wavelengths of road roughness (measured as LPV at 3 and 10 metres wavelengths) and macro-texture (measured as SMTD) have a relatively small impact on the fuel consumption of the considered fleet of heavy trucks (see Table 29). The impact of larger wavelengths of roughness, such as those of LPV at 30 metres seems to be more significant. Overall, analysis of the Dataset 1 indicates that the Big Data approach is feasible and, considering a larger amount of data from more diverse conditions, could enable one to evaluate the effect the characteristics of the road surface have on truck fleet fuel consumption.

Data in the Dataset 2 were collected in a larger quantity than that of the Dataset 1 and was selected to allow investigation of a wider range of vehicle speeds and road surface conditions. Unfortunately, this was only partially achieved as inclusion of the A roads under the management of HE and part of the A41 managed by the Oxfordshire County Council did not allow investigation of very high levels of road roughness and macro-texture (between 1.58 m/km and 2.56 m/km IRI, and between 0.55 mm and 1.70 mm MPD). This limitation of the present study could be linked to two factors: 1) road surface conditions are quite good on roads that are part of the SRN and that are primary roads in the UK, and 2) the low frequency of data collection from the SAE J1939 truck sensors makes the average of road surface conditions over a mile (or more) poorly affected by the extreme conditions of the road surface, which might be characteristic of relatively short road segments. Although Benbow *et al.* (2011) showed that over 90% of the SRN in the UK is in relatively good

conditions, analysis of the data available evidenced that it is possible to find relatively short segments of road in bad conditions (see Appendix C). For this reason, as a recommendation for future research work conducted in the field, one should try to increase the frequency of collection of data from the SAE J1939 truck sensors to something between 10m and maximum 100m (which is the standard frequency of data collection of road surface characteristics in the UK). This should enable one to correlate the truck fuel consumption to road characteristics averaged over smaller road segments and therefore more clearly show how road roughness and macro-texture of the road pavement affect consumption. Results from the variable selection analysis performed on the Dataset 2 indicate that similar variables affect the fuel consumption of light, medium and heavy trucks and that some of them may introduce non-linearity in the problem. Consideration of additional variables, to those considered for the Dataset 1, such as additional truck sensors readings and weather conditions, together with the larger amount of data fed to the ML algorithms allowed the development of models that achieve better performance in terms of R^2, RMSE and MAE. Among the ML models, the RF model is the one able to achieve the highest level of performance, better than the SVR and the ANN. However, this could be due also to some overfitting (i.e. the ANN1 in Figure 38) produced by the ML algorithms and, in terms of interpretability of the results of the developed models, the LR remains the best. The fact that all the developed models identify variables such as the *Gross.Weight_kg*, *Geom.Gradient*, *Torque.AVG* and *Revs.AVG* to be among the most important variables to model the fuel consumption of the considered fleet of trucks improves confidence in the adopted approach, which nevertheless requires optimisation.

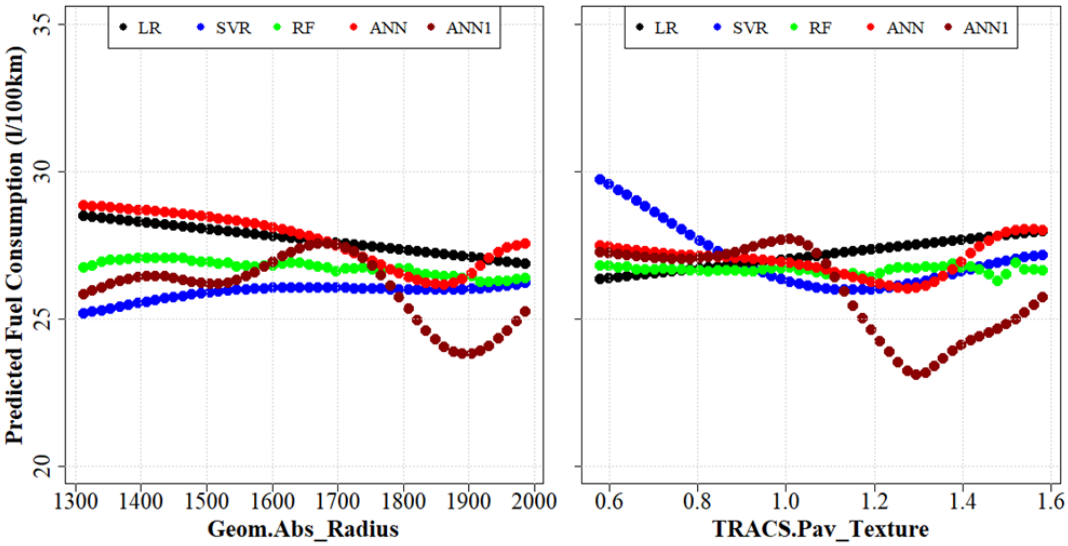


Figure 38. Example of overfitting by the ML models on the *Geom.Abs_Radius* and *TRACS.Pav_Texture* for the case of heavy trucks App 2.

The fact that the *Gross.Weight_kg* is estimated from engine performance parameters by Microlise might introduce collinearity issues and increase variance of the estimates made by the developed models. In this regard, further improvements to the Microlise algorithm, or future introduction of standard sensors to directly measure the weight of trucks, could help at improving robustness and reliability of the developed models. The parametric analysis seems to highlight a decrease of fuel consumption with higher temperature and an increase with higher wind speed at the front of trucks. This is physically explainable and improves confidence that the developed models work well. However, since it was not within the scope of the study to target the effect of weather conditions on truck fleet fuel consumption, the author suggests that further research should be conducted in the area, to improve the adopted method and possibly reliability of the obtained estimates in this regard.

Regarding the effect of road surface characteristics on truck fleet fuel consumption, it is possible to say that neither analysis of the Dataset 2 allows to quantify that clearly. Each of the developed models detect for these variables different

trends, which only sometimes match the results of studies conducted in the past. This is probably due to the small range of road surface characteristics investigated (yet characteristic of the SRN of England) and because of the low frequency of data collection of the SAE J1939 truck telematic data available, which is not designed to conduct the attempted analysis. As mentioned earlier, collection of truck telematic data with a frequency close to that of road surface conditions stored in the HAPMS (e.g. 10m or 100m), should enable one to better correlate the quantity of fuel consumed by the considered vehicles to the characteristics of the road surface. In addition, refinement and optimisation of the developed method could be the key to improve the interpretability and accuracy of the developed models.

Table 27, 28 and 29 summarise the results of the present study with focus on the estimation of the effect of road surface characteristics typical of the SRN of England on truck fleet fuel consumption. In order to reduce variance, and for clarity, averages of the obtained estimates made by the SVR, RF and ANN between the 5^{th} and 95^{th} percentiles are reported in the tables. Because previous studies (see Table 8) investigated the effect of unevenness using measurements of IRI, which in England was found not to be correlated with LPV at 30 metres wavelengths (Benbow *et al.*, 2011), only results to LPV at 3 and 10 metres wavelengths are compared to an equivalent IRI. However, it is important to note that, in the present study, some of the algorithms applied for variable selection found LPV30 to be a variable that, statistically, can impact the fuel consumption of the considered fleets of truck significantly and, for this reason, its effect was included in each of the developed models (see sect. 5.1).

Table 27 shows the results of the study obtained for light trucks.

Table 27 - Summary of the results obtained for light trucks (at 87.5 km/h).

Vehicle	App	Explored range of LPV (mm^2)	Effect of LPV	Explored range of SMTD (mm)	Effect of SMTD
light truck	D2 App 1 (A roads)	0.15 to 0.70 LPV 3 m (1.86 to 2.68 m/km IRI)	+6.27%	0.55 to 1.70 (0.86 to 2.22 mm MPD)	+1.04%
		0.70 to 2.70 LPV 10 m (1.76 to 2.56 m/km IRI)	+0.25%		
		3 to 20 LPV 30 m	-2.91%		
	D2 App 2 (M and A roads)	0.08 to 0.36 LPV 3 m (1.67 to 2.25 m/km IRI)	+0.04%	0.80 to 1.55 (1.18 to 2.05 mm MPD)	+1.21%
		0.50 to 1.70 LPV 10 m (1.63 to 2.22 m/km IRI)	-0.28%		
		3 to 12 LPV 30 m	+2.26%		

To give some context, for this truck type, 9,633 records were used to train the developed models in the App 1 and 85,199 records in the App 2 (when analysis of the learning curves, see Figure 123 and 124, showed that a minimum of about 150k is necessary to obtain stable models). The results are given by an average of the estimates made by the developed RF, SVR and ANN models. From the table, it is possible to see that when only primary roads are considered (App 1) the effect of LPV at 3 and 10 metres wavelength is estimated at +6.27% and +0.25% over a range between 1.76 and 2.68 m/km IRI. The fact that, when motorways are included in the analysis (App 2), the effect of IRI is reduced to +0.04% and -0.28% (for LPV at 3 and 10 metres respectively) might be due to the fact that data from motorways (typically in better conditions) represent the vast majority, which makes the developed MLmodels unable to detect any clear trend through the data. Chatti *et al.* (2012) estimated that, in the US, for similar vehicles and similar vehicle speed, the effect of IRI within a similar range is of approximately +1%, which partially agrees with the results obtained in the present study, for the UK. Regarding the effect of LPV at 30 metres wavelengths on the fuel consumption of light trucks, it is possible to see (Table 27) that this seems to

counterbalance the effect of LPV at 3 and 10 metres wavelengths reducing the overall effect of road roughness on fuel consumption. Regarding the effect of MPD on the fuel consumption of light trucks, it is possible to see that (Table 27) this is estimated to be +1.04% when considering primary roads only (App 1), and +1.21% including motorways in the conducted analysis (App 2). This highlights that, although including motorways restricts the considered range of MPD from 0.86 mm - 2.22 mm to 1.18 mm - 2.08 mm, the effect of macro-texture seems to not change much on primary roads and motorways in England. Conversely, Chatti *et al.* (2012) found that for light trucks the effect of macro-texture on fuel consumption is negligible at any speed in US. However, it must be noted that large part of the road stretches investigated by Chatti *et al.* (2012) present some low macro-texture (with min. 0.2 - 0.3 mm MPD), typical of US conditions.

Table 28 - Summary of the results obtained for medium trucks (at 87.5 km/h).

Vehicle	App	Explored range of LPV (mm^2)	Effect of LPV	Explored range of SMTD (mm)	Effect of SMTD
medium truck	D2 App 1 (A roads)	0.15 to 0.48 LPV 3 m (1.86 to 2.42 m/km IRI)	+3.30%	0.55 to 1.70 (0.86 to 2.22 mm MPD)	+5.98%
		0.55 to 2.60 LPV 10 m (1.66 to 2.53 m/km IRI)	+3.97%		
		3 to 18 LPV 30 m	-8.56%		
	D2 App 2 (M and A roads)	0.09 to 0.48 LPV 3 m (1.70 - 2.42 m/km IRI)	+1.14%	0.60 to 1.60 (0.92 to 2.11 mm MPD)	-1.21%
		0.50 to 2.50 LPV 10 m (1.63 to 2.50 m/km IRI)	-1.00%		
		3 to 18 LPV 30 m	-3.31%		

Table 28 shows the results of the study obtained for medium trucks. To give some context, for this truck type, 118,990 records were used to train the developed models in the App 1 and 137,845 records in the App 2. Also in this case, results are

given by an average of the estimates made by the developed RF, SVR and ANN models. Table 28 shows that when only primary roads are considered (App 1) the effect of LPV at 3 and 10 metres wavelength is estimated at +3.30% and +3.97% over a range between 1.86 and 2,42 and 1.66 and 2.53 m/km IRI. Again, when motorways are included in the analysis (App 2), the effect of IRI is reduced to +1.14% and -1.00% (for LPV at 3 and 10 metres respectively), which reflects the impact of a smaller 95th percentile unevenness range. This might be due to the fact that data from motorways include more cases of roads in better conditions than those from primary roads, which makes the developed ML models unable to detect any clear trend through the data as discussed for light trucks. Regarding the effect of LPV at 30 metres wavelength, it seems that this has the opposite effect to LPV 3 and 10. Regarding the effect of MPD (investigated between 0.86 mm - 2.22 mm), it is possible to see that (Table 28) its effect is estimated to be +5.98% when considering primary roads only (App 1), and -1.21% including motorways (App 2). Again, reduction of the estimates made in the App 2 might be due to changes in the distribution of the data analysed which does not allow the developed ML models to detect the patterns buried in the data correctly. Although the opposite sign of the obtained estimates and the comparable ranges of MPD considered makes it hard to interpret the results, it seems that the effect of MPD could be sometimes higher for medium trucks than for light trucks. In the literature, the only study which focused on the effect of road surface conditions on the fuel consumption of medium trucks (du Plessis *et al.*, 1990) estimated an effect of up to 20% on the fuel consumption of this vehicle type and buses for road unevenness between 1.0 to 16 m/km IRI, while they found null the effect of the macro-texture. However, because the study of du Plessis *et al.* (1990) is quite old and uses non-standardised measurements (i.e. QI in counts/km to characterise the road roughness),

it is difficult to make any accurate comparison to the results of the present work, as differences might be due to improvements in tire and engine technology more than anything else.

Table 29 shows the results of the study obtained for heavy trucks.

Table 29 - Summary of the results obtained for heavy trucks (at 87.5 km/h).

Vehicle	App	Explored range of LPV (mm^2)	Effect of LPV	Explored range of SMTD (mm)	Effect of SMTD
heavy truck	D1	0.08 to 0.31 LPV 3 m (1.67 to 2.15 m/km IRI)	+0.82%	0.95 to 1.46 (mm MPD)	-0.51%
		0.4 to 1.55 LPV 10 m (1.58 to 2.07 m/km IRI)	-1.52%		
		3 to 12 LPV 30 m	+3.34%		
	D2 App 1 (A roads)	0.15 to 0.48 LPV 3 m (1.86 to 2.42 m/km IRI)	-0.77%	0.55 to 1.70 (0.86 to 2.22 mm MPD)	+7.84%
		0.60 to 2.60 LPV 10 m (1.70 to 2.53 m/km IRI)	-0.25%		
		3 to 18 LPV 30 m	-6.55%		
	D2 App 2 (M and A roads)	0.10 to 0.53 LPV 3 m (1.73 to 2.49 m/km IRI)	+6.28%	0.60 to 1.60 (0.92 to 2.11 mm MPD)	-2.33%
		0.50 to 3.50 LPV 10 m (1.63 to 2.79 m/km IRI)	+1.19%		
		3 to 27 LPV 30 m	-4.32%		

In this case, 198,653 records were used to train the developed models in the App 1 and 208,082 records in the App 2. Again, results are given by an average of the estimates made by the developed RF, SVR and ANN models. Table 29 shows for the Dataset 1 an overall increase of fuel consumption due to the contribution of LPV at 3, 10 and 30 metres wavelengths. The same analysis showed that the effect of MPD between 0.95 mm and 1.46 mm is a slight reduction of fuel consumption which is not of easy interpretability from a physical perspective. Regarding the analysis of the data from heavy trucks from the Dataset 2, when only primary roads are considered (App 1) the effect of LPV at 3 and 10 metres wavelength was estimated to be -0.77% and -

0.25% between 1.70 and 2.53 m/km of IRI. Again, results appear of difficult interpretability. The fact that, when motorways are included in the analysis (App 2), the effect of IRI increases to +6.28% and +1.19% (for LPV at 3 and 10 metres respectively) does not respect the trends seen for light and medium trucks, might have a physical explanation and is in line with the results obtained by Chatti *et al.* (2012). In fact, within a similar range of IRI, Chatti *et al.* (2012) found the impact of IRI on the fuel consumption of heavy trucks to be of about +1.5%. Regarding the effect of LPV at 30 metres on the fuel consumption of heavy trucks, it is possible to see (Table 29) that, while from the analysis of the Dataset 1 this seems to increase the effect of road roughness, on the Dataset 2 this counterbalances the effect of LPV at 3 and 10 metres wavelengths, reducing its overall effect. Also in this case, it is difficult to give a clear interpretation of the physical phenomenon behind the identified behaviour, however, it seems clear that different wavelengths of road roughness seem to affect the fuel consumption of different truck types differently. This would confirm the findings of Cebon (1985a) and may point out the inadequacy of IRI to characterise the road surface unevennes from a fuel economy perspective, and the advantages of using LPV measurements instead. Regarding the effect of MPD on the fuel consumption of heavy trucks, it is possible to see that (Table 29) this is estimated to be +7.84% when considering primary roads only (App 1), and -2.33% when including motorways in the conducted analysis (App 2). This highlights that the effect of macro-texture on the fuel consumption of heavy trucks can be in certain situations higher than for light and medium trucks, and confirms that, inclusion of more roads in good conditions such as motorways reduces the ability of SVR, RF and ANN to detect clear patterns through the data. Conversely, Chatti *et al.* (2012) found that for heavy trucks the effect of macro-texture (MPD) on fuel consumption is negligible at high speed (88 km/h) in

US. Recently, performing RR measurements using the trailer method, Ejsmont *et al.* (2017) found that the effect of MPD in a range between 0.40 and 1.30 mm MPD can be estimated to be of approximately +6%. This matches the findings obtained in the App 1 giving more confidence in utilising that result. However, due to the high variance that the estimates made by the developed models show, further investigations should be made before these can be applied in pavement LCA studies. In fact, this high variance in the obtained estimates could be due to the heterogeneity of the considered fleet of trucks, which includes vehicles of slightly different characteristics. Nonetheless, also Hammarström *et al.* (2008), found difficulties at estimating the impact of road roughness and macro-texture on the fuel consumption of heavy trucks, and defined their results for this vehicle type to be unreliable. Therefore, variance of the obtained estimates might depend from variables which neither the present study, nor the study conducted by Hammarström *et al.* (2008), considered (e.g. shape of the truck, performance of the engine for loaded and unloaded trucks, age of the engine, tire pressure, etc.). This requires further investigation.

6 Conclusions and Recommendations

In this thesis, a novel approach for estimating truck fleet fuel consumption and the impact of road roughness and macro-texture, based on Big Data, was introduced and tested. The method was applied to several cases and its performance evaluated on datasets of various size for different truck types.

6.1 Conclusions

Analysis of the existing literature highlighted that results from existing studies do not always agree regarding the exact impacts of road surface roughness and macro-texture on vehicle fuel consumption (see Table 8). This can be because of the different sets of assumptions made in each of the considered studies, but it is certainly due also to the fact that a number of external factors, such as weather conditions (e.g. air temperature, air pressure, film of water on the road surface, wind speed and direction, etc.), affect vehicle fuel consumption and it is difficult to consider all of them in experimental set ups or physical/mechanistic based models. This makes conclusions of experimental and theoretical approaches limited to specific sets of assumptions, and not always applicable for real driving conditions at network level. For this reason, a new approach, which considers data from all vehicles driving across a certain road network of interest, such as that presented in this thesis, could represent

a solution to reduce uncertainties and give better insights into the effect of RR and related parameters on vehicle fuel economy. This could help in obtaining results representative of real driving and support road managers in making informed decisions towards a re-evaluation of the current road maintenance strategies.

Results of the study show that, although optimisation is required, there is great potential for the Big Data approach and ML techniques to evaluate the impact of road surface characteristics and other variables, such as weather conditions, on truck fleet fuel consumption.

Results showed that the developed LR models are too simplistic even to model the fuel consumption of the considered vehicles driving at relatively constant speed in real driving conditions (e.g. subject to the effect of traffic, wind speed, air temperature, etc.). In this regard, analysis of the available datasets showed that the records collected from trucks driving at relatively constant speed on motorways and A roads in England represent only 10% of the available records (see sect. 4.2). This implies that the analysed conditions are not very common and are expected to be even rarer in other circumstances, such as in an urban environment. This certainly reduces applicability of the result of the present study, but reduces also the representativeness of the results of studies that performed tests under carefully controlled conditions (e.g. Laganier *et al.*, 1990; Sandberg, 1990; Hammarström *et al.*, 2009; Karlsson *et al.*, 2011; Chatti *et al.*, 2012). In fact, it is expected that, consideration of the effect of acceleration and deceleration, start and stop, traffic, wind speed, and other variables can sometimes overcome the effect of road surface characteristics on vehicle fuel consumption, which may become more apparent when these external factors are controlled such as in the experimental data. These more general conditions were not investigated either in the past or the present study. However, while data for application

of the Big Data approach are already available and further development of the methodology would allow investigation of these conditions, this will remain a limitation for experimental set-ups.

In all cases considered, the LR models showed lower performance than any of the other techniques applied (i.e. SVR, RF and ANN). This is probably due to the effect of non-linearity, which was shown to be significant even under relatively controlled conditions such as those investigated in the present study (e.g. constant speed). Conversely, SVR, RF and ANN were able to achieve high performance also for small datasets (e.g. Dataset 1). Generally, performance of all the ML algorithms applied increased with the amount of data used to train the models. These were always in line or superior with the performance of the physical/mechanistic models recently developed and currently available in the literature (e.g. Chatti *et al.*, 2012; Almér, 2015) (see Table 19).

In regard to the application of SVR, RF and ANN it was seen that in the past a similar method was applied by Almér (2015). However, in the present study it was seen how utilisation of a larger dataset of minimum size (determined using the learning curves, see sect. 5.2), together with the optimisation of the hyperparameters and structure of the developed models and application of ten-fold cross-validation allowed to achieve performance between 2.5 and 3.9 times better in terms of RMSE (see Table 19 in sect. 5.3). In this regard, the study conducted by Almér (2015) considered only a specific truck model from a specific and known manufacturer and a wide range of variables (see sect. 2.8). The present study instead does not consider any specific vehicle model (as this information is unknown) and the number of variables considered was carefully reduced to include only those features that allow the developed models to improve their performance. This should emphasize the obtained

improvement. Also, analysis of the performed PA allowed the present study to estimate how each of the developed models consider each of the cosnisered variables affecting the fuel consumption of the considered fleet of trucks (see Appendix E.1.3Appendix E.2.3), and this was not pert of the method presented by Almér (2015). This allowed the effect of the road surface conditions and other parameters to be analysed, which was not possible to be evauated in the study by Almér (2015).

Among the developed models, the RF outperformed all the other regression techniques used in the study, demonstrating to be more precise than SVR and ANN, in all applications in terms of R^2, RMSE and MAE. A current limitation for the the RF models to be applied in practice is represented by the high computational time required to cross-validate (train) the model. Application of deep neural networks (e.g. ANN1) to very large datasets (e.g. heavy trucks App 2) showed that, although this technique can achieve high levels of precision, comparable to those achieved by the RF, it has a high risk of overfitting the single variables considered and thus reduce interpretability of the developed model (e.g. see Figure 38). Application of deep neural networks (e.g. ANN1) to small datasets is not recommended as it is very prone to overfitting.

A major criticism often addressed to ML models such as SVR, RF, ANN and ANN1, is that they are poorly interpretable. In this study, use of cross-validation and the parametric analysis allowed improvement of the reliability and robustness of the developed models. Nevertheless, results of the conducted analysis showed high variance and often proved to be of poor interpretability. This confirms the general criticism of the method, valid also for the RF model which shows the best performance overall. This might be due to overfitting of some of the variables considered in the study, which affect the estimates made by the models on the other variables. In these terms, the model, which seems to be able to identify non-linearity and yet avoid

overfitting is the SVR, which rarely shows complex trends and is able to approximate all patterns with smooth curves. That highlights how for the available data simplicity of the developed model can represent an advantage in comparison to the higher R^2, and low error, achieved by RF and ANN, which sometimes have a risk of extrapolating complex trends, making interpretation difficult. In addition to this, it was seen that ensemble averaging helped at reducing the variance of the obtained estimates and might represent a solution for optimisation of the method in future applications.

With focus on the effect of road surface characteristics on truck fuel consumption it is fair to say that, at the current state of technology, it is difficult to estimate the effect of road surface conditions on truck fleet fuel consumption using the Big Data approach.

For light trucks, it was seen that for conditions typical of the SRN of England, the impact of road unevenness on fuel consumption was estimated between -2.91% and +6.27%, depending on the wavelength of LPV considered, while for the macro-texture, measured as SMTD, this was between +1.04% and +1.21%.

For medium trucks, it was seen that for conditions typical of the SRN of England, the impact of different wavelength of road unveness, as LPV, varies between -8.56% and +3.97%, while for the macrotexture, as SMTD, this is between -1.21% and +5.98%.

Finally, for heavy trucks in conditions characteristic of the SRN of England, the impact of road unevenness, as LPV, was estimated to be between -6.55% and +6.28%, while it was between -2.33% and +7.84% for the macrotexture, measured as SMTD.

In terms of order of magnitude, the identified impacts are quite similar to those identified in the past (see Table 8). In addition, according to e.g. Hammarström

et al. (2008) and Chatti *et al.*, (2012) it was expected that the impact of road unevnees and macro-texture could change for different vehicle types. However, estimates from the developed models present high variance and different applications of the method seem to determine different effects of roughness and macro-texture on fuel consumption, for the considered fleets of truck and limited range of road conditions. This does not help in the interpretation of the results from a physical perspective and assessment of the impact of road surface characteristics on truck fleet fuel consumption certainly requires further investigation before, for instance, this can be included in the pavement LCA framework, or maintenance planning, with confidence. However, the method and results of the present study seem to be promising, and future applications of the method in road asset management and pavement LCA studies seem feasible. The main advantages of the presented methodology are its repeatability and adaptability. The fact that truck sensors are standard (SAE, 2016) and weather condition data are available worldwide, together with the need for road agencies to monitor the conditions of their infrastructure periodically, offers the opportunity to apply and develop the presented method anywhere.

6.2 Limitations

Limitations of the study can be listed as follows:

- It was not within the aim of this study to investigate the effect of road surface characteristics on cars or vehicles other than trucks.
- It was not within the aim of this study to investigate the effect of pavement stiffness, pavement deflection and road materials on truck fleet fuel consumption.
- It was not within the scope of this study to investigate variation of the performance of the developed models with variation of the resolution and

frequency of data collection from truck sensors or road conditions. Only standard measurements (Highways Agency, 2008a; SAE, 2016) were considered.

- Results of the study are representative only of trucks driving at relatively constant and high speed, along motorways and A roads in England.
- The parametric analysis performed considers only variation of one variable at a time. This does not allow investigation of the effect of covariance of the considered variables. In this regard, performing a sensitivity analysis or analysis of covariance on the developed ML models (e.g. Cortez *et al.*, 2013; Cortez, 2016) might help to quantify the uncertainty and assess the reliability of the obtained estimates.
- Accuracy and reliability of the results of the study are likely affected by the collection frequency of the truck telematic data, which is standard for fleet management applications and was not specifically designed for this study. This might have reduced the range of road surface condition explored in the study, due to the averaging of roughness and macro-texture of a mile or more of road, preventing the regression techniques applied in the study to identify a clear relationship between these and the fuel consumption of the considered trucks.
- Most of trucks are currently equipped with Diesel engines, and this is true for all the trucks considered in the present study. This makes results of the present study representative of Diesel fuelled trucks only, and results may change for other fuel types. This somewhat reduces the applicability of the results of the study, but future introduction of new fuel types will not limit the applicability of the presented methodology, which could be easily adapted to estimate the

consumption of petrol, electricity, hydrogen or any other fuel type that in the future might be introduced.

- In the present study it was not possible to investigate the effect of some variables, such as tire pressure, on vehicle fuel consumption. In fact, although this information is contemplated by the SAE J1939 standards, this measurement is not required by the customers of Microlise who, for this reason, do not collect the data. Future work should therefore investigate this and find out what other measurements could be interesting to include.
- As it was scope of this study to investigate the effect of road roughnees and macrotexture on vehicle fuel consumption, no other pavement characteristics were considered. However, the HAPMS stores information about several other parameters, from pavement deflection, to pavement materials, skid-resistance, and others. Therefore, future work should also attempt to apply the presented methodology to explore the effect of road pavement characteristics other than roughness and macro-texture on vehicle fuel consumption.
- Although included in the second part of the study, it was not within the scope of the present study to investigate the effect of weather on truck fleet fuel consumption. This should be investigated further in future work.
- Another limitation of the current study lies in the fact that currently the database of truck fleet managers are not available to road agencies and managers of the infrastructure, who, for this reason are currently not able to apply the presented method. In fact, this might be limited by the willingness of vehicle fleet managers to share their data, which might have to be anonymised as it was done in this study. However, this does not constitute an insuperable limitation. Indeed, when with the advent of connected and autonomous

vehicles open data will be required and become standard in transportation full applicability of the presented methodology will be possible at global level.

6.3 Recommendations

In order to extend applicability of the results of the study, future work should investigate a wider range of vehicle speeds and road conditions. This could be done first by analysing data from trucks driving along a few road segments immediately before and after maintenance and compare the data distributions. Then, data from trucks driving in urban environments, acceleration and deceleration, and other effects which might be predominant sometimes over that of road surface characteristics, should be considered. Although it was not within the scope of the present study, current availability of information such as road pavement materials and layers stiffness in the HAPMS could immediately allow investigation of the effect of these and other road pavement characteristics on truck fleet fuel consumption. In addition, future availability of standardised sensor measurements for other vehicles than trucks, such as cars, will help at extending applicability of the presented approach.

The default reporting frequency of data from truck sensors (established for truck fleet management applications) represented a major limitation for the study. In fact, results show that averaging road surface characteristics over a minute/mile of travel (or more) does not allow to investigate extreme road conditions, which would have been of major interest for the study. Therefore, future studies should increase the frequency of data collection from trucks, from a reading every minute/mile, to one every 10m or 100m, which are the standard frequency of data collection of road surface characteristics. This should allow consideration of the effect of a wider range of road conditions and the effect of extreme values on truck fleet fuel consumption. Additionally, this should help ML algorithms to identify clearer trends.

Also, refinement of the developed methodology is required, and optimisation of the data cleaning algorithm together with application of more advanced techniques, might help at improving the reliability of the obtained estimates and applicability of these into road asset management and pavement LCA studies. For example, the utilised parametric analysis has a limitation in the fact that variation of one variable at a time is considered, while in real-world scenarios multiple parameters can vary at the same time. In this regard, performing a sensitivity analysis together with an analysis of covariance on the developed models, should help at understanding which variables affect the fuel consumption of the considered fleet of trucks the most and how variation of these affects the impact of road surface roughness and macro-texture. In this regard, for road asset management and pavement LCA applications, it would be of particular interest to see how the effect of road surface characteristics can change with the vehicle speed.

Due to the large quantity of variables, that ML algorithms allow to be considered, and that may increase in the future thanks to the introduction of new standard sensors on trucks, higher computational power might be required. Although in the future this may become routinely available at some point, use of high-performance computers (HPC) should offer a temporary solution to keep testing and improving the method meanwhile. Therefore, further investigation could also make use a larger variety of more frequently collected data (see hypotheses made earlier in the section), which in the future might become standardised on trucks or other vehicle types and might give ML algorithms the power to elaborate more robust responses, allowing to quantify the identified impacts more accurately.

Addressing all the above, would give more confidence at including the effect of road surface characteristics on vehicle fuel consumption in the GHG emissions

inventory of road pavement LCA studies, and further development of the presented method might be used to quantify that. In turn, the results of LCA studies could be used in maintenance planning to reduce the lifecycle direct and indirect costs from road transport.

References

AASHTO (2011) A Policy on Geometric Design of Highways and Streets - Green Book. 6th edn. - American Association of State and Highway Transportation Officials (AASHTO), Washington, DC.

ACEA (2018) Vehicles in Use, European Automobile Manufacturers Association. Available at: https://www.acea.be/statistics/tag/category/vehicles-in-use (Accessed: 13 December 2018).

AFDC (2018) Vehicle Weight Classes & Categories, Alternative Fuels Data Centre, U.S. Department of Energy. Available at: https://www.afdc.energy.gov/data/-10380 (Accessed: 7 November 2018).

Almér, H. (2015) Machine learning and statistical analysis in fuel consumption prediction for heavy vehicles, Master of Science Thesis. KTH, Sweden.

Andersen, L. G., Larsen, J. K., Fraser, E. S., Schmidt, B. and Dire, J. C. (2014) 'RR Measurement and Model Development', Journal of Transportation Engineering, 141(2), pp. 1–10. doi: 10.1061/(ASCE)TE.1943-5436.0000673.

André, M., Keller, M., Sjödin, Å., Gadrat, M., Crae, I. and Dilara, P. (2009) 'The ARTEMIS European tools for estimating the transport pollutant emissions', in 18th International Emission Inventories Conference, April 14–17. Baltimore, MD.

Araújo, J. P. C., Oliveira, J. R. M. and Silva, H. M. R. D. (2014) 'The importance of the use phase on the LCA of environmentally friendly solutions for asphalt road pavements',

Transportation Research Part D: Transport and Environment, 32, pp. 97–110. doi: 10.1016/j.trd.2014.07.006.

Azari Jafari, H., Yahia, A. and Ben Amor, M. (2016) 'Life cycle assessment of pavements: Reviewing research challenges and opportunities', Journal of Cleaner Production. Elsevier Ltd, 112, pp. 2187–2197. doi: 10.1016/j.jclepro.2015.09.080.

Ben-Chaim, M., Shmerling, E. and Kuperman, A. (2013) 'Analytic modeling of vehicle fuel consumption', Energies, 6(1), pp. 117–127. doi: http://dx.doi.org/10.3390/en6010117.

Benbow, E., Brittain, S. and Viner, H. (2013) MIRAVEC D3.1: Potential for NRAs to provide energy reducing road infrastructure. Modeling Infrastructure Influence on Road Vehicle Energy Consumption (MIRAVEC). Deliverable D 3.1.

Benbow, E., Iaquinta, J., Lodge, R. and Wright, A. (2007) Investigation of the Effects of Pavement Stiffness on Fuel Consumption - PPR253. Crowthorne, UK.

Benbow, E., Nesnas, K. and Wright, A. (2006) PPR131 - Shape (surface form) of Local Roads. Crowthorne, UK.

Benbow, E., Wright, D., Nesnas, K. and Wright, A. (2011) Measures for Assessing Ride Quality on Trunk Roads - CPR1553. Crowthorne, UK.

Bennett, C. R. and Greenwood, I. D. (2000a) Modelling Road User and Environmental Effects in HDM-4, Volume 7, The Highway Development and Management Series, International Study of Highway Development and Management (ISOHDM). World Roads Association (PIARC). Paris, France.

Bennett, C. R. and Paterson, W. D. O. (2000b) HDM 4: A Guide to Calibration and Adaption, Volume 5, The Highway Development and Management Series, International Study of Highway Development and Management (ISOHDM). Paris: World Roads Association (PIARC). Paris, France.

Bergiers, A., Goubert, L., Anfosso Lédée, F., Dujardin, N., Ejsmont, J., Sandberg, U. and Zöller, M. (2011) 'MIRIAM SP1 D.3: Comparison of RR Measuring Equipment - Pilot Study', Report from Models for RR In Road Infrastructure Asset Management systems (MIRIAM).

Berry, I. M. (2010) The effects of driving style on vehicle performance on the real-world fuel consumption of US light-duty vehicles. Virginia Polytechnic Institute and State University. Blacksburg, VA.

Beuving, E., De Jonghe, T., Goos, D., Lindahl, T. and Stawiarski, A. (2004) 'Environmental Impacts and Fuel Efficiency of Road Pavements', (March).

Bielaczyc, P., Woodburn, J., Szczotka, A. and Pajdowski, P. (2015) 'The Impact of Alternative Fuels on Fuel Consumption and Exhaust Emissions of Greenhouse Gases from Vehicles Featuring SI Engines', Physics Procedia. Elsevier B.V., 66, pp. 21–24. doi: 10.1016/j.egypro.2015.02.011.

Bifulco, G., Galante, F., Pariota, L. and Spena, M. (2015) 'A Linear Model for the Estimation of Fuel Consumption and the Impact Evaluation of Advanced Driving Assistance Systems', Sustainability, 7(10), pp. 14326–14343. doi: 10.3390/su71014326.

Birkner, C. (2016) 'ROSANNE WP2 Delivarable D2.6: Report on the development of the procedure for characterisation of noise properties of road surfaces including the updated draft standard', 43(0), pp. 1–11.

Birkner, C., Adesiyun, A., Greene, M., Conter, M., Schmidt, B., Sandberg, U. and Scharnigg, K. (2016) 'ROSANNE Final Summary Report'.

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., Rouault, E., Wrmerdam, F., Ooms, J. and Rundel, C. (2018) 'R Package: rgdal, Ver. 1.3-4', CRAN. Available at: https://cran.r-project.org/web/packages/rgdal-/index.html.

Boriboonsomsin, K. and Barth, M. (2009) 'Impacts of Road Grade on Fuel Consumption and Carbon Dioxide Emissions Evidenced by Use of Advanced Navigation Systems', Transportation Research Record: Journal of the Transportation Research Board, 2139, pp. 21–30. doi: 10.3141/2139-03.

Breiman, L. (2001) 'Random forests', Machine Learning, 45(1), pp. 5–32. doi: 10.1023/A:1010933404324.

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984) 'Classification and Regression Trees', The Wadsworth Statistics/Probability Series, 19, p. 368. doi:

10.1371/journal.pone.0015807.

Bryce, J., Katicha, S., Flintsch, G., Sivaneswaran, N. and Santos, J. (2014a) 'Probabilistic Life-Cycle Assessment as Network-Level Evaluation Tool for Use and Maintenance Phases of Pavements', Transportation Research Record: Journal of the Transportation Research Board, 2455, pp. 44–53. doi: 10.3141/2455-06.

Bryce, J. M. (2014) Applying LCA to enhance sustainable pavement management decision making, PhD Thesis. Virginia Polytechnic Institute and State University. Blacksburg, VA.

Bryce, J., Santos, J., Flintsch, G., Katicha, S., McGhee, K. and Ferreira, A. (2014b) 'Analysis of RR models to analyse vehicle fuel consumption as a function of pavement properties', in International Conference on Asphalt Pavements. Raleigh, North Carolina, USA: CRC Press/Balkema, pp. 263–273. Available at: http://books.google.com/books?hl=en&lr=&id=p-ZhBBAAAQBAJ&oi=fnd&pg=PA263&dq=Analysis+of+rolling+resistance+models+to+analyse+vehicle+fuel+consumption+as+a+function+of+pavement+properties&ots=fnu6p_6S2Y&sig=Y6txe2ZVCD_Yh1d77E7qNe1Ha9Y.

BSI (2006) 'BS 7941-1:2006 - Methods for measuring the skid resistance of pavement surfaces. Sideway-force coefficient routine investigation machine', The British Standards Institution.

Burkov, A. (2019) The Hundred-Page Machine Learning Book. Draft. Montreal, Canada. Available at: http://themlbook.com.

Çapraz, A. G., Özel, P., Şevkli, M. and Beyca, Ö. F. (2016) 'Fuel Consumption Models Applied to Automobiles Using Real-time Data: A Comparison of Statistical Models', Procedia Computer Science, 83, pp. 774–781. doi: 10.1016/j.procs.2016.04.166.

Carlson, A. (2011) Life cycle assessment of roads and pavements - Studies made in Europe, VTI rapport 736A. Linköping, Sweden.

Carlsson, G. (1986) 'VEjstandard og Transport-Omkostninger (VETO)', VTI rapport 307. Linköping, Sweden.

Carrese, S., Gemma, A. and La Spada, S. (2013) 'Impacts of Driving Behaviours, Slope and Vehicle Load Factor on Bus Fuel Consumption and Emissions: A Real Case Study in the City of Rome', Procedia - Social and Behavioral Sciences. Elsevier B.V., 87, pp. 211–221. doi: 10.1016/j.sbspro.2013.10.605.

Cebon, D. (1985a) An investigation of the dynamic interaction between wheeled vehicles and road surfaces. University of Cambridge. Cambridge, UK.

Cebon, D. (1985b) 'Heavy Vehicle Vibration - A Case Study', Vehicle System Dynamics, 14(1–3), pp. 27–32. doi: 10.1080/00423118508968791.

Cebon, D. (1988) 'Road Damaging Effects of Dynamic Axle Loads', International Symposium on Heavy Vehicle Weights and Dimensions, pp. 12–15. Available at: http://road-transport-technology.org/Proceedings/1-ISHWD/Road Damaging Effects of Dynamic Axle Loads - Cebon.pdf.

Cenek, P. D. (1994) 'RR Characteristics of New Zealand Road Surfaces', in Kulakowski, B. T. (ed.) Vehicle-Road Interaction. Philadelphia: American Society for Testing and Materials, pp. 248–262. doi: https://doi.org/10.1520/STP13260S.

Cenek, P. D., Jamieson, N. J. and Ball, G. (1996) 'Effect of Pavement Deflection on RR of Commercial Vehicle Tires', in 3rd International Symposium on Pavement Surface Characteristics. Christchurch, New Zealand.

Chatti, K. and Zaabar, I. (2012) Estimating the Effects of Pavement Condition on Vehicle Operating Costs, NCHRP Report 720. Washington, DC.

Chen, X., Shan, X., Ye, J., Yi, F. and Wang, Y. (2017a) 'Evaluating the Effects of Traffic Congestion and Passenger Load on Feeder Bus Fuel and Emissions Compared with Passenger Car', Transportation Research Procedia. Elsevier B.V., 25, pp. 616–626. doi: 10.1016/j.trpro.2017.05.446.

Chen, Y., Zhu, L., Gonder, J., Young, S. and Walkowicz, K. (2017b) 'Data-driven fuel consumption estimation: A multivariate adaptive regression spline approach', Transportation Research Part C: Emerging Technologies. Elsevier Ltd, 83(October), pp. 134–145. doi: 10.1016/j.trc.2017.08.003.

Ciklum (2017) Pros and Cons of Big Data, Ciklum - Empowering Collaboration. Blog. Available at: https://www.ciklum.com/blog/pros-and-cons-of-big-data/ (Accessed: 30 April 2018).

Clark, N., Thompson, G. and Delgado, O. (2009) Modeling Heavy-duty Vehicle Fuel Economy Based on Cycle Properties. West Virginia University. Morgantown, WV.

Cole, D. J. and Cebon, D. (1989) 'Simulation and measurement of dynamic tire forces', Cambridge University Engineering Department. Cambridge, UK.

Cortes, C. and Vapnik, V. (1995) 'Support-Vector Networks', Machine Learning, 20(3), pp. 273–297. doi: 10.1023/A:1022627411411.

Cortez, P. (2016) 'R Package: rminer, Ver. 1.4.2', CRAN. Available at: https://cran.r-project.org/web/packages/rminer/index.html.

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.

Cristianini, N. and Schölkopf, B. (2002) 'Support Vector Machines and Kernel Methods: The New Generation of Learning Machines', Artificial Intelligence Magazine, 23(3), pp. 31–42. doi: http://dx.doi.org/10.1609/aimag.v23i3.1655.

Cristianini, N. and Shawe-Taylor, J. (2000) 'An Introduction to Support Vector Machines and other kernel based learning methods', Ai Magazine, p. 190. doi: citeulike-article-id:114719.

Dávila, A. and Nombela, M. (2010) 'SARTRE: SAfe Road TRains for the Environment', Project Report.

DEFRA and DfT (2017) 'UK plan for tackling roadside nitrogen dioxide concentrations: Technical report', Department for Environment Food & Rural Affairs together with Department for Transport., (July), pp. 1–11. Available at: www.nationalarchives.gov.uk/doc/open-government-licence/version/3-/%0Awww.gov.uk/government/publications.

Delanne, Y. (1994) 'Influence of pavement evenness and macrotexture on fuel consumption',

in Kulakowski, B. T. (ed.) Vehicle-Road Interaction. Philadelphia: American Society for Testing and Materials, pp. 240–247. doi: 10.1520/stp13259s.

Delgado, O. F., Clark, N. N. and Thompson, G. J. (2011) 'Modeling Transit Bus Fuel Consumption on the Basis of Cycle Properties', Journal of the Air & Waste Management Association, 61(4), pp. 443–452. doi: http://dx.doi.org/10.3155/1047-3289.61.4.443.

Deloncle, A., Berk, R., D'Andrea, F. and Ghil, M. (2007) 'Weather Regime Prediction Using Statistical Learning', J. Atmos. Sci., 64(5), pp. 1619–1635. doi: http://dx.doi.org/10.1175/JAS3918.1.

Descornet, G. (1990) 'Road-surface influence on tire RR', in Meyer, W. E. and Reichert, J. (eds) Surface Characteristics of Roadways: International Research and Technologies. Philadelphia: American Society for Testing and Materials, pp. 401–415. doi: https://doi.org/10.1520/STP23377S.

DfBEIS (2017) 'Energy Consumption in the UK', Department for Business, Energy & Industrial Strategy, (July). Available at: https://assets.publishing.service.-gov.uk/government/uploads/system/uploads/attachment_data/file/633503/ECUK_2017.pdf.

DfBEIS (2018) '2017 UK Greenhouse Gas Emissions, Provisional Figures: Statistical release', Department for Business, Energy & Industrial Strategy, (March). Available at: https://www.gov.uk/government/uploads/system/uploads/-attachment_-data/file/295968/20140327_2013_UK_Greenhouse_Gas_Emissions_Provisional_Figures.pdf.

DfT (2003) 'A simplified guide to lorry types and weights', Driver and Vehicle Standards Agency, Department for Transport. London, UK.

DfT (2018a) 'Road Conditions in England 2017', Department for Transport. Statistical Release. London, UK.

DfT (2018b) 'Road Lengths in Great Britain 2017', Department for Transport. Statistical Release. London, UK.

DfT (2018c) The Highway Code, Deparment for Transport. London, UK.

Ding, F. and Jin, H. (2018) 'On the Optimal Speed Profile for Eco-Driving on Curved Roads', IEEE Transactions on Intelligent Transportation Systems, pp. 1–11. doi: 10.1109/-TITS.2018.2795602.

EC (2011) Roadmap to a Single European Transport Area - Towards a competitive and resource efficient transport system. European Commission. doi: 10.2832/30955.

EC (2016) 'A European Strategy for Low-Emission Mobility', European Commission - Working Document, p. 107. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016SC0244&from=en.

EC (2017a) Europe on the Move, European Commission. Available at: https://ec.europa.eu/transport/modes/road/news/2017-05-31-europe-on-the-move_en (Accessed: 22 March 2018).

EC (2017b) Passenger cars in the EU, European Commission, eurostat - Statistics Explained. Available at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Passenger_-cars_in_the_EU (Accessed: 26 March 2018).

EEA (2016a) Explaining road transport emissions. A non-technical guide, European Environment Agency. Copenhagen, Denmark. doi: 10.2800/71804.

EEA (2016b) 'Transport in Europe: key facts and trends', European Environment Agency. Copenhagen, Denmark. Available at: https://www.eea.europa.eu-/signals/signals-2016/articles/transport-in-europe-key-facts-trends.

EEA (2017a) Final energy consumption by sector and fuel, European Environment Agency. Copenhagen, Denmark.

EEA (2017b) 'Total greenhouse gas (GHG) emission trends and projections', European Environment Agency - Report. Available at: http://www.eea.europa.eu/data-and-maps/indicators/greenhouse-gas-emission-trends-5/assessment-1.

EERE (2015) 'Advanced Combustion Engines - 2015 Annual Report', Energy Efficiency & Renewable Energy, U.S. Department of Energy. Washington, DC.

EIA (2018a) FAQ – 'How much energy is consumed in U.S. residential and commercial

buildings?', U.S. Energy Information Administration. Washington, DC.

EIA (2018b) U.S. Energy Facts Explained, U.S. Energy Information Administration. Washington, DC.

Ejsmont, J. A., Ronowski, G., Świeczko-Żurek, B. and Sommer, S. (2017) 'Road texture influence on tire RR', Road Materials and Pavement Design, 18(1), pp. 181–198. doi: 10.1080/14680629.2016.1160835.

Ejsmont, J., Sjögren, L., Świeczko-Żurek, B. and Ronowski, G. (2015) 'Influence of Road Wetness on Tire-Pavement RR', Journal of Civil Engineering and Architecture, 9(11), pp. 1302–1310. doi: 10.17265/1934-7359/2015.11.004.

Ejsmont, J., Taryma, S., Ronowski, G. and Swieczko-Zurek, B. (2016) 'Influence of load and inflation pressure on the tire RR', International Journal of Automotive Technology, 17(2), pp. 237–244. doi: 10.1007/s12239-016-0023-z.

Ejsmont, J., Taryma, S., Ronowski, G. and Swieczko-Zurek, B. (2018) 'Influence of temperature on the tire RR', International Journal of Automotive Technology, 19(1), pp. 45–54. doi: 10.1007/s12239-018-0005-4.

Ekins, J. D. K. and Hawker, L. G. (2003) TRACS Type Surveys for Local Roads, Final Report. UK Roads Liason Group. London, UK. Available at: http://www.ukpms.com/owner_forum/shared_files/dft_roads_023919.pdf.

Ekman, P., Gardhagen, R., Virdung, T. and Karlsson, M. (2016) 'Aerodynamics of an Unloaded Timber Truck - A CFD Investigation', SAE International Journal of Commerical Vehicles, (2016-01-8022), p. 9(2):217-223. doi: 10.4271/2016-01-8022.

ENERGY.GOV (2012) 54.5 MPG and Beyond: Materials Lighten the Load for Fuel Economy, U.S. Department of Energy. Available at: https://www.energy.gov/-articles/545-mpg-and-beyond-materials-lighten-load-fuel-economy (Accessed: 14 May 2018).

EP (2006) 'Directive 2006/126/EC on driving licenses', European Parliament, the Council of The European Union, (7), pp. 18–60.

EP (2007) 'Regulation (EC) No 715/2007', European Parliament the Council of The European

Union - Official journal of the European Union, L171(December 2006), pp. 1–16. doi: OJEU 29.6.2007 L171.

EP (2009a) 'Regulation (EC) No 443/2009', European Parliament the Council of The European Union - Official journal of the European Union, 140(1), pp. 1–15. doi: 10.1524/zkri.2009.1105.

EP (2009b) 'Regulation (EC) No 595/2009', European Parliament the Council of The European Union - Official journal of the European Union, L 188(June), pp. 1–13.

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.

EP (2017) 'Research for TRAN Committee - Decarbonisation of EU transport', Policy Department for Structural and Cohesion Policies, European Parliament. doi: 10.2861/840058.

EP (2018) 'Proposal for amending Regulation (EC) No 715/2007 (recast)', European Parliament The Council Of The European Union, 0293(715).

EPA (2017a) 'Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2015', U.S. Environmental Protection Agency. Federal Report EPA 430-R-13-001., 82(30), p. 10767.

EPA (2017b) Vehicle Weight Classifications for the Emission Standards Reference Guide, U.S. Environmental Protection Agency. Available at: https://www.epa.gov/emission-standards-reference-guide/vehicle-weight-classifications-emission-standards-reference-guide (Accessed: 13 December 2017).

EPA (2018a) Air Pollution: Current and Future Challenges, United States Environmental Protection Agnecy. Available at: https://www.epa.gov/clean-air-act-overview/air-pollution-current-and-future-challenges#limiting (Accessed: 27 June 2018).

EPA (2018b) Energy Efficient Technologies, U.S. Environmental Protection Agency, Web Page. Available at: https://www.fueleconomy.gov/feg/tech_adv.shtml (Accessed: 2 March 2018).

EPA (2018c) Fuel Economy Guide - Model Year 2018. Washington, DC.

EPA (2018d) Many Factors Affect Fuel Economy - www.fueleconomy.gov, United States Environmental Protection Agnecy. Available at: https://www.fueleconomy.gov/-feg/factors.shtml (Accessed: 21 March 2018).

EPA (2018e) Shared MPG estimates - www.fueleconomy.gov, United States Environmental Protection Agnecy. Available at: https://www.fueleconomy.gov/mpg/MPG.do?-action=browseList (Accessed: 14 May 2018).

ERF (2014) 'Road Asset Management - An ERF position paper for maintaining and improving a sustainable and efficient road network', European Union Road Federation. Brussels, Belgium.

Evans, L. (1979) 'Driver Behavior Effects on Fuel Consumption in Urban Driving', Human Factors: The Journal of the Human Factors and Ergonomics Society, 21(4), pp. 389–398. doi: 10.1177/001872087902100401.

Fawcett, J. and Robinson, P. (2000) 'Adaptive routing for road traffic', IEEE Computer Graphics and Applications, 20(3), pp. 46–53. doi: 10.1109/38.844372.

Feillet, D., Dejax, P., Gendreau, M. and Gueguen, C. (2004) 'An exact algorithm for the elementary shortest path problem with resource constraints: Application to some vehicle routing problems', Networks, 44(3), pp. 216–229. doi: 10.1002/net.20033.

Ferreira, L. J. A. (1982) Car Fuel Consumption in Urban Traffic. The Results of a Survey in Leeds using Instrumented Vehicles, White Rose University Consortium. 162. Leeds, UK. doi: 10.1016/S1366-5545(02)00012-1.

FHWA (2014) Verification, Refinement, and Applicability of Long-Term Pavement Performance Vehicle Classification Rules - Chapter 2. Introduction to Vehicle Classification, Federal Highway Administration, FHWA-HRT-13-091. Available at: https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/-13091/002.cfm (Accessed: 24 July 2018).

FHWA (2016) 'Pavement Life Cycle Assessment Framework', U.S. Department of Transport, Federal Highway Administration (FHWA). doi: 10.1007/978-3-662-44719-2_1.

Figueredo, G. P., Agrawal, U., Mase, J. M., Mesgarpour, M., Wagner, C., Soria, D., Garibaldi, J. M., Siebers, P. and John, R. I. (2017) 'Identifying Heavy Goods Vehicle Driving Styles in the United Kingdom', IEEE Transactions on Intelligent Transport Systems, p. (accepted).

Figueredo, G. P., Quinlan, P. R., Mesgarpour, M., Garibaldi, J. M. and John, R. I. (2015) 'A Data Analysis Framework to Rank HGV Drivers', IEEE Conference on Intelligent Transportation Systems, ITSC, October, 2(Section II), pp. 2001–2006. doi: http://dx.doi.org/10.1109/ITSC.2015.324.

Friedman, J., Hastie, T., Simon, N., Qian, J. and Tibshirani, R. (2017) 'R Package: glmnet, Ver. 2.0-13', CRAN. Available at: https://cran.r-project.org/web/packages/glmnet-/index.html.

Fritsch, S., Guenther, F., Suling, M. and Mueller, S. M. (2016) 'R Package: neuralnet, Ver. 1.33', CRAN. Available at: https://cran.r-project.org/web/packages/neuralnet-/index.html.

Gartner Inc. (2018) Big Data, Gartner, Inc., IT - Glossary. Available at: https://www.gartner.com/it-glossary/big-data (Accessed: 30 April 2018).

Gillespie, T. D. (1992) Fundamentals of Vehicle Dynamics, R-114. SAE International.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning, Adaptive computation and machine learning series. Cambridge, MA: The MIT Press, Massachussets Institute of Technology. Available at: http://www.deeplearningbook.org/.

Goodyear (2008) 'Factors Affecting Truck Fuel Economy', The Goodyear Tire & Rubber Company. Akron, OH.

GOV.UK (2017) Average heavy goods vehicle fuel consumption, www.gov.uk. Available at: https://www.gov.uk/government/uploads/system/uploads/-attachment_data/file/661675/env0104.ods (Accessed: 14 May 2018).

GOV.UK (2018) Moving goods by road, www.gov.uk. Available at: https://www.gov.uk/guidance/moving-goods-by-road (Accessed: 14 May 2018).

Greene, M. (2014) ROSANNE WP1 Deliverable D1.1 'Definition of boundaries and

requirements for the common scale for harmonisation of skid resistance measurements'.

Greenwood, I. D., Dunn, R. C. and Raine, R. R. (2007) 'Estimating the Effects of Traffic Congestion on Fuel Consumption and Vehicle Emissions Based on Acceleration Noise', Journal of Transportation Engineering, 133(2), pp. 96–104. doi: 10.1061/(ASCE)0733-947X(2007)133:2(96).

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.

Guzzella, L. and Sciaretta, A. (2015) Vehicle Propulsion Systems, Introduction to Modeling and Optimization, Springer. Zurich: Springer Science+Business Media. doi: http://dx.doi.org/10.1007/978-3-642-35913-2.

Haider, M. and Conter, M. (2012) MIRAVEC D1.1: Report on the road infrastructure effects contributing to road vehicle energy consumption and their governing parameters Modelling Infrastructure Influence on RoAd Vehicle Energy Consumption. Modeling Infrastructure Influence on Road Vehicle Energy Consumption (MIRAVEC). Deliverable D1.1.

Haider, M., Conter, M. and Glaeser, K.-P. (2011) 'MIRIAM SP2 D.5: What are RR and other influencing parameters on energy consumption in road transport', Report from Models for RR In Road Infrastructure Asset Management systems (MIRIAM).

Haider, M. and Kriegisch, M. (2013) MIRAVEC D5.3: Modeling Infrastructure Influence on Road Vehicle Energy Consumption. Modeling Infrastructure Influence on Road Vehicle Energy Consumption (MIRAVEC). Deliverable D5.3, Final report.

Hall, J. N. (2011) 'Evolution of Advanced High Strength Steels in Automotive Applications', p. 2.

Hammache, M., Michaelian, M. and Browand, F. (2001) Aerodynamic forces on truck models, including two trucks in tandem, Institute of Transportation Studies University of California, Berkley. California PATH Program. Report UCB-ITS-PRR-2001-27. doi:

10.4271/2002-01-0530.

Hammarström, U., Eriksson, J. and Karlsson, R. (2012) RR model, fuel consumption model and the traffic energy saving potential from changed road surface conditions. Linköping, Sweden. Available at: http://miriam-co2.net/Publications/modeller_for_-rullmotstand_och_bransleforbrukning_samt_energisparpotential_som_foljd_av_andrade_vagyteforhallanden.pdf.

Hammarström, U. and Karlsson, B. (1987) 'VETO - A computer program for calculation of transport costs as a function of road standard', VTI/MEDDELANDE-501

Hammarström, U., Karlsson, R. and Sörensen, H. (2008) 'Road surface effects on RR - coastdown measurements with uncertainty analysis in focus', Energy Conservation in Road Pavement Design, Maintenance and Utilisation (ECRPD). Deliverable D5(a). Linköping, Sweden.

Hammarström, U., Karlsson, R. and Sörensen, H. (2009) Road surface effects on RR - Coastdown measurements with uncer- tainty analysis in focus. Linköping, Sweden.

Harvey, J., Kendall, A., Lee, I.-S., Santero, N. J., Van Dam, T. and Wang, T. (2010) Pavement Life Cycle Assessment Workshop: Discussion Summary and Guidelines. Davis, California, May 5-7, 2010. Available at: http://link.springer.com/10.1007/978-3-662-44719-2_1.

Hecht-Nielsen, R. (1989) Neurocomputing. HNC Inc. a. San Diego, CA: Addison-Wesley Publishing Co.

Hernandez, J. A. and Al-Qadi, I. L. (2017) 'Tire–pavement interaction modelling: hyperelastic tire and elastic pavement', Road Materials and Pavement Design, 18(5), pp. 1067–1083. doi: 10.1080/14680629.2016.1206485.

Highways Agency (2002) 'TD 9/93 Road Geometry', Design Manual for Roads and Bridges (DMRB), 6(1). Available at: http://www.standardsforhighways.-co.uk/ha/standards-/dmrb/vol6/section1/td993.pdf.

Highways Agency (2006a) 'HD 24/06 Pavement Design and Maintenance', Design Manual for Roads and Bridges (DMRB), 7(2).

Highways Agency (2006b) ‘HD 28/15 Pavement design and maintenance’, Design Manual for Roads and Bridges (DMRB), 7(February 2006).

Highways Agency (2008a) ‘HD 29/08 Data for Pavement Assessment’, Design Manual for Roads and Bridges (DMRB), Vol 7.

Highways Agency (2008b) ‘HD 30/08 Maintenance Assessment Procedure’, Design Manual for Roads and Bridges (DMRB), 7(3).

Hooghwerff, J., van Gils, E. W. and Reinink, H. F. (2013) Influence of road surface type on RR: results of the measurements.

Horvath, C. D. and Cannon, M. (2012) Future Material Opportunities and Direction for Lightweighting Automotive Body Structures. General Motors Company.

House of Commons (2014) ‘Maintaining strategic infrastructure: roads’, House of Commons, Committee of Public Accounts. 15th report of session 2014/15., (September).

Hu, J., Gao, X., Wang, R. and Sun, S. (2017) ‘Research on Comfort and Safety Threshold of Pavement Roughness’, Transportation Research Record: Journal of the Transportation Research Board, 2641. doi: 10.3141/2641-17.

Huang, X. and Peng, H. (2018) ‘Eco-Routing based on a Data Driven Fuel Consumption Model’, arXiv preprint arXiv:1801.08602.

IBM (2016) ‘10 Key Marketing Trends For 2017’, IBM Marketing Cloud. Available at: https://www.ibm.com/downloads/cas/XKBEABLN.

ISO (2005) ‘ISO 18164:2005 - Passenger car, truck, bus and motorcycle tires - Methods of measuring RR’, International Standards Organization. Available at: https://www.iso.org/standard/33328.html.

ISO (2006) ‘EN ISO 14040:2006 - Environmental Management: Life cycle assessment; Principles and framework’, International Standards Organization.

ISO (2009) ‘ISO 28580:2009 - Passenger car, truck and bus tires - Methods of measuring RR - Single point test and correlation of measurement results’, International Standards Organization. Available at: https://www.iso.org/standard/44770.html.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (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.

Jamieson, N. J. and Cenek, P. D. (1997) 'Effects of Pavement Construction on the Fuel Consumption of Trucks', Opus Central Laboratories.

Jaynes, E. T. (2014) Probability Theory: The Logic of Science. 1st edn. Cambridge, UK: Cambridge University Press.

Jianhua, L., Muench, S. T., Mahoney, J. P., White, G. C., Pierce, L. M. and Sivaneswaran, N. (2004) Application of HDM-4 in the WSDOT Highway System. Washington, DC

Jiao, X. (2015) Effect of Pavement-Vehicle Interaction on Highway Fuel Consumption and Emission. Florida International University. doi: 10.25148/etd.FIDC000142.

Jiao, X. and Bienvenu, M. (2016) 'The influence of pavement–vehicle interaction on highway fuel consumption by field measurement', Transport, 31(2), pp. 202–210. doi: 10.3846/16484142.2016.1193048.

Kahle, D. and Wickham, H. (2016) 'R Package: ggmap, Ver. 2.6.1', CRAN. Available at: https://cran.r-project.org/web/packages/ggmap/index.html.

Kan, Z., Tang, L., Kwan, M.-P. and Zhang, X. (2018) 'Estimating Vehicle Fuel Consumption and Emissions Using GPS Big Data', International Journal of Environmental Research and Public Health, 15(4). doi: 10.3390/ijerph15040566.

Kane, M., Scharnigg, K. and Schwalbe, G. (2009) 'TYROSAFE D10: Report on different parameters influencing skid resistance, RR and noise emissions', Tire and Road Surface Optimisation for Skid Resistance and Further Effects - Coordination Action FP7-217920 - WP3.

Karlsson, M., Gardhagen, R., Ekman, P., Söderblom, D. and et al. (2015) 'Aerodynamics of Timber Trucks - a Wind Tunnel Investigation', SAE Technical Paper, (2015-01-1562). doi: 10.4271/2015-01-1562.

Karlsson, R., Carlson, A. and Dolk, E. (2012) Energy use generated by traffic and pavement maintenance - Decision support for optimization of low RR maintenance treatments (MIRIAM), VTI notat 36A–2012. Linköping, Sweden.

Karlsson, R., Hammarström, U., Sörensen, H. and Eriksson, O. (2011) Road surface influence on RR Coastdown measurements for a car and an HGV. VTI notat 24A-2011. Linköping, Sweden.

Kerali, H. R., Odoki, J. B. and Stannard, E. E. (2006) Overview of HDM-4, Volume 1, The Highway Development and Management Series, International Study of Highway Development and Management (ISOHDM). World Roads Association (PIARC). Paris, France.

Khaidem, L., Saha, S. and Dey, S. R. (2016) 'Predicting the direction of stock market prices using random fores', CoRR. Available at: http://arxiv.org/abs/1605.00003.

Khavassefat, P. (2014) Vehicle-Pavement Interaction. KTH Royal Institute of Technology. Stockholm, Sweden.

Khavassefat, P., Jelagin, D. and Birgisson, B. (2015) 'Dynamic response of flexible pavements at vehicle–road interaction', Road Materials and Pavement Design, 16(2), pp. 256–276. doi: 10.1080/14680629.2014.990402.

Klein, P. and Hamet, J. F. (2004) 'Road texture and rolling noise: an envelopment procedure for tire-road contact', HAL, archives-ouvertes.fr - Rapport de recherche: hal-00546120, p. 17.

Klingbeil, W. W. (1980) 'Theoretical Prediction of Test Variable Effects, Including Twin-Rolls, on RR'. SAE International. doi: 10.4271/800088.

Ko, M. (2011a) 'Incorporating Vehicle Emissions Models into the Geometric Highway Design Process', Transportation Research Record: Journal of the Transportation Research Board, 2503(December), pp. 1–9. doi: 10.3141/2503-01.

Ko, M. (2011b) 'Incorporating Vehicle Emissions Models into the Geometric Highway Design Process', Transportation Research Record: Journal of the Transportation Research Board, 2503(December), pp. 1–9. doi: 10.3141/2503-01.

Ko, M. (2014) 'Environmental Conscious Highway Design for Horizontal Curves by', in Transportation Research Board 93rd Annual Meeting. January 12-16. Washington, DC.

Kuhn, M. (2018) ‘R Package: caret, Ver. 6.0-80’, CRAN. Available at: https://cran.r-project.org/web/packages/caret/index.html.

Laganier, R. and Lucas, J. (1990) ‘Influence of pavement evenness and macrotexture on fuel consumption’, in Meyer, W. E. and Reichert, J. (eds) Surface Characteristics of Roadways: International Research and Technologies. Philadelphia: American Society for Testing and Materials, pp. 454–459. doi: https://doi.org/10.1520/STP23381S.

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.

Lee, H. W., Cho, J. R., Jeong, W. B., Jeong, K. M. and Kim, K. W. (2014) ‘Mesh generation and hysteretic loss prediction of 3-D periodic patterned tire’, International Journal of Automotive Technology, 15(3), pp. 411–417. doi: 10.1007/s12239-014-0043-5.

Liaw, A., Wiener, M., Breimann, L. and Cutler, A. (2017) ‘R Package: randomForest, Ver. 4.6-12’, CRAN. doi: 10.5244/C.22.54.

Loecher, M. (2018) ‘R Package: RgoogleMaps, Ver. 1.4.2’, CRAN. Available at: https://cran.r-project.org/web/packages/RgoogleMaps/index.html.

Louhghalam, A., Akbarian, M. and Ulm, F.-J. (2015) ‘Roughness-Induced Pavement–Vehicle Interactions’, Transportation Research Record: Journal of the Transportation Research Board. Transportation Research Board, 2525(October), pp. 62–70. doi: 10.3141/2525-07.

Ma, H., Xie, H., Huang, D. and Xiong, S. (2015) ‘Effects of driving style on the fuel consumption of city buses under different road conditions and vehicle masses’, Transportation Research Part D: Transport and Environment. Elsevier Ltd, 41, pp. 205–216. doi: http://dx.doi.org/10.1016/j.trd.2015.10.003.

Mansura, D. A., Thom, N. H. and Beckedahl, H. J. (2017) ‘A Novel Multiscale Numerical Model for Prediction of Texture-Related Impacts on Fuel Consumption’, Tire Science and Technology, 45(1), pp. 55–70. doi: 10.2346/tire.17.450104.

Mansura, D. A., Thom, N. H. and Beckedahl, H. J. (2018a) ‘Numerical and Experimental

Predictions of Texture-Related Influences on RR', Transportation Research Record: Journal of the Transportation Research Board. doi: 10.1177/0361198118776114.

Mansura, D. and Thom, N. (2018b) 'Numerical and Experimental Predictions of Pavement Texture Effects on Vehicle Fuel Consumption', in 7th Transport Research Arena TRA 2018, April 16-19, 2018. Vienna, Austria.

McCulloch, W. S. and Pitts, W. (1943) 'A logical calculus of the ideas immanent in nervous activity', The Bulletin of Mathematical Biophysics, 5(4), pp. 115–133. doi: 10.1007/BF02478259.

McRobbie, S., Wright, A. and Willett, M. (2002) 'The effect of Road Geometry on the interpretation of Longitudinal Profile', TRL Published Report PR/CSN/08/02.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C. and Lin, C.-C. (2017) 'R Package: e1071, Ver. 1.6-8', CRAN. Available at: https://cran.r-project.org/web/packages/e1071/index.html.

Michelin (2003) 'The tire: RR and fuel savings', Report. Clermont-Ferrand, France.

Miege, A. J. P. and Popov, A. A. (2005) 'The RR of truck tires under a dynamic vertical load', Vehicle System Dynamics, 43(SUPPL.), pp. 135–144. doi: 10.1080/00423110500140039.

Mitchell, T. M. (1997a) 'Decision Tree Learning', Machine Learning, pp. 52–80. doi: 10.1007/978-3-642-27737-5_117-2.

Mitchell, T. M. (1997b) Machine Learning, McGraw-Hill Science/Engineering/Math. New York, NY, USA: WCB/McGraw-Hill. doi: 10.1145/242224.242229.

NAPA (2017) 'Product Category Rules for Asphalt Mixtures', National Asphalt Pavement Association (NAPA).

NAS (2010) Technologies and approaches to reducing fuel consumption of medium and heavy-duty vehicles. National Academy of Sciences. doi: 10.17226/12845.

Ng, A. (2018) Machine Learning Yearning - Technical Strategy for AI Engineers in the Era of Deep Learning. Stamford, CA, USA: Draft.

Nitsche, P. and Spielhofer, R. (2009) 'TYROSAFE - D06 - Report on policies and standards

concerning skid resistance, RR and noise emissions', Report on policies and standards concerning skid resistance, RR and noise emissions - WP1.

Noshadravan, A., Kirchain, R. and Gregory, J. R. (2012) 'Key Drivers of Uncertainty in Pavement LCA - Life Cycle Assessment Research Brief', Concrete Sustainability Hub@MIT.

NRC (2006) 'Tire and Passenger Vehicle Fuel Economy', National Research Council of the National Academies, Transportation Research Board, Special Report 286.

Odoki, J. B. (2016) 'Case Study: HDM-4 adaptation for analyzing Kenya roads', International Conference on Transport and Road Research. Mombasa, Kenya.

Odoki, J. B., Anyala, M. and Bunting, E. (2013) 'HDM-4 adaptation for strategic analysis of UK local roads', Transport, Institution of Civil Engineering (ICE), 166(2), pp. 65–78. doi: 10.1680/tran.9.00026.

Pacejka, H. B. (2002) Tire and Vehicle Dynamics, Tire and Vehicle Dynamics. doi: 10.1016/B978-0-7506-6918-4.X5000-X.

Perrotta, F., Parry, T. and Neves, L. C. (2017a) 'Application of Machine Learning for Fuel Consumption Modelling of Trucks', in 2017 IEEE International Conference on Big Data. December 11-14, 2017. Boston, MA.

Perrotta, F., Parry, T. and Neves, L. C. (2017b) '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., Trupia, L., Parry, T. and Neves, L. C. (2017c) '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.

du Plessis, H. W., Visser, A. T. and Curtayne, P. C. (1990) 'Fuel Consumption of Vehicles as Affected by Road-surface Characteristics', in Meyer, W. E. and Reichert, J. (eds)

Surface Characteristics of Roadways: International Research and Technologies. Philadelphia: American Society for Testing and Materials, pp. 480–496. doi: https://doi.org/10.1520/STP23383S.

Prasad, C., Swamy, A. K. and Tiwari, G. (2013) 'Calibration of HDM-4 Emission Models for Indian Conditions', Procedia - Social and Behavioral Sciences. Elsevier B.V., 104, pp. 274–281. doi: 10.1016/j.sbspro.2013.11.120.

Rakha, H. A., Ahn, K., Moran, K., Saerens, B. and Bulck, E. Van den (2011) 'Virginia Tech Comprehensive Power-Based Fuel Consumption Model: Model development and testing', Transportation Research Part D: Transport and Environment, 16(7), pp. 492–503. doi: http://dx.doi.org/10.1016/j.trd.2011.05.008.

Rebitzer, G., Ekvall, T., Frischknecht, R., Hunkeler, D., Norris, G., Rydberg, T., Schmidt, W., Suh, S., Weidema, B. and Pennington, D. W. (2004) 'Life Cycle Assessment: Framework, goal and scope definition, inventory analysis, and applications.', Environment International, 30(5), pp. 701–720. doi: 10.1016/j.envint.2003.11.005.

Riedmiller, M. and Braun, H. (1993) 'A direct adaptive method for faster backpropagation learning: The RPROP algorithm', in IEEE International Conference on Neural Networks. San Francisco, pp. 586–591. doi: 10.1109/ICNN.1993.298623.

Ripley, B. and Lapsley, M. (2017) 'R Package: RODBC, Ver. 1.3-15', CRAN. Available at: https://cran.r-project.org/web/packages/RODBC/index.html.

Roeth, M., Rondini, D., Schaller, D. A. and Wang, Z. (2018) '2018 Annual Fleet Fuel Study', North American Council for Freight Efficiency (NACFE). Technical Report., (August). doi: 10.13140/RG.2.2.35621.35047.

SAE (2006) 'SAE J1269:2006 - RR Measurement Procedure for Passenger Car, Light Truck, and Highway Truck and Bus Tires', Society of Automotive Engineering. Available at: https://www.sae.org/standards/content-/j1269_200609/.

SAE (2016) 'SAE J1939-71, Vehicle Application Layer - Surface Vehicle Recommended Practice', Society of Automotive Engineering. SAE International. Available at: http://standards.sae.org/j1939/71_201610/.

Sakhaeifar, M., Banihashemrad, A., Liao, G. and Waller, B. (2017) 'Tire–pavement interaction noise levels related to pavement surface characteristics', Road Materials and Pavement Design, 0629, pp. 1–13. doi: 10.1080/14680629.2017.1287770.

Sakhaeifar, M., Liao, G. and Waller, B. (2012) 'Relationship between Tire-Pavement Interaction Noise and Surface Characteristics in 2009 Pavement Test Track Cycle', Transportation research board 92nd annual meeting, 500, pp. 1–15.

Saltzman, E. J. and Meyer, R. R. (1999) 'A Reassessment of Heavy-Duty Truck Aerodynamic Design Features and Priorities', NASA, National Aeronautics and Space Administration. Report TP-1999-206574, (June).

Samuel, A. L. (1959) 'Some Studies in Machine Learning Using the Game of Checkers', IBM Journal of Research and Development, 3(3), pp. 210–229. doi: 10.1147/rd.33.0210.

Sandberg, T. (2001a) Heavy Truck Modeling for Fuel Consumption Simulations and Measurements. Linköping University. Linköping, Sweden. doi: ISSN 0280-7971.

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.

Sandberg, U. (1997) Influence of Road Surface Texture on Traffic Characteristics Related to Environment, Economy and Safety. A state of the art Study Regarding Measures and Measuring Methods, VTI notat 53A–1997. Linköping, Sweden.

Sandberg, U. (2001b) 'Tire/ road noise - Myths and realities', in the 2001 International Congress and Exhibition on Noise Control Engineering, p. 22.

Sandberg, U., Bergiers, A., Ejsmont, J. A., Goubert, L. and Karlsson, R. (2011a) 'MIRIAM SP1 D.4: Road surface influence on tire/road RR', Report from Models for RR In Road Infrastructure Asset Management systems (MIRIAM)

Sandberg, U., Bergiers, A., Ejsmont, J. A., Goubert, L. and Zöller, M. (2012) 'MIRIAM SP1 D.2: RR – Measurement Methods for Studies of Road Surface Effects', Report from Models for RR In Road Infrastructure Asset Management systems (MIRIAM). Available at: http://miriam-co2.net/Publications/MIRIAM_SP1_Meas_Method_-

Report-_%28SP1_Deliverable_2%29.pdf.

Sandberg, U. and Ejsmont, J. A. (2002) Tire/Road Noise Reference Book. Kisa, Sweden: INFORMEX.

Sandberg, U., Haider, M., Conter, M., Goubert, L., Bergiers, A., Glaeser, K.-P., Schwalbe, G., Zöller, M., Boujard, O., Hammarström, U., Karlsson, R., Ejsmont, J., Wang, T. and Harvey, J. (2011b) 'MIRIAM SP1 D.1: RR-Basic information and State-of-the-Art on measurement methods', Report from Models for RR In Road Infrastructure Asset Management systems (MIRIAM). Available at: http://miriam-co2.net/Publications-/MIRIAM_SoA_Report_Final_110601.pdf.

Santero, N. J., Harvey, J. and Horvath, A. (2011a) 'Environmental policy for long-life pavements', Transportation Research Part D: Transport and Environment. Elsevier Ltd, 16(2), pp. 129–136. doi: 10.1016/j.trd.2010.09.005.

Santero, N. J. and Horvath, A. (2009) 'Global warming potential of pavements', Environmental Research Letters, 4(3), p. 034011. doi: 10.1088/1748-9326/4/3/034011.

Santero, N. J., Masanet, E. and Horvath, A. (2011b) 'Life-Cycle Assessment of Pavements: A Critical Review of Existing Literature and Research', SN3119a. Skokie, Illinois, USA: Portland Cement Association, p. 81. doi: 10.1016/j.resconrec.2011.03.010.

Santero, N. J., Masanet, E. and Horvath, A. (2011c) 'Life-cycle assessment of pavements Part II: Filling the research gaps', Resources, Conservation and Recycling, pp. 810–818. doi: 10.1016/j.resconrec.2011.03.009.

Santos, J., Bryce, J., Flintsch, G., Ferreira, A. and Diefenderfer, B. (2015a) 'A life cycle assessment of in-place recycling and conventional pavement construction and maintenance practices', Structure and Infrastructure Engineering. Taylor & Francis, 11(9), pp. 1199–1217. doi: 10.1080/15732479.2014.945095.

Santos, J., Ferreira, A. and Flintsch, G. (2015b) 'A life cycle assessment model for pavement management: Methodology and computational framework', International Journal of Pavement Engineering. Taylor & Francis, 16(3), pp. 315–336. doi:

10.1080/10298436.2014.942862.

Santos, J., Flintsch, G. and Ferreira, A. (2017) 'Environmental and economic assessment of pavement construction and management practices for enhancing pavement sustainability', Resources, Conservation and Recycling, 116(2), pp. 15–31. doi: 10.1016/j.resconrec.2016.08.025.

Sayers, M. (1989) 'Two Quarter-Car Models for Defining Road Roughness: IRI and HRI', Transportation Research Record: Journal of the Transportation Research Board, (6), pp. 165–172.

Schilling, G. D. (1997) Modeling Aircraft Fuel Consumption with a Neural Network. Virginia Polytechnic Institute and State University, Blacksburg, VA.

Scottish Enterprise (2003) 'Review of Timber Haulage and Forest Roads - Solutions for Cost-effective Transport and Strategic Benefits in Scotland', Final Report, (March).

Sjödin, Å., Jerksjö, M., Sandström, C., Erlandsson, L., Almén, J., Ericsson, E., Larsson, H., Hammarström, U., Mohammed-Reza, Y. and Johansson, H. (2009) Implementering av ARTEMIS road model i Sverige - Implementation of the ARTEMIS road model in Sweden - IVL Rapport B1831. Göteborg, Sweden.

Smith, M. C. and Swift, S. J. (2011) 'Power dissipation in automotive suspensions', Vehicle System Dynamics, 49(1–2), pp. 59–74. doi: 10.1080/00423110903427421.

Sohaney, R. C. and Rasmussen, R. O. (2013) 'Pavement Texture Evaluation and Relationships to RR at MnROAD', MnRoad Technical Report, MN/RC 2013-16 2, (June).

Spencer-Bickle, A., Aduan, J., Gleave, J. and Halden, D. (2017) 'International Journey Time Benchmarking: Strategic Road Networks Analysis Report', Office of Rail and Roads (ORR).

Stergiou, C. and Siganos, D. (1995) Neural Networks. Available at: https://www.doc.ic.ac-.uk/~nd/surprise_96/journal/vol4/cs11/report.html#What is a Neural Network (Accessed: 4 October 2016).

Stone, R. (2012) Introduction to Internal Combustion Engines, Basingstoke: Palgrave Macmillan. doi: 10.1017/CBO9781107415324.004.

Sudin, M. and Abdullah, M. (2014) 'Review of Research on Vehicles Aerodynamic Drag Reduction Methods', International Journal of Mechanical & Mechatronics Engineering, (02), pp. 35–47. Available at: http://www.ijens.org/Vol_14_I_02/-145302-6868-IJMME-IJENS.pdf.

Svenson, G. and Fjeld, D. (2012) 'The influence of road characteristics on fuel consumption for logging trucks', HVTT12: 12th International Symposium on Heavy Vehicle Transport Technology, 31(5), pp. 526–536. doi: 10.1080/02827581.2015.1092574.

Svenson, G. and Fjeld, D. (2016) 'The impact of road geometry and surface roughness on fuel consumption of logging trucks', Scandinavian Journal of Forest Research. Taylor & Francis, 31(5), pp. 526–536. doi: 10.1080/02827581.2015.1092574.

Tao, L. (2010) The Influence of Pavement Stiffness on Vehicle Fuel Consumption. University of Nottingham. Nottingham, UK.

Thom, N. H., Lu, T. and Parry, T. (2010) 'Fuel Consumption Due to Pavement Deflection Under Load', in Claisse, P., Ganjian, E., Canpolat, F., and Naik, T. (eds) 2nd International Conference on sustainable Construction Materials and Technologies. Ancona, Italy: Special Technical Proceedings.

Thube, D. T. (2011) 'Highway Development and Management Model (HDM-4): calibration and adoption for low-volume roads in local conditions', International Journal of Pavement Engineering, 14(1), pp. 50–59. doi: http://dx.doi.org/10.1080/-10298436.2011.606320.

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.

Trani, A. A., Schilling, G., Baik, H. and Seshadri, A. (2004) 'A Neural Network Model to Estimate Aircraft Fuel Consumption', in AIAA 4th Aviation Technology, Integration and Operations (ATIO) Forum. Chicago, Illinois: American Institute of Aeronautics and Astronautics, p. 24.

Trupia, L. (2017) System Boundary Expansion in Road Pavement Life Cycle Analysis.

University of Nottingham. Nottingham, UK.

Trupia, L., Parry, T., Neves, L. C. and Lo Presti, D. (2016) 'RR contribution to a road pavement life cycle carbon footprint analysis', International Journal of Life Cycle Assessment, pp. 972–985. doi: 10.1007/s11367-016-1203-9.

UN (2015) 'Paris Agreement', United Nations Framework Convention on Climate Change - Conference of the Parties on its 21st session. doi: FCCC/CP/2015/L.9/Rev.1.

Valdés, M., Hidalgo, P., Chamorro, A. and Gutiérrez, D. (2011) 'Monitoring Asphalt Pavements to Calibrate HDM-4', in 8th International Conference on Managing Pavement Assets (ICMPA8). Santiago, Chile, p. 10.

Viner, H., Abbott, P., Dunford, A., Dhillon, N., Parsley, L. and Read, C. (2006) PPR148 - Surface Texture Measurement on Local Roads. Crowthorne, UK.

Walnum, H. J. and Simonsen, M. (2015) 'Does driving behavior matter? An analysis of fuel consumption data from heavy-duty trucks', Transportation Research Part D: Transport and Environment. Elsevier Ltd, 36, pp. 107–120. doi: 10.1016/j.trd.2015.02.016.

Wang, H. and Al-Qadi, I. L. (2009) 'Combined Effect of Moving Wheel Loading and Three-Dimensional Contact Stresses on Perpetual Pavement Responses', Transportation Research Record: Journal of the Transportation Research Board, 2095, pp. 53–61. doi: 10.3141/2095-06.

Wang, H., Al-Qadi, I. L. and Stanciulescu, I. (2010) 'Effect of Friciton on Rolling Tire – Pavement Interaction', USDOT Region V Regional University Transportation Center Final Report - NEXTRANS Project No 049IY02.

Wang, H., Al-Qadi, I. L. and Stanciulescu, I. (2012a) 'Simulation of tire-pavement interaction for predicting contact stresses at static and various rolling conditions', International Journal of Pavement Engineering, 13(4), pp. 310–321. doi: 10.1080/10298436.2011.565767.

Wang, H., Al-Qadi, I. L. and Stanciulescu, I. (2014a) 'Effect of Surface Friction on Tire – Pavement Contact Stresses during Vehicle Maneuvering', Journal of Engineering Mechanics (ASCE), 140(4), pp. 1–8. doi: 10.1061/(ASCE)EM.1943-

7889.0000691.

Wang, J. and Rakha, H. A. (2017) 'Fuel consumption model for heavy duty diesel trucks: model development and testing', Transportation Research Part D: Transport and Environment. Elsevier Ltd, 55, pp. 127–141. doi: 10.1016/j.trd.2017.06.011.

Wang, T., Harvey, J. and Kendall, A. (2014b) 'Reducing greenhouse gas emissions through strategic management of highway pavement roughness', Environmental Research Letters, 9(3). doi: 10.1088/1748-9326/9/3/034007.

Wang, T., Harvey, J., Lea, J. and Kim, C. (2014c) 'Impact of Pavement Roughness on Vehicle Free-Flow Speed', Journal of Transportation Engineering, 140(9), p. 04014039. doi: 10.1061/(ASCE)TE.1943-5436.0000689.

Wang, T., Lee, I.-S., Harvey, J., Kendall, A., Lee, E. B. and Kim, C. (2012b) UCPRC Life Cycle Assessment Methodology and Initial Case Studies on Energy Consumption and GHG Emissions for Pavement Preservation Treatments with Different RR. Davis, California, USA.

Wang, T., Lee, I. S., Kendall, A., Harvey, J., Lee, E. B. and Kim, C. (2012c) 'Life cycle energy consumption and GHG emission from pavement rehabilitation with different RR', Journal of Cleaner Production. Elsevier Ltd, 33, pp. 86–96. doi: 10.1016/j.jclepro.2012.05.001.

Watkins, S. (2005) 'Recent Developments in Road Vehicle Aerodynamics', International Conference on Mechanical Engineering 2005, (December), p. ICME-KEY-01. Available at: http://buet.ac.bd/me/icme/icme2005/Proc-eedings/PDF/ICME05-KEY-01.pdf.

Webb, C. R. (1952) 'The Effect of Gradient on Fuel Consumption and Speed of a Road Vehicle', Proceedings of the Institution of Mechanical Engineers, 89(March), pp. 104–117. doi: 10.1243/PIME.

Werbos, P. (1974) Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences, Foundations. Harvard University, Washington (DC). Available at: http://www.citeulike.org/group/1938/article/1055600.

Werner, K. W. D., Smith, J., Fox, J., Werner, E., Dunkle, R. and Childers, D. (2013) 'Driver Behavior and Fuel Efficiency'. Available at: https://deepblue.lib.umich.edu/-bitstream/handle/2027.42/98892/karldw.pdf?sequence=1.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C. and Woo, K. (2018a) 'R Package: ggplot2, Ver. 3.0.0', CRAN. Available at: https://cran.r-project.org/web/packages/ggplot2/index.html.

Wickham, H., François, R., Henry, L. and Müller, K. (2018b) 'R Package: dplyr, Ver. 0.7.6', CRAN. Available at: https://cran.r-project.org/web/packages-/dplyr/index.html.

Wolpert, D. H. and Macready, W. G. (1997) 'No free lunch theorems for optimization', IEEE Transactions on Evolutionary Computation, 1(1), pp. 67–82. doi: 10.1109/4235.585893.

Xu, X., Gregory, J. and Kirchain, R. (2015) 'Role of the use phase and pavement-vehicle interaction in comparative life cycle assessment', in Transportation Research Board 94th Annual Meeting. Washington, DC.

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.

Yamada, M., Jitkrittum, W., Sigal, L., Xing, E. P. and Sugiyama, M. (2013) 'High-Dimensional Feature Selection by Feature-Wise Non-Linear Lasso', pp. 1–18. Available at: http://arxiv.org/abs/1202.0515.

Yang, S., Li, S. and Lu, Y. (2010) 'Investigation on dynamical interaction between a heavy vehicle and road pavement', Vehicle System Dynamics, 48(8), pp. 923–944. doi: 10.1080/00423110903243166.

Zaabar, I. and Chatti, K. (2010) 'Calibration of HDM-4 models for estimating the effect of pavement roughness on fuel consumption for U. S. conditions', Transportation Research Record, (2155), pp. 105–116. doi: 10.3141/2155-12.

Zaabar, I. and Chatti, K. (2014) 'Estimating Vehicle Operating Costs due to Pavement Surface

Conditions', in Transportation Research Board 93rd Annual Meeting. January 12-16. Washington, DC.

Zabat, M., Stabile, N., Frascaroli, S. and Browand, F. (1995) 'The Aerodynamic Performance of Platoons: Final Report', California PATH Research Report, UCB-ITS-PRR-95-35, (October).

Zaniewski, J. P. (1989) Effect of Pavement Surface Type on Fuel Consumption.

Zaniewski, J. P., Butler, B. C., Cunningham, G., Elkins, G. E., Paggi, M. S. and Machemehl, R. (1982) 'Vehicle Operating Costs, Fuel Consumption, and Pavement Type and Condition Factors', FHWA-PL-82-001, Texas Research and Development Foundation.

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.

Zhang, Y., Guo, W. and Edu, C. (2016) 'On the Consistency of Feature Selection with Lasso for Non-linear Targets', in Proceedings of the 33rd International Conference on Machine Learning. New York, NY, USA: JMLR: W&CP.

Zhou, M. and Jin, H. (2017) 'Development of a transient fuel consumption model', Transportation Research Part D: Transport and Environment. Elsevier Ltd, 51, pp. 82–93. doi: 10.1016/j.trd.2016.12.001.

Zhou, M., Jin, H. and Wang, W. (2016) 'A review of vehicle fuel consumption models to evaluate eco-driving and eco-routing', Transportation Research Part D: Transport and Environment. Elsevier Ltd, 49(5), pp. 203–218. doi: 10.1016/j.trd.2016.09.008.

Zöller, M. (2014) 'ROSANNE WP3 Deliverable D3.1: State of the art on RR measurement devices'.



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

Appendices

Appendix A Hardware and software

This annex of the thesis is dedicated to the description of devices and tools used to perform the data analysis. Details are provided in regard to the specifications of the hardware and software used in the study to clarify performance and allow comparison with previous and future studies.

Appendix A.1 Computer used

The computer used for running the analysis in this study was a Lenovo Think Station (P720) provided with Windows 10 64-bit. The machine was equipped with Intel® Xeon® CPU E5-1620 v.3 with 8 cores of 3.50 Ghz clock frequency. That allows parallelization of processes and speed up the analysis of the large quantity of information available. Due to the large quantity of data to deal with the computer has been equipped with 32 Gb of RAM. No specific graphic processing unit was installed on the computer which is therefore equipped with the standard provided by Lenovo and that corresponds to a Nvidia NVS 310 with 512 Mb of dedicated RAM.

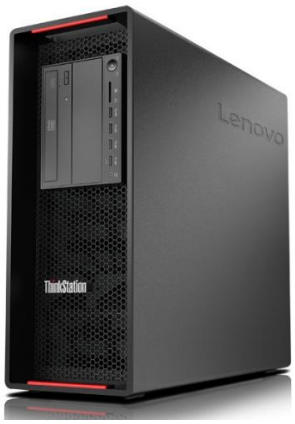


Figure A-1. The Lenovo Think Station P720 used in the study to carry out the data analysis.

At the current state of knowledge, future development of the methods may want to look into using GPUs to train ML models and use of CC and HPC to perform data cleaning. That should help in speeding up the whole data analysis.

Appendix A.2 R & RStudio

R is the main software used in this study to carry out the analysis of data. In the following sections details in regards R and R Studio are listed.

Appendix A.2.1 R

R is a free software for statistical analysis and graphics. It compiles and run on any kind of operating systems from UNIX to MacOS and obviously on Windows.

Major advantage of R is that everybody can upload their libraries of functions to the CRAN, the Comprehensive R Archive Network. That allows scientists from anywhere in the world to share their work and use or improve libraries created by others without having to implement the methods on their own. That also means that new methods may be released quicker on the CRAN than everywhere else, which means that scientists can access to already implemented versions of the newest methods available. That represents a significant advantage and may give possibility to

experiment new algorithms or test the latest implemented by other researchers to optimise performance.

For this study R version 3.4.1 “Single Candle” (2017) was used. Different versions of the software may work differently. It is therefore recommended that, in order to ensure repeatability of the study, the mentioned version of the software is used or, if a different version of the software is available, it is responsibility of the user to update the code provided in Annex C following the method described in the dedicated chapter (see sect. 4).

It is possible to download the version 3.4.1 (or other) versions of the software from the links available on the CRAN: https://cran.r-project.org/mirrors.html.

Appendix A.2.2 RStudio

RStudio is an integrated development environment (IDE) for R. Thanks to its graphical interface it makes R easier to use allowing direct code execution, script editing and visualization of outputs.

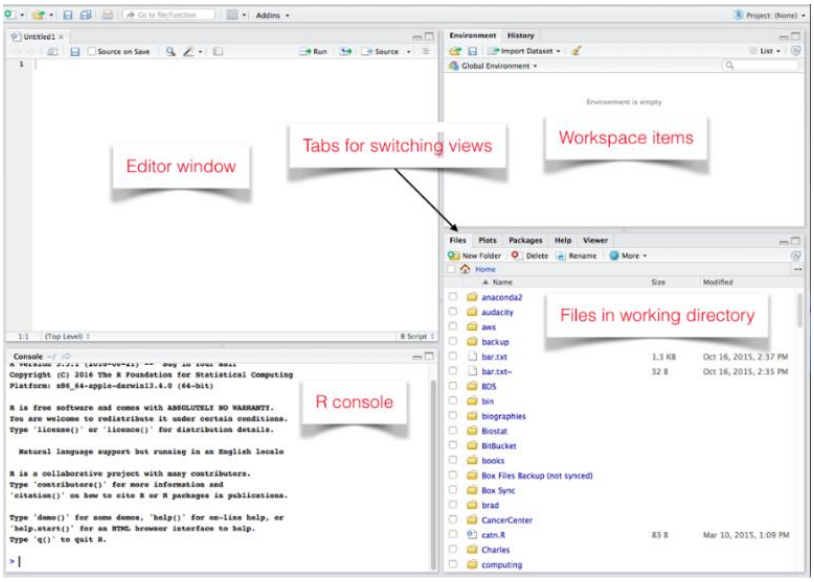


Figure A-2. The R Studio IDE.

RStudio is available in open source and commercial edition and runs on desktop (Windows, Mac or Linux) or through browser connected to the RStudio servers. For this study the open source version of the software has been used. In particular the software has been periodically updated with latest version installed that is the 1.1.447. Since RStudio provides just a nicer graphical interface to R it should not have effect on functionality of the R script developed and used in this thesis to analyse the data available.

Appendix A.2.3 R functions & packages used

R Studio is a very versatile software provided with an R code compiler. In order to analyse data, it is possible to use already developed packages, by creating one or by downloading them from the CRAN. Alternatively it is possible to write personalised functions to perform analysis of data, implement new methods or test new algorithms.

In this study, both already developed packages and personalised functions have been used to perform the analysis of data. It is possible to find full details about the developed functions and code in the Annex D.

In the following, full list and details in regards to the major packages used in the study to carry out the data analysis is given.

- RODBC (Ripley *et al.*, 2017): this package allows connection and interaction with MS SQL databases or other ODBC based databases.
- ggplot2 (Wickham *et al.*, 2018a): this package is one of the most complete tools for creating elegant data visualizations currently available. It represents a valid alternative to commercial software, such as Excel or Tableau (see Appendix A.5), and sometimes it demonstrates necessary to personalise plots adequately.

- ggmap (Kahle *et al.*, 2016): this package represents a useful add-in to the ggplot2 package which allows visualization of spatial data and models on top of static maps from various online sources including Google Maps.
- RgoogleMaps (Loecher, 2018): this package provides with comfortable interface to query Google servers for static maps.
- rgdal (Bivand *et al.*, 2018): this package provides bindings to the GDAL (Geospatial Data Abstraction Library) and PROJ.4 libraries which allows projection/transformation of geospatial data into different coordinate systems.
- dplyr (Wickham *et al.*, 2018b): this package is a complete, fast and consistent data manipulation tool to work with data frame objects in and out of memory.
- glmnet (Friedman *et al.*, 2017): this package implements extremely efficient procedures for fitting the lasso, elastic-net regularization path for LR, logistic and multinomial regression models, among others.
- caret (Kuhn, 2018): R package explicitly developed to perform classification and regression training. It also allows to create useful plots which allow optimisation of models.
- e1071 (Meyer *et al.*, 2017): this package is designed to train and develop support vector machines. It allows also use of other methods such as fuzzy clustering, naïve Bayes classification, shortest path computation, etc.
- randomForest (Liaw *et al.*, 2017): package designed to perform classification and regression analysis based on a forest of decision trees based on random inputs. Algorithm developed by Breiman (2001).
- neuralnet (Fritsch *et al.*, 2016): package specifically designed for training of neural networks. It implements various training algorithms such as the traditional

backpropagation, the resilient backpropagation, or the modified globally convergent algorithm among others.

Further information on each package, tools and functions implemented in each of them can be found in their reference manuals downloadable from the CRAN.

Appendix A.3 Microsoft SQL Server

Microsoft SQL Server (MS SQL) Management Studio is a software application for relational database management developed by Microsoft. Primary functions of MS SQL are to store and retrieving data as requested by the user from database servers (instances) which may run on the same computer or another, also remotely based or from the Internet.



Figure A-3. Logo of MS SQL software.

In this project, MS SQL was used to deal with Microlise data. In fact, Microlise stores and secures data from fleet managers in MS SQL databases which can be queried to retrieve any information needed. However, for security and privacy reasons no direct access to Microlise databases was guaranteed for this project and a dummy SQL database containing anonymised data from some fleet managers and specific time frames was created by Microlise and restored locally at the University of

Nottingham. From there data where queried and retrieved to be analysed then using R.

Appendix A.4 ArcGIS

ArcGIS is a geographic information system (GIS) software developed by ESRI (the Environmental Systems Research Institute). The software allows users to create and work with maps, analyse geographical information, share maps and geographic information with external applications and manage geographic information databases.



Figure A-4. ArcGIS logo.

In this study, ArcGIS, and specifically the ArcMap tool implemented in ArcGIS, was used to visualize and analyse data from the HAPMS of HE.

Appendix A.5 Tableau

Tableau is a data visualization (dataviz) software developed by Tableau Software and focused on business intelligence. It offers a very intuitive interface which allows the user to build informative and nicely represented plots within seconds. Tableau allows direct connection and visualization of different data in various formats

from the most disparate sources (including GIS) and implements tools to perform basic and advanced analytics interactively.



Figure A-5. Tableau logo.

In this study Tableau was used to improve data visualization and enhance interpretation of the analysis carried out in R.

Appendix B Selection of case study

This annex provides the reader with information regarding the data collected and in particular to the choice of the case study analysed in the study. Data distributions from the whole SRN network and those from the considered roads are compared for each case study and considerations are made about representability of road surface conditions typical of the SRN of England.

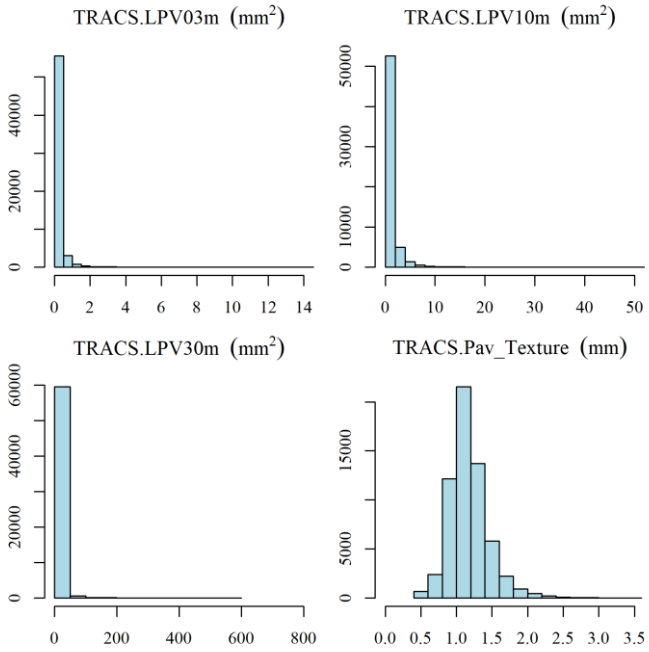


Figure B-1. Distribution of road surface conditions for the M18 and part of M1 considered in the study (for 10 m sections).

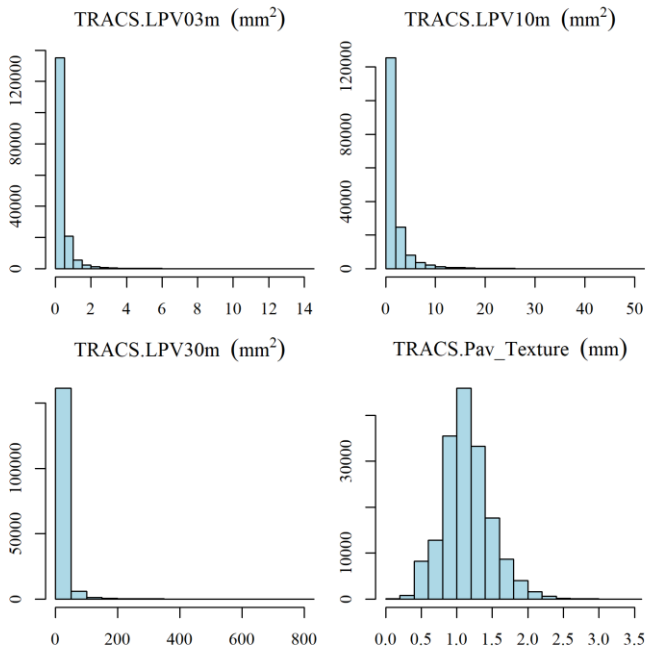


Figure B-2. Distribution of road surface conditions for the parts of A21, A41, A47, A49, A50, A259, M1 and M18 considered in the study (for 10 m sections).

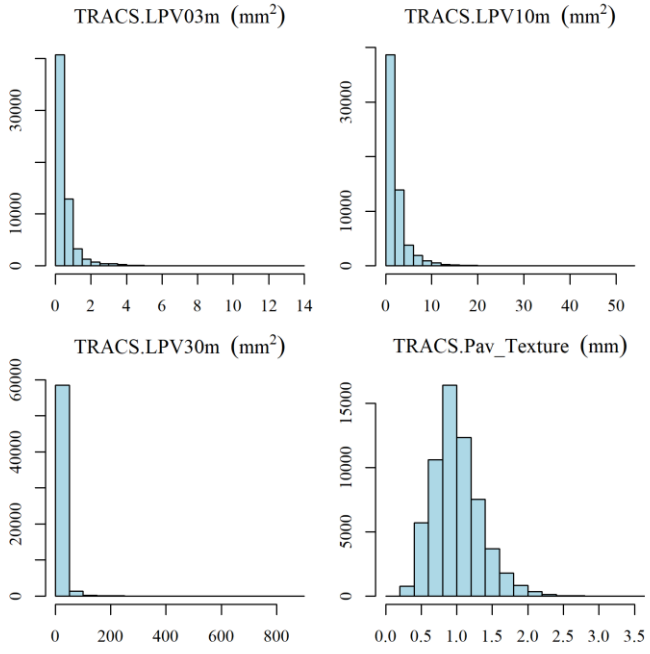


Figure B-3. Distribution of road surface conditions for the SRN of England (for 100 m sections).

From a comparison of Figure B-1, B-2 and B-3, it is possible to see that, consideration of parts of the A21, A47, A49, and A259 roads managed by HE, and the

part of A41 managed by the council of Oxfordshire, allows to consider a wide range of road conditions. That could not be explored by considering the A50, M18 and M1 only. However, in consideration of the Table 2A of the Annex 2 of the Vol. 7 of the DMRB , this range of speeds cannot be considered exhaustive of conditions representative of the UK. This reduces validity and applicability of the present study to conditions typical of major roads managed by HE and local authorities.

Appendix C Data distributions

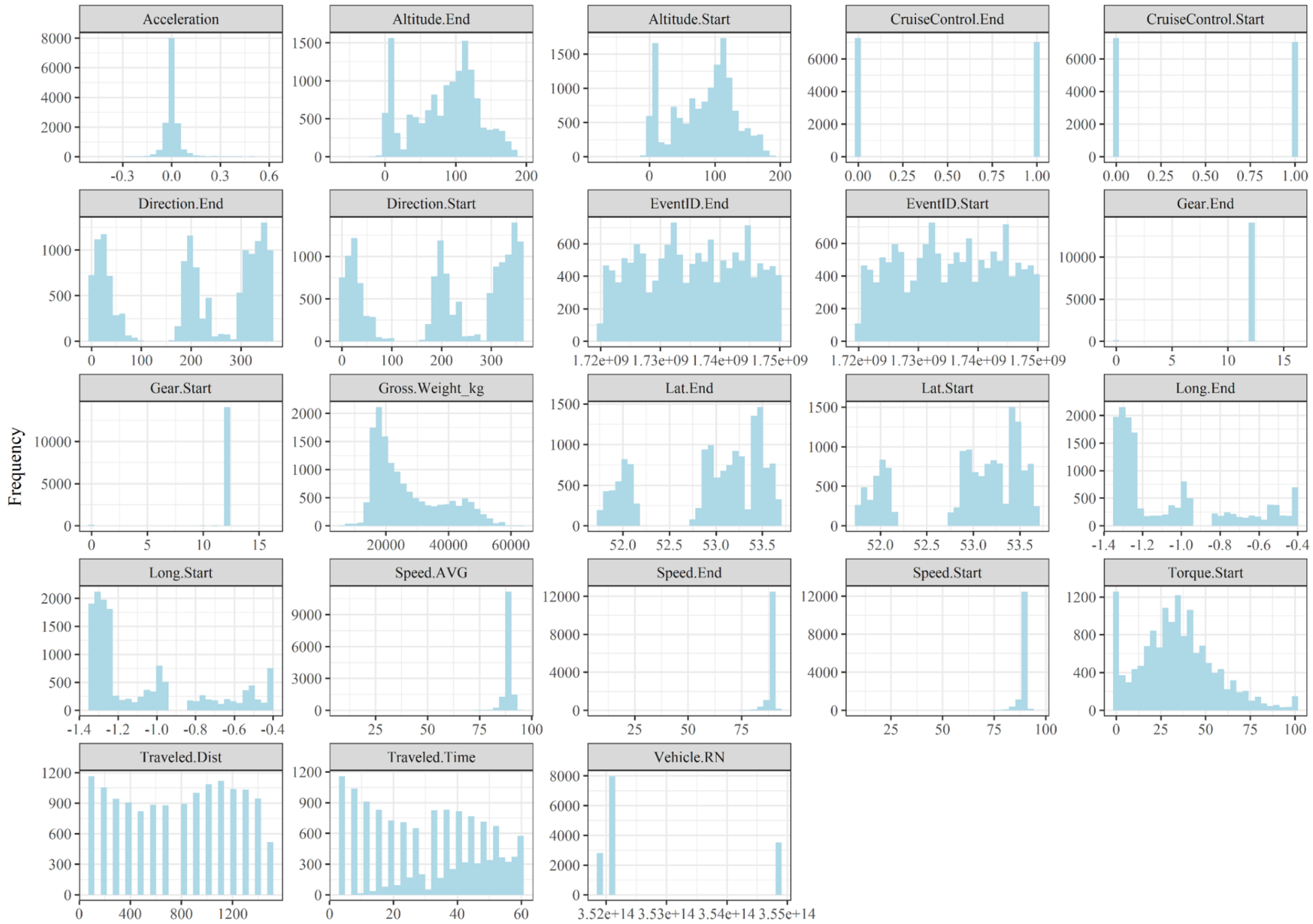


Figure C-1. Data distributions for the Dataset 1 (heavy trucks driving along motorways), Part 1.

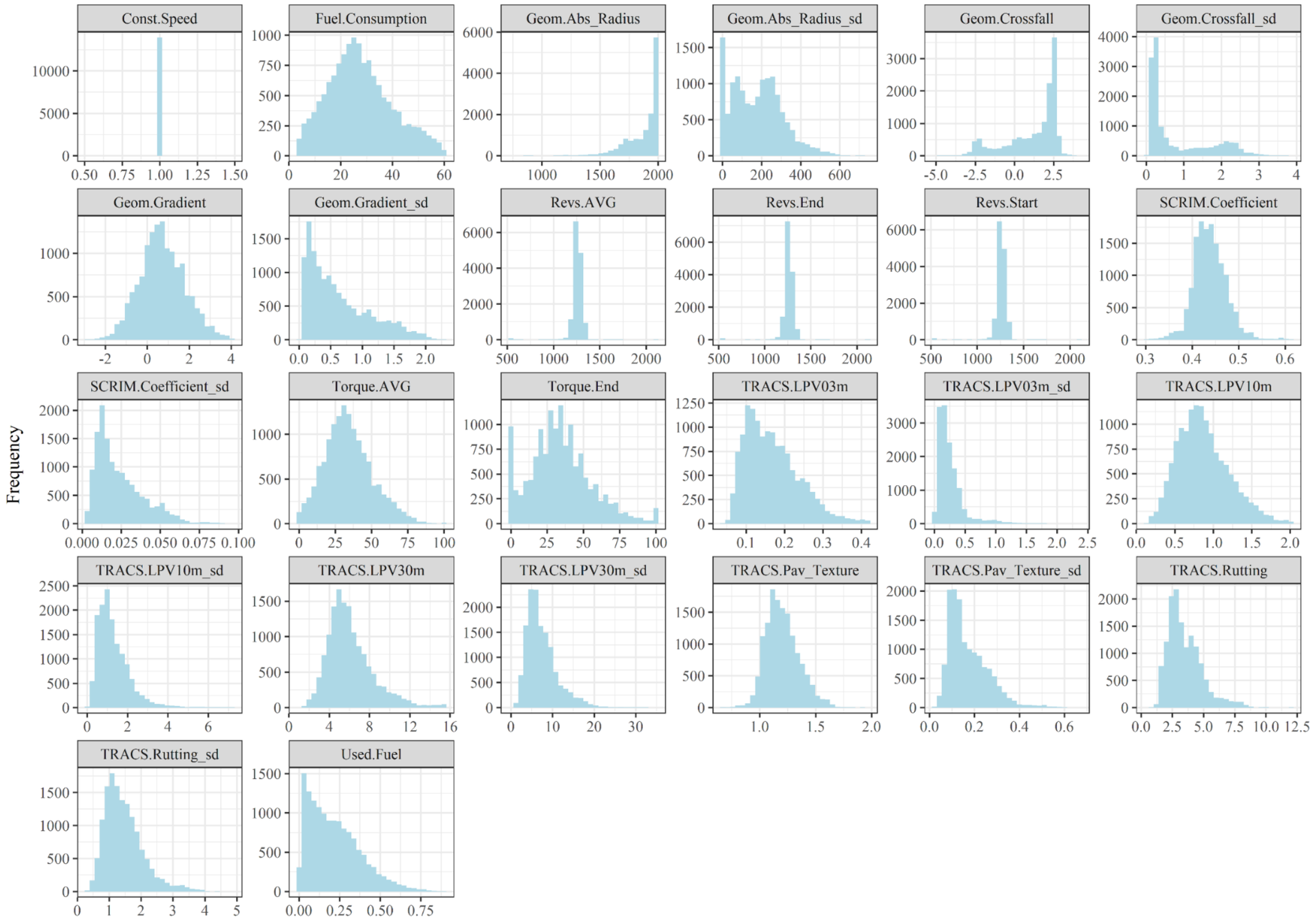


Figure C-2. Data distributions for the Dataset 1 (heavy trucks driving along motorways), Part 2.

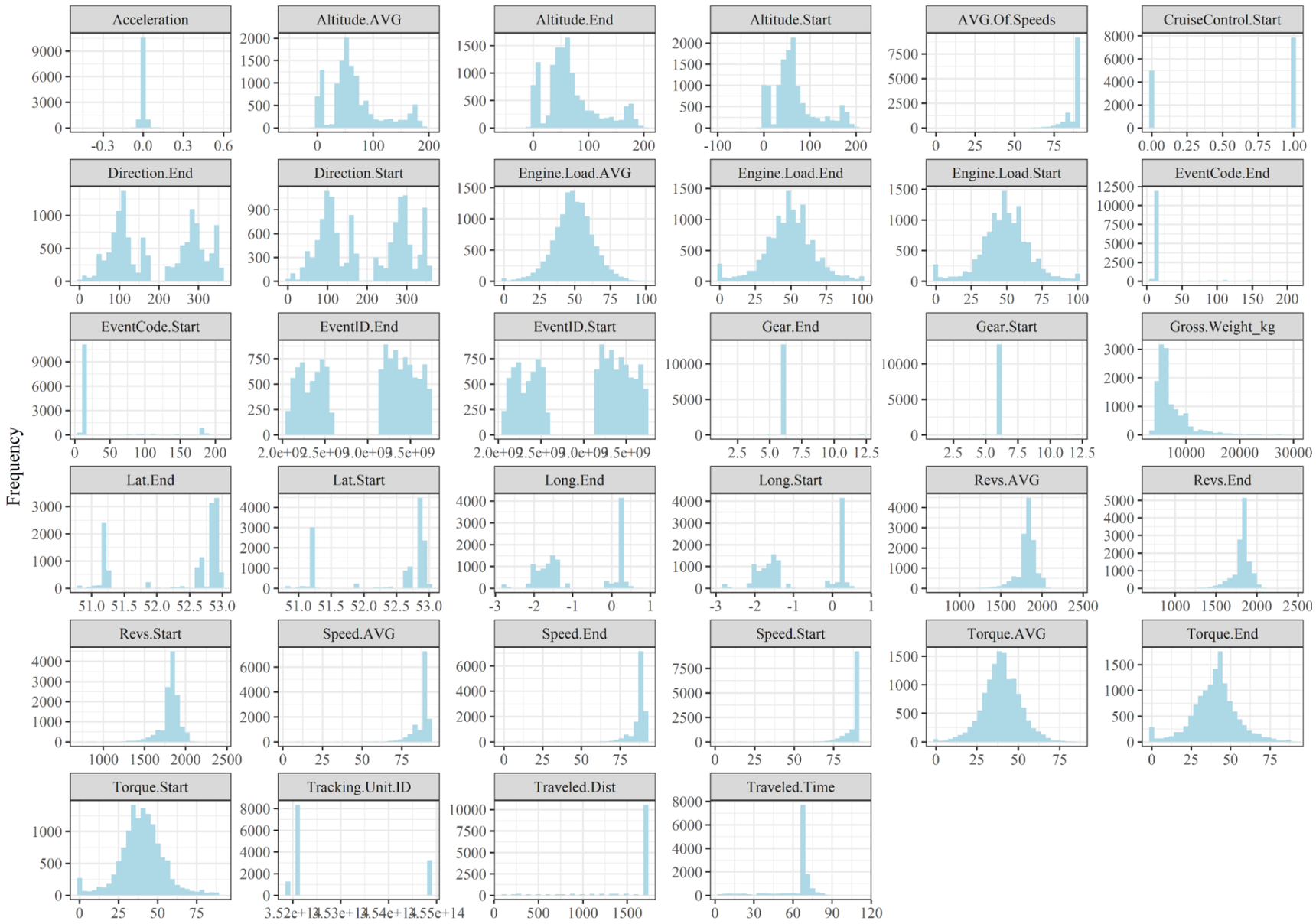


Figure C-3. Data distributions for the light trucks driving along A roads (App 1), Dataset 2, Part 1.



Figure C-4. Data distributions for the light trucks driving along A roads (App 1), Dataset 2, Part 2.

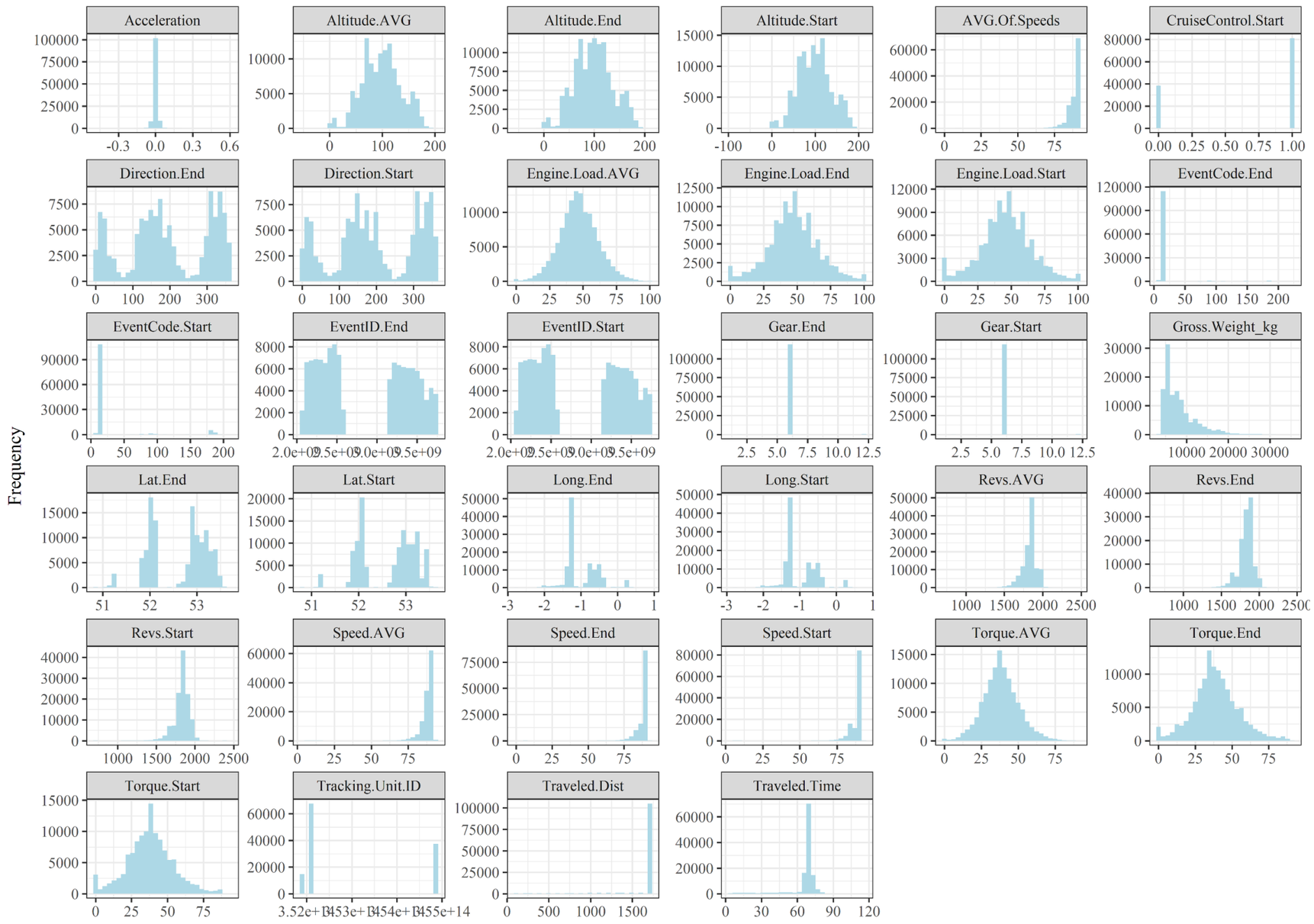


Figure C-5. Data distributions for the light trucks driving along M and A roads (App 2), Dataset 2, Part 1.

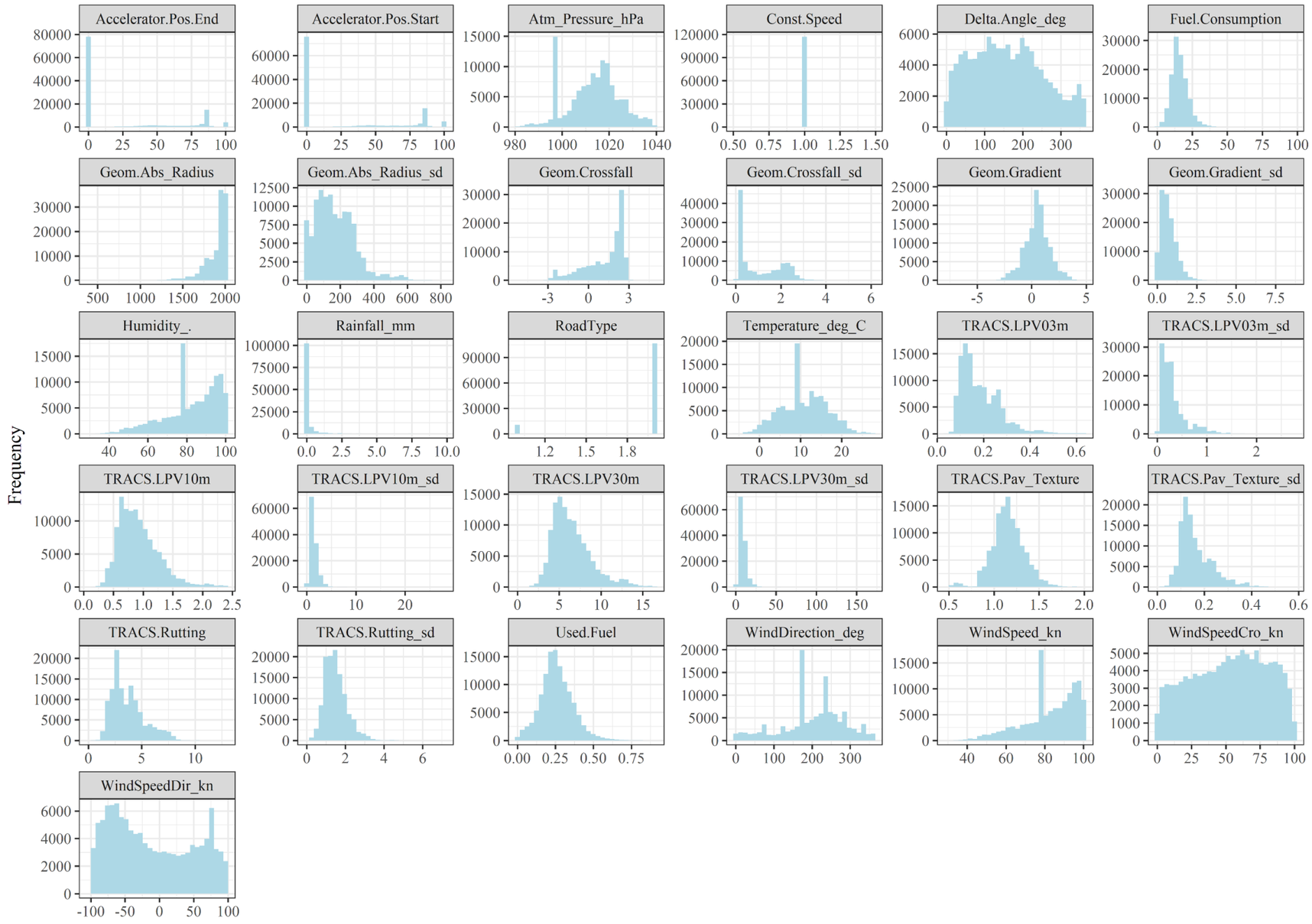


Figure C-6. Data distributions for the light trucks driving along M and A roads (App 2), Dataset 2, Part 2.

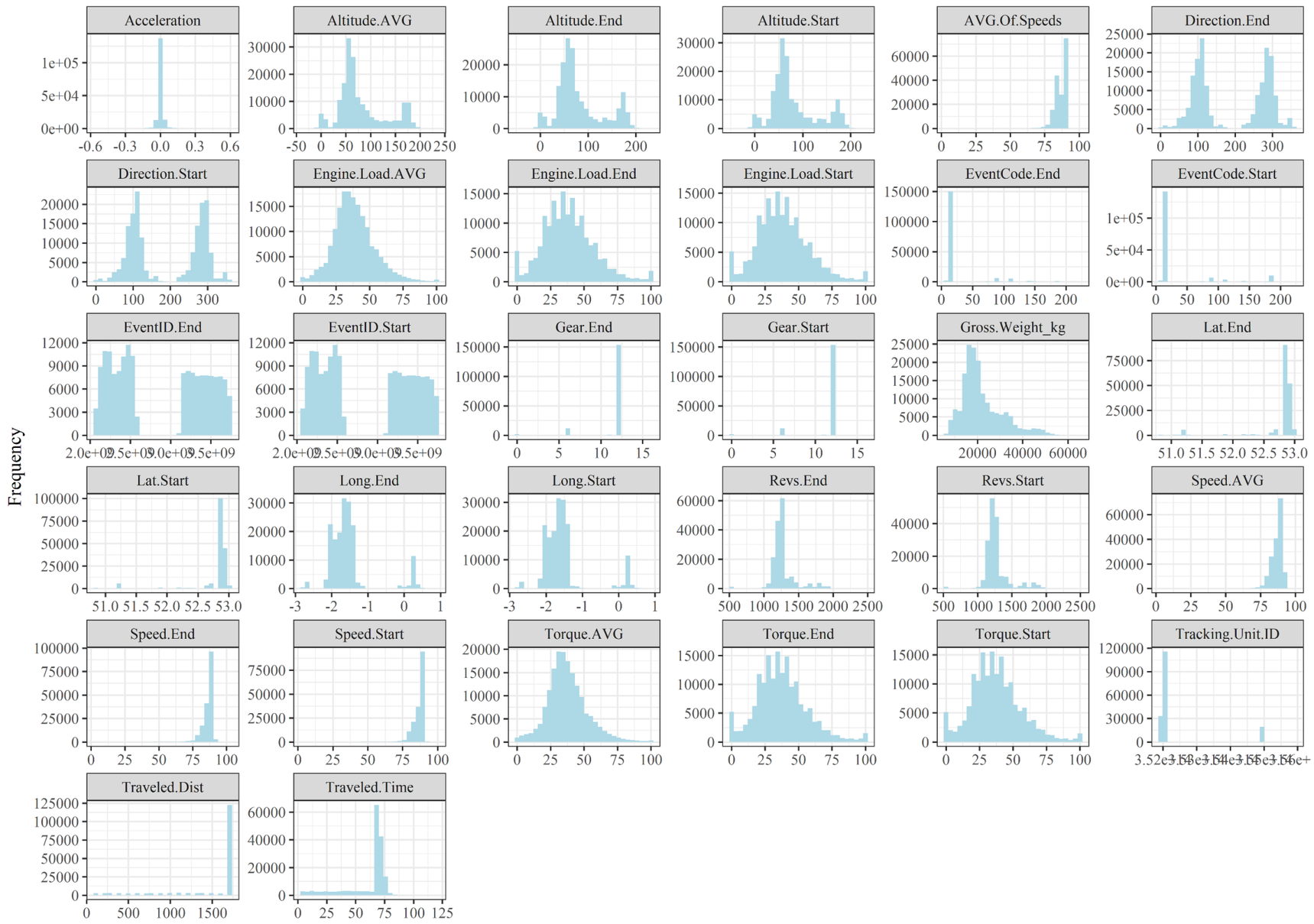


Figure C-7. Data distributions for the medium trucks driving along A roads (App 1), Dataset 2, Part 1.

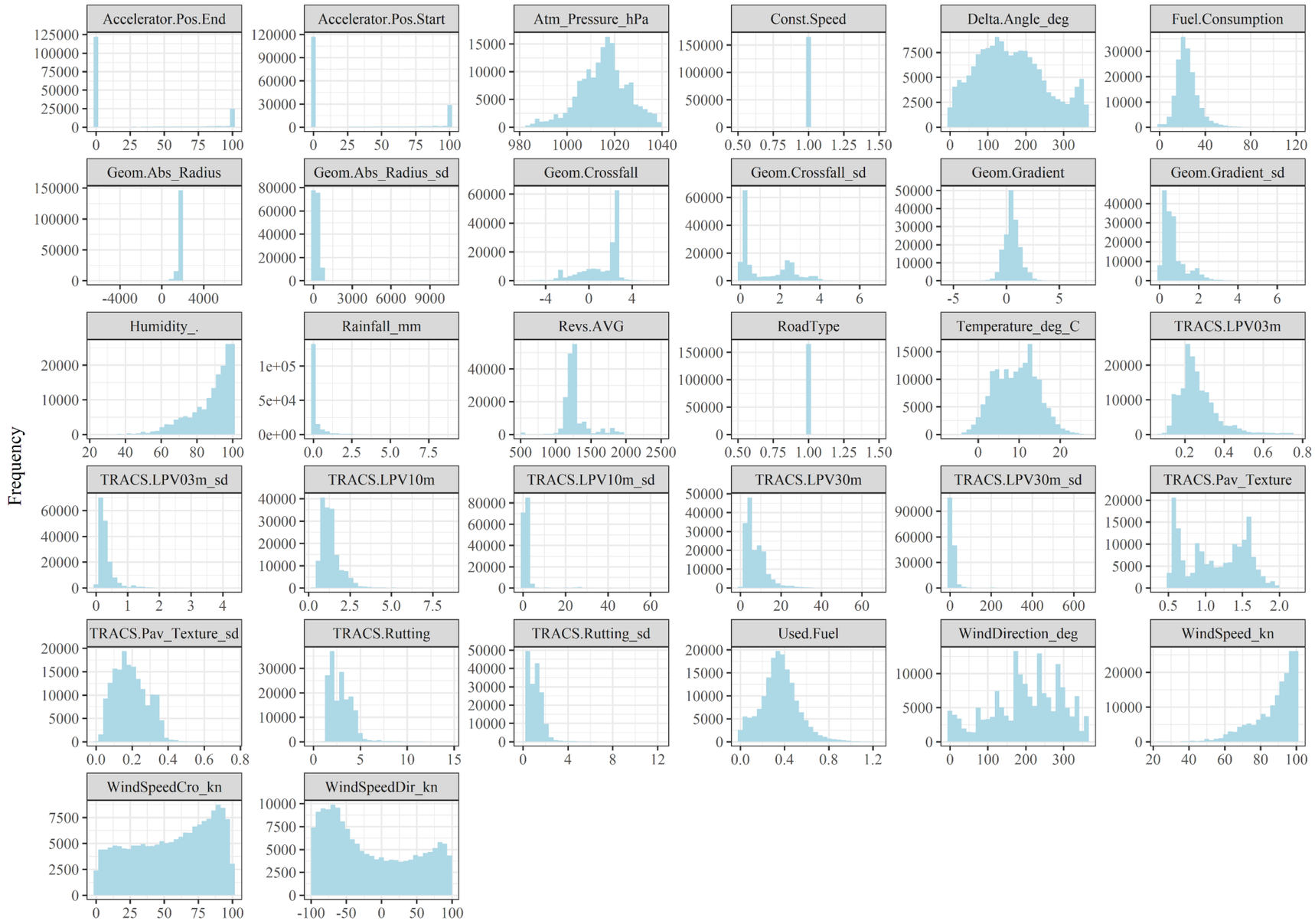


Figure C-8. Data distributions for the medium trucks driving along A roads (App 1), Dataset 2, Part 2.

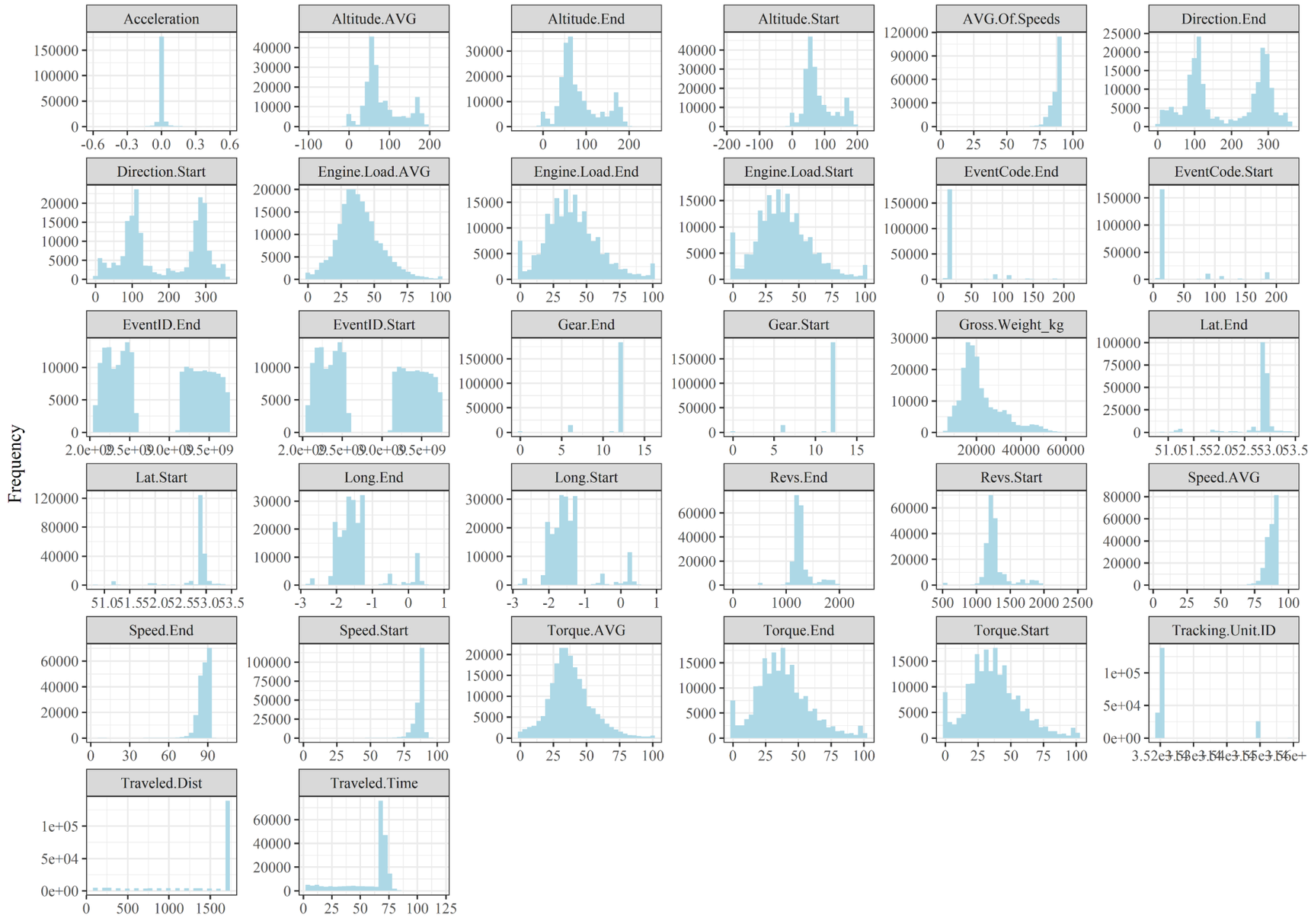


Figure C-9. Data distributions for the medium trucks driving along M and A roads (App 2), Dataset 2, Part 1.

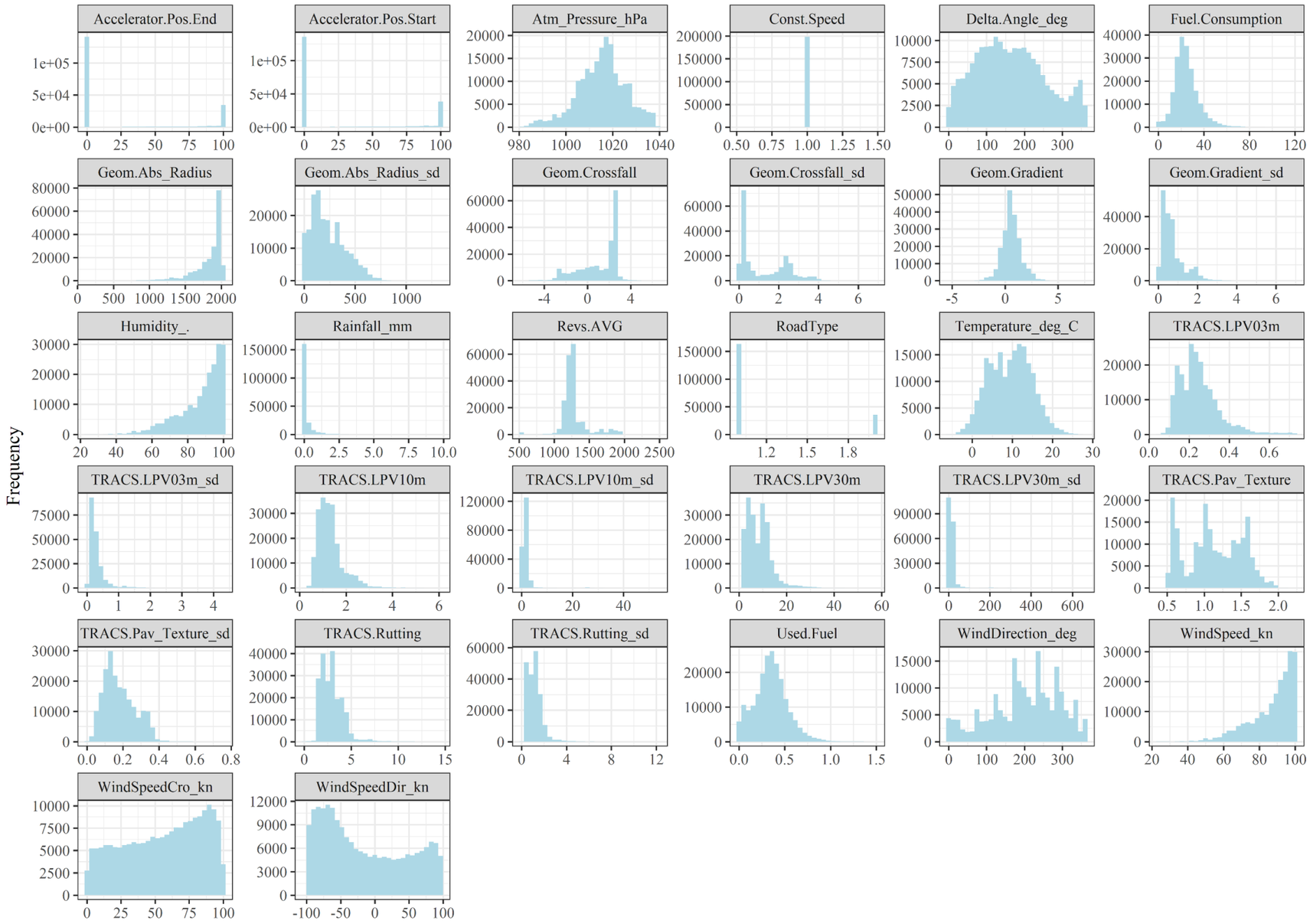


Figure C-10. Data distributions for the medium trucks driving along M and A roads (App 2), Dataset 2, Part 2.

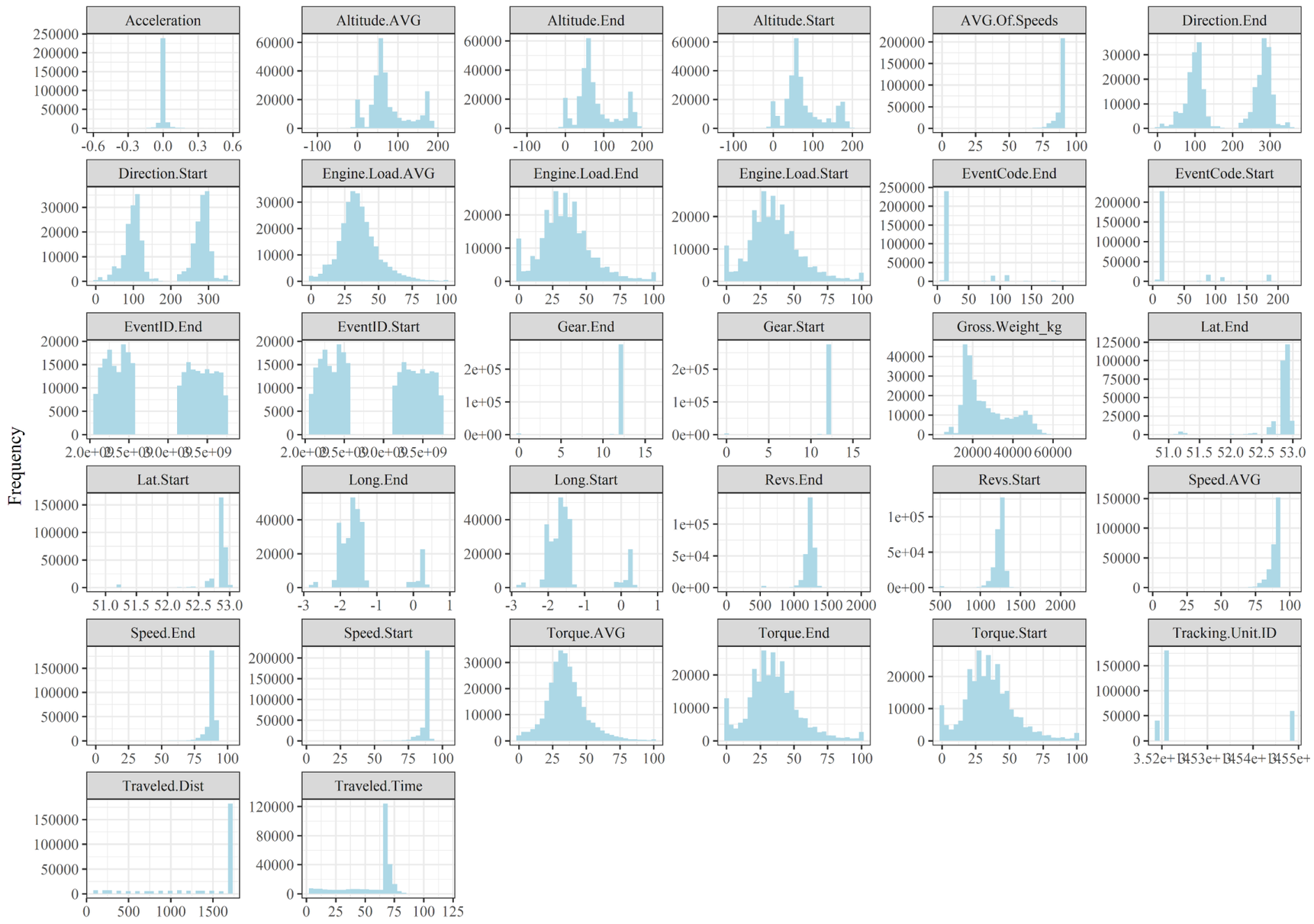


Figure C-11. Data distributions for the heavy trucks driving along A roads (App 1), Dataset 2, Part 1.

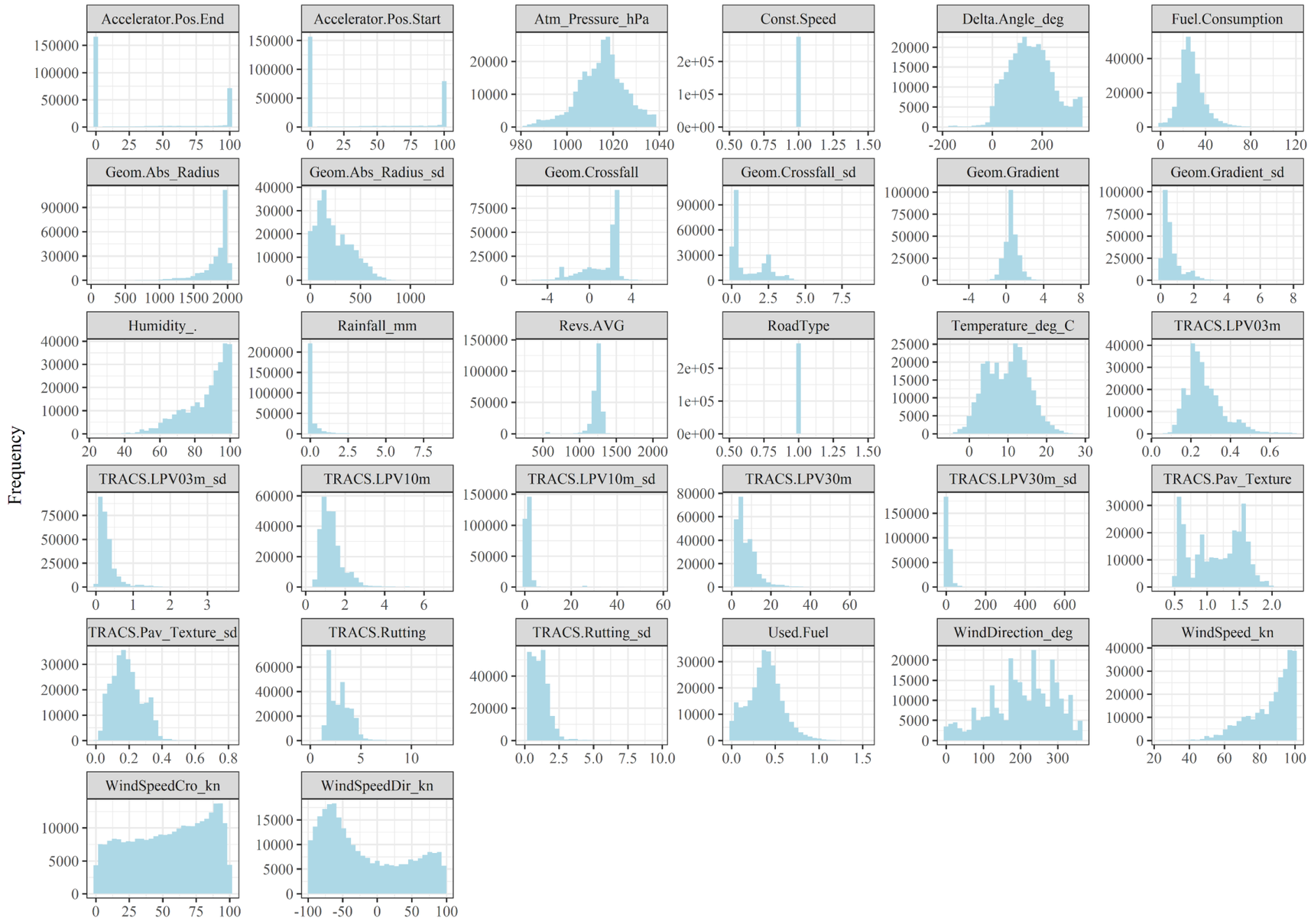


Figure C-12. Data distributions for the heavy trucks driving along A roads (App 1), Dataset 2, Part 2.

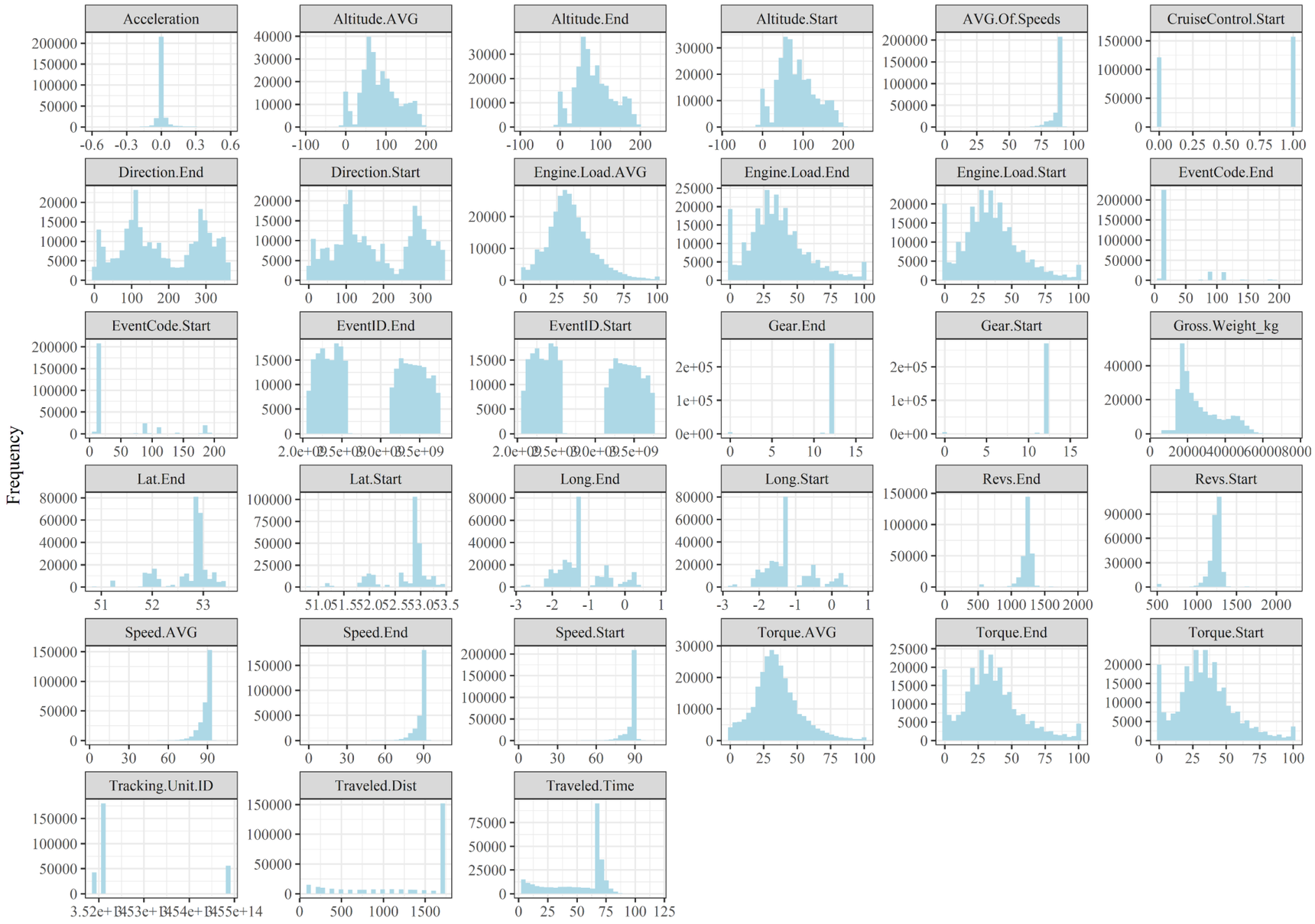


Figure C-13. Data distributions for the heavy trucks driving along M and A roads (App 2), Dataset 2, Part 1.

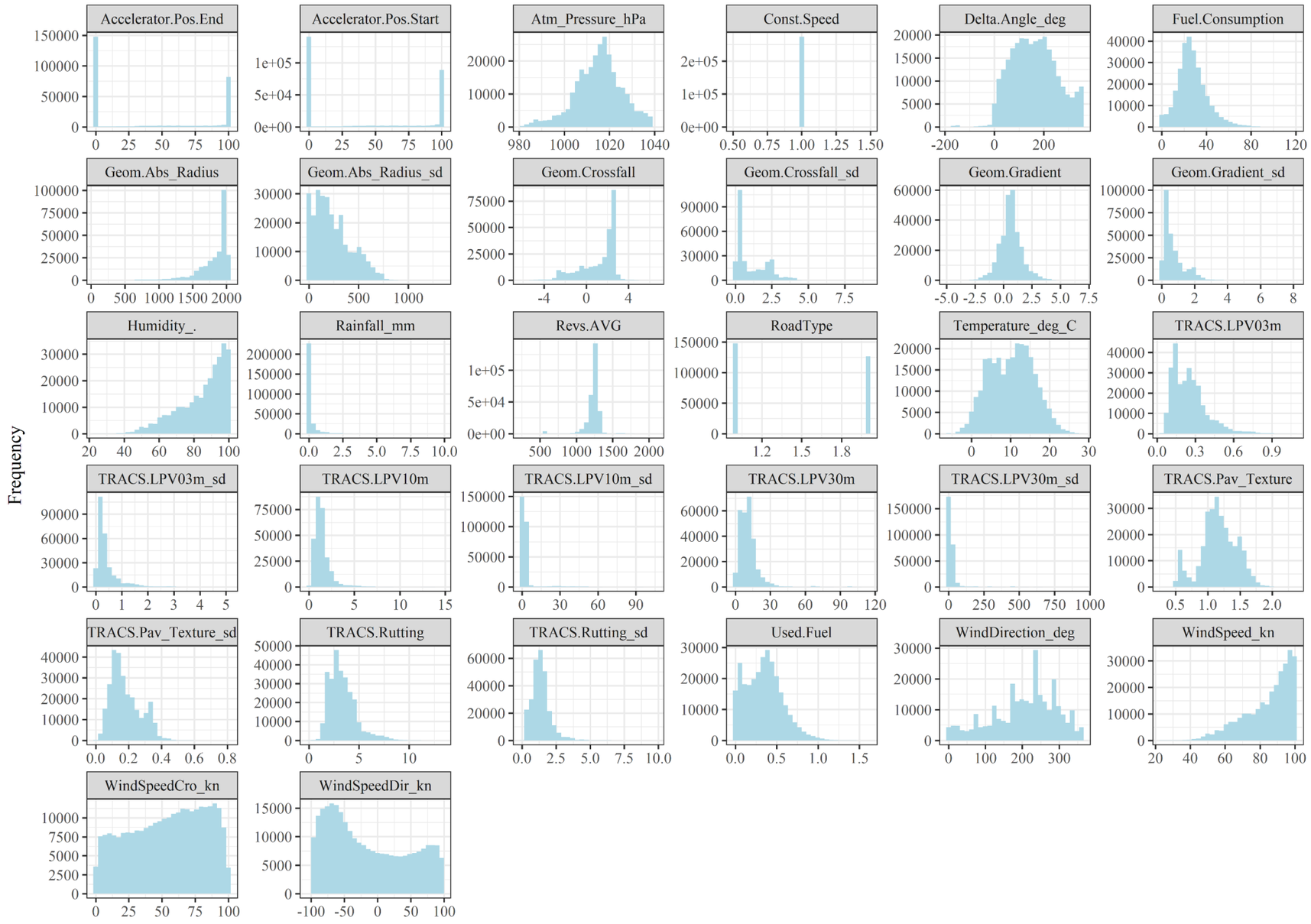


Figure C-14. Data distributions for the heavy trucks driving along M and A roads (App 2), Dataset 2, Part 2.

Appendix D R code

In this annex of the thesis details in regards to the code developed to carry out the data analysis are given. The annex is divided in five subsections in which each step of the used code is reported and commented. That should help future users in applying the methodology described in section 4 of the thesis ensuring reproducibility of the study.

Each bit of code presented in this annex has been used in the study directly however not all the code has been reported as small variations had to be made in different occasions based on a number of factors. Therefore, readers should consider this code not as a reference but as an example only.

Appendix D.1 Data procurement

As explained in the thesis (see sect. 3) data are provided in different formats by Microlise and TRL (on behalf of HE). Therefore, in order to make the data uniform and suitable to be analysed the following is done.

Appendix D.1.1 Truck telematic data

This part of the code extracts sensors measurements from trucks from MS SQL databases stored on the computer. That is to simulate extraction of the data from Microlise's databases.

```
# Extract Truck Performance Data - SQL version

## Connect to the MS SQL Local Server

#library(RODBC)

Con <- odbcDriverConnect("driver={SQL Server}; server= PEZP63099; database=

Trucks_2017; trusted_connection= TRUE")


## Query data from the MS SQL Local Server (specifying location of the road segments

to consider)

FC.Data1 <- sqlQuery(Con, 'SELECT [ResourceName]

                      ,[tbl_TrackerEvents4].[TrackerEventID]

                      ,[TrackerID]

                      ,[EventDateTime]

                      ,[Latitude]

                      ,[Longitude]

                      ,[LocationDescription]

                      ,[Speed]

                      ,[Bearing]

                      ,[Distance]

                      ,[EventCodeID]

                      ,[EventInfo1]

                      ,[EventInfo2]

                      ,[EventInfo3]

                      ,[EventInfo4]

                      ,[CBSpeed]

                      ,[CruiseControl]

                      ,[CBDistance]

                      ,[FuelUsed]

                      ,[FuelLevel]

                      ,[AvgFuelEconomy]

                      ,[CoolantTemp]

                      ,[FuelTemp]

                      ,[EngineOilTemp]

                      ,[AmbientAirTempCelcius]

                      ,[EngineRunState]

                      ,[Hacc]

                      ,[CalibratedRawFuelUsed]

                      ,[HighResFuelUsed]
```

```
                    ,[DistanceFromPreviousEventInMetres]
                    ,[CBDistanceFromPreviousEventInMetres]
                    ,[GrossCombinationVehicleWeightInKg]
                    ,[IntakeManifoldTemperatureInDegC]
                    ,[EngineFuelDeliveryPressureInKPa]
                    ,[AcceleratorPedalPositionPercentage]
                    ,[BrakePedalSwitch]
                    ,[CurrentGear]
                    ,[EngineSpeed]
                    ,[EngineLoadPercentage]
                    ,[EnginePercentageTorque]
                    FROM [Trucks_2017].[dbo].[tbl_TrackerEvents4] LEFT JOIN
[Trucks_2017].[dbo].[tbl_FleetCanData4]
                    ON ([DAF_2017].[dbo].[tbl_TrackerEvents4].[TrackerEventId] =
[Trucks_2017].[dbo].[tbl_FleetCanData4].[TrackerEventId])
                    WHERE (Latitude > 53.379984 and Latitude < 53.400494 and
Longitude > -1.278470 and Longitude < -1.265220) OR
                    (Latitude > 53.380717 and Latitude <= 53.400529 and Longitude >
-1.279019 and Longitude <= -1.265317) OR
                    (Latitude > 53.399915 and Latitude <= 53.423747 and Longitude >
-1.265895 and Longitude <= -1.251377) OR
                    (Latitude > 53.422242 and Latitude <= 53.441109 and Longitude >
-1.252272 and Longitude <= -1.238540) OR
                    (Latitude > 53.439669 and Latitude <= 53.459099 and Longitude >
-1.239298 and Longitude <= -1.201196) OR
                    (Latitude > 53.458309 and Latitude <= 53.475886 and Longitude >
-1.203238 and Longitude <= -1.156675) OR
                    (Latitude > 53.474864 and Latitude <= 53.479384 and Longitude >
-1.158820 and Longitude <= -1.148678) OR
                    (Latitude > 53.478914 and Latitude <= 53.488829 and Longitude >
-1.148769 and Longitude <= -1.115028) OR
                    (Latitude > 53.487309 and Latitude <= 53.488923 and Longitude >
-1.116587 and Longitude <= -1.085350) OR
                    (Latitude > 53.487309 and Latitude <= 53.494596 and Longitude >
-1.085350 and Longitude <= -1.063713) OR
                    (Latitude > 53.494300 and Latitude <= 53.516000 and Longitude >
-1.064082 and Longitude <= -1.045058) OR
                    (Latitude > 53.514773 and Latitude <= 53.537286 and Longitude >
-1.045908 and Longitude <= -1.024666) OR
```

```
                    (Latitude > 53.536131 and Latitude <= 53.559451 and Longitude >
-1.025880 and Longitude <= -1.002868) OR
                    (Latitude > 53.558118 and Latitude <= 53.586672 and Longitude >
-1.004093 and Longitude <= -0.984913) OR
                    (Latitude > 53.585959 and Latitude <= 53.603844 and Longitude >
-0.991470 and Longitude <= -0.984913) OR
                    (Latitude > 53.603844 and Latitude <= 53.619113 and Longitude >
-0.991470 and Longitude <= -0.977439) OR
                    (Latitude > 53.618235 and Latitude <= 53.688149 and Longitude >
-0.977872 and Longitude <= -0.959207);')

# repeat the data extraction for all the tables you need and then bind them
FC.Data <- rbind(FC.Data1, FC.Data2)
FC.Data <- rbind(FC.Data, FC.Data3)
FC.Data <- rbind(FC.Data, FC.Data4)
## You can clean memory from temporary variables, just uncomment the line below
#rm(FC.Data1, FC.Data2, FC.Data3, FC.Data4)

# erease spaces from resource names and make data consistents
FC.Data$ResourceName <- gsub(" ", "", FC.Data$ResourceName)

#extract vehicles characteristics data
#Vehicles.Data <- sqlQuery(Con, 'SELECT * FROM [MAN_2015].[dbo].[Vehicles];')
Path <- ("C:/Data/Trucks/160505_VehicleLookup_MAN.xlsx")
Vehicles.Data <- read_excel(Path, sheet = "Sheet1", col_names = TRUE, col_types =
NULL, na = "NULL", skip = 0)
rm(Path)

##Subset data so that the dataframe contains just data we are interested in and sort
it per vehicle and time
Vehicles.Data <- subset(Vehicles.Data, (Vehicles.Data$smmt_euro_status == 3 |
Vehicles.Data$smmt_euro_status == 4 | Vehicles.Data$smmt_euro_status == 5 |
Vehicles.Data$smmt_euro_status == 6))

##You can obtain FC_data using FC Vehicles Data CSV Import.R
Initial.FC.Data <- FC.Data
#order data by plate number and event time
FC.Data <- FC.Data[order(FC.Data$ResourceName, FC.Data$EventDateTime),]
#consider records measured just in 2017
```

```
FC.Data <- subset(FC.Data, grepl("2017", FC.Data$EventDateTime))
#rm(FC.Data)
```

Other variants of the code reported above can include direct extraction of clean data from csv or xlsx files.

Appendix D.1.2 Road conditions data

This part of the code extracts road conditions from the shp files from the HAPMS provided by TRL on behalf of HE.

```
#example of the A50 road
Geometry.Shape <- readOGR("D:/Data/Roads/A50", "A50 Geometry") #this retrieves the
Geometry data from the HDD (or SSD)
proj4string(Geometry.Shape) <- CRS("+proj=tmerc +lat_0=49 +lon_0=-2 +k=0.9996012717
+x_0=400000 +y_0=-100000 +datum=OSGB36 +units=m +no_defs +ellps=airy
+towgs84=446.448,-125.157,542.060,0.1502,0.2470,0.8421,-20.4894")
Geometry.Shape <- spTransform(Geometry.Shape, CRS("+proj=longlat +datum=WGS84"))
#this changes the reference system of the data to WGS84

TRACS.Shape <- readOGR("D:/Data/Roads/A50", "A50 TRACS") #this retrieves the TRACS
data from the HDD (or SSD)
proj4string(TRACS.Shape) <- CRS("+proj=tmerc +lat_0=49 +lon_0=-2 +k=0.9996012717
+x_0=400000 +y_0=-100000 +datum=OSGB36 +units=m +no_defs +ellps=airy
+towgs84=446.448,-125.157,542.060,0.1502,0.2470,0.8421,-20.4894")
TRACS.Shape <- spTransform(TRACS.Shape, CRS("+proj=longlat +datum=WGS84"))
#this changes the reference system of the data to WGS84
#initialization of variables needed (repeat if you consider a different road)
Geometry.GPS_lat <- c()
Geometry.GPS_long <- c()
SCRIM.GPS_lat <- c()
SCRIM.GPS_long <- c()
TRACS.GPS_lat <- c()
TRACS.GPS_long <- c()
```

```
# then, for each record in Geometry.Shape extract latitude and longitude associated
with it
for (r in 1:nrow(Geometry.Shape@data)) {
  Geometry.GPS_lat[r] <- Geometry.Shape@lines[[r]]@Lines[[1]]@coords[1,][2]
  Geometry.GPS_long[r] <- Geometry.Shape@lines[[r]]@Lines[[1]]@coords[1,][1]
}

#transform data in a data frame and assign to each record latitude and longitude
extracted
Geometry.Shape_DF <- as.data.frame(Geometry.Shape)
Geometry.Shape_DF$Latitude <- Geometry.GPS_lat
Geometry.Shape_DF$Longitude <- Geometry.GPS_long

rm(Geometry.Shape) #free some space in memory
r <- 0

#repeat a similar procedure for TRACS
for (r in 1:nrow(TRACS.Shape@data)) {
  TRACS.GPS_lat[r] <- TRACS.Shape@lines[[r]]@Lines[[1]]@coords[1,][2]
  TRACS.GPS_long[r] <- TRACS.Shape@lines[[r]]@Lines[[1]]@coords[1,][1]
}

TRACS.Shape_DF <- as.data.frame(TRACS.Shape)
TRACS.Shape_DF$Latitude <- TRACS.GPS_lat
TRACS.Shape_DF$Longitude <- TRACS.GPS_long

rm(TRACS.Shape)
r <- 0

Geometry.Shape_DF <- subset(Geometry.Shape_DF, (Geometry.Shape_DF$sect_label) %in%
(TRACS.Shape_DF$sect_label)) # subset data that are in Geometry so that these are
included in TRACS as well

j <- 0
i <- 0
dir <- c()

#assign a direction of traffic to TRACS data (road conditions)
for (i in 1:length(unique(TRACS.Shape_DF$sect_label))) {
```

```
  print(paste0("processing section number ",
as.character(unique(TRACS.Shape_DF$sect_label)[i]),"..."))
  Test <- subset(TRACS.Shape_DF, TRACS.Shape_DF$sect_label ==
as.character(unique(TRACS.Shape_DF$sect_label)[i]))
  if (as.character(Test$direc_code[i]) == "WB") {
    dir[i] <- "WB"
  } else if (as.character(Test$direc_code[i]) == "EB") {
    dir[i] <- "EB"
  } else {
    dir[i] <- "CW"
  }
}

j <- 0
i <- 0
check <- 0
direc <- c()

#assign a direction of traffic to Geometry data (based on what has been assigned to
TRACS)
for (j in 1:nrow(Geometry.Shape_DF)) {
  print(paste0("processing section number ", j, " ..."))
  for (i in 1:length(dir)) {
    if (toString(Geometry.Shape_DF$sect_label[j]) ==
as.character(unique(TRACS.Shape_DF$sect_label)[i])) {
      direc[j] <- dir[i]
    }
  }
}

Geometry.Shape_DF$direction <- direc
j <- 0
i <- 0
rm(direc, dir) #free up some space in memory

#add to each dataframe a reference road id
Geometry_A50$road <- 'A50'
TRACS_A50$road <- 'A50'
```

```
#save the generated data frames to csv
write.csv(Geometry_A50, 'Geometry_A50.csv', row.names = FALSE)
write.csv(TRACS_A50, 'TRACS_A50.csv', row.names = FALSE)
```

Appendix D.1.3 Creation of Data.T

This part of the code is to create Data.T. This data frme contains records from trucks which have been driven along selected segments of motorways (M) and 'A' type roads in UK.

```
#apply Road.ID to FC.Data to identify from which road the data come from
for (i in 1:(nrow(FC.Data))) {
  print(i)
  if (FC.Data$Latitude[i] >= 51.765 & FC.Data$Latitude[i] <= 52.150 &
FC.Data$Longitude[i] >= -0.843 & FC.Data$Longitude[i] <= -0.415) {
    FC.Data$Road.ID[i] <- "M1b"
  } else if (FC.Data$Latitude[i] >= 53.387 & FC.Data$Latitude[i] <= 53.675 &
FC.Data$Longitude[i] >= -1.278 & FC.Data$Longitude[i] <= -0.965) {
    FC.Data$Road.ID[i] <- "M18"
  } else if (FC.Data$Latitude[i] >= 50.905 & FC.Data$Latitude[i] <= 51.007 &
FC.Data$Longitude[i] >= 0.619 & FC.Data$Longitude[i] <= 0.857) {
    FC.Data$Road.ID[i] <- "A259"
  } else if (FC.Data$Latitude[i] >= 50.820 & FC.Data$Latitude[i] <= 50.850 &
FC.Data$Longitude[i] >= 0.280 & FC.Data$Longitude[i] <= 0.520) {
    FC.Data$Road.ID[i] <- "A259"
  } else if (FC.Data$Latitude[i] >= 52.857 & FC.Data$Latitude[i] <= 52.992 &
FC.Data$Longitude[i] >= -2.172 & FC.Data$Longitude[i] <= -1.328) {
    FC.Data$Road.ID[i] <- "A50"
  } else if (FC.Data$Latitude[i] >= 50.896 & FC.Data$Latitude[i] <= 51.270 &
FC.Data$Longitude[i] >= 0.164 & FC.Data$Longitude[i] <= 0.531) {
    FC.Data$Road.ID[i] <- "A21"
  } else if (FC.Data$Latitude[i] >= 52.614 & FC.Data$Latitude[i] <= 52.737 &
FC.Data$Longitude[i] >= -0.164 & FC.Data$Longitude[i] <= 0.370) {
    FC.Data$Road.ID[i] <- "A47"
  } else if (FC.Data$Latitude[i] >= 51.931 & FC.Data$Latitude[i] <= 52.460 &
FC.Data$Longitude[i] >= -2.840 & FC.Data$Longitude[i] <= -2.677) {
    FC.Data$Road.ID[i] <- "A49"
```

```
  } else if ((FC.Data$Latitude[i] >= 52.696 & FC.Data$Latitude[i] <= 53.387) &
((FC.Data$Longitude[i] >= -1.3386 & FC.Data$Longitude[i] <= -1.229) |
(FC.Data$Longitude[i] >= -1.311 & FC.Data$Longitude[i] <= -1.272))) {
    FC.Data$Road.ID[i] <- "M1a"
  } else if (FC.Data$Latitude[i] >= 50.699 & FC.Data$Latitude[i] <= 50.803 &
FC.Data$Longitude[i] >= -3.175 & FC.Data$Longitude[i] <= -2.447) {
    FC.Data$Road.ID[i] <- "A35"
  } else {
    FC.Data$Road.ID[i] <- "NA"
  }
}

#sort data by TrackingUnitID and eventDateTime
FC.Data <- FC.Data[order(FC.Data$Allocated.Tracking.Unit.ID, FC.Data$Vehicle.ID,
FC.Data$Event.Date.Time),] #order data by plate number and event time

#uniform Tracking Unit ID
#for (i in 1:nrow(FC.Data)) {print(i); FC.Data$Allocated.Tracking.Unit.ID[i] <-
as.character(paste0("0", as.character(FC.Data$Allocated.Tracking.Unit.ID[i]))) }

t1 <- Sys.time() #check time spent for calculation
i <- 0
j <- 0
EuroTyp <- c()
TestDT <- c()

#use function strhead to get a function able to extract the first n characters out
from a string
strhead <- function(s,n) {
  if(n < 0)
    substr(s, 1, nchar(s) + n)
  else
    substr(s, 1, n)
}

#build the datatable for the data analysis
for (i in 1:(nrow(FC.Data)-1)) {
  if ((FC.Data$Allocated.Tracking.Unit.ID[i] ==
FC.Data$Allocated.Tracking.Unit.ID[i+1]) &
```

```
      (FC.Data$Vehicle.ID[i] == FC.Data$Vehicle.ID[i+1])) {
    print(paste0("now processing row ", i, "..."))
    j <- j+1
    TestDT$Vehicle.RN[j] <- as.character(FC.Data$Vehicle.ID[i])
    TestDT$Tracking.Unit.ID[j] <- as.character(FC.Data$Allocated.Tracking.Unit.ID[i])
    TestDT$Make[j] <- as.character(FC.Data$Make[i])
    TestDT$Type[j] <- as.character(substr(FC.Data$Model[i], 5, 9))
    TestDT$Wheel.Base[j] <- as.character(strhead(FC.Data$Model[i], 3))
    TestDT$Model[j] <- as.character(FC.Data$Range[i])
    TestDT$Series[j] <- as.character(FC.Data$Series[i])
    TestDT$EventID.Start[j] <- as.character(FC.Data$ID[i])
    TestDT$EventCode.Start[j] <- as.numeric(FC.Data$Event.Code[i])
    TestDT$EventID.End[j] <- as.character(FC.Data$ID[i+1])
    TestDT$EventCode.End[j] <- as.numeric(FC.Data$Event.Code[i+1])
    TestDT$Date.Start[j] <-
as.character(as.POSIXct(as.character(FC.Data$Event.Date.Time[i]), format = "%d/%m/%Y
%H:%M:%S", tz = "GMT"))
    TestDT$Date.End[j] <-
as.character(as.POSIXct(as.character(FC.Data$Event.Date.Time[i+1]), format =
"%d/%m/%Y %H:%M:%S", tz = "GMT"))
    TestDT$Lat.Start[j] <- FC.Data$Latitude[i]
    TestDT$Long.Start[j] <- FC.Data$Longitude[i]
    TestDT$Lat.End[j] <- FC.Data$Latitude[i+1]
    TestDT$Long.End[j] <- FC.Data$Longitude[i+1]
    TestDT$Road.ID[j] <- FC.Data$Road.ID[i]
    TestDT$Direction.Start[j] <-
as.numeric(as.character(FC.Data$Course.In.Degrees[i]))
    TestDT$Direction.End[j] <-
as.numeric(as.character(FC.Data$Course.In.Degrees[i+1]))
    TestDT$Altitude.Start[j] <- as.numeric(as.character(FC.Data$Altitude[i]))
    TestDT$Altitude.End[j] <- as.numeric(as.character(FC.Data$Altitude[i+1]))
    TestDT$Altitude.AVG[j] <- (TestDT$Altitude.Start[j] + TestDT$Altitude.End[j])/2
    TestDT$Traveled.Dist[j] <-
as.numeric(as.character(FC.Data$Can.Distance.In.Metres[i+1])) -
as.numeric(as.character(FC.Data$Can.Distance.In.Metres[i]))
    TestDT$Traveled.Time[j] <-
difftime(as.POSIXct(as.character(FC.Data$Event.Date.Time[i+1]), format = "%d/%m/%Y
%H:%M:%S", tz = "GMT"), as.POSIXct(as.character(FC.Data$Event.Date.Time[i]), format =
"%d/%m/%Y %H:%M:%S", tz = "GMT"), units = "secs")
```

```
    TestDT$Gross.Weight_kg[j] <-
as.numeric(as.character(FC.Data$Gross.Combination.Vehicle.Weight.In.Kg[i]))
    TestDT$Speed.Start[j] <-
round(3.6*as.numeric(as.character(FC.Data$Speed.In.Metres.Per.Second[i])), 2)
    TestDT$Speed.End[j] <-
round(3.6*as.numeric(as.character(FC.Data$Speed.In.Metres.Per.Second[i+1])), 2)
    TestDT$AVG.Of.Speeds[j] <- round((TestDT$Speed.Start[j] + TestDT$Speed.End[j])/2,
2)
    TestDT$Speed.AVG[j] <- round(3.6*TestDT$Traveled.Dist[j]/TestDT$Traveled.Time[j],
2)
    TestDT$Acceleration[j] <- round((TestDT$Speed.Start[j] -
TestDT$Speed.End[j])/TestDT$Traveled.Time[j], 2)

    ##separate directions (North bound, NB, South bound, SB, WB or EB) based on the
considered trunk of road

    if (TestDT$Road.ID[j] == "M1a" | TestDT$Road.ID[j] == "M1b" |
        TestDT$Road.ID[j] == "A21" | TestDT$Road.ID[j] == "A49") {

      if (TestDT$Lat.End[j] > TestDT$Lat.Start[j]) {
        TestDT$Travel.Direction[j] <- "NB"
      } else {
        TestDT$Travel.Direction[j] <- "SB"
      }

    } else if (TestDT$Road.ID[j] == "M18") {

      if ((TestDT$Long.Start[j] < -1.111 | TestDT$Long.Start[j] > -1.086) &
(TestDT$Long.End[j] < -1.111 | TestDT$Long.End[j] > -1.086)) {
        if (TestDT$Lat.End[j] > TestDT$Lat.Start[j]) {
          TestDT$Travel.Direction[j] <- "NB"
        } else {
          TestDT$Travel.Direction[j] <- "SB"
        }
      } else {
        if (TestDT$Lat.End[j] > TestDT$Lat.Start[j]) {
          TestDT$Travel.Direction[j] <- "SB"
        } else {
          TestDT$Travel.Direction[j] <- "NB"
```

```
        }
      }

    } else { #in this case Road should be the A50, A27, A259, A49 or A47

      if (TestDT$Long.End[j] > TestDT$Long.Start[j]) {
        TestDT$Travel.Direction[j] <- "EB"
      } else {
        TestDT$Travel.Direction[j] <- "WB"
      }

    }

    TestDT$Gear.Start[j] <- as.numeric(as.character(FC.Data$Current.Gear[i]))
    TestDT$Gear.End[j] <- as.numeric(as.character(FC.Data$Current.Gear[i+1]))
    TestDT$CruiseControl.Start[j] <- as.character(FC.Data$Cruise.Control[i])
    TestDT$CruiseControl.End[j] <- as.character(FC.Data$Cruise.Control[i+1])
    TestDT$Torque.Start[j] <-
as.numeric(as.character(FC.Data$Engine.Percentage.Torque[i]))
    TestDT$Torque.End[j] <-
as.numeric(as.character(FC.Data$Engine.Percentage.Torque[i+1]))
    TestDT$Torque.AVG[j] <- (TestDT$Torque.Start[j] + TestDT$Torque.End[j])/2
    TestDT$Engine.Load.Start[j] <-
as.numeric(as.character(FC.Data$Engine.Load.Percentage[i]))
    TestDT$Engine.Load.End[j] <-
as.numeric(as.character(FC.Data$Engine.Load.Percentage[i+1]))
    TestDT$Engine.Load.AVG[j] <- (TestDT$Engine.Load.Start[j] +
TestDT$Engine.Load.End[j])/2
    TestDT$Revs.Start[j] <- as.numeric(as.character(FC.Data$Engine.Speed.In.RPM[i]))
    TestDT$Revs.End[j] <- as.numeric(as.character(FC.Data$Engine.Speed.In.RPM[i+1]))
    TestDT$Revs.AVG[j] <- (TestDT$Revs.Start[j] + TestDT$Revs.End[j])/2
    TestDT$Accelerator.Pos.Start[j] <-
as.numeric(FC.Data$ï..Accelerator.Pedal.Position.Percentage[i])
    TestDT$Accelerator.Pos.End[j] <-
as.numeric(FC.Data$ï..Accelerator.Pedal.Position.Percentage[i+1])
    TestDT$Used.Fuel[j] <-
as.numeric(as.character(FC.Data$CalculatedFuelUsedInPoint00001OfALitre[i+1]))/100000
- as.numeric(as.character(FC.Data$CalculatedFuelUsedInPoint00001OfALitre[i]))/100000
    TestDT$Fuel.Consumption[j] <- 100000*TestDT$Used.Fuel[j]/TestDT$Traveled.Dist[j]
```

```
    #use a different algorithms to attach road and weather conditions (later)
  } else {
    next
  }
}

TestDT <- as.data.frame(TestDT)
t2 <- Sys.time()
difftime(t2, t1, units = "mins")

##save these data
write.csv(TestDT, "D:/Data/Trucks/R results/20180912 TestDT Heavy Jan-Mar 2017
Part4.csv")

#visualize data distributions
library(purrr)
library(tidyr)
library(ggplot2)

TestDT %>%
  keep(is.numeric) %>% # Keep only numeric values
  gather() %>% # Convert to key value pairs
  ggplot(aes(value)) + #Plot
  facet_wrap(~ key, scales = "free") + #In separate panels
  geom_histogram(fill = "lightblue") # As histograms

#consider only reliable data (avoid twins and couple of events that happen with an
event in between)
#Data.T <- c() #initialize Data.T
Data.T <- subset(TestDT, Speed.AVG > 0 & Speed.AVG < 150 & Traveled.Time > 0 &
Traveled.Time <= 120 & Traveled.Dist > 0 & Traveled.Dist <= 1700 & Used.Fuel > 0 &
Fuel.Consumption <= 120)

#rm(TestDT)

#consider only journeys that can be considered performed at constant speed
for (i in 1:nrow(Data.T)) {
  print(paste0("now processing row ", i, "..."))
```

```
  if (abs(Data.T$Speed.Start[i] - Data.T$Speed.End[i]) < 2.5) {
    if (abs(mean(Data.T$Speed.Start[i], Data.T$Speed.End[i]) - Data.T$Speed.AVG[i]) <
2.5) {
      Data.T$Const.Speed[i] <- 1
    } else {
      Data.T$Const.Speed[i] <- 0
    }
  } else {
    Data.T$Const.Speed[i] <- 0
  }
}

##save these data (non-controlled speed)
#write.csv(Data.T, "D:/Data/Trucks/20180701 DataT Medium Jan-Mar 2017.csv")
#write.csv(FC.Data, "D:/Data/Trucks/20180701 FCData Medium Jan-Mar 2017.csv")
#rm(FC.Data)

##you can now consider and analyze only journey performed at constant speed
Data.T <- subset(Data.T, Const.Speed == 1 & Gear.Start == Gear.End &
CruiseControl.Start == CruiseControl.End)
#Data.T.2.5 <- Data.T
##save final data (constant speed)
write.csv(Data.T, "20180914 DataT Heavy 25kmh Jan-Mar 2017 const speed Part4.csv")

t2 <- Sys.time()
difftime(t2, t1, units = "mins")

#visualize data distributions
Data.T %>%
  keep(is.numeric) %>% # Keep only numeric values
  gather() %>% # Convert to key value pairs
  ggplot(aes(value)) + #Plot
  facet_wrap(~ key, scales = "free") + #In separate panels
  geom_histogram(fill = "lightblue") # As histograms
```

Appendix D.1.4 Append road conditions to Data.T

This part of the code is to assign to each record in Data.T road conditions.

```
# Append RC data to Data.T
##assign to each fuel consumption data in Data.T the corresponding average road
condition that causes it
# filters are applied based on the geographical position of data

## Initialization of data vectors
Geometry.GPS_lat <- c()
Geometry.GPS_long <- c()
Geometry.Radius <- c()
Geometry.Radius_sd <- c()
Geometry.Gradient <- c()
Geometry.Gradient_sd <- c()
Geometry.Crossfall <- c()
Geometry.Crossfall_sd <- c()

SCRIM.GPS_lat <- c()
SCRIM.GPS_long <- c()
SCRIM.SC <- c()
SCRIM.SC_sd <- c()

TRACS.GPS_lat <- c()
TRACS.GPS_long <- c()
TRACS.Rut <- c()
TRACS.Rut_sd <- c()
TRACS.LPV3 <- c()
TRACS.LPV3_sd <- c()
TRACS.LPV10 <- c()
TRACS.LPV10_sd <- c()
TRACS.LPV30 <- c()
TRACS.LPV30_sd <- c()
TRACS.Texture <- c()
TRACS.Texture_sd <- c()

t1 <- Sys.time()
r <- 0
l <- 0

for (r in 1:nrow(Data.T)) {
  print(paste0("now processing row ", r, "..."))
```

```
  Lat1 <- Data.T$Lat.Start[r]
  Lon1 <- Data.T$Long.Start[r]
  Lat2 <- Data.T$Lat.End[r]
  Lon2 <- Data.T$Long.End[r]
  Dir <- Data.T$Travel.Direction[r]

  if (Lat1 > Lat2) {

    if (Lon1 > Lon2) {
      l <- l + 1
      Geometry.Radius[l] <- mean(subset(Geometry.Shape_DF$abs_radius,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude > Lon2) &
(Geometry.Shape_DF$Longitude < Lon1) & (Geometry.Shape_DF$Latitude > Lat2) &
(Geometry.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      Geometry.Radius_sd[l] <- sd(subset(Geometry.Shape_DF$abs_radius,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude > Lon2) &
(Geometry.Shape_DF$Longitude < Lon1) & (Geometry.Shape_DF$Latitude > Lat2) &
(Geometry.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      Geometry.Gradient[l] <- mean(subset(Geometry.Shape_DF$gradient,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude > Lon2) &
(Geometry.Shape_DF$Longitude < Lon1) & (Geometry.Shape_DF$Latitude > Lat2) &
(Geometry.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      Geometry.Gradient_sd[l] <- sd(subset(Geometry.Shape_DF$gradient,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude > Lon2) &
(Geometry.Shape_DF$Longitude < Lon1) & (Geometry.Shape_DF$Latitude > Lat2) &
(Geometry.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      Geometry.Crossfall[l] <- mean(subset(Geometry.Shape_DF$crossfall,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude > Lon2) &
(Geometry.Shape_DF$Longitude < Lon1) & (Geometry.Shape_DF$Latitude > Lat2) &
(Geometry.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      Geometry.Crossfall_sd[l] <- sd(subset(Geometry.Shape_DF$crossfall,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude > Lon2) &
(Geometry.Shape_DF$Longitude < Lon1) & (Geometry.Shape_DF$Latitude > Lat2) &
(Geometry.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      SCRIM.SC[l] <- mean(subset(SCRIM.Shape_DF$coefficien, (SCRIM.Shape_DF$direction
== Dir) & (SCRIM.Shape_DF$Longitude > Lon2) & (SCRIM.Shape_DF$Longitude < Lon1) &
(SCRIM.Shape_DF$Latitude > Lat2) & (SCRIM.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      SCRIM.SC_sd[l] <- sd(subset(SCRIM.Shape_DF$coefficien,
(SCRIM.Shape_DF$direction == Dir) & (SCRIM.Shape_DF$Longitude > Lon2) &
```

```
(SCRIM.Shape_DF$Longitude < Lon1) & (SCRIM.Shape_DF$Latitude > Lat2) &
(SCRIM.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      TRACS.Rut[l] <- mean(cbind(subset(TRACS.Shape_DF$left_rut,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$right_rut,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.Rut_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$left_rut,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$right_rut,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.LPV3[l] <- mean(cbind(subset(TRACS.Shape_DF$lef_lpv_3m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$rig_lpv_3m,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.LPV3_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$lef_lpv_3m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$rig_lpv_3m,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.LPV10[l] <- mean(cbind(subset(TRACS.Shape_DF$le_lpv_10m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$ri_lpv_10m,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.LPV10_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$le_lpv_10m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$ri_lpv_10m,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
```

```
      TRACS.LPV30[l] <- mean(cbind(subset(TRACS.Shape_DF$le_lpv_30m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$ri_lpv_30m,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.LPV30_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$le_lpv_30m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$ri_lpv_30m,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.Texture[l] <- mean(subset(TRACS.Shape_DF$texture,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      TRACS.Texture_sd[l] <- sd(subset(TRACS.Shape_DF$texture,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), na.rm=TRUE)


    } else {
      l <- l + 1
      Geometry.Radius[l] <- mean(subset(Geometry.Shape_DF$abs_radius,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude < Lon2) &
(Geometry.Shape_DF$Longitude > Lon1) & (Geometry.Shape_DF$Latitude > Lat2) &
(Geometry.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      Geometry.Radius_sd[l] <- sd(subset(Geometry.Shape_DF$abs_radius,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude < Lon2) &
(Geometry.Shape_DF$Longitude > Lon1) & (Geometry.Shape_DF$Latitude > Lat2) &
(Geometry.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      Geometry.Gradient[l] <- mean(subset(Geometry.Shape_DF$gradient,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude < Lon2) &
(Geometry.Shape_DF$Longitude > Lon1) & (Geometry.Shape_DF$Latitude > Lat2) &
(Geometry.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      Geometry.Gradient_sd[l] <- sd(subset(Geometry.Shape_DF$gradient,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude < Lon2) &
(Geometry.Shape_DF$Longitude > Lon1) & (Geometry.Shape_DF$Latitude > Lat2) &
(Geometry.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
```

```
      Geometry.Crossfall[l] <- mean(subset(Geometry.Shape_DF$crossfall,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude < Lon2) &
(Geometry.Shape_DF$Longitude > Lon1) & (Geometry.Shape_DF$Latitude > Lat2) &
(Geometry.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      Geometry.Crossfall_sd[l] <- sd(subset(Geometry.Shape_DF$crossfall,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude < Lon2) &
(Geometry.Shape_DF$Longitude > Lon1) & (Geometry.Shape_DF$Latitude > Lat2) &
(Geometry.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      SCRIM.SC[l] <- mean(subset(SCRIM.Shape_DF$coefficien, (SCRIM.Shape_DF$direction
== Dir) & (SCRIM.Shape_DF$Longitude < Lon2) & (SCRIM.Shape_DF$Longitude > Lon1) &
(SCRIM.Shape_DF$Latitude > Lat2) & (SCRIM.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      SCRIM.SC_sd[l] <- sd(subset(SCRIM.Shape_DF$coefficien,
(SCRIM.Shape_DF$direction == Dir) & (SCRIM.Shape_DF$Longitude < Lon2) &
(SCRIM.Shape_DF$Longitude > Lon1) & (SCRIM.Shape_DF$Latitude > Lat2) &
(SCRIM.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      TRACS.Rut[l] <- mean(cbind(subset(TRACS.Shape_DF$left_rut,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$right_rut,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.Rut_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$left_rut,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$right_rut,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.LPV3[l] <- mean(cbind(subset(TRACS.Shape_DF$lef_lpv_3m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$rig_lpv_3m,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.LPV3_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$lef_lpv_3m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$rig_lpv_3m,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
```

```
      TRACS.LPV10[l] <- mean(cbind(subset(TRACS.Shape_DF$le_lpv_10m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$ri_lpv_10m,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.LPV10_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$le_lpv_10m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$ri_lpv_10m,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.LPV30[l] <- mean(cbind(subset(TRACS.Shape_DF$le_lpv_30m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$ri_lpv_30m,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.LPV30_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$le_lpv_30m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), subset(TRACS.Shape_DF$ri_lpv_30m,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude > Lat2) & (TRACS.Shape_DF$Latitude < Lat1))), na.rm=TRUE)
      TRACS.Texture[l] <- mean(subset(TRACS.Shape_DF$texture,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), na.rm=TRUE)
      TRACS.Texture_sd[l] <- sd(subset(TRACS.Shape_DF$texture,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude > Lat2) &
(TRACS.Shape_DF$Latitude < Lat1)), na.rm=TRUE)

    }

  } else {

    if (Lon1 > Lon2) {
      l <- l + 1
```

```
      Geometry.Radius[l] <- mean(subset(Geometry.Shape_DF$abs_radius,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude > Lon2) &
(Geometry.Shape_DF$Longitude < Lon1) & (Geometry.Shape_DF$Latitude < Lat2) &
(Geometry.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      Geometry.Radius_sd[l] <- sd(subset(Geometry.Shape_DF$abs_radius,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude > Lon2) &
(Geometry.Shape_DF$Longitude < Lon1) & (Geometry.Shape_DF$Latitude < Lat2) &
(Geometry.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      Geometry.Gradient[l] <- mean(subset(Geometry.Shape_DF$gradient,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude > Lon2) &
(Geometry.Shape_DF$Longitude < Lon1) & (Geometry.Shape_DF$Latitude < Lat2) &
(Geometry.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      Geometry.Gradient_sd[l] <- sd(subset(Geometry.Shape_DF$gradient,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude > Lon2) &
(Geometry.Shape_DF$Longitude < Lon1) & (Geometry.Shape_DF$Latitude < Lat2) &
(Geometry.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      Geometry.Crossfall[l] <- mean(subset(Geometry.Shape_DF$crossfall,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude > Lon2) &
(Geometry.Shape_DF$Longitude < Lon1) & (Geometry.Shape_DF$Latitude < Lat2) &
(Geometry.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      Geometry.Crossfall_sd[l] <- sd(subset(Geometry.Shape_DF$crossfall,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude > Lon2) &
(Geometry.Shape_DF$Longitude < Lon1) & (Geometry.Shape_DF$Latitude < Lat2) &
(Geometry.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      SCRIM.SC[l] <- mean(subset(SCRIM.Shape_DF$coefficien, (SCRIM.Shape_DF$direction
== Dir) & (SCRIM.Shape_DF$Longitude > Lon2) & (SCRIM.Shape_DF$Longitude < Lon1) &
(SCRIM.Shape_DF$Latitude < Lat2) & (SCRIM.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      SCRIM.SC_sd[l] <- sd(subset(SCRIM.Shape_DF$coefficien,
(SCRIM.Shape_DF$direction == Dir) & (SCRIM.Shape_DF$Longitude > Lon2) &
(SCRIM.Shape_DF$Longitude < Lon1) & (SCRIM.Shape_DF$Latitude < Lat2) &
(SCRIM.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      TRACS.Rut[l] <- mean(cbind(subset(TRACS.Shape_DF$left_rut,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$right_rut,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.Rut_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$left_rut,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
```

```
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$right_rut,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.LPV3[l] <- mean(cbind(subset(TRACS.Shape_DF$lef_lpv_3m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$rig_lpv_3m,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.LPV3_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$lef_lpv_3m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$rig_lpv_3m,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.LPV10[l] <- mean(cbind(subset(TRACS.Shape_DF$le_lpv_10m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$ri_lpv_10m,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.LPV10_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$le_lpv_10m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$ri_lpv_10m,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.LPV30[l] <- mean(cbind(subset(TRACS.Shape_DF$le_lpv_30m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$ri_lpv_30m,
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.LPV30_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$le_lpv_30m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$ri_lpv_30m,
```

```
(TRACS.Shape_DF$Longitude > Lon2) & (TRACS.Shape_DF$Longitude < Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.Texture[l] <- mean(subset(TRACS.Shape_DF$texture,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      TRACS.Texture_sd[l] <- sd(subset(TRACS.Shape_DF$texture,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude > Lon2) &
(TRACS.Shape_DF$Longitude < Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), na.rm=TRUE)

    } else {
      l <- l + 1
      Geometry.Radius[l] <- mean(subset(Geometry.Shape_DF$abs_radius,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude < Lon2) &
(Geometry.Shape_DF$Longitude > Lon1) & (Geometry.Shape_DF$Latitude < Lat2) &
(Geometry.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      Geometry.Radius_sd[l] <- sd(subset(Geometry.Shape_DF$abs_radius,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude < Lon2) &
(Geometry.Shape_DF$Longitude > Lon1) & (Geometry.Shape_DF$Latitude < Lat2) &
(Geometry.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      Geometry.Gradient[l] <- mean(subset(Geometry.Shape_DF$gradient,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude < Lon2) &
(Geometry.Shape_DF$Longitude > Lon1) & (Geometry.Shape_DF$Latitude < Lat2) &
(Geometry.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      Geometry.Gradient_sd[l] <- sd(subset(Geometry.Shape_DF$gradient,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude < Lon2) &
(Geometry.Shape_DF$Longitude > Lon1) & (Geometry.Shape_DF$Latitude < Lat2) &
(Geometry.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      Geometry.Crossfall[l] <- mean(subset(Geometry.Shape_DF$crossfall,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude < Lon2) &
(Geometry.Shape_DF$Longitude > Lon1) & (Geometry.Shape_DF$Latitude < Lat2) &
(Geometry.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      Geometry.Crossfall_sd[l] <- sd(subset(Geometry.Shape_DF$crossfall,
(Geometry.Shape_DF$direction == Dir) & (Geometry.Shape_DF$Longitude < Lon2) &
(Geometry.Shape_DF$Longitude > Lon1) & (Geometry.Shape_DF$Latitude < Lat2) &
(Geometry.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
```

```
      SCRIM.SC[l] <- mean(subset(SCRIM.Shape_DF$coefficien, (SCRIM.Shape_DF$direction
== Dir) & (SCRIM.Shape_DF$Longitude < Lon2) & (SCRIM.Shape_DF$Longitude > Lon1) &
(SCRIM.Shape_DF$Latitude < Lat2) & (SCRIM.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      SCRIM.SC_sd[l] <- sd(subset(SCRIM.Shape_DF$coefficien,
(SCRIM.Shape_DF$direction == Dir) & (SCRIM.Shape_DF$Longitude < Lon2) &
(SCRIM.Shape_DF$Longitude > Lon1) & (SCRIM.Shape_DF$Latitude < Lat2) &
(SCRIM.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      TRACS.Rut[l] <- mean(cbind(subset(TRACS.Shape_DF$left_rut,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$right_rut,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.Rut_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$left_rut,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$right_rut,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.LPV3[l] <- mean(cbind(subset(TRACS.Shape_DF$lef_lpv_3m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$rig_lpv_3m,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.LPV3_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$lef_lpv_3m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$rig_lpv_3m,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.LPV10[l] <- mean(cbind(subset(TRACS.Shape_DF$le_lpv_10m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$ri_lpv_10m,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.LPV10_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$le_lpv_10m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
```

```
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$ri_lpv_10m,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.LPV30[l] <- mean(cbind(subset(TRACS.Shape_DF$le_lpv_30m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$ri_lpv_30m,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.LPV30_sd[l] <- sd(cbind(subset(TRACS.Shape_DF$le_lpv_30m,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), subset(TRACS.Shape_DF$ri_lpv_30m,
(TRACS.Shape_DF$Longitude < Lon2) & (TRACS.Shape_DF$Longitude > Lon1) &
(TRACS.Shape_DF$Latitude < Lat2) & (TRACS.Shape_DF$Latitude > Lat1))), na.rm=TRUE)
      TRACS.Texture[l] <- mean(subset(TRACS.Shape_DF$texture,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), na.rm=TRUE)
      TRACS.Texture_sd[l] <- sd(subset(TRACS.Shape_DF$texture,
(TRACS.Shape_DF$direction == Dir) & (TRACS.Shape_DF$Longitude < Lon2) &
(TRACS.Shape_DF$Longitude > Lon1) & (TRACS.Shape_DF$Latitude < Lat2) &
(TRACS.Shape_DF$Latitude > Lat1)), na.rm=TRUE)

    }
  }
}

r <- 0

#rm(Geometry_Shape_DF, SCRIM_Shape_DF, TRACS_Shape_DF) # free up some memory

####add road pavement conditions to Data.T#####

Data.T$Geom.Abs_Radius <- Geometry.Radius
Data.T$Geom.Abs_Radius_sd <- Geometry.Radius_sd
Data.T$Geom.Gradient <- Geometry.Gradient
Data.T$Geom.Gradient_sd <- Geometry.Gradient_sd
```

```
Data.T$Geom.Crossfall <- Geometry.Crossfall
Data.T$Geom.Crossfall_sd <- Geometry.Crossfall_sd
#SCRIM is not in A50 and therefore it is better to neglect
#Data.T$SCRIM.Coefficient <- SCRIM.SC #uncomment if SCRIM is to be considered
#Data.T$SCRIM.Coefficient_sd <- SCRIM.SC_sd #uncomment if SCRIM is to be considered
Data.T$TRACS.Rutting <- TRACS.Rut
Data.T$TRACS.Rutting_sd <- TRACS.Rut_sd
Data.T$TRACS.LPV03m <- TRACS.LPV3
Data.T$TRACS.LPV03m_sd <- TRACS.LPV3_sd
Data.T$TRACS.LPV10m <- TRACS.LPV10
Data.T$TRACS.LPV10m_sd <- TRACS.LPV10_sd
Data.T$TRACS.LPV30m <- TRACS.LPV30
Data.T$TRACS.LPV30m_sd <- TRACS.LPV30_sd
Data.T$TRACS.Pav_Texture <- TRACS.Texture
Data.T$TRACS.Pav_Texture_sd <- TRACS.Texture_sd

Data.T <- Data.T[complete.cases(Data.T),] # hold only complete cases

t2 <- Sys.time()
difftime(t2, t1, units = "mins")

# save results
write.csv(Data.T, "20180913 DataT Heavy Jan Mar 2017 25 kmh with Road Conditions
Part4.csv")
```

Appendix D.1.5 Append weather conditions to Data.T

This part of the code is to assign weather conditions to each record in Data.T.

```
# Append Weather Conditions Data to DataT
# create a function to transform latitude and longitude from deg to rad
degrees.to.radians <- function(degrees=45,minutes=30)
{
  if(!is.numeric(minutes)) stop("Please enter a numeric value for minutes!\n")
  if(!is.numeric(degrees)) stop("Please enter a numeric value for degrees!\n")
  decimal<-minutes/60
  c.num<-degrees+decimal
  radians<-c.num*pi/180
```

```
  radians
}

t1 <- Sys.time() #take track of how long the code takes to run: Start
i <- 0
Lat.Meteo.St <- unique(Weather$Latitude)
Lon.Meteo.St <- unique(Weather$Longitude)
m <- c()
Lat.St <- c()
Lon.St <- c()
Dist <- c()

for (i in (1:nrow(Data.T))) {

  print(paste0("processing row ", i))

  # define the measurement in Weather which is closer in time and distance to the
truck
  # start is reference

  for (j in 1:length(Lat.Meteo.St)) {

    Dist[j] <- acos(sin(degrees.to.radians(Lat.Meteo.St[j])) *
sin(degrees.to.radians(Data.T$Lat.Start[i])) +
cos(degrees.to.radians(Lat.Meteo.St[j])) *
cos(degrees.to.radians(Data.T$Lat.Start[i])) *
cos(degrees.to.radians(Data.T$Long.Start[i])-degrees.to.radians(Lon.Meteo.St[j]))) *
6373

  }

  Lat.St[i] <- Lat.Meteo.St[which.min(Dist)]
  Lon.St[i] <- Lon.Meteo.St[which.min(Dist)]

  #extract data from the closest weather station
  Meteo <- subset(Weather, Latitude == Lat.St[i] & Longitude == Lon.St[i])

  # select the closest data in time
  for (r in 1:nrow(Meteo)) {
```

```
    # calculate the difference in time
    Meteo$Diff.Time[r] <- abs(difftime(as.POSIXct(Data.T$Date.Start[i], format = "%Y-
%m-%d %H:%M:%OS"), format(as.POSIXct(Meteo$DateTime[r], format = "%d/%m/%Y
%H:%M:%OS"), "%Y-%m-%d %H:%M:%OS"), units = "secs"))

  }

  m[i] <- as.numeric(which.min(Meteo$Diff.Time))

  # append meteo data to Data.T from the closest weather station to the considered
truck
  Data.T$Weather_Station[i] <- Meteo$Weather_Station[m[i]]
  Data.T$Temperature_deg_C[i] <- Meteo$Temperature_deg_C[m[i]]
  Data.T$`Humidity_%`[i] <- Meteo$`Humidity_%`[m[i]]
  Data.T$Atm_Pressure_hPa[i] <- Meteo$Atm_Pressure_hPa[m[i]]
  Data.T$Rainfall_mm[i] <- mean(c(Meteo$Rainfall_mm[m[i]-1], Meteo$Rainfall_mm[m[i]],
Meteo$Rainfall_mm[m[i]+1]), na.rm = TRUE)
  Data.T$WindSpeed_kn[i] <- Meteo$WindSpeed_kn[m[i]]
  Data.T$WindDirection_deg[i] <- Meteo$WindDirection_deg[m[i]]

}

t2 <- Sys.time() #take track of how long the code takes to run: End
difftime(t2, t1) #show how long took the code to run

write.csv(Data.T, "20180913 DataT Heavy 25 kmh Jan Sep 2017 Weather Part2.csv") #save
results
save.image("~/Heavy Jan Sep 2017 25 kmh Weather Part2.RData") #save workspace
```

Appendix D.1.6 Complete Data.T

This part of the code is to transform certain variables (e.g. wind speed into its two components in direction and transversal to the direction of travel of the considered truck) and complete the set of measurements included in Data.T.

```
## Complete Data.T ##
```

```
library(maptools)

# create a function that transforms angles in consideration of the deviation from the
azimuth
trackAngle <- function(x) { #this function needs the library(maptools) to use
trackAzimuth()
  angles <- abs(c(trackAzimuth(x), 0) -
                  c(0, rev(trackAzimuth(x[nrow(x):1, ]))))
  angles <- ifelse(angles > 180, 360 - angles, angles)
  angles[is.na(angles)] <- 180
  angles[-c(1, length(angles))]
}

for (i in 1:nrow(Data.T)) {

  print(paste0("working on row nr ", i))

  angle1 <- as.matrix(data.frame("longitude" = c(Data.T$Long.Start[i],
Data.T$Long.End[i], 0),
                                 "latitude" = c(Data.T$Lat.Start[i],
Data.T$Lat.End[i], 0))) #get the direction of travel
  angle1 <- trackAngle(angle1) # calculate the angle between the direction of travel
and the azimuth (geographic north)

  if (angle1 < 0){
    angle1 <- angle1 + 360 #transform the result to a 0-360 deg angle
  }

  angle2 <- as.numeric(Data.T$WindDirection_deg[i]) #direction of wind

  delta.angle <- angle2 - angle1 #difference of direction between truck and wind
direction

  Data.T$RoadType[i] <- substr(Data.T[i,]$Road.ID, 1, 1) #assign a road type
  Data.T$Delta.Angle_deg[i] <- delta.angle #assigne the difference in angle between
truck and wind direction
  Data.T$WindSpeedDir_kn[i] <- Data.T$WindSpeed_kn[i] * cos(delta.angle*pi/180) #wind
speed in direction of travel of the considered vehicle (positive if same direction)
```

```
  Data.T$WindSpeedCro_kn[i] <- abs(Data.T$WindSpeed_kn[i] * sin(delta.angle*pi/180))
#absolute value of cross windspeed to the direction of travel of the considered
vehicle

}

# finally, save results
write.csv(Data.T, "20180918 DataT Heavy Jan Mar 2017 Complete Part2.csv", row.names =
FALSE)
```

Appendix D.2 Variable selection

The following sections detail the bits of code used to select variables to help the developed models in giving the best possible performance.

Appendix D.2.1 p-values

This part of the code tests significance (p-value) of each of the considered variables in the development of an LR.

```
# fit an LR of all measured variables
lm.fit <- lm(Fuel.Consumption ~ ., data = data.t)
# summarize results of the LR and visualize p-values
summary(lm.fit)
```

Appendix D.2.2 Adjusted-R^2

This part of the code compares how each of the considered variables influence the resulting adjusted-R^2 of an LR.

```
# calculate a summary of values for a LR of the measured variables
regfit.full<-regsubsets(Fuel.Consumption~., data.t)
# summarize results of the LR and visualize p-values
reg.summary<-summary(regfit.full)

#Plot results
```

```
#png(filename = "adjustedR-squared XF Series.png", width = 800, height = 800, units =
"px")
par(oma=c(1,2,1,1), mar=c(4, 5, 3, 2)) # uncomment to save as a png file
plot(reg.summary$adjr2, xlab = "Number of Variables", ylab = "Adjusted R^2", cex.main
= 2, cex.lab = 1.8, cex.axis = 1.7, type= "l")
points(which.max(reg.summary$adjr2), reg.summary$adjr2[which.max(reg.summary$adjr2)],
col = "red", cex=5, pch=20)
#dev.off()

#png(filename = "adjustedR-squared_table XF Series.png", width = 800, height = 800,
units = "px") # uncomment to save as a png file
plot(regfit.full, scale = "adjr2", main = "Adjusted R^2")
#dev.off()
```

Appendix D.2.3 Lasso

This part of the code is to apply the Lasso to regularize the developed models.

```
#### Lasso for variable selection ####

grid <- 10^seq(10, -2, length = 100)

x <- as.matrix(scaled)
y <- as.matrix(scaled$Fuel.Consumption)

x.train <- as.matrix(train)
y.train <- as.matrix(train$Fuel.Consumption)

x.test <- as.matrix(validation)
y.test <- as.matrix(validation$Fuel.Consumption)

lasso.mod <- glmnet(x, y, alpha=1, lambda=grid)

#png(filename = "Lasso mod.png", width = 800, height = 800, units = "px")
plot(lasso.mod)
#dev.off()
```

```
set.seed(774) #set a random seed to ensure repeatability of the results
cv.out <- cv.glmnet(x, y, alpha=1)

#png(filename = "Lasso crossvalidation.png", width = 800, height = 800, units = "px")
plot(cv.out)
#dev.off()

bestlam <- cv.out$lambda.min
lasso.pred <- predict(lasso.mod, s=bestlam, newx=x)
lsRMSE <- sqrt(mean(((lasso.pred*s+c)-(y*s+c))^2))
lsMAE <- mean(abs((lasso.pred*s+c)-(y*s+c)))
print(paste("The RMSE for the lasso linear model is", round(lsRMSE, 2), sep = " "))
print(paste("The MAE for the lasso linear model is", round(lsMAE, 2), sep = " "))

out <- glmnet(x*s+c, y*s+c, alpha=1, lambda=grid)
lasso.coef <- predict(out, type= "coefficients", s=bestlam)
lasso.coef
# values of lasso.coeff = 0 means that that variable does not allow the developed
model to explore a wider range of variance and therefore it can be removed from the
analysis
```

Appendix D.2.4 Random Forest

This part of the code is apply the RF algorithm to rank the variables considered in the study based on how important their influence is on truck fleet fuel consumption (the target variable for this study).

```
#lets generate a random forest
#library(randomForest)
#to plot square plots use: par(pty="s")

set.seed(48) #set a random seed to ensure repeatability
modelrf <- randomForest(Fuel.Consumption ~ ., data = train, ntree = 1000, mtry = 8)
VarImpPlot(modelrf) #this shows how each variable influence the target
(Fuel.Consumption)
```

Appendix D.3 Cross-validation

The following section show the code used to split the available datasets and how different splits of data are used to train the model. That is to ensure that performance do not change for different splits of the available dataset.

Appendix D.3.1 Data split

This part of the code is to show how data is split. This represents an example and may vary case by case for different truck types.

```
#extract a sample of approximately 5% of data including all records from certain
trucks only and exclude these trucks from the rest of the data
validation <- subset(scaled, scaled$Vehicle.RN == "352108072455483" |
scaled$Vehicle.RN == "352108072455954" | scaled$Vehicle.RN == "352108072436202" |
scaled$Vehicle.RN == "352108071523505" | scaled$Vehicle.RN == "352108071421015" |
scaled$Vehicle.RN == "352108070884361" | scaled$Vehicle.RN == "351934048688910" |
scaled$Vehicle.RN == "351934047792770" | scaled$Vehicle.RN == "354898045606489" |
scaled$Vehicle.RN == "354898043824399" | scaled$Vehicle.RN == "352108072512895" |
scaled$Vehicle.RN == "352054059180605" | scaled$Vehicle.RN == "352108071658210")

#exclude records from trucks in validation from the training/testing procedure
scaled <- subset(scaled, scaled$Vehicle.RN != "352108072455483" & scaled$Vehicle.RN
!= "352108072455954" & scaled$Vehicle.RN != "352108072436202" & scaled$Vehicle.RN !=
"352108071523505" & scaled$Vehicle.RN != "352108071421015" & scaled$Vehicle.RN !=
"352108070884361" & scaled$Vehicle.RN != "351934048688910" & scaled$Vehicle.RN !=
"351934047792770" & scaled$Vehicle.RN != "354898045606489" & scaled$Vehicle.RN !=
"354898043824399" & scaled$Vehicle.RN != "352108072512895" & scaled$Vehicle.RN !=
"352054059180605" & scaled$Vehicle.RN != "352108071658210")

test$Vehicle.RN <- c()
scaled$Vehicle.RN <- c()
data.t$Vehicle.RN <- c()
```

Appendix D.3.2 10-fold cross-validation

This part of the code shows how 10-fold cross-validation is performed. This example shows how cross-validation is used to train the ANN but similar code is used to train RF and SVR.

```
## 10-fold crossvalidation Artificial Neural Network ##
#init some variables needed
cv.RMSE <- c()
cv.MAE <- c()
cv.MAPE <- c()
cv.Rsq <- c()
tr.Time <- c()
tt.Time <- 0
k <- 10

#init some libraries needed
library(neuralnet)
library(gridExtra)
library(plyr)
pbar <- plyr::create_progress_bar('text')
pbar$init(k)

#save results of the cv process in a figure
png("10-fold cross-validation ANN Heavy All 2017 All Roads.png", width = 30, height =
12, units = "cm", res = 100)
par(mfrow = c(2, 5), oma = c(3, 3, 0, 0), mar=c(2,2,1,1), mgp=c(3,1,0), font.lab = 2,
family = "serif")

#start the 10-fold cv process
for(i in 1:k){
  print(paste("trial no ", i))
  print("split the data...")
  set.seed(1523+i)
  index <- sample(1:nrow(scaled),round(0.75*nrow(scaled)))
  train.cv <- scaled[index,]
  test.cv <- scaled[-index,]
```

```
  print("build the model")
  t1 <- Sys.time()
  nnetwork <- neuralnet(formula = Fuel.Consumption ~ Altitude.AVG + Traveled.Time +
Gross.Weight_kg + Speed.AVG + Acceleration + Torque.AVG + Revs.AVG + Geom.Abs_Radius
+ Geom.Gradient + Geom.Crossfall +TRACS.LPV03m + TRACS.LPV10m + TRACS.LPV30m +
TRACS.Pav_Texture + Temperature_deg_C +Humidity_. + Atm_Pressure_hPa +
WindSpeedDir_kn, data = train.cv, hidden = c(14, 4), threshold = 8, stepmax = 10^6,
rep = 1, learningrate.factor = list(minus = 0.5, plus = 1.2), lifesign.step = 5000,
lifesign = "full",algorithm = "rprop-", err.fct = "sse",  act.fct = "logistic",
linear.output = TRUE, likelihood = FALSE)
  t2 <- Sys.time()
  tr.Time[i] <- difftime(t2, t1, units = "mins")
  tt.Time <- tt.Time + tr.Time[i]

  print("make predictions...")
  pr.nn <- neuralnet::compute(nnetwork, test.cv[,2:19])
  pr.nn <- as.data.frame(pr.nn$net.result)
  pr.nn <- pr.nn*s[1]+c[1]

  test.cv.r <- (test.cv$Fuel.Consumption)*s[1]+c[1]

  print("plot graph...")
  plot(test.cv.r, pr.nn[1]$V1, pch = 1, col = "darkred", cex = 1, cex.axis = 1,
cex.lab = 1, ylim = c(-10, 100), xlim = c(0, 100), ylab = "", xlab = "")
  grid(nx = NULL, ny = NULL, col = "lightgrey", lty = "dotted")
  legend(x = 78, y = 5, inset = 0, legend = c("test"), col = c("darkred"), pch = 1,
cex = 1, horiz = TRUE)

  print("compute errors...")
  cv.RMSE[i] <- sqrt(mean(((test.cv.r - pr.nn[1])$V1)^2))
  cv.MAE[i] <- mean(abs((test.cv.r - pr.nn[1])$V1), na.omit = TRUE)
  cv.MAPE[i] <- mean(abs((test.cv.r - pr.nn[1])$V1)/(test.cv.r))*100
  cv.Rsq[i] <- cor(pr.nn[1]$V1, test.cv.r)^2

  print(pbar$step())

}
```

```
mtext("Measured Fuel Consumption (l/100km)", side = 1, line = 1, outer = TRUE, font =
2)
mtext("Predicted Fuel Consumption (l/100km)", side = 2, line = 1, outer = TRUE, las =
0, font = 2)

dev.off()

cv.NN.RMSE <- cv.RMSE
cv.NN.MAE <- cv.MAE
cv.NN.MAPE <- cv.MAPE
cv.NN.Rsq <- cv.Rsq
cv.NN.Time <- tr.Time
tt.NN.Time <- tt.Time

#uncomment the following to see distributions of RMSE and MAE from the cv
#boxplot(cv.RF.RMSE, ylab = "RMSE (l/100km)", ylim = c(0,5))
#boxplot(cv.RF.MAE, ylab= "MAE (l/100km)", ylim = c(0,5))

#print out results of the cv process
print(paste("the calculated RMSEs for the 10 ANN models are:"))
print(round(cv.NN.RMSE, 2))
print(paste("which average is ", round(mean(cv.NN.RMSE), 2)))
print(paste("the MAEs for the 10 ANN models are:"))
print(round(cv.NN.MAE, 2))
print(paste("which average is ", round(mean(cv.NN.MAE), 2)))
print(paste("the MAPEs for the 10 ANN models are:"))
print(round(cv.NN.MAPE, 2))
print(paste("which average is ", round(mean(cv.NN.MAPE), 2)))
print(paste("the Rsqs for the 10 ANN models are:"))
print(round(cv.NN.Rsq, 2))
print(paste("which average Rsq is ", round(mean(cv.NN.Rsq), 2)))
print(paste("the average training time for the ANN model is ",
round(mean(cv.NN.Time), 2), " mins"))
print(paste("the total crossvalidation time for the ANN model is ", round(tt.NN.Time,
2), " mins"))
```

Appendix D.4 Modelling

In the following sections details in regards how each of the developed models is build are given. Piece of code reported show only an example and differences may occur when building a model for different types of truck or if a different set of variables is considered.

Appendix D.4.1 LR

This part of the code is to build the MLR model.

```
# fit an LR of all measured variables
lm.fit <- lm(Fuel.Consumption ~ ., data = data.t)
```

Appendix D.4.2 SVR

This part of the code is to build the SVR model.

```
#let us fit a support vector machine regression model
#library(e1071)
#test the developed model
set.seed(34) #set a random seed to ensure repeatability
modelsv <- svm(Fuel.Consumption ~ ., train)
Sys.time()

#test the developed model
tunedModelY <- predict(modelsv, train)
tunedModelY_test <- predict(modelsv, test)
tunedModelY_val <- predict(modelsv, validation)
Sys.time()

#create data frame for plotting data...
svm.train <- data.frame(Predicted = tunedModelY*s+c, Measured =
train$Fuel.Consumption*s+c)
```

```
svm.train$Group <- "train"
svm.test <- data.frame(Predicted = tunedModelY_test*s+c, Measured =
test$Fuel.Consumption*s+c)
svm.test$Group <- "test"
svm.val <- data.frame(Predicted = tunedModelY_val*s+c, Measured =
validation$Fuel.Consumption*s+c)
svm.val$Group <- "validation"
svm.data <- rbind(svm.train, svm.test, svm.val)

# finally, estimate performance of the developed model (for the validation set)
svmRMSE <- sqrt(mean(((svm.val$Measured)-(svm.val$Predicted))^2))
svmMAE <- mean(abs((svm.val$Measured)-(svm.val$Predicted)))
svmMAPE <- mean(abs((svm.val$Measured)-(svm.val$Predicted))/(ann.test$Measured))*100
svmRsq <- cor(svm.val$Predicted, svm.val$Measured)^2
print(paste("The RMSE for the support vector regression model is", round(svmRMSE, 2),
sep = " "))
print(paste("The MAE for the support vector regression model is", round(svmMAE, 2),
sep = " "))
print(paste("The MAPE for the random forest model is", round(svmMAPE, 2), sep = " "))
print(paste("The Rsq for the support vector regression model is", round(svmRsq, 3),
sep = " "))
Sys.time()
```

Appendix D.4.3 RF

This part of the code is to build the RF model.

```
#let us fit a random forest regression model
#library(randomForest)
#train the developed model
set.seed(48) #set a random seed to ensure repeatability
modelrf <- randomForest(Fuel.Consumption ~ ., data = train, ntree = 1000, mtry = 8)

#test the developed model
predictedYrf <- predict(modelrf, train)
predictedYrf_test <- predict(modelrf, test)
predictedYrf_val <- predict(modelrf, validation)
Sys.time()
```

```
#create dataframe for plotting data...
rf.train <- data.frame(Predicted = predictedYrf*s+c, Measured =
train$Fuel.Consumption*s+c)
rf.train$Group <- "train"
rf.test <- data.frame(Predicted = predictedYrf_test*s+c, Measured =
test$Fuel.Consumption*s+c)
rf.test$Group <- "test"
rf.val <- data.frame(Predicted = predictedYrf_val*s+c, Measured =
validation$Fuel.Consumption*s+c)
rf.val$Group <- "validation"
rf.data <- rbind(rf.train, rf.test, rf.val)
Sys.time()

# finally, estimate performance of the developed model (for the validation set)
rfRMSE <- sqrt(mean(((rf.val$Measured) - (rf.val$Predicted))^2))
rfMAE <- mean(abs((rf.val$Measured) - (rf.val$Predicted)))
rfMAPE <- mean(abs((rf.val$Measured) - (rf.val$Predicted))/(ann.test$Measured))*100
rfRsq <- cor(rf.val$Predicted, rf.val$Measured)^2
print(paste("The RMSE for the random forest model is", round(rfRMSE, 2), sep = " "))
print(paste("The MAE for the random forest model is", round(rfMAE, 2), sep = " "))
print(paste("The MAPE for the random forest model is", round(rfMAPE, 2), sep = " "))
print(paste("The Rsq for the random forest model is", round(rfRsq, 3), sep = " "))
Sys.time()
```

Appendix D.4.4 ANN

This part of the code is to build the ANN model.

```
#let us generate an artificial neural network
#library(neuralnet)
#train the developed model
set.seed(49)
nnetwork <- neuralnet(formula = Fuel.Consumption ~ Gross.Weight + Geom.Gradient +
Speed.AVG + Torque.Start + Torque.End + Revs.Start + Acceleration + Gear, data =
train.cv, hidden = c(10,3), threshold = 1.5, stepmax = 10^6, rep = 1,
learningrate.factor = list(minus = 0.5, plus = 1.2), lifesign.step = 5000, lifesign =
```

```
"full", algorithm = "rprop-", err.fct = "sse",  act.fct = "logistic", linear.output =
TRUE, likelihood = FALSE)
plot(nnetwork) #show structure of the developed ANN

#test the developed model
Sys.time()
predict.nnr <- as.data.frame(nnetwork$net.result)
predict.nn_test <- compute(nnetwork, test)
predict.nnr_test <- as.data.frame(predict.nn_test$net.result)
predict.nn_val <- compute(nnetwork, validation)
predict.nnr_val <- as.data.frame(predict.nn_val$net.result)

#create dataframe for plotting data...
ann.train <- data.frame(Predicted = predict.nnr[,1]*s+c, Measured =
train.cv$Fuel.Consumption*s+c)
ann.train$Group <- "train"
ann.test <- data.frame(Predicted = predict.nnr_test[,1]*s+c, Measured =
test$Fuel.Consumption*s+c)
ann.test$Group <- "test"
ann.val <- data.frame(Predicted = predict.nnr_val[,1]*s+c, Measured =
validation$Fuel.Consumption*s+c)
ann.val$Group <- "validation"
ann.data = rbind(ann.train, ann.test, ann.val)
Sys.time()

# finally, estimate performance of the developed model (for the validation set)
nnRMSE <- sqrt(mean(((ann.val$Measured)-(ann.val$Predicted))^2))
nnMAE <- mean(abs((ann.val$Measured)-(ann.val$Predicted)))
nnMAPE <- mean(abs((ann.val$Measured)-(ann.val$Predicted))/(ann.val$Measured))*100
nnRsq <- cor(ann.val$Predicted, ann.val$Measured)^2
print(paste("The RMSE for the neural network model is", round(nnRMSE, 2), sep = " "))
print(paste("The MAE for the neural network model is", round(nnMAE, 2), sep = " "))
print(paste("The MAPE for the neural network model is", round(nnMAPE, 2), sep = " "))
print(paste("The Rsq for the neural network model is", round(nnRsq, 3), sep = " "))
Sys.time()
```

Appendix D.5 Parametric analysis

This part of the code performs a parametric analysis to test reliability and robustness of the developed ML models. This version of the code is to perform an automatic parametric analysis which does not consider the type of the considered variables. Better plots can be obtained by identifying Boolean or discontinuous variables and plotting parametric values only for actual values of the variables.

```
c <- c() #initialise a vector for the mean value of each variable considered
s <- c() #initialise a vector for the value of standard deviation each variable
considered

# calculate mean and sd of each variable considered in the study, that is to re-scale
results of machine learning algorithms that are trained on scaled values
for (i in 1:ncol(data.t)) { #calculate mean and sd for each of the considered
variables

  c[i] <- mean(data.t[,i])
  s[i] <- sd(data.t[,i])

}

# initialise vectors and dataframes to perform the parametric analysis
scaled.par <- c()
datat.par <-c()
scaled.par1 <- c()
datat.par1 <- c()
v <- c()
v.t <- c()
Test <- c()

for (i in 1:(ncol(scaled)-1)) {

  name <- colnames(scaled[i+1])
  name.t <- paste0(colnames(data.t[i+1]), "t")
  name1 <- paste0(colnames(scaled[i+1]), ".Par")
```

```
  name.t1 <- paste0(colnames(data.t[i+1]), "t.Par")
  assign(name, rep(mean(scaled[, i+1]), 50))
  assign(name.t, rep(mean(data.t[, i+1]), 50))

  #We need to calculate the intervals where we are confident to have data enough to
make conclusions
  #for normal distr. This corresponds to 2 times the standard deviation
  #for non-normal distributions we instead evaluate the intervals where ~90% of data
is available (between the 5th and 95th percentiles)

  v[1] <- as.numeric(as.character(quantile(scaled[, i+1], .05))) #5th percentile
  d <- (as.numeric(as.character(quantile(scaled[, i+1], .95))) - v[1])/50 #95th
percentile

  v.t[1] <- as.numeric(as.character(quantile(data.t[, i+1], .05))) #5th percentile
  d.t <- (as.numeric(as.character(quantile(data.t[, i+1], .95))) - v.t[1])/50 #95th
percentile

  for (j in 2:50) {
    v[j] <- v[j-1] + d
    v.t[j] <- v.t[j-1] + d.t
  }

  assign(name1, v)
  assign(name.t1, v.t)
}

#Create first scaled data frame
scaled.par <- data.frame(Altitude.AVG, Traveled.Time, Gross.Weight_kg, Speed.AVG,
Acceleration, Gear.Start, CruiseControl.Start, Torque.AVG, Revs.AVG, Geom.Abs_Radius,
Geom.Gradient, Geom.Crossfall, TRACS.LPV03m, TRACS.LPV10m, TRACS.LPV30m,
TRACS.Pav_Texture, Temperature_deg_C, Humidity_., Atm_Pressure_hPa, Rainfall_mm,
WindDirection_deg, WindSpeedDir_kn, RoadType, WindSpeedCro_kn)
# Create first data.t data frame
datat.par <- data.frame(Altitude.AVGt, Traveled.Timet, Gross.Weight_kgt, Speed.AVGt,
Accelerationt, Gear.Startt, CruiseControl.Startt, Torque.AVGt, Revs.AVGt,
Geom.Abs_Radiust, Geom.Gradientt, Geom.Crossfallt, TRACS.LPV03mt, TRACS.LPV10mt,
TRACS.LPV30mt, TRACS.Pav_Texturet, Temperature_deg_Ct, Humidity_.t,
```

```
Atm_Pressure_hPat, Rainfall_mmt, WindDirection_degt, WindSpeedDir_knt, RoadTypet,
WindSpeedCro_knt)

# Create second scaled data frame
scaled.par1 <- data.frame(Altitude.AVG.Par, Traveled.Time.Par, Gross.Weight_kg.Par,
Speed.AVG.Par, Acceleration.Par, Gear.Start.Par, CruiseControl.Start.Par,
Torque.AVG.Par, Revs.AVG.Par, Geom.Abs_Radius.Par, Geom.Gradient.Par,
Geom.Crossfall.Par, TRACS.LPV03m.Par, TRACS.LPV10m.Par, TRACS.LPV30m.Par,
TRACS.Pav_Texture.Par, Temperature_deg_C.Par, Humidity_..Par, Atm_Pressure_hPa.Par,
Rainfall_mm.Par, WindDirection_deg.Par, RoadType.Par, WindSpeedDir_kn.Par,
WindSpeedCro_kn.Par)

# Create second data frame
datat.par1 <- data.frame(Altitude.AVGt.Par, Traveled.Timet.Par, Gross.Weight_kgt.Par,
Speed.AVGt.Par, Acceleration t.Par, Gear.Startt.Par, CruiseControl.Startt.Par,
Torque.AVGt.Par, Revs.AVGt.Par, Geom.Abs_Radiust.Par, Geom.Gradientt.Par,
Geom.Crossfallt.Par, TRACS.LPV03mt.Par, TRACS.LPV10mt.Par, TRACS.LPV30mt.Par,
TRACS.Pav_Texturet.Par, Temperature_deg_Ct.Par, Humidity_.t.Par,
Atm_Pressure_hPat.Par, Rainfall_mmt.Par, WindDirection_degt.Par, RoadTypet.Par,
WindSpeedDir_knt.Par, WindSpeedCro_knt.Par)

# Uniform column names of all dataframes
colnames(datat.par) <- colnames(scaled.par)
colnames(scaled.par1) <- colnames(scaled.par)
colnames(datat.par1) <- colnames(scaled.par)

library(e1071)
library(randomForest)
library(neuralnet)

png("Parametric All.png", width = 2000, height = 3000, res = 100)
par(mfrow = c(6, 4), oma = c(0, 0, 0, 0), mar=c(5,5,2,2), mgp=c(3,1,0), font.lab = 2)
for (i in 1:ncol(scaled.par)) {

  Test <- scaled.par[i]
  Testt <- datat.par[i]
  scaled.par[i] <- scaled.par1[i]
  datat.par[i] <- datat.par1[i]
```

```
  #Fuel.Cons.MLR <- predict(lm.fit, newdata = datat.par) #uncomment for comparing the
MLR
  Fuel.Cons.SVM <- predict(modelsv, scaled.par)*s[1]+c[1]
  Fuel.Cons.RF <- predict(modelrf, scaled.par)*s[1]+c[1]
  Fuel.Cons.ANN <- c(neuralnet::compute(nnetwork, scaled.par)$net.result*s[1]+c[1])

  plot(scaled.par[, i]*s[i+1]+c[i+1], Fuel.Cons.SVM, pch = 19, col = "blue", cex = 2,
cex.axis = 2, cex.lab = 2, ylim = c(10, 30),
       ylab = "Predicted Fuel Consumption (l/100km)", xlab = colnames(scaled.par)[i])
  #points(scaled.par[, i]*s[i+1]+c[i+1], Fuel.Cons.MLR, pch = 19, col = "black", cex
= 2)
  points(scaled.par[, i]*s[i+1]+c[i+1], Fuel.Cons.RF, pch = 19, col = "green", cex =
2)
  points(scaled.par[, i]*s[i+1]+c[i+1], Fuel.Cons.ANN, pch = 19, col = "red", cex =
2)
  grid(nx = NULL, ny = NULL, col = "lightgrey", lty = "dotted")
  legend(x = "top", inset = 0, legend = c("SVM", "RF", "ANN"),
         col=c("blue", "green", "red"), pch = 19, cex=2, horiz = TRUE)

  #reset the data frames used for the parametric analysis
  scaled.par[i] <- Test
  datat.par[i] <- Testt

}
dev.off()
```

Appendix E Results

This annex is dedicated to the results of the study, evaluation of the developed methodology and comparison of performance of the developed models.

Appendix E.1 Analysis of the Dataset 1

The main objective of analysing the Dataset 1 was to verify feasibility of the study, to define a new methodology for modelling truck fleet fuel consumption using Big Data, and to test possibility of estimating the influence of each of the considered variables on truck fleet fuel economy using the developed models. This included definition of a variable selection procedure, development of an initial model to define baseline performance, improvement of the performance of the developed model, and definition of a new method to evaluate impacts of each of the considered variables on truck fleet fuel consumption.

As already explained in sect. 3.4.1, for this initial case study, Microlise provided 14,281 records from 1,110 Euro 6 heavy trucks. The data, collected in a week in October 2016, come from trucks, which travelled at relatively constant speed (see sect. 4.2.2), along two segments of M1 and the whole M18, two motorways part of the SRN of England. That makes up for about 300 km of road considered.

From these, data from vehicles which had any technical issues or that triggered multiple events in short time intervals such as journeys shorter than 150 m or that produced fuel consumption lower than 3.90 l/100 km or higher than 60 l/100 km were excluded from the data analysis. In fact, these are very rare in the data (< 0.2 %) and, for the considered dataset, can be considered outliers.

The following sections provide details about each phase of the analysis of the Dataset 1.

Appendix E.1.1 Variable Selection 1

Initially, the linear correlation between all data available was investigated by reporting the Pearson's correlation coefficients for all the available measurements and in a correlation matrix (see Figure 39).

Figure 39 report only results on variables which have been finally included in the developed models. This is for clarity of the figures, and includes only variables which show some linear correlation with *Fuel.Consumption*, or that, from a physical perspective (see sect. 2.4 and 2.4), can be considered to be significantly impacting the fuel consumption of trucks and road vehicles in general (such as road surface conditions, see sect. 2.6). Additionally, although some of the variables included in Figure 39, do not show a strong linear correlation with *Fuel.Consumption*, these are shown to be important to model the fuel consumption of the considered fleet of trucks by other statistics applied and other case studies analysed in the following. In fact, from an analysis of the correlation matrix (see Figure 39), only six out of 56 available measurements show strong linear correlation with *Fuel.Consumption* ($|r| \geq 0.10$). These are the gross vehicle weight (*Gross.Weight_kg*), the road gradient (*Geom.Gradient*), the torque % at the start of the travel (*Torque.Start*), the torque % at the end of the travel (*Torque.End*) and the average acceleration (*Acceleration*).

Distributions of these and those of the remaining variables finally considered in the developed models are reported in the chapter dedicated to Data (see Figure 14 in sect, 3.4.1).

	Fuel.Consumption	Gross.Weight	Geom.Gradient	Speed.AVG	Torque.Start	Torque.End	Revs.Start	Acceleration	Cruise.Control	Geom.Abs_Radius	TRACS.LPV03m	TRACS.LPV10m	TRACS.LPV30m	TRACS.Pav_Texture
Fuel.Consumption	1	0.19	0.76	-0.07	0.58	0.59	0.03	-0.19	0.06	0.01	-0.04	-0.05	0.01	-0.04
Gross.Weight	0.19	1	-0.05	-0.03	0.12	0.16	-0.02	0.04	-0.05	-0.01	0.03	0.02	-0.01	0.03
Geom.Gradient	0.76	-0.05	1	-0.03	0.44	0.48	0.02	-0.02	0.07	0	-0.05	-0.05	0.01	-0.06
Speed.AVG	-0.07	-0.03	-0.03	1	-0.02	-0.01	0.24	0.07	0.02	0	-0.01	-0.01	0.01	-0.02
Torque.Start	0.58	0.12	0.44	-0.02	1	0.18	0.05	-0.22	0.05	0.02	-0.03	-0.02	0	-0.04
Torque.End	0.59	0.16	0.48	-0.01	0.18	1	0.04	0.07	0.03	-0.02	-0.03	-0.07	0.01	-0.03
Revs.Start	0.03	-0.02	0.02	0.24	0.05	0.04	1	0.06	-0.06	0	-0.01	0	0.01	-0.01
Acceleration	-0.19	0.04	-0.02	0.07	-0.22	0.07	0.06	1	-0.01	-0.02	-0.03	-0.03	0	-0.01
Cruise.Control	0.06	-0.05	0.07	0.02	0.05	0.03	-0.06	-0.01	1	0.01	0.01	0	0.01	-0.01
Geom.Abs_Radius	0.01	-0.01	0	0	0.02	-0.02	0	-0.02	0.01	1	0.09	-0.02	-0.19	0.02
TRACS.LPV03m	-0.04	0.03	-0.05	-0.01	-0.03	-0.03	-0.01	-0.03	0.01	0.09	1	0.66	0.25	0.39
TRACS.LPV10m	-0.05	0.02	-0.05	-0.01	-0.02	-0.07	0	-0.03	0	-0.02	0.66	1	0.56	0.1
TRACS.LPV30m	0.01	-0.01	0.01	0.01	0	0.01	0.01	0	0.01	-0.19	0.25	0.56	1	-0.15
TRACS.Pav_Texture	-0.04	0.03	-0.06	-0.02	-0.04	-0.03	-0.01	-0.01	-0.01	0.02	0.39	0.1	-0.15	1

Figure 39. Correlation matrix for data in the Dataset 1.

From Figure 14, it is possible to see that generally, the distributions of the continuous variables have shapes typical of normal distributions and can, therefore, be used in the LR analysis. However, some of the variables which result to be non-linearly correlated ($|r| < 0.10$) to *Fuel.Consumption*, such as the radius of curvature (*Geom.Abs_Radius*), or the vehicle speed (*Speed.AVG*), have right-skewed distributions. For *Geom.Abs_Radius*, this is because the recording software calculates this in absolute value and assigns all road sections with radius of curvature above 2000 m a value equal to 2000 m. Therefore, being motorways (such as the considered M1

and M18) typically straight, the resulting distribution of the *Geom.Abs_Radius* results right skewed.

Similarly, only a limited range of *Speed.AVG* is considered, from around 70 km/h to 95 km/h, which is typical of trucks driving on motorways. This makes the distribution of *Speed.AVG* to result also right skewed. Consequently, only a few gears are used (*Gear*), mainly gear 12, producing a narrow range of engine revolutions (*Revs.Start*) (from 1000 to 1500 rpm). For the same reasons, and because of the fact that only journeys of trucks performed at relatively constant speed is considered in the study, a low level of acceleration (*Acceleration*) results.

From an analysis of the plotted distributions, it is also possible to see that both *Torque.Start* and *Torque.End* have a high number of values towards zero. These data are possibly due to faulty sensors. However, this cannot be confirmed and, for this reason, these measurements cannot be excluded from the analysis as outliers.

Then, *p-values* are computed for all variables to test whether or not they are significant for predicting the fuel consumption of the fleet of trucks considered, using LR. Table 30 reports the *p-values* for the variables, which are finally included in the developed model. From this, only ten out of the twelve variables show a significant correlation with fuel consumption (*p-value* < 0.05). However, since it is aim of this study to estimate the impact of *TRACS.LPV03m* and *TRACS.Pav_Texture* on *Fuel.Consumption* and including these in a LR does not produce any significant decay of the adjusted-R^2 (see Figure 40), these are not excluded from the following data analysis. Conversely, other variables such as the geographical location (*Lat.Start, Long.Start, Lat.End and Long.End*), *Gear* and *Cruise.Control* can be considered of low significance, and therefore excluded from the developed models.

Table 30 - Summary of the computed p-values for the highest correlated variables for the Dataset 1.

Variable	p-value	Significance level
Gross.Weight_kg	< 2.00 e^{-16}	99.9 %
Geom.Gradient	< 2.00 e^{-16}	99.9 %
Torque.Start	2.64 e^{-15}	99.9 %
Torque.End	< 2.00 e^{-16}	99.9 %
Revs.Start	< 2.00 e^{-16}	99.9 %
Speed.AVG	< 2.00 e^{-16}	99.9 %
Acceleration	3.73 e^{-5}	99.9 %
Geom.Abs_Radius	< 2.00 e^{-16}	99.9 %
TRACS.LPV03m	0.891	0 %
TRACS.LPV10m	0.001	99.9 %
TRACS.LPV30m	6.80 e^{-6}	99.9 %
TRACS.Pav_Texture	0.179	0 %

However, again, looking at Figure 40, it is possible to see that the inclusion of *Cruise.Control* (which would be the 13th variable in Figure 40) can increase the adjusted-R^2. This means that inclusion of this can help at improving the accuracy of the developed regression model. Nevertheless, in order to simplify the data analysis, reduce complexity of the developed models the *Cruise.Control* was excluded from the development of regression models.

Due to the discrepancies of results between analysis of the *p-values* and adjusted-R^2, a Lasso regression was performed on the variables to confirm that the correlation between these and fuel consumption is actually significant for the considered case study.

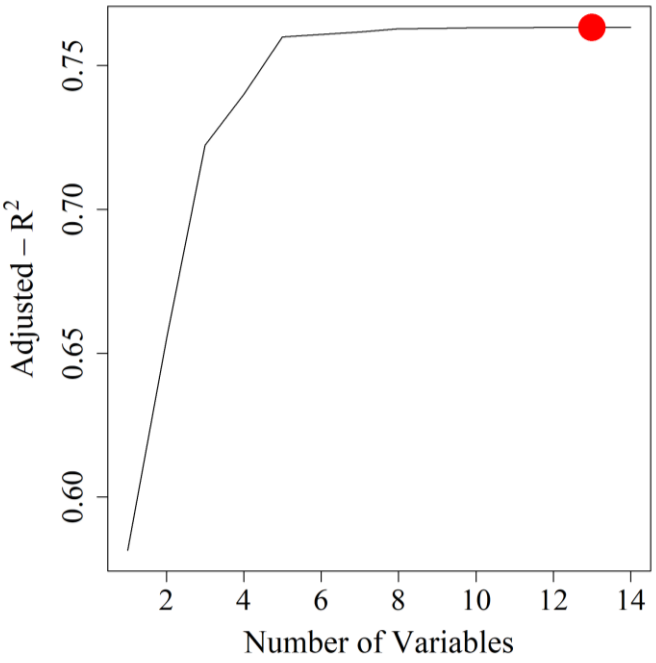


Figure 40. Analysis of the adjusted-R^2 for the Dataset 1.

Table 31 reports the regression coefficients estimated by the Lasso for each of the variables, which later will be included in the developed regression models. From an analysis of these (see Table 31), since none of the coefficients estimated by the Lasso are completely reduce to zero, it is possible to conclude that all the considered variables should be included in the developed regression models. This should help the developed models in explaining a larger portion of variance and improve their accuracy.

Finally, the RF algorithm was used for a final test and to rank the variables by importance. As an increase in MSE (defined in Equation (19)) makes the developed models able to make more accurate estimates (Breiman, 2001; Liaw *et al.*, 2017), it is possible to see (in Figure 41) that all the considered variables contribute in improving the MSE value of the resulting regression model. It is also important to notice that, for this case study, the algorithm ranks measurements of macro-texture and roughness of the road surface to be more important than other variables, such as vehicle speed

(*Speed.AVG*) and revolutions of the engine (*Revs.Start*). Although, other studies showed these to be quite important to determine the fuel consumption of road vehicles accurately (e.g. Ben-Chaim *et al.*, 2013; Benbow *et al.*, 2013; Werner *et al.*, 2013; Guzzella *et al.*, 2015; Zeng *et al.*, 2015; Kan *et al.*, 2018), this might not be important for the present study due to the restricted range of vehicle speeds considered and, therefore, this might be specific of the considered case study, which correspond to heavy trucks driving at high speed on motorways. Although the fact that road surface characteristic impact the fuel consumption of trucks at high speed matches the hypothesis made by Sandberg (1990), that does not agree with the results recently obtained by Chatti *et al.* (2012). For this reason, before drawing conclusions, further investigation is required to validate this result.

Table 31 - Summary of the Lasso coefficients computed for the selected variables, for the Dataset 1.

Variable	Lasso Coefficient
Gross.Weight_kg	0.149
Geom.Gradient	0.525
Speed.AVG	-0.0319
Torque.Start	0.245
Torque.End	0.269
Revs.Start	0.0156
Acceleration	-0.143
Geom.Abs_Radius	0.0091
TRACS.LPV03m	-0.0001
TRACS.LPV10m	-0.0170
TRACS.LPV30m	0.0209
TRACS.Pav_Texture	0.0025

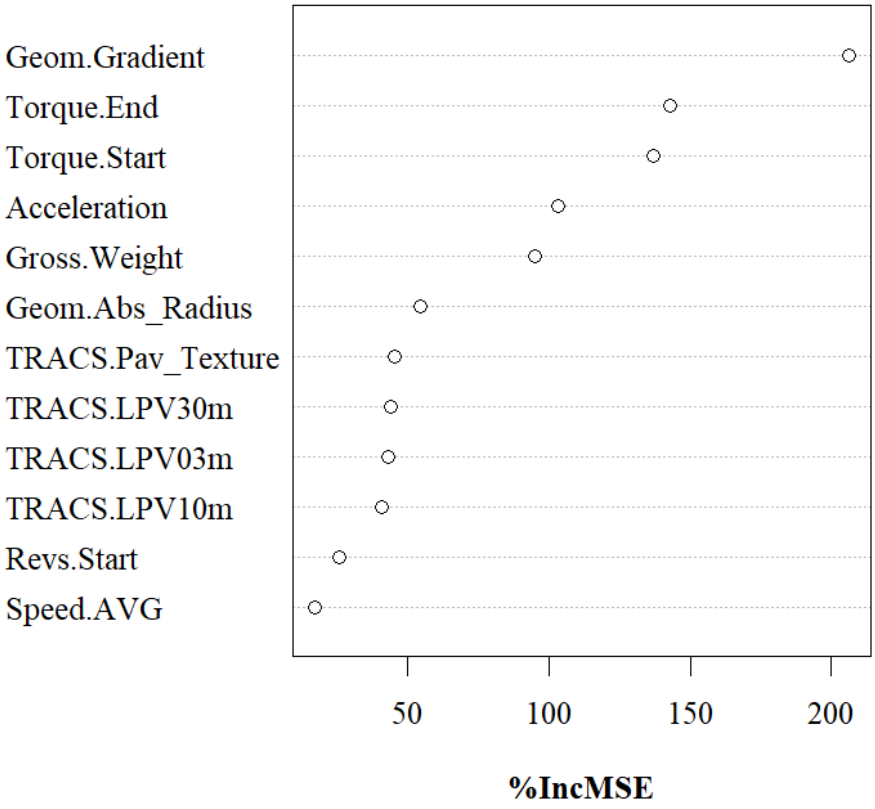


Figure 41. Rank of variable importance as increase in MSE on the Dataset 1, as computed by the RF algorithm.

Appendix E.1.2 Modelling 1

From an LR of the selected variables, the generated model takes the form:

$$\begin{aligned} Fuel.Consumption &= 16.01 + 1.6\times10^{-4}\times Gross.Weight + 6.06\times Geom.Gradient - 0.14\times Speed.AVG + \\ &+0.14\times\ Torque.Start + 0.16\times Torque.End + 2.6\times10^{-3}\times Revs.Start + 34.09\times Acceleration + \\ &+8\times10^{-4}\times Geom.Abs_Radius - 0.13\times TRACS.LPV03m - 0.69\times TRACS.LPV10m + \\ &+0.11\times TRACS.LPV30m + 0.52\times TRACS.Pav_Texture \end{aligned} \quad (30)$$

In comparison to SVR, RF and ANN (used in the following), the LR represents a very interpretable model, which allows users to perform calculation of *Fuel.Consumption* and to evaluate the impact of each of the considered variables from Equation (30) directly. This represents a significant advantage and, together with the lower amount of data available from experiments, it represents the reason why studies

conducted in the past on the impact of the road infrastructure on vehicle fuel consumption (e.g. Carlsson, 1986; Kerali *et al.*, 2006; Chatti *et al.*, 2012; Benbow *et al.*, 2013) preferred this type of technique to more advanced methods.

Using the Equation (30) to calculate the impact of road surface characteristics on vehicle fuel consumption, it is possible to see that, because positive and negative signs are assigned to the coefficients of LPV 3, 10 and 30, their total impact might be close to zero, which, considering that in the UK IRI is correlated to LPV (see Equation (1) at sect. 2.1.1), that would match the results obtained by Chatti *et al.* (2012). However, Equation (30) also highlights a possible significant effect of macro-texture (measured as SMTD) on the fuel consumption of heavy trucks driven at high speed in the UK, which (measured as MPD) was not considered to be significant by Chatti *et al.* (2012), in their study conducted in the US.

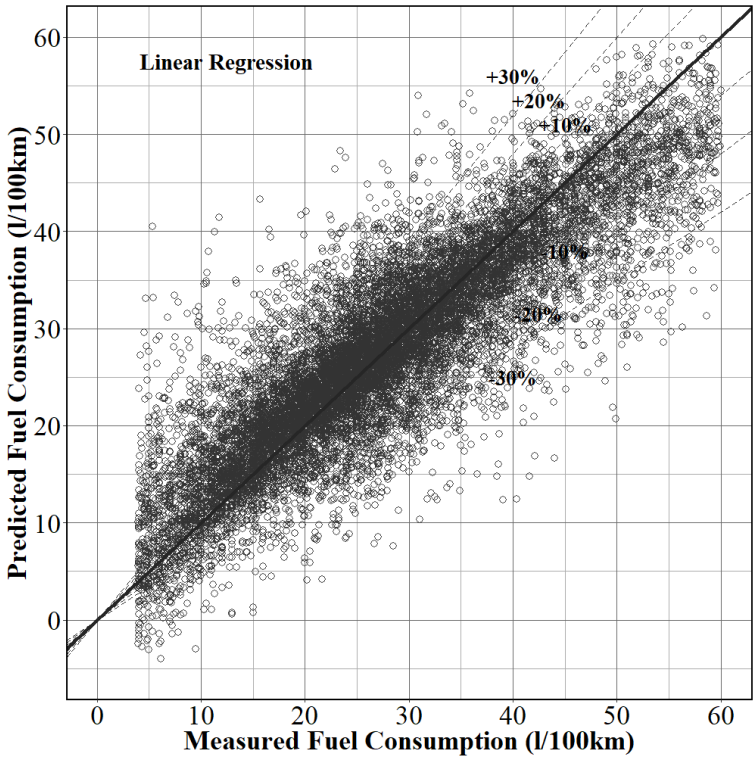


Figure 42. Fit of the developed LR to the Dataset 1.

Figure 42 shows the fit of the developed LR to the Dataset 1. The figure shows that the LR detects a trend in the data, but it is not able to fit extreme values of fuel consumption very well. In particular, while the model tends to fit well the values of fuel consumption towards the median, it underestimates the high fuel consumption and overestimate the low values. This is likely due also to the effect of non-linearity.

Regarding the ML methods, it is difficult to make a pictographic representation of the developed models. For this reason and for clarity, only a brief description of the structure of each of the developed models is given in the following. Instead, a general representation of how each of the utilised models is usually structured can be found at section 4.5.1 (for the SVR), 4.5.2 (for the RF), and 4.5.3 (for the ANN).

For this initial case study, the SVR model was a regression model, with RBF kernel function and tolerance set to 0.001. Figure 43 shows the fit of the SVR model on the test set over the cross-validation process. Although Figure 43 already shows how performance of the SVR model result quite stable over the ten-fold cross-validation process, these are also reported in Table 32 for reference.

Table 32 - Summary of the performance of the SVR model over the ten-fold cross-validation process on the Dataset 1.

SVR	1	2	3	4	5	6	7	8	9	10	AVG
R^2	0.85	0.84	0.85	0.83	0.84	0.84	0.84	0.84	0.84	0.84	0.84
RMSE (l/100 km)	4.90	4.88	4.86	5.00	4.82	4.88	4.99	4.95	4.87	5.06	4.92
MAE (l/100 km)	3.40	3.35	3.38	3.44	3.36	3.40	3.41	3.41	3.36	3.44	3.39
Time	0.22	0.24	0.23	0.23	0.23	0.23	0.23	0.23	0.24	0.23	0.23

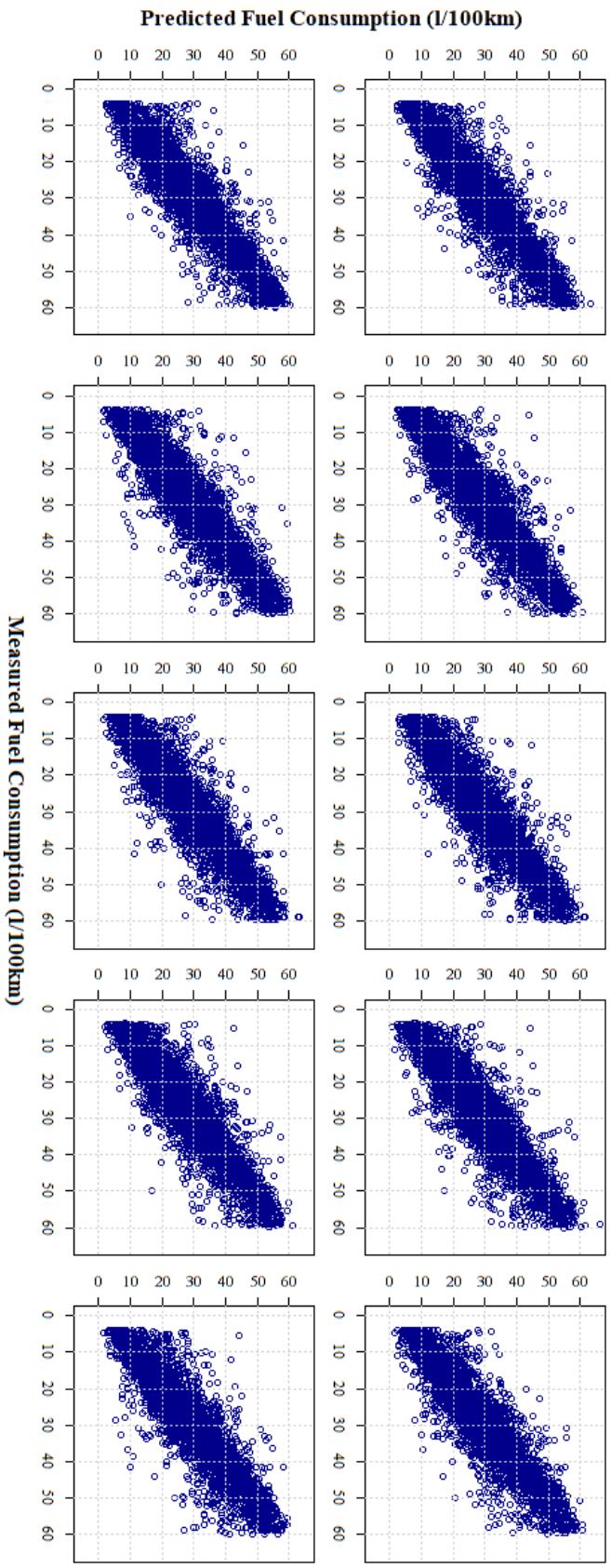


Figure 43. Ten-fold cross-validation of the SVR model for the Dataset 1.

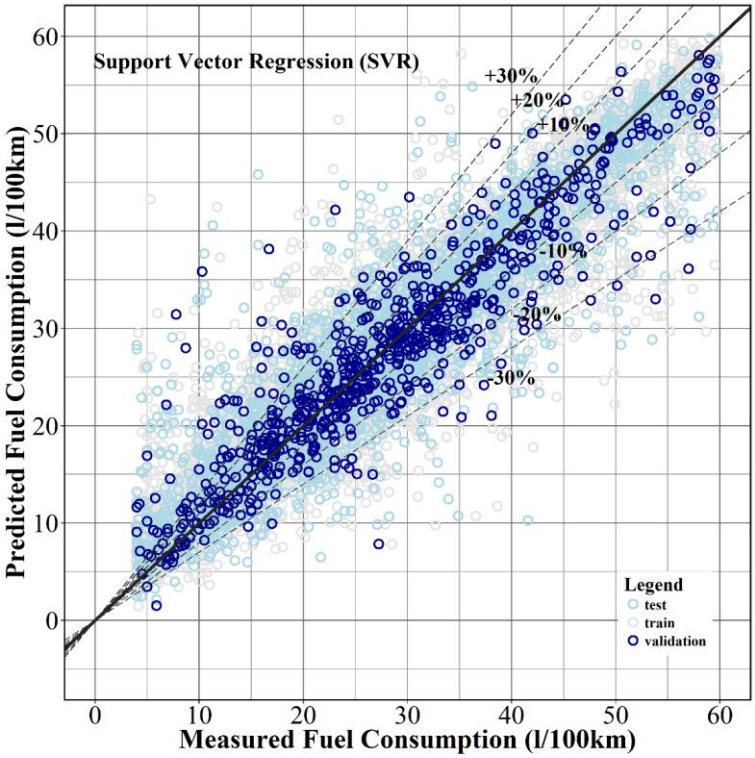


Figure 44. Fit of the developed SVR model to the Dataset 1.

Figure 44 shows the fit of the SVR model to the Dataset 1. The figure shows already an improvement in comparison to the LR model but, also in this case, the model seems to be prone in overestimating low fuel consumption and underestimating high fuel consumption.

Table 33 - Summary of the performance of the RF model over the ten-fold cross-validation process on the Dataset 1.

RF	1	2	3	4	5	6	7	8	9	10	AVG
R^2	0.86	0.86	0.87	0.85	0.86	0.86	0.85	0.86	0.86	0.86	0.86
RMSE (l/100 km)	4.70	4.70	4.57	4.76	4.57	4.62	4.73	4.75	4.59	4.78	4.68
MAE (l/100 km)	3.25	3.24	3.22	3.30	3.22	3.25	3.25	3.29	3.21	3.28	3.25
Time	1.31	1.34	1.23	1.21	1.22	1.23	1.22	1.22	1.24	1.25	1.25

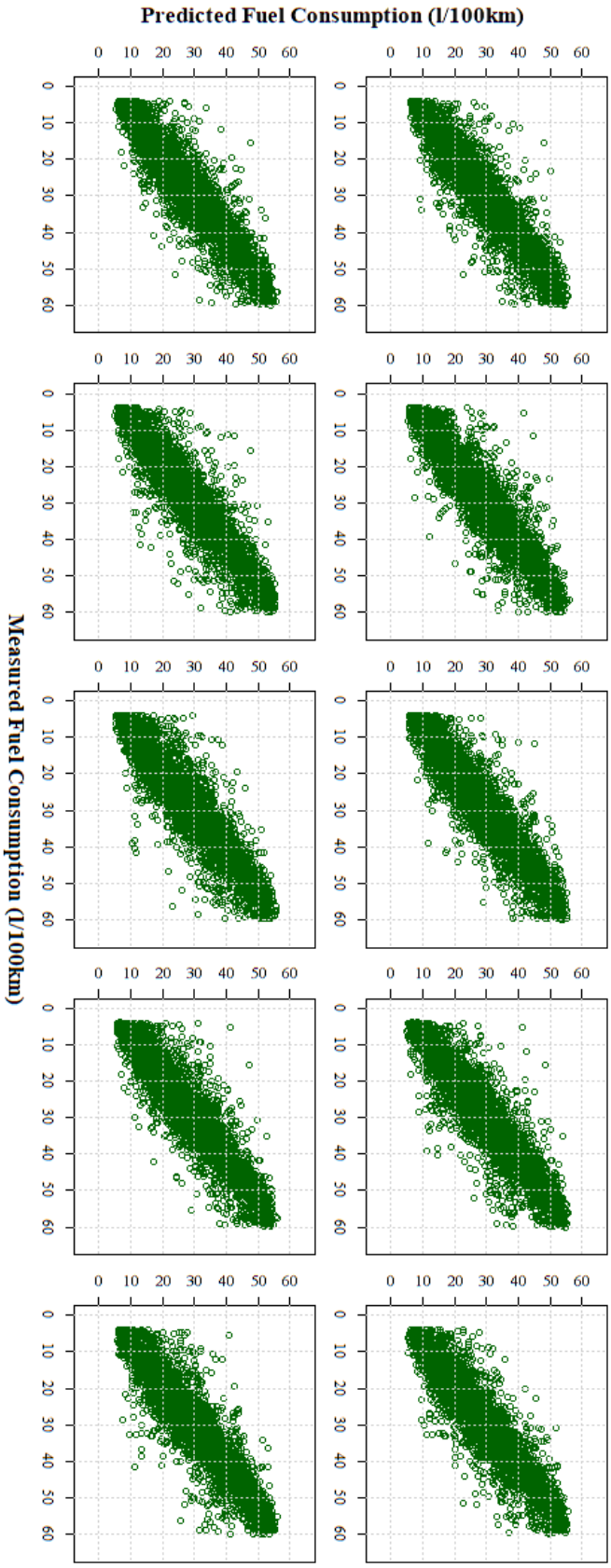


Figure 45. Ten-fold cross-validation of the RF model for the Dataset 1.

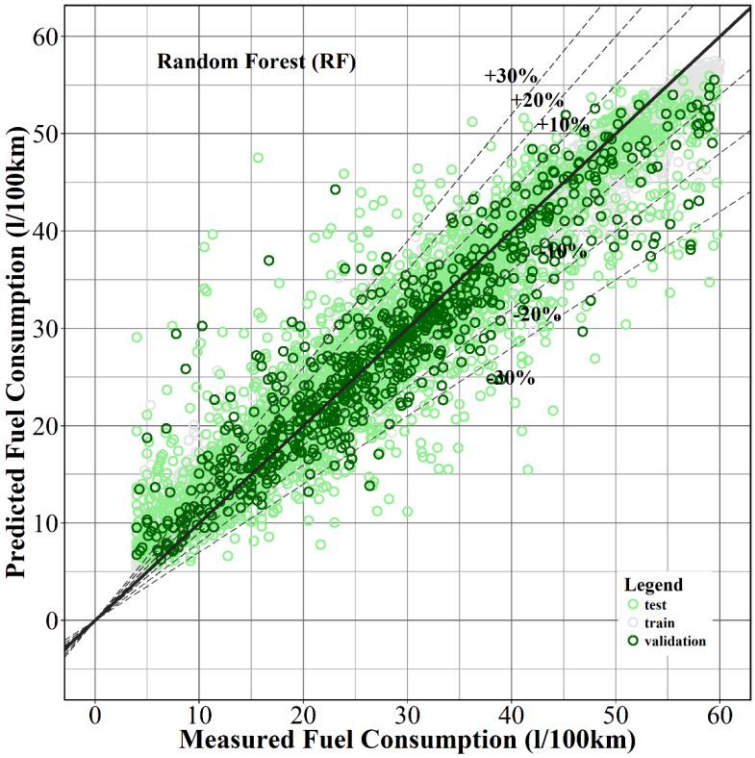


Figure 46. Fit of the developed RF model to the Dataset 1.

The developed RF is composed of 500 trees and in each of these it uses 4 randomly sampled variables to make predictions. Figure 45 shows the fit of the RF model on the test set over the cross-validation process. Although the figure (Figure 45) already shows how performance of the RF model result quite stable over the ten-fold cross-validation process, these are also reported in Table 33 for reference.

Figure 46 shows instead the fit of the developed RF model to the Dataset 1. The figure shows an improvement in comparison to both the SVR and LR models but, also in this case, the model seems to be prone in overestimating low fuel consumption and underestimating high fuel consumption.

Finally, the developed ANN and its performance are considered. The ANN is structured on two layers of 10 and 3 neurons respectively. The model was trained using the rprop- algorithm (Riedmiller *et al.*, 1993), and it used the logistic activation function to identify patterns through the data. Figure 47 shows the fit of the ANN

model on the test set over the cross-validation process. Although the figure (Figure 47) already shows how performance of the ANN model result quite stable over the ten-fold cross-validation process, these are also reported in Table 34 for reference.

Figure 48 shows instead the fit of the developed ANN model to the Dataset 1. The figure shows performance similar to those of the developed SVR and, again, the ANN seems to be prone at overestimating low fuel consumption and underestimating high fuel consumption.

Interestingly, from a comparison of the three ML models and their fits to the Dataset 1 (see Figure 44, 46, and 48) there are some data points that neither SVR, RF or ANN, can fit properly. This could be due to the reduce size of the considered dataset (and thus lack of examples for the algorithms in the training set) or, alternatively, this may be due to faulty sensors, which reporting erroneous data introduced bias in the data. Another possibility is that the unfitted datapoints might be depending on variables that have not been considered in the present study, such as the presence of a vehicle in front of the considered one, the tire pressure or the pavement stiffness/deflection.

Table 34 - Summary of the performance of the ANN model over the ten-fold cross-validation process on the Dataset 1.

ANN	1	2	3	4	5	6	7	8	9	10	AVG
R^2	0.80	0.77	0.81	0.79	0.81	0.77	0.81	0.82	0.82	0.79	0.80
RMSE (l/100 km)	5.53	5.91	5.31	5.60	5.30	5.90	5.39	5.36	5.24	5.85	5.54
MAE (l/100 km)	3.92	4.18	3.81	3.98	3.82	4.26	3.79	3.85	3.75	4.10	3.95
Time	3.53	2.19	0.60	1.26	0.90	1.02	1.67	1.25	2.38	1.57	1.64

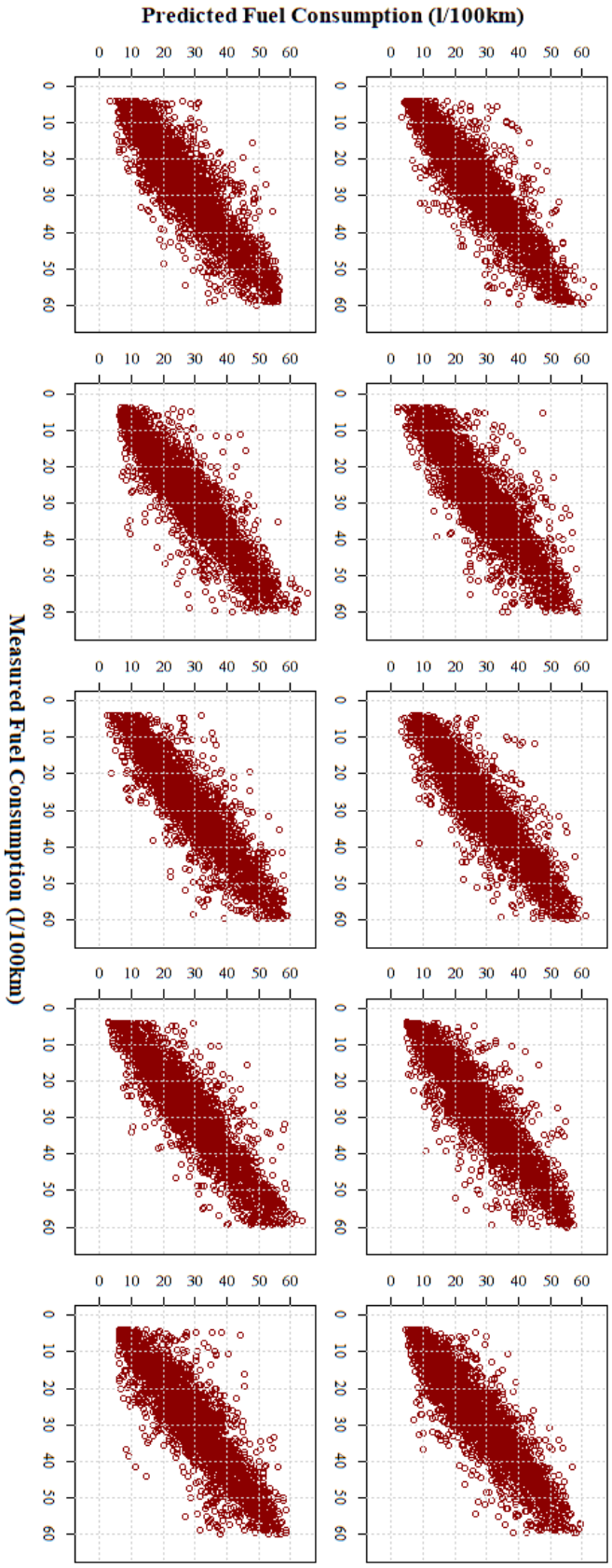


Figure 47. Ten-fold cross-validation of the ANN model for the Dataset 1.

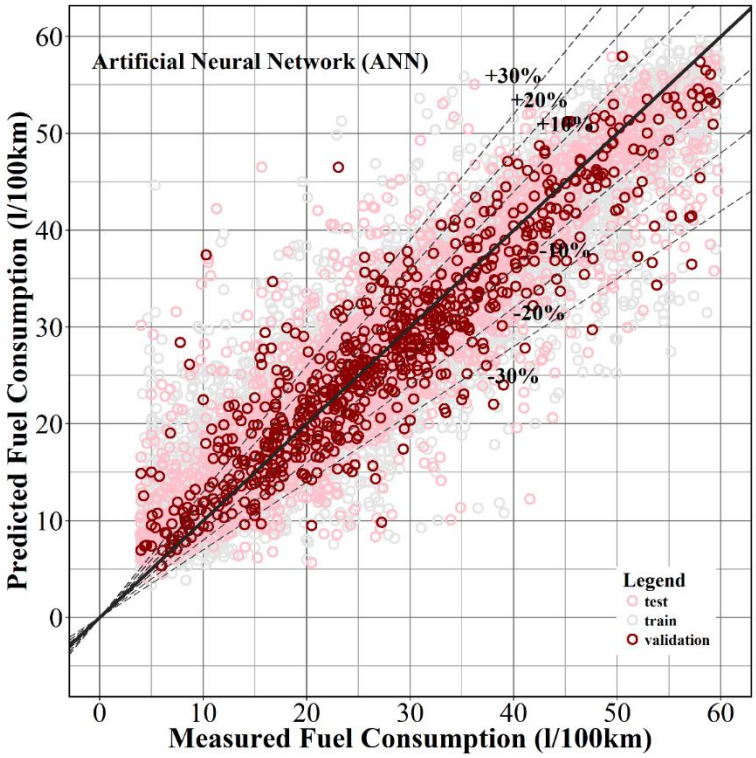


Figure 48. Fit of the developed ANN model to the Dataset 1.

As specified in the chapter dedicate to the Methodology of the thesis (see sect. 4.6), performance of the developed models, are compared in terms of R^2, RMSE and MAE for the cross-validation and validation sets. Table 35 summarises results of this investigation.

Table 35 - Summary of performance of the regression models developed using the Dataset 1.

Model	R^2_{cv}	$RMSE_{cv}$	MAE_{cv}	$Time_{cv}$	R^2_{val}	$RMSE_{val}$	MAE_{val}
	-	*(l/100 km)*	*(l/100 km)*	*(mins)*	*(-)*	*(l/100 km)*	*(l/100 km)*
LR	0.76	6.04	4.44	-	-	-	-
SVR	0.84	4.92	3.39	2.31	0.84	4.99	3.43
RF	0.86	4.68	3.25	12.49	0.86	4.70	3.26
ANN	0.80	5.11	3.50	16.39	0.83	5.15	3.68
N.B.: The subscript 'val' stays for validation set, and 'cv' for cross-validation.							

Results in Table 35 show that, although the RF model takes a bit longer to cross-validate than the SVR, the model achieves the highest level of accuracy, in terms of R^2, RMSE and MAE at estimating the overall fuel consumption of the considered fleet of trucks, both during the cross-validation process that on the validation set.

Since, from the analysis of Figure 43, 45, and 47, together with that of Table 32, 33, and 34, all algorithms used in the study demonstrate to be quite robust and reliable showing quite constant performance all throughout the cross-validation process, also for the reduced amount of records available in the Dataset 1. For brevity, only summary of the average performance of the developed models (e.g. Table 35), and comparison of this to performance of the models on the validation set, is reported for all the applications of the algorithms on the Dataset 2.

Learning curves were calculated for the ANN. This was to understand if the amount of data available in the Dataset 1 is actually enough or more information is required to develop robust and reliable models.

From an analysis of the learning curves showed in Figure 49 and 50 it is possible to see two things. Firstly, it seems like the MAE tend to converge quicker than the RMSE. Therefore, in order to estimate the minimum amount of data necessary to fit robust ML fuel consumption models, preferably the RMSE learning curve should be analysed, to ensure the error to converge. Secondly, it can be seen that both RMSE and MAE seem close but still not close enough to convergence.

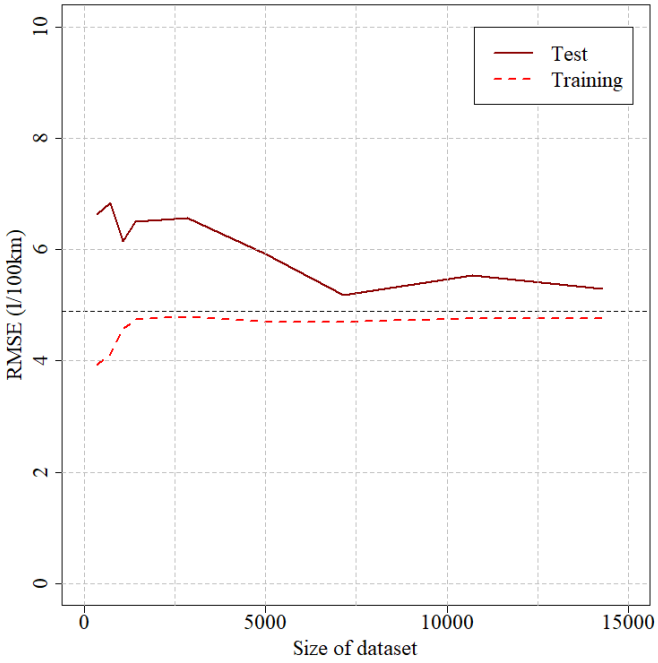


Figure 49 – RMSE learning curve for the ANN fitted on the Dataset 1.

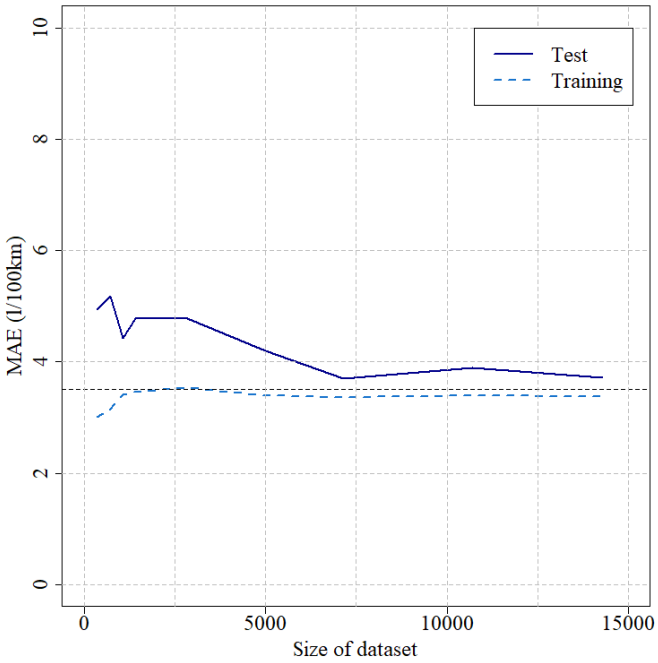


Figure 50. MAE learning curves for the ANN fitted to the Dataset 1.

This might mean that double or triple of the size of Dataset 1 should allow SVR, RF and ANN to strengthen reliability of their estimates. However, when the focus of the analysis will move to the Dataset 2, which includes data from a larger

road network, a wider range of vehicle speeds, worse road conditions, and generally more variables, such as meteorological conditions, variance might increase and the size of the dataset needed with it.

Finally, a larger neural network ANN1 was fitted to the data and its performance compared to those of the ANN. The ANN1 uses the same learning algorithm used by the ANN (rprop-), and same activation function (logistic), but it adopts three layers of 22, 30 and 22 neurons respectively. This should help the ANN1 to detect more complex patterns through the data, but it may also cause overfitting (Goodfellow *et al.*, 2016; Ng, 2018).

Comparison of performance between the ANN1 and ANN are reported in Table 36. In this, it is possible to see that ANN1 tends to overfit the training set, showing a significant reduction of R^2 and increment of both RMSE and MAE on the validation set (see Table 36). On the other hand, the ANN shows more stable performance, comparable in both the training and validation set, and requires a lower training time, about four times lower, than the ANN1 (see Table 36). For all these reasons, and to enhance interpretability of the developed model, the ANN was preferred to the ANN1 in this part of the study.

Table 36 - Comparison of performance between the small (ANN) and larger (ANN1) neural networks fitted to the Dataset 1.

Model	R^2_{train}	$RMSE_{train}$	MAE_{train}	$Time_{train}$	R^2_{val}	$RMSE_{val}$	MAE_{val}
	-	*(l/100 km)*	*(l/100 km)*	*(mins)*	-	*(l/100 km)*	*(l/100 km)*
ANN	0.84	5.08	3.47	1.59	0.83	5.15	3.68
ANN1	0.89	4.96	3.37	6.12	0.77	5.43	3.88
N.B.: The subscript 'train' stays for training set (part of cross-validation), and 'val' for validation.							

Appendix E.1.3 Parametric Analysis 1

Finally, a parametric analysis was performed to see how each of the developed models addresses the impacts of each of the considered variables on the fuel consumption of the fleet of trucks monitored in the Dataset 1.

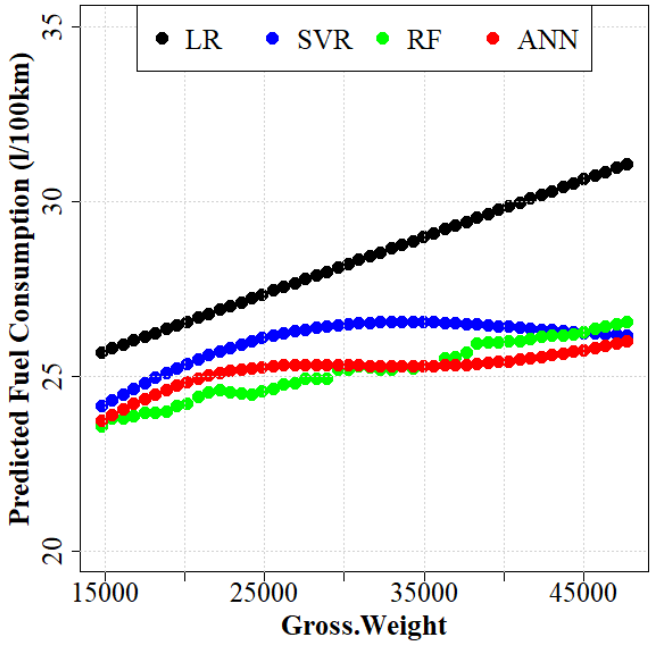


Figure 51. Parametric analysis of *Gross.Weight_kg* (kg) for the Dataset 1.

In Figure 51 variations of fuel consumption in consideration of changes in the *Gross.Weight_kg* are considered. The figure shows that all models detects an increase of fuel consumption with the *Gross.Weight_kg*. This is reasonable, but it is possible to see that while SVR, RF and ANN seem to agree on the detected trends, the LR identifies a slightly different behaviour.

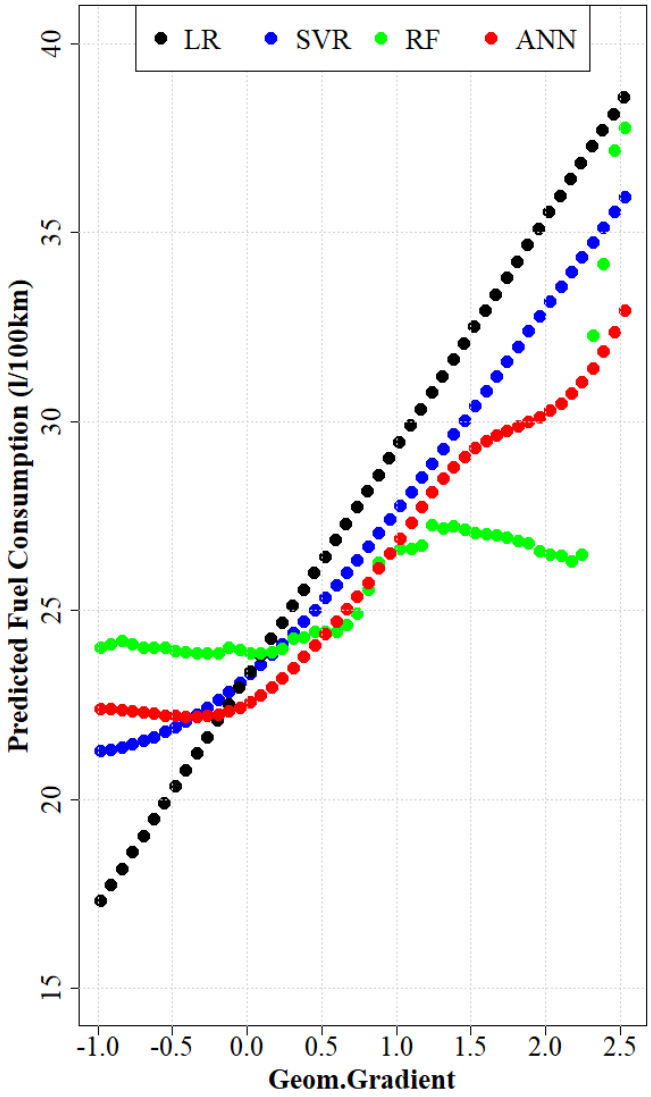


Figure 52. Parametric analysis of *Geom.Gradient* (%) for the Dataset 1.

Figure 52 shows instead the trends of fuel consumption estimated by the developed models in consideration of variation of *Geom.Gradient* (%). The figure shows that all models identify an increase of fuel consumption due to steeper inclinations of the road pavement. However, again, while the trends identified by SVR, RF and ANN are to a certain extent similar, some differences can be noticed with the developed LR model. For instance, it can be noticed that while SVR, RF and ANN identify an asymptotic behaviour in correspondence to the case of flat or downhill roads, that is not the case for the LR model. Because it is reasonable to expect that on

a flat road or going downhill a vehicle might consume some fuel (due to idle at least), it is possible to say that the trend identified by SVR, RF and ANN might be closer to reality than that identified by the LR. Additionally, it is possible to notice that the RF model tends to produce unreliable estimates for very steep roads. This might be due to some overfitting or to a lack of data for this kind of situations.

Figure 53 and 54 show variation of fuel consumption due to the percent of torque used by the engine at the beginning and end of the considered journey. As the two considered variables represent two very similar measurements, highly linearly correlated one each other (see Figure 39), all models tend to estimate the effect of each of them similarly in both plots.

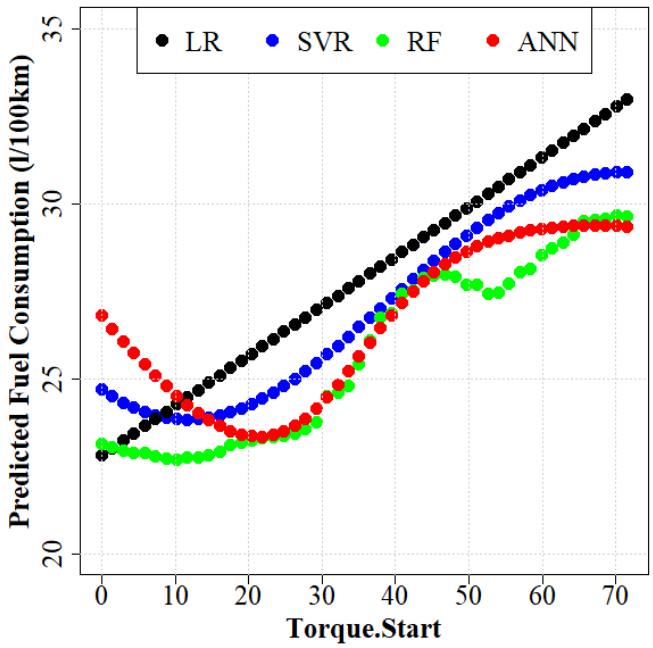


Figure 53. Parametric analysis of *Torque.Start* (%) for the Dataset 1.

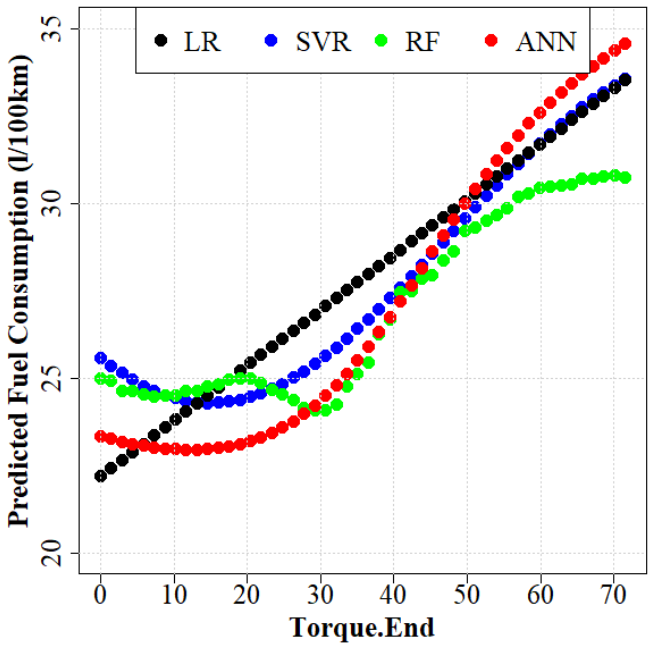


Figure 54. Parametric analysis of *Torque.End* (%) for the Dataset 1.

However, it is also possible to notice that while SVR, RF and ANN all detect a sort of sigmoidal trend across the data; the LR approximates this linearly. This, in consideration of the results obtained in the parametric analysis of *Gross.Weight_kg* (see Figure 51) and *Geom.Gradient* (see Figure 52), might highlight a remarkable non-linear effect of these variables, which cannot be neglected, and that cannot be considered using a simple LR model.

Figure 55 reports changes in fuel consumption due to the different use of engine revolutions (*Revs.Start*). From the figure, it is possible to see that all models detect an increase in fuel consumption for higher *Revs.Start*. However, it is possible to see also that, again, while SVR, RF and ANN tend to agree on the identified pattern, this is different for the LR that detects a much smaller increase in fuel consumption in comparison to the others.

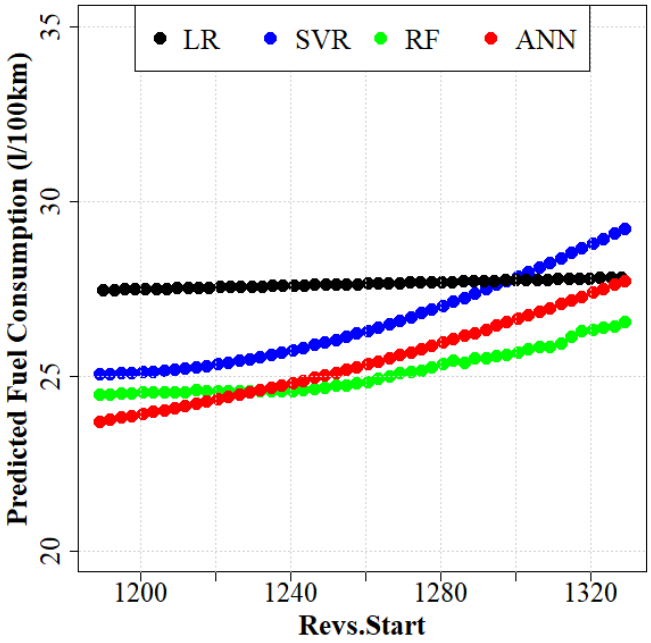


Figure 55. Parametric analysis of *Revs.Start* (rpm) for the Dataset 1.

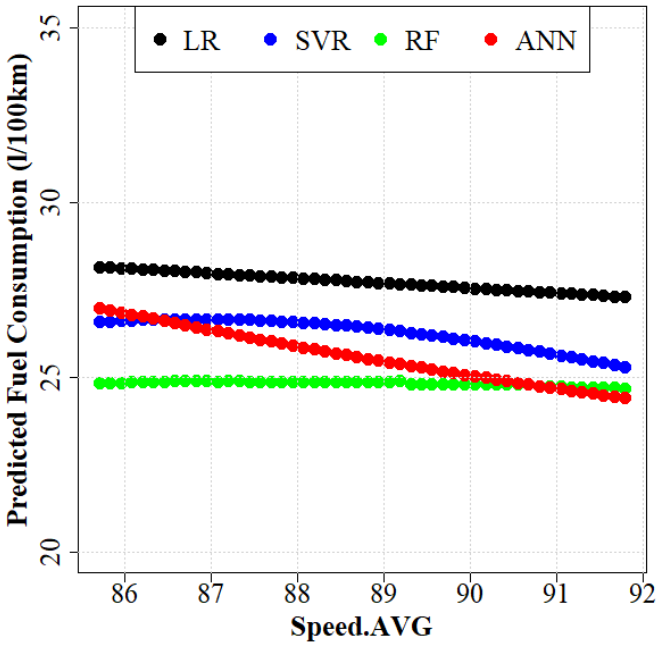


Figure 56. Parametric analysis of *Speed.AVG* (km/h) for the Dataset 1.

In Figure 56, estimates of fuel consumption for different vehicle speed (Speed.AVG) is reported. In this case, it is possible to see that generally, LR, SVR and

ANN tend to agree on the fact that higher fuel consumption is produced for lower speed. This is not the case for the RF, which shows almost constant fuel consumption and no significant difference for the considered range of speeds. In fact, it is reasonable to assume that, for the considered restricted range of speeds and, in consideration of the fact that all considered journeys are performed at relatively constant speed, close to 90 km/h (see Figure 14), no significant difference or a small increase in fuel consumption should be detected.

This is the reason why the RF algorithm ranks the *Speed.AVG* amongst the variables that influence the fuel consumption of the considered fleet of trucks the least. Therefore, before drawing conclusions, quantity and distribution of the considered data might be the reason why this is not the case, and further investigation, in consideration of a wider range of speeds, should be performed. In fact, as anticipated in the chapter dedicated to the description of the data available (see sect. 3.4), the Dataset 2 was collected to be larger than the Dataset 1 and to include a wider range of vehicle speeds.

Figure 57 shows variation of fuel consumption due to acceleration of the considered truck (*Acceleration*). Considering Figure 56 and 57 together, it is possible to see that all the developed models associate lower fuel consumption with higher speed and acceleration. Although at first, these findings may not be intuitive, it must be considered that the ranges of speed and acceleration considered in the study are very narrow and that the effect of the road gradient may be prevailing.

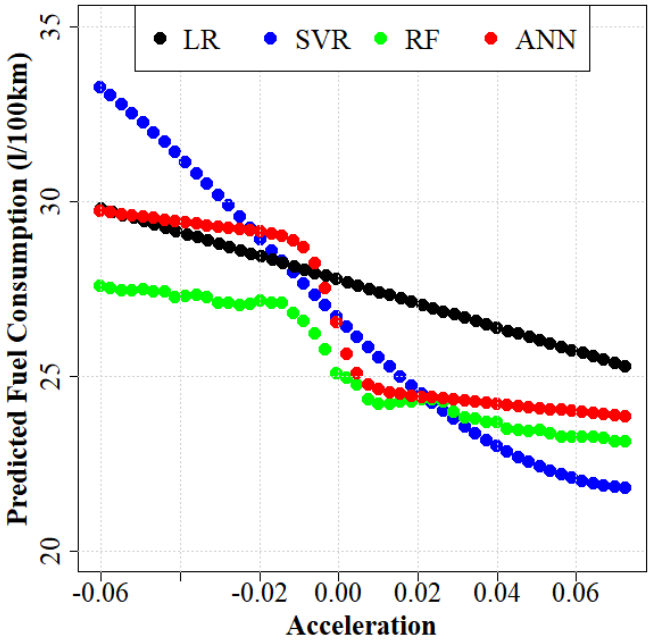


Figure 57. Parametric analysis of *Acceleration* (m/s^2) for the Dataset 1.

In particular, considering a truck travelling at relatively constant speed, a downhill road segment (characterised by negative gradient) would lead it to accelerate and produce yet lower fuel consumption. Instead, when the truck drives uphill (or road segments characterised by positive gradient), that would make the truck to decelerate and cause higher fuel consumption. This might also explain the discontinuity detected by the RF and the ANN between negative and positive accelerations (see Figure 57), as the model tries to explain two separate phenomena; however, in order to confirm this, further investigation on a larger dataset might help.

In Figure 58, variation of fuel consumption in consideration of changes of the road geometry is considered. In particular, the figure focuses on the radius of curvature of the road (*Geom.Abs_Radius*) and shows how the fuel consumption of the considered fleet of trucks changes with variations of this. As explained earlier on in this chapter (see sect.), a radius of 2,000 m is assigned by the road agency software to roads that have a radius of over 2,000 m as considered fairly straight road segments.

Bearing this in mind, in Figure 58 it is possible to see that all models seem to detect larger fuel consumption for straight roads than others.

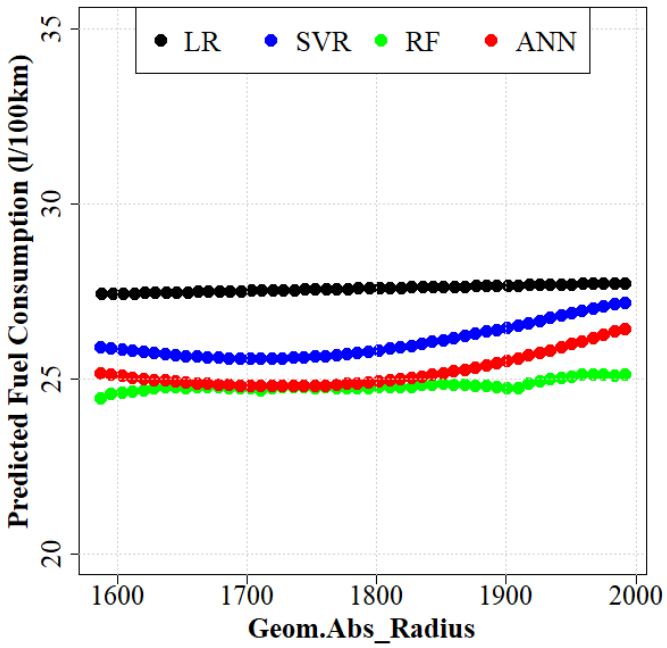


Figure 58. Parametric analysis of *Geom.Abs_Radius* (m) for the Dataset 1.

In particular, SVR, RF, and ANN show very similar trends, while the LR probably tends to overfit the data available for fairly straight roads, which on motorways represents the vast majority of cases (see Figure 14).

Finally, in the following figures (Figure 59, 60, 61, and 62), variation of the fuel consumption of the considered fleet of trucks due to different characteristics of the road surface is considered.

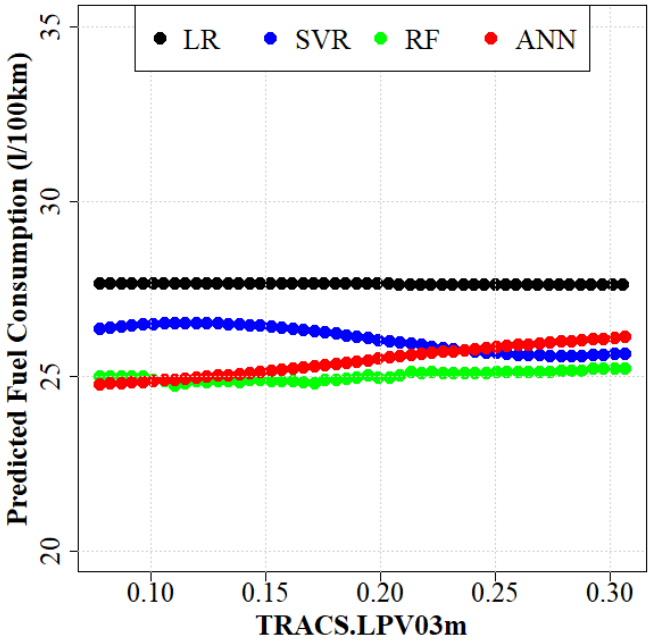


Figure 59. Parametric analysis of *TRACS.LPV03m* (mm^2) for the Dataset 1.

For instance, in Figure 59 variation of fuel consumption due to changes of LPV at 3 m wavelength (*TRACS.LPV03m*) is reported. The first thing to notice from the graph is that a very restricted range of *TRACS.LPV03m*, which goes from just above zero to approximately 0.3 mm^2 is considered. In this range, it is possible to see that RF and ANN detect both similar trends, which indicate an increase of fuel consumption for larger *TRACS.LPV03m* considered. In regard to the SVR, this seems to disagree with the RF and ANN, showing a reduction of fuel consumption for larger *TRACS.LPV03m* considered. However, it is possible to see that from a certain level of *TRACS.LPV03m* and above the trend shown by the SVR tend to converge to that shown by the RF and ANN. This might indicate an overfitting of the data by the SVR for *TRACS.LPV03m* lower than 0.20 mm^2, which values might be quite rare in practice (see Figure 14). The LR model shows no significant impact of *TRACS.LPV03m* on fuel consumption for the considered case study.

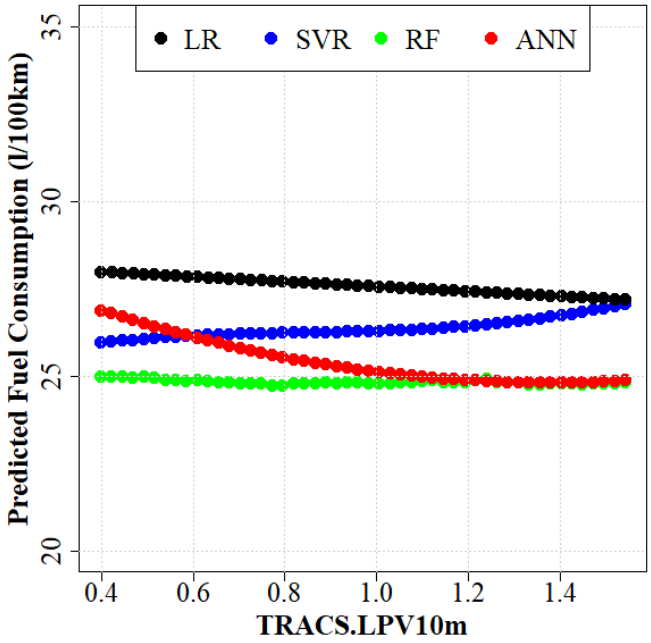


Figure 60. Parametric analysis of *TRACS.LPV10m* (mm^2) for the Dataset 1.

Figure 60 reports variation of fuel consumption due to LPV at 10 m wavelength (*TRACS.LPV10m*) as estimated by each of the developed models. The figure shows that none of the developed models substantially agree on how fuel consumption varies due to changes in *TRACS.LPV10m.* This might be due to the restricted range of conditions considered in the study or because of the small set of data analysed, which probably produces an overall overfitting. Therefore, no actual conclusion can be done from looking at Figure 60, and further investigation on a larger dataset might be required before drawing conclusions.

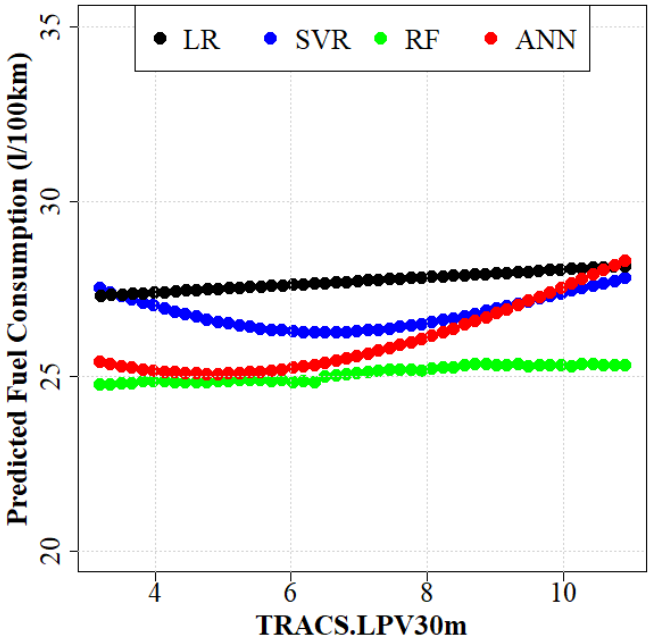


Figure 61. Parametric analysis of *TRACS.LPV30m* (mm^2) for the Dataset 1.

Figure 61 shows how fuel consumption is estimated to vary by the developed models for changes in LPV at 30 m wavelength (*TRACS.LPV30m*). Although the restricted range of considered conditions, from Figure 61 it is possible to see that all models seem to agree on the fact that higher fuel consumption can be associated to higher *TRACS.LPV30m.* Although consistency of these results increases confidence in the identified trend, further investigation might be required before conclusions can be made. In particular, consideration of the Dataset 2, larger and inclusive of a wider range of road surface conditions than the Dataset 1, could help in confirming or rejecting these results and lead to more robust conclusions.

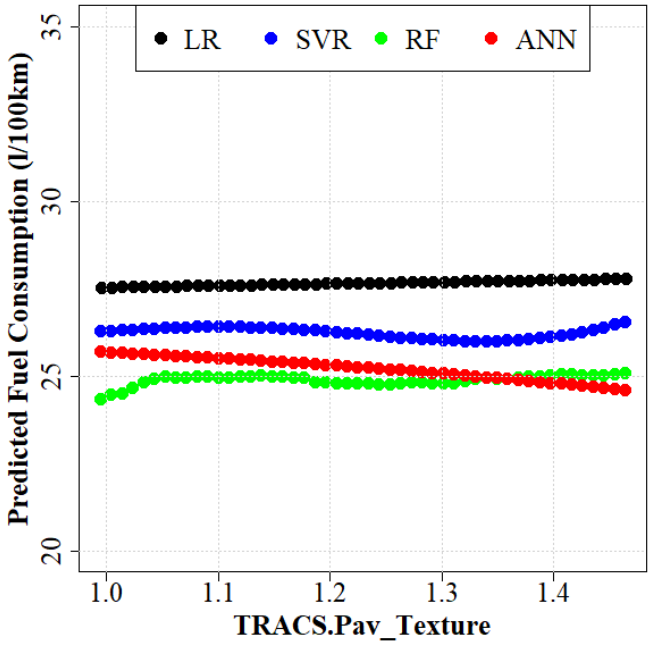


Figure 62. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the Dataset 1.

Finally, Figure 62 shows how each of the developed models estimates variations of fuel consumption for different levels of road surface macro-texture (*TRACS.Pav_Texture*). Again, it must be considered that the range of *TRACS.Pav_Texture* considered is quite narrow (Viner *et al.*, 2006) (see Figure 14) and, for the quantity of data available, it might be difficult to establish variations of fuel consumption within this range. Bearing this in mind, Figure 62 shows that, the LR model detects an increase of fuel consumption with larger *TRACS.Pav_Texture* within the considered range. This is the case also for the RF, but opposite behaviour is shown by the ANN. By looking at the figure, nothing is possible to say about the SVR. Therefore, as in the case of the *TRACS.LPV10m* no actual conclusions can be made in regard to the *TRACS.Pav_Texture* which seems to produce only overfitting in the developed models.

Finally, results of the parametric analysis are compared from a numerical perspective within the 25th and 75th percentiles and within the 5th and 95th percentiles of the considered distributions (5th – 95th percentiles of the observed distributions). This is to compare results of the developed models from a numerical perspective, and it might help to interpret the figures listed and described earlier on in the present section of the thesis. In particular, while Table 37, reports variations of fuel consumption estimated by each of the developed models, within the 25th and 75th percentiles, Table 38 reports variations of fuel consumption between the 5th and 95th percentiles, in percentage to the average fuel consumption of the considered fleet of trucks (which for the Dataset 1 is 27.65 l/100 km).

Table 37 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the Dataset 1.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	2.74	9.93	0.74	2.69	1.38	4.98	0.29	1.05
Geom.Gradient	10.85	39.25	8.18	29.58	3.16	11.42	7.22	26.12
Speed.AVG	-0.44	-1.59	-0.68	-2.46	-0.07	-0.25	-1.33	-4.81
Torque.Start	5.18	18.74	5.57	20.15	4.49	16.26	5.26	19.01
Torque.End	5.78	20.91	5.90	21.35	4.69	16.95	7.83	28.33
Revs.Start	0.18	0.65	2.14	7.75	0.94	3.40	2.01	7.26
Acceleration	-2.30	-8.32	-7.01	-25.34	-3.43	-12.41	-5.04	-18.23
Geom.Abs_Radius	0.16	0.58	0.74	2.69	0.02	0.09	0.55	2.00
TRACS.LPV03m	-0.02	-0.05	-0.82	-2.97	0.23	0.82	0.80	2.89
TRACS.LPV10m	-0.40	-1.45	0.28	1.03	0.10	0.34	-1.08	-3.92
TRACS.LPV30m	0.43	1.55	0.29	1.06	0.47	1.70	1.62	5.84
TRACS.Pav_ Texture	0.12	0.45	-0.41	-1.48	0.02	0.07	-0.55	-1.99

Table 38 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the Dataset 1.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	4.94	17.87	1.90	6.87	2.56	9.26	1.92	6.95
Geom.Gradient	19.53	70.65	13.42	48.54	8.18	29.58	9.01	32.59
Speed.AVG	-0.79	-2.85	-1.14	-4.11	-0.14	-0.51	-2.39	-8.65
Torque.Start	9.32	33.73	6.33	22.91	6.49	23.48	2.94	10.64
Torque.End	10.41	37.64	7.61	27.52	5.79	20.95	10.67	38.58
Revs.Start	0.33	1.18	3.72	13.47	1.84	6.64	3.63	13.13
Acceleration	-4.14	-14.98	-11.08	-40.08	-4.26	-15.39	-5.75	-20.79
Geom.Abs_Radius	0.29	1.05	1.12	4.07	0.57	2.06	1.04	3.77
TRACS.LPV03m	-0.03	-0.10	-0.80	-2.89	0.22	0.80	1.26	4.56
TRACS.LPV10m	-0.72	-2.62	0.90	3.27	-0.22	-0.78	-1.95	-7.05
TRACS.LPV30m	0.77	2.79	0.19	0.67	0.57	2.06	2.56	9.28
TRACS.Pav_ Texture	0.22	0.81	0.02	0.09	0.54	1.96	-0.99	-3.59

As specified in section 4.11, being the 25^{th} to 75^{th} percentiles a narrower range than 5^{th} to 95^{th} percentiles, it is expected that more robust conclusions can be made on the first than on the latter.

Appendix E.2 Analysis of the Dataset 2

The main objective of analysing the Dataset 2 was to apply the methodology developed in the first part of the study to a much larger dataset. This includes verification of the adopted data cleaning procedures, validation of the selection of variables identified to play a significant role in determining the fuel consumption of the considered fleet of trucks, and comparison of performance between models developed for different vehicle types. Again, a parametric analysis was performed on this second set of data (Dataset 2) to investigate how each of the considered variables

influence the fuel consumption of the considered truck fleets at network level. Because of the inclusion of weather condition data in the Dataset 2, also variation of fuel consumption due to meteorological parameters was investigated in this phase.

As already detailed in sect. 3.4.2, for this second case study, Microlise provided hundreds of thousand records from light, medium and heavy trucks (see sect. 3.4.2) driving at relatively constant speed, between January and March 2017, and between July and September 2017, along the roads of interest for the study. Of these, a selection of records from two segments of the M1, the whole M18, A21 and A50, and part of the A41, A47, A49, and A259 were considered in the study. This allows to consider a road network that, for condition of the road surface, can be considered representative of the whole SRN and primary roads at local level in England (see Annex B).

Because the Dataset 2 includes data from different roads, which are characterised by different traffic composition and speed limits (DfT, 2018c), two different applications of the developed methodology ('App 1', for A type roads only, and 'App 2', for M and A type roads together) are performed for each of the considered truck types (see sect. 3.4.2 for further details).

The following sections provide details about each phase of the analysis of the Dataset 2. Separate discussions are made for each vehicle type considered in the study, and for different applications of the developed method.

Appendix E.2.1 Variable Selection 2

In this section, results of the variable selection process are reported, and discussed per vehicle type and different sets of data, part of the Dataset 2.

E.2.1.1 Light trucks on A roads (App 1)

Initially, the linear correlation between variables for the light trucks App 1, was analysed. This was done by comparing the values of Pearson's correlation coefficient reported in the correlation matrix (see Figure 63). From an analysis of Figure 63, it is possible to see that the variables, which show the highest linear correlation with *Fuel.Consumption* are, in order, *Geom.Gradient*, *Gross.Weight_kg* and *Torque.AVG*, followed by *WindSpeedDir_kn*, *Revs.AVG* and *Acceleration*. This is reasonable and reflects both physics and vehicle dynamics principles (see sect. 2.4 and 2.5). The figure also shows some correlation of *Fuel.Consumption* with *TRACS.Pav_Texture* (SMTD). LPV, at any wavelength, seems instead to have very poor linear correlation with *Fuel.Consumption*. Additionally, all measurements of LPV seems to be mutually correlated one to each other. This is reasonable, since degradation of the road surface should follow similar trends at all wavelengths, however, in principle, *TRACS.LPV03m*, *TRACS.LPV10m* and *TRACS.LPV30m* might be also unrelated. Interesting is to see that, measurements of LPV seem to be negatively correlated to *Speed.AVG*, meaning that truck drivers tend to drive slower on highly deteriorated roads, which again is reasonable.

Other important observations to point out are the high correlations existing between *Speed.AVG* and *Revs.AVG*, that between *Temperature_deg_C* and *Humidity_.*, and that between *Temperature_deg_C* and *Altitude*, which again are linked to physics principles. Additionally, it is possible to see that *TRACS.Pav_Texture* seems to be highly correlated to *Altitude* negatively. Although intuition suggests that these two parameters should be completely unrelated, the detected correlation might be due to the different materials used in different

geographical areas, with aggregate sizes or asphalt mixtures which might depend on local availability of the used materials.

	Fuel.Consumption	Altitude.AVG	Traveled.Time	Gross.Weight_kg	Speed.AVG	Acceleration	Torque.AVG	Revs.AVG	Geom.Abs_Radius	Geom.Gradient	Geom.Crossfall	TRACS.LPV03m	TRACS.LPV10m	TRACS.LPV30m	TRACS.Pav_Texture	Temperature_deg_C	Humidity_.	Atm_Pressure_hPa	WindSpeedDir_kn
Fuel.Consumption	1	-0.09	0.06	0.43	0.04	-0.14	0.52	0.16	0.02	0.65	0.08	0	-0.01	-0.02	0.11	-0.03	-0.07	0	0.21
Altitude.AVG	-0.09	1	-0.03	0.03	0.05	0	-0.1	0.02	-0.1	-0.03	-0.28	-0.03	0	-0.03	-0.55	-0.17	0.15	0.03	-0.01
Traveled.Time	0.06	-0.03	1	0.04	-0.01	-0.1	0.05	0	0.13	0.07	0.01	-0.02	-0.03	-0.04	0	-0.07	0.04	-0.01	0
Gross.Weight_kg	0.43	0.03	0.04	1	-0.12	0	-0.01	0	-0.01	0.02	-0.01	0	0	0	-0.05	-0.04	0.06	-0.01	-0.03
Speed.AVG	0.04	0.05	-0.01	-0.12	1	0.01	0.19	0.76	0.26	0.01	-0.05	-0.18	-0.2	-0.23	0.06	-0.03	0.06	0.03	0.01
Acceleration	-0.14	0	-0.1	0	0.01	1	-0.05	0	-0.05	-0.1	0.02	0.01	0	0.01	-0.01	0.01	0	0.01	-0.01
Torque.AVG	0.52	-0.1	0.05	-0.01	0.19	-0.05	1	0.16	0.06	0.48	0.14	-0.07	-0.07	-0.08	0.13	-0.04	-0.03	0.02	0.2
Revs.AVG	0.16	0.02	0	0	0.76	0	0.16	1	0.25	0.02	-0.04	-0.16	-0.17	-0.2	0.08	-0.04	0.03	0.03	0.03
Geom.Abs_Radius	0.02	-0.1	0.13	-0.01	0.26	-0.05	0.06	0.25	1	0.07	0.05	-0.04	-0.09	-0.09	0.17	0.01	0.01	-0.01	0.02
Geom.Gradient	0.65	-0.03	0.07	0.02	0.01	-0.1	0.48	0.02	0.07	1	0.13	-0.01	-0.02	-0.04	0.07	0.07	-0.09	0.01	0.25
Geom.Crossfall	0.08	-0.28	0.01	-0.01	-0.05	0.02	0.14	-0.04	0.05	0.13	1	0	0	0	0.23	0.01	-0.01	0.01	0.05
TRACS.LPV03m	0	-0.03	-0.02	0	-0.18	0.01	-0.07	-0.16	-0.04	-0.01	0	1	0.98	0.91	-0.01	0.03	-0.03	-0.02	0.01
TRACS.LPV10m	-0.01	0	-0.03	0	-0.2	0	-0.07	-0.17	-0.09	-0.02	0	0.98	1	0.95	-0.03	0.02	-0.03	-0.01	0.01
TRACS.LPV30m	-0.02	-0.03	-0.04	0	-0.23	0.01	-0.08	-0.2	-0.09	-0.04	0	0.91	0.95	1	0.02	0.04	-0.04	-0.02	0.01
TRACS.Pav_Texture	0.11	-0.55	0	-0.05	0.06	-0.01	0.13	0.08	0.17	0.07	0.23	-0.01	-0.03	0.02	1	0.11	-0.1	-0.01	0.08
Temperature_deg_C	-0.03	-0.17	-0.07	-0.04	-0.03	0.01	-0.04	-0.04	0.01	0.07	0.01	0.03	0.02	0.04	0.11	1	-0.51	-0.08	-0.01
Humidity_.	-0.07	0.15	0.04	0.06	0.06	0	-0.03	0.03	0.01	-0.09	-0.01	-0.03	-0.03	-0.04	-0.1	-0.51	1	-0.03	-0.07
Atm_Pressure_hPa	0	0.03	-0.01	-0.01	0.03	0.01	0.02	0.03	-0.01	0.01	0.01	-0.02	-0.01	-0.02	-0.01	-0.08	-0.03	1	0.16
WindSpeedDir_kn	0.21	-0.01	0	-0.03	0.01	-0.01	0.2	0.03	0.02	0.25	0.05	0.01	0.01	0.01	0.08	-0.01	-0.07	0.16	1

Figure 63. Correlation matrix for the data from light trucks App 1, from the Dataset 2.

Table 39 reports a summary of some of the calculated p-values, estimated for each regression coefficient, when all the available data are fed to an LR. From the table, it is possible to see that all variables, apart from the *Traveled.Time*, *WindSpeed.Cro_kn* and LPV at 3, 10 and 30 metres wavelengths are considered to help the developed model in explaining a larger variance across the data at a significance level of 99.9%. This indicates that for light trucks driving on A roads, the

effect of LPV at 3, 10 and 30 m wavelengths is insignificant or non-linearly correlated to *Fuel.Consumption*.

Table 39 - Summary of the computed p-values of the correlated variables for light trucks App 1, from the Dataset 2.

Variable	p-value	Significance level
Gross.Weight_kg	< 2.00 e^{-16}	99.9 %
Geom.Gradient	< 2.00 e^{-16}	99.9 %
Torque.AVG	< 2.00 e^{-16}	99.9 %
Revs.AVG	< 2.00 e^{-16}	99.9 %
Speed.AVG	< 2.00 e^{-16}	99.9 %
Acceleration	< 2.00 e^{-16}	99.9 %
Traveled.Time	0.833	0 %
Geom.Abs_Radius	< 2.00 e^{-16}	99.9 %
Geom.Crossfall	4.22 e^{-07}	99.9 %
TRACS.LPV03m	0.573	0 %
TRACS.LPV10m	0.610	0 %
TRACS.LPV30m	0.769	0 %
TRACS.Pav_Texture	1.27 e^{-12}	99.9 %
Altitude.AVG	5.76 e^{-14}	99.9 %
Atm_Pressure_hPa	3.64 e^{-07}	99.9 %
Temperature_deg_C	< 2.00 e^{-16}	99.9 %
Humidity_.	< 2.00 e^{-16}	99.9 %
Delta.Angle_deg	3.27 e^{-04}	99.9 %
WindSpeedDir_kn	4.41e^{-13}	99.9 %
WindSpeed.Cro_kn	0.645	0 %

Figure 64 reports results of the analysis of the adjusted-R^2. The figure shows that, by feeding to an LR information from up to 21 variables out of the 77 available

in total, it is possible to improve accuracy of the developed model, and this includes measurements of LPV at 3, 10 and 30 metres wavelength. The fact that results of the analysis of *p-values* and that of the adjusted-R^2 disagree, shows the complexity of the relationships existing between *Fuel.Consumption* and the selection of variables considered in the study. For this reason, Lasso and RFs were interrogated and included as part of the variable selection process (see sect. 4.3).

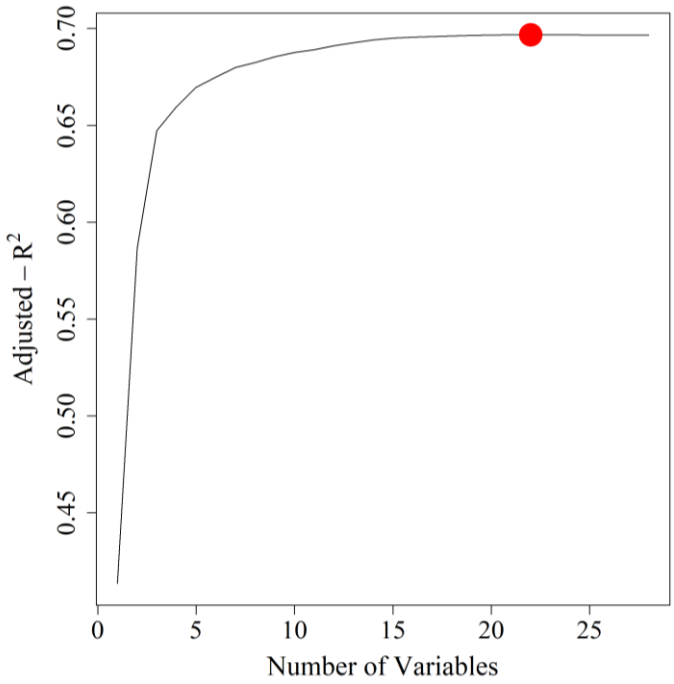


Figure 64. Analysis of the adjusted-R^2 for the light trucks App 1, from the Dataset 2.

Table 40 summarises results of the Lasso regression analysis. As it is possible to see from the table, the only variables, which the algorithm is able to exclude from the modelling of *Fuel.Consumption* for the considered case study, are the *Traveled.Time* and the *TRACS.LPV10m*. Since, the Lasso is able to detect also non-linear correlations between variables, this might point out that LPV at 3 and 30 metres

wavelength might have non-linear correlation with the fuel consumption of the considered fleet of trucks.

Table 40 - Summary of the Lasso coefficients for the light trucks App1, from the Dataset 2.

Variable	Lasso Coefficient
Gross.Weight_kg	0.401
Geom.Gradient	0.496
Torque.AVG	0.344
Revs.AVG	0.208
Speed.AVG	-0.083
Acceleration	-0.071
Traveled.Time	0
Geom.Abs_Radius	-0.044
Geom.Crossfall	-0.018
TRACS.LPV03m	0.025
TRACS.LPV10m	0
TRACS.LPV30m	0.0015
TRACS.Pav_Texture	0.037
Altitude.AVG	-0.0016
Atm_Pressure_hPa	-0.0041
Temperature_deg_C	-0.051
Humidity_.	-0.052
Delta.Angle_deg	0.0110
WindSpeedDir_kn	0.046
WindSpeed.Cro_kn	0.0039

Finally, the RF algorithm was used to rank the variables which, from the analysis of *p-values*, adjusted-R^2 and Lasso, showed to have the highest correlation

with the fuel consumption of the considered fleet of trucks. Figure 65 provides a summary of the results.

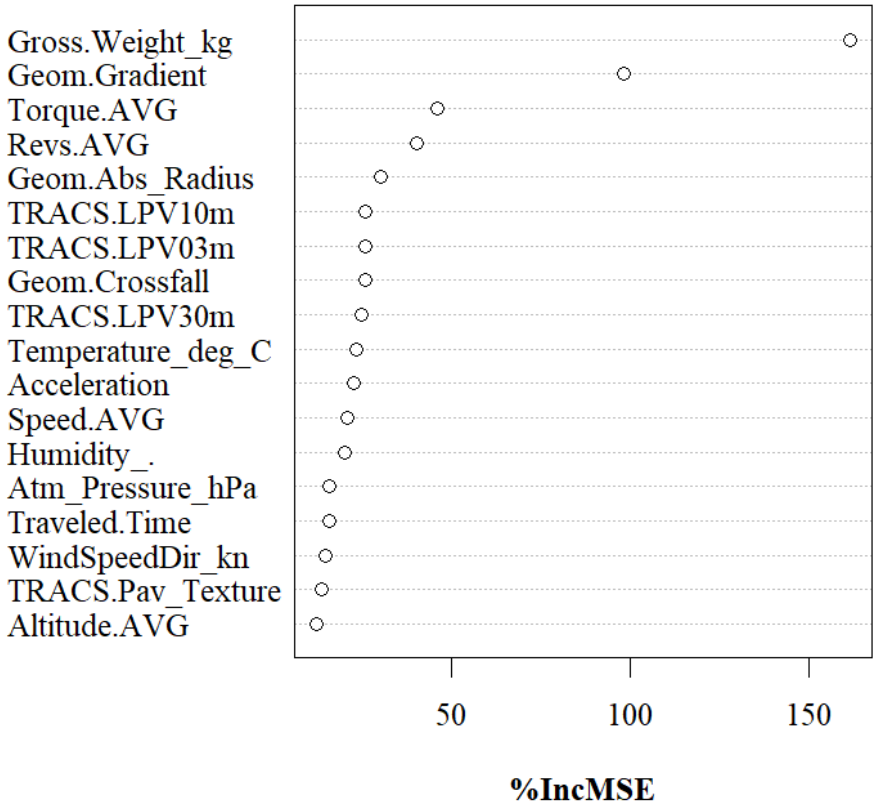


Figure 65. Rank of variable importance as increase in MSE for the light trucks App1, from the Dataset 2, as computed by the RF algorithm.

In particular, as explained in the Methodology (see sect. 4.3.4), the algorithm tests the increase in MSE given by inclusion or exclusion of each single variable from an RF model developed using the remaining variables. From the figure, it is possible to see that the algorithm identifies *Gross.Weight_kg*, *Geom.Gradient*, *Torque.AVG* and *Revs.AVG* as the top four variables that influence *Fuel.Consumption*. This matches the results of the analysis of *p-values*, adjusted-R^2 and Lasso, which gives more confidence in the used method. Another observation which can be made by looking at Figure 65 is that variables such as *Geom.Abs_Radius*, LPV at 3, 10 and 30 metres

wavelength, and *Geom.Crossfall*, characteristics of the road infrastructure, are ranked to be more important than *Speed.AVG* and *Acceleration*, which generally affect the fuel consumption of road vehicles from the physical point of view. Although this might be characteristic of light trucks driving at constant speed along A roads only, this could confirm that, in these conditions, correlation between road characteristics and fuel consumption might be non-linear.

E.2.1.2 Light trucks on Motorways and A roads (App 2)

Figure 66 shows results of the analysis of the Pearson's correlation coefficients for the light trucks App 2.

	Fuel.Consumption	Altitude.AVG	Traveled.Time	Gross.Weight_kg	Speed.AVG	Acceleration	Torque.AVG	Revs.AVG	Geom.Abs_Radius	Geom.Gradient	Geom.Crossfall	TRACS.LPV03m	TRACS.LPV10m	TRACS.LPV30m	TRACS.Pav_Texture	Temperature_deg_C	Humidity_.	Atm_Pressure_hPa	WindSpeedDir_kn
Fuel.Consumption	1	-0.01	0.03	0.44	0.04	-0.12	0.56	0.13	0	0.71	0.09	0	0	-0.01	0.02	-0.08	0.02	-0.01	0.05
Altitude.AVG	-0.01	1	-0.02	0.04	0.09	-0.01	-0.02	0.09	0.01	-0.02	-0.06	-0.05	-0.04	-0.08	-0.17	0.01	-0.05	0.01	0.01
Traveled.Time	0.03	-0.02	1	0.02	-0.11	-0.07	0.05	-0.05	0.04	0.03	0.02	-0.01	-0.02	-0.02	0	-0.02	0.02	0.01	0
Gross.Weight_kg	0.44	0.04	0.02	1	-0.08	0	0.04	0.05	0.02	-0.01	0.02	-0.01	-0.01	-0.02	-0.07	-0.03	0.1	0.04	0.02
Speed.AVG	0.04	0.09	-0.11	-0.08	1	0.02	0.13	0.64	0.09	-0.01	-0.02	-0.09	-0.09	-0.11	-0.02	-0.04	-0.02	-0.05	0
Acceleration	-0.12	-0.01	-0.07	0	0.02	1	-0.04	0	-0.02	-0.1	-0.01	0	0	0.01	0	0	0	0	0
Torque.AVG	0.56	-0.02	0.05	0.04	0.13	-0.04	1	0.04	-0.02	0.56	0.07	-0.02	-0.02	-0.02	0.06	-0.09	0.03	-0.03	0.06
Revs.AVG	0.13	0.09	-0.05	0.05	0.64	0	0.04	1	0.08	0	-0.01	-0.07	-0.07	-0.09	-0.02	-0.07	-0.01	0.02	0
Geom.Abs_Radius	0	0.01	0.04	0.02	0.09	-0.02	-0.02	0.08	1	0	0.27	-0.04	-0.07	-0.17	0.03	0.01	-0.04	-0.02	0.01
Geom.Gradient	0.71	-0.02	0.03	-0.01	-0.01	-0.1	0.56	0	0	1	0.14	0	0	-0.01	0.03	0.01	-0.02	0	-0.01
Geom.Crossfall	0.09	-0.06	0.02	0.02	-0.02	-0.01	0.07	-0.01	0.27	0.14	1	0.01	0	-0.03	0.08	0.02	-0.04	-0.01	-0.01
TRACS.LPV03m	0	-0.05	-0.01	-0.01	-0.09	0	-0.02	-0.07	-0.04	0	0.01	1	0.98	0.89	0.01	0.01	-0.01	0	0
TRACS.LPV10m	0	-0.04	-0.02	-0.01	-0.09	0	-0.02	-0.07	-0.07	0	0	0.98	1	0.94	0.01	0.01	0	0	0
TRACS.LPV30m	-0.01	-0.08	-0.02	-0.02	-0.11	0.01	-0.02	-0.09	-0.17	-0.01	-0.03	0.89	0.94	1	0.03	0.01	0.01	0.01	0
TRACS.Pav_Texture	0.02	-0.17	0	-0.07	-0.02	0	0.06	-0.02	0.03	0.03	0.08	0.01	0.01	0.03	1	0.05	-0.08	0.01	-0.01
Temperature_deg_C	-0.08	0.01	-0.02	-0.03	-0.04	0	-0.09	-0.07	0.01	0.01	0.02	0.01	0.01	0.01	0.05	1	-0.5	-0.1	-0.05
Humidity_.	0.02	-0.05	0.02	0.1	-0.02	0	0.03	-0.01	-0.04	-0.02	-0.04	-0.01	0	0.01	-0.08	-0.5	1	0.07	0.02
Atm_Pressure_hPa	-0.01	0.01	0.01	0.04	-0.05	0	-0.03	0.02	-0.02	0	-0.01	0	0	0.01	0.01	-0.1	0.07	1	0.02
WindSpeedDir_kn	0.05	0.01	0	0.02	0	0	0.06	0	0.01	-0.01	-0.01	0	0	0	-0.01	-0.05	0.02	0.02	1

Figure 66. Correlation matrix for the data from light trucks App 2, from the Dataset 2.

From the figure, it is possible to see that, as it was happening for the light trucks App 1, *Geom.Gradient*, *Torque.AVG* and *Gross.Weight_kg* are the variables, which show the highest linear correlation with *Fuel.Consumption*. Again, it is possible to see a quite high correlation existing between *Speed.AVG* and *Revs.AVG*, and between *Speed.AVG* and the different wavelengths of LPV considered in the study. This could mean that drivers tend to decelerate when driving along uneven roads, or, alternatively, this could be due to the heavier traffic, which causes congestion and makes road pavements to deteriorate quicker.

However, these can be considered only assumptions, which require further investigation to be confirmed. Again, the figure shows also how the three wavelengths of LPV considered in the study are mutually correlated with each other. As expected, also in this case, *Humidity_*. shows some correlation with *Temperature_deg_C*. Additionally, as it happened for the light trucks App 1, *Altitude.AVG* reveals to be somewhat correlated with *TRACS.Pav_Texture*, which might depend on the different construction techniques used in different areas of the UK.

Table 41 shows a summary of the calculated p-values for a selection of variables from those considered in the study. This shows that for the light trucks App 2, all the regression coefficient of a LR including all variables would be statistically significant at 99.9% significance level apart from one. That includes measurements of LPV at 3, 10 and 30 metres wavelength, and measurements of SMTD, which characterise the surface conditions of the considered roads, and excludes the calculated cross-wind (*WindSpeed.Cro_kn*). This points out that, in different conditions, road surface characteristics might influence the fuel consumption of road vehicles differently.

Table 41 - Summary of the computed p-values of the correlated variables for light trucks App 2, from the Dataset 2.

Variable	p-value	Significance level
Gross.Weight_kg	< 2.00 e^{-16}	99.9 %
Geom.Gradient	< 2.00 e^{-16}	99.9 %
Torque.AVG	< 2.00 e^{-16}	99.9 %
Revs.AVG	< 2.00 e^{-16}	99.9 %
Speed.AVG	< 2.00 e^{-16}	99.9 %
Acceleration	< 2.00 e^{-16}	99.9 %
Traveled.Time	1.04 e^{-10}	99.9 %
Geom.Abs_Radius	1.97 e^{-06}	99.9 %
Geom.Crossfall	1.02 e^{-12}	99.9 %
TRACS.LPV03m	4.95 e^{-06}	99.9 %
TRACS.LPV10m	4.48 e^{-14}	99.9 %
TRACS.LPV30m	< 2.00 e^{-16}	99.9 %
TRACS.Pav_Texture	< 2.00 e^{-16}	99.9 %
Altitude.AVG	3.87 e^{-15}	99.9 %
Atm_Pressure_hPa	< 2.00 e^{-16}	99.9 %
Temperature_deg_C	< 2.00 e^{-16}	99.9 %
Humidity_.	< 2.00 e^{-16}	99.9 %
Delta.Angle_deg	1.12 e^{-13}	99.9 %
WindSpeedDir_kn	< 2.00 e^{-16}	99.9 %
WindSpeed.Cro_kn	0.583	0 %

Figure 67 shows the results of the analysis of the adjusted-R^2. This shows that inclusion of 30 out of 77 type of measurements available allows to improve the accuracy of an LR. In particular, also inclusion of the *WindSpeed.Cro_kn* should help. In detail though, it is possible to see also that (see Figure 67) the adjusted-R^2 does not increase significantly after about 20 variables are considered in the model. This, since

a smaller model might be better interpretable, should raise questions about the actual need of including all variables.

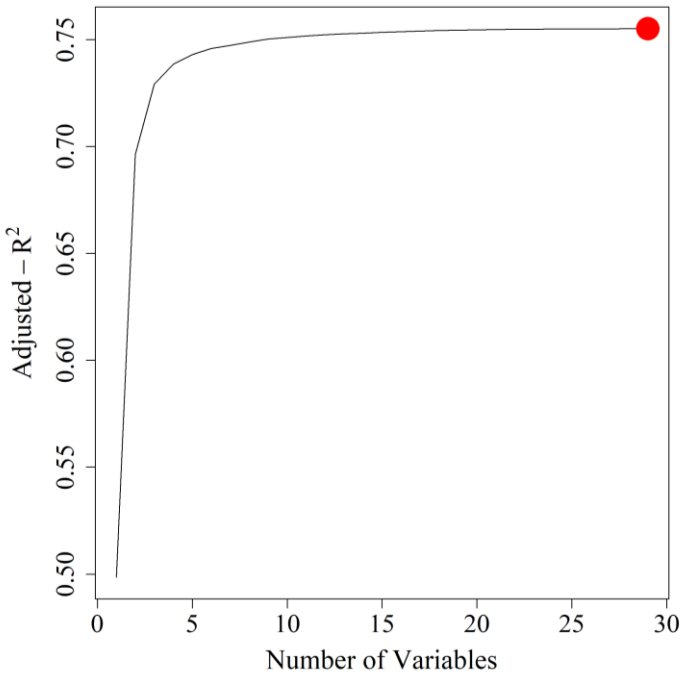


Figure 67. Analysis of the adjusted-R^2 for light trucks App 2, from the Dataset 2.

Results of the Lasso regression analysis (see Table 42) show that, among the variables, which show high correlation with *Fuel.Consumption*, only the coefficient related to *TRACS.LPV03m* is reduced to zero by algorithm. This, again, highlights the fact that in different situations different characteristics of the road surface can impact the fuel consumption of light trucks differently, which might also affect results of variable selection algorithms.

Table 42 - Summary of the Lasso coefficients for light trucks App2, from the Dataset 2.

Variable	Lasso Coefficient
Gross.Weight_kg	0.432
Geom.Gradient	0.594
Torque.AVG	0.267
Revs.AVG	0.117
Speed.AVG	-0.023
Acceleration	-0.053
Traveled.Time	-0.0047
Geom.Abs_Radius	-0.0091
Geom.Crossfall	-0.016
TRACS.LPV03m	0
TRACS.LPV10m	-0.0066
TRACS.LPV30m	0.019
TRACS.Pav_Texture	0.015
Altitude.AVG	-0.0055
Atm_Pressure_hPa	-0.016
Temperature_deg_C	-0.034
Humidity_.	-0.052
Delta.Angle_deg	-0.0097
WindSpeedDir_kn	0.040
WindSpeed.Cro_kn	-0.0097

Finally, the RF was used to rank the considered variables for their importance and see how these affect the accuracy of a model developed including them (see Figure 68). From Figure 68, it is possible to see that, also in this case, after the top four variables, which again are Gross.Weight_kg, Geom.Gradient, Revs.AVG and Torque.AVG, *Geom.Abs_Radius* and road surface characteristics are ranked as

playing a significant role for the model to estimate the fuel consumption of the light trucks App 2 accurately. Although the figure shows that contribution of these variables is much lower compared to others such as *Gross.Weight_kg* or *Geom.Gradient*, these are again ranked higher than some others such as *Speed.AVG* and *Acceleration*, which traditionally are considered to play an important role in determining the fuel consumption of road vehicles.

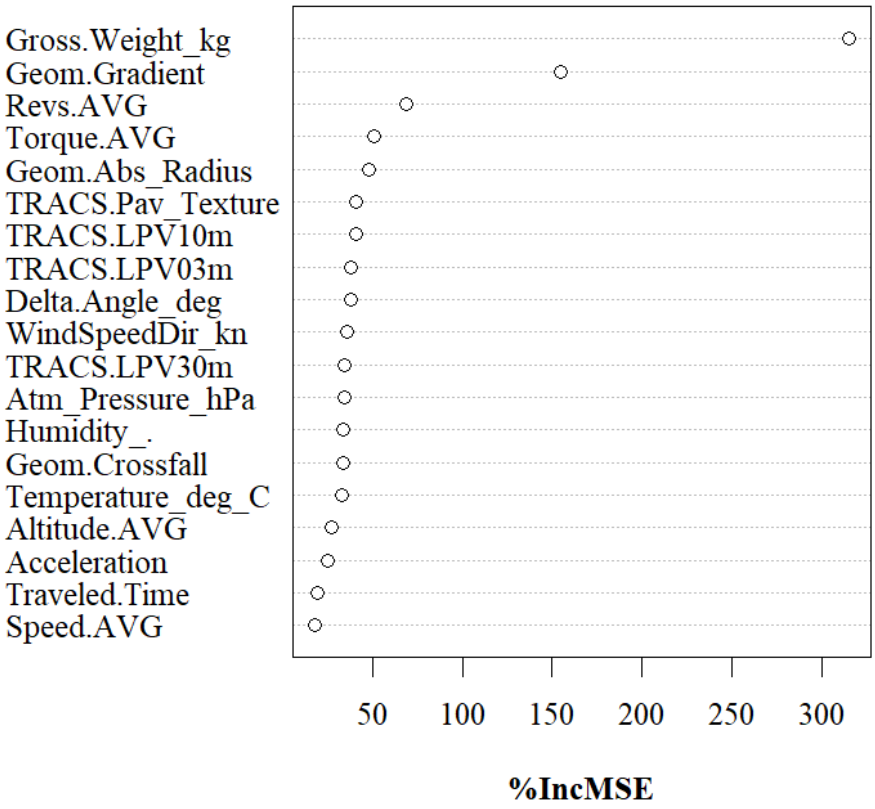


Figure 68. Rank of variable importance as increase in MSE for light trucks App 2, from the Dataset 2, as computed by the RF algorithm.

E.2.1.3 Medium trucks on A roads (App 1)

In Figure 69, results of the analysis of Pearson's correlation coefficients are reported for the medium trucks App 1, from the Dataset 2. In particular, the figure reports the correlation matrices calculated for the considered application. Again, it is

possible to see how the major trends highlighted in the previous sections are confirmed for light trucks. For example, also in this case, *Geom.Gradient*, *Gross.Weight_kg* and *Torque.AVG* are identified as the variables, which correlate with *Fuel.Consumption* the most.

	Fuel.Consumption	Altitude.AVG	Traveled.Time	Gross.Weight_kg	Speed.AVG	Acceleration	Torque.AVG	Revs.AVG	Geom.Abs_Radius	Geom.Gradient	Geom.Crossfall	TRACS.LPV03m	TRACS.LPV10m	TRACS.LPV30m	TRACS.Pav_Texture	Temperature_deg_C	Humidity_.	Atm_Pressure_hPa	WindSpeedDir_kn
Fuel.Consumption	1	0.03	-0.06	0.38	0.09	-0.17	0.62	-0.01	-0.03	0.66	-0.05	0	0	0	-0.03	-0.09	0.07	-0.01	0.02
Altitude.AVG	0.03	1	-0.03	0.02	-0.04	0	0.02	-0.01	-0.12	0.04	-0.32	-0.01	0.01	0	-0.61	-0.03	0.07	0	-0.01
Traveled.Time	-0.06	-0.03	1	-0.12	-0.24	-0.13	0	0	0.1	0.01	0.04	-0.01	-0.01	-0.03	0.02	-0.02	0.02	0	0.01
Gross.Weight_kg	0.38	0.02	-0.12	1	0.15	0.02	0.23	-0.28	0.01	0.01	0	0.01	0.01	0.01	-0.02	-0.03	0.07	-0.01	-0.02
Speed.AVG	0.09	-0.04	-0.24	0.15	1	0.05	0.08	0.22	0.19	0	0	-0.08	-0.1	-0.14	0.01	-0.03	0.02	0	0
Acceleration	-0.17	0	-0.13	0.02	0.05	1	-0.08	-0.04	-0.01	-0.07	0.01	0	0	0.01	0	0	0	0.01	0
Torque.AVG	0.62	0.02	0	0.23	0.08	-0.08	1	0.12	-0.01	0.45	-0.05	-0.01	-0.01	-0.01	-0.04	-0.09	0.06	0	0.01
Revs.AVG	-0.01	-0.01	0	-0.28	0.22	-0.04	0.12	1	0.04	0.03	-0.02	-0.01	-0.02	-0.03	-0.02	-0.01	-0.01	0.02	0.01
Geom.Abs_Radius	-0.03	-0.12	0.1	0.01	0.19	-0.01	-0.01	0.04	1	0.04	0.17	-0.03	-0.06	-0.17	0.1	0	0	0	-0.01
Geom.Gradient	0.66	0.04	0.01	0.01	0	-0.07	0.45	0.03	0.04	1	-0.05	0	0	-0.03	-0.09	0.02	-0.01	0.01	-0.01
Geom.Crossfall	-0.05	-0.32	0.04	0	0	0.01	-0.05	-0.02	0.17	-0.05	1	0	0	0.01	0.39	0.01	-0.01	0	0
TRACS.LPV03m	0	-0.01	-0.01	0.01	-0.08	0	-0.01	-0.01	-0.03	0	0	1	0.98	0.89	0	0	0	0	0
TRACS.LPV10m	0	0.01	-0.01	0.01	-0.1	0	-0.01	-0.02	-0.06	0	0	0.98	1	0.94	-0.01	0	0	0	0
TRACS.LPV30m	0	0	-0.03	0.01	-0.14	0.01	-0.01	-0.03	-0.17	-0.03	0.01	0.89	0.94	1	0.03	0	-0.01	0	0
TRACS.Pav_Texture	-0.03	-0.61	0.02	-0.02	0.01	0	-0.04	-0.02	0.1	-0.09	0.39	0	-0.01	0.03	1	0.01	-0.04	0	0.01
Temperature_deg_C	-0.09	-0.03	-0.02	-0.03	-0.03	0	-0.09	-0.01	0	0.02	0.01	0	0	0	0.01	1	-0.41	-0.17	-0.13
Humidity_.	0.07	0.07	0.02	0.07	0.02	0	0.06	-0.01	0	-0.01	-0.01	0	0	-0.01	-0.04	-0.41	1	-0.08	0.01
Atm_Pressure_hPa	-0.01	0	0	-0.01	0	0.01	0	0.02	0	0.01	0	0	0	0	0	-0.17	-0.08	1	0.21
WindSpeedDir_kn	0.02	-0.01	0.01	-0.02	0	0	0.01	0.01	-0.01	-0.01	0	0	0	0	0.01	-0.13	0.01	0.21	1

Figure 69. Correlation matrix for the data for the medium trucks App 1, from the Dataset 2.

Also in this case, it is possible to see how road LPV characteristics seem to correlate quite well with *Speed.AVG*, while again *TRACS.Pav_Texture* seems to correlate with *Altitude.AVG*.

Table 43 reports results of the analysis of *p-values* for the coefficients of a linear model fitted to the data from the medium trucks App 1. From the analysis of the values reported in the table, it is possible to see that the only variable to exclude from the development of fuel consumption models would be the *Traveled.Time,* for this vehicle type.

Table 43 - Summary of the computed p-values of the correlated variables for the medium trucks App 1, from the Dataset 2.

Variable	p-value	Significance level
Gross.Weight_kg	< 2.00 e^{-16}	99.9 %
Geom.Gradient	< 2.00 e^{-16}	99.9 %
Torque.AVG	< 2.00 e^{-16}	99.9 %
Revs.AVG	< 2.00 e^{-16}	99.9 %
Speed.AVG	< 2.00 e^{-16}	99.9 %
Acceleration	< 2.00 e^{-16}	99.9 %
Traveled.Time	< 2.00 e^{-16}	0 %
Geom.Abs_Radius	< 2.00 e^{-16}	99.9 %
Geom.Crossfall	< 2.00 e^{-16}	99.9 %
TRACS.LPV03m	3.28 e^{-04}	99.9 %
TRACS.LPV10m	1.19 e^{-10}	99.9 %
TRACS.LPV30m	< 2.00 e^{-16}	99.9 %
TRACS.Pav_Texture	< 2.00 e^{-16}	99.9 %
Altitude.AVG	< 2.00 e^{-16}	99.9 %
Atm_Pressure_hPa	< 2.00 e^{-16}	99.9 %
Temperature_deg_C	< 2.00 e^{-16}	99.9 %
Humidity_.	0.0065	99.9 %
Delta.Angle_deg	0.921	0 %
WindSpeedDir_kn	< 2.00 e^{-16}	99.9 %
WindSpeed.Cro_kn	5.31 e^{-10}	99.9 %

Figure 70 reports results of the analysis of the adjusted-R^2 for the considered vehicle fleet. Again, as in the case of light trucks App 1, the results show that inclusion of 30 out of 77 variable help in increasing accuracy of an LR. However, also in this case, it is possible to see that after approximately 20 variables, no significant improvement is obtained.

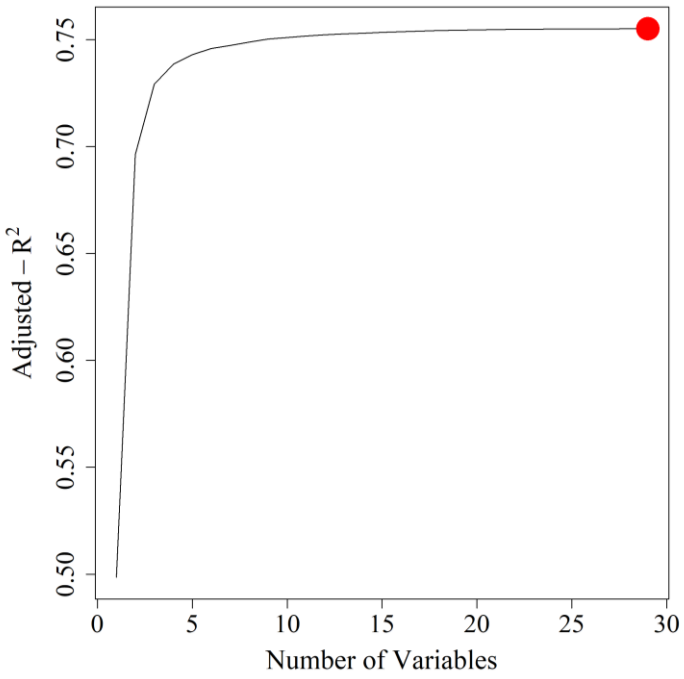


Figure 70. Analysis of the adjusted-R^2 the medium trucks App 1, from the Dataset 2.

In Table 44, results of the Lasso analysis are reported. The table shows the estimated regression coefficients using a Lasso regression technique. From this, it is possible to see that *Humidity_.* and *Delta.Angle_deg* are the only variables, which the algorithm establishes to have poor influence on the fuel consumption of the considered fleet of trucks. Therefore, also in this case, in lieu of the results of the Lasso, it is not possible to exclude that characteristics of the road surface measured as LPV at 3, 10

and 30 metres wavelength and TRACS.Pav_Texture (SMTD) have a significant influence on the fuel consumption of the considered fleet of trucks. Fact that they result insignificant sometimes, might be due to non-linearities in the problem (see sect. 2.4), which make traditional statistics, such as the *p-values*, ineffective in the identification of their actual impacts in certain situation.

Table 44 - Summary of the Lasso coefficients the medium trucks App 1, from the Dataset 2.

Variable	Lasso Coefficient
Gross.Weight_kg	0.303
Geom.Gradient	0.528
Torque.AVG	0.301
Revs.AVG	0.020
Speed.AVG	0.022
Acceleration	-0.119
Traveled.Time	-0.029
Geom.Abs_Radius	-0.049
Geom.Crossfall	-0.020
TRACS.LPV03m	-0.0205
TRACS.LPV10m	-0.0026
TRACS.LPV30m	0.0245
TRACS.Pav_Texture	0.066
Altitude.AVG	-0.0045
Atm_Pressure_hPa	-0.013
Temperature_deg_C	-0.057
Humidity_.	0
Delta.Angle_deg	0
WindSpeedDir_kn	0.027
WindSpeed.Cro_kn	0.0053

Finally, Figure 71 shows the rank of variables performed by the RF algorithm for the medium trucks App 1. Again, the figure shows that, after the most important variables in the problem, the algorithm ranks road geometry (radius and crossfall) and road surface characteristics (LPV and SMTD) among the variables which influence the most the fuel consumption of the considered fleet of trucks.

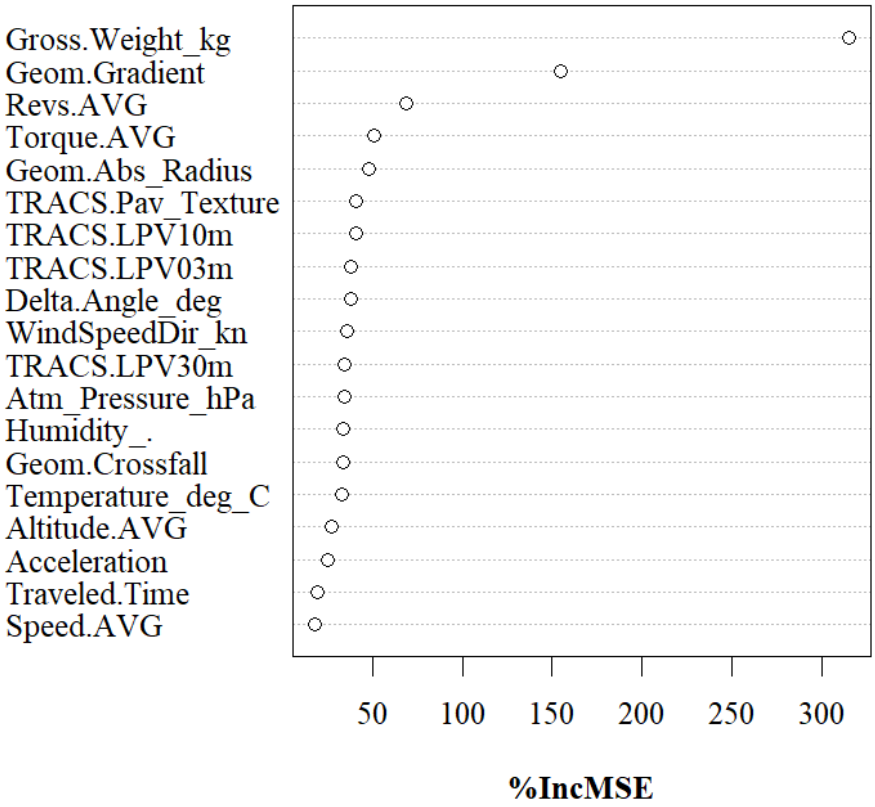


Figure 71. Rank of variable importance as increase in MSE for the medium trucks App 1, from the Dataset 2, as computed by the RF algorithm.

E.2.1.4 Medium trucks on Motorways and A roads (App 2)

In Figure 72, the correlation matrix for the medium trucks App 2 is reported. From the figure, trends similar to those showed by the other vehicles previously considered are highlighted. For instance, the three variables, which show the highest

linear correlation with *Fuel.Consumption*, are again *Geom.Gradient*, *Gross.Weight_kg* and *Torque.AVG*. Again, correlation between the different wavelengths of LPV considered and *Speed.AVG* is shown together with a correlation between *Altitude.AVG* and *TRACS.Pav_Texture*. However, also in this case, it looks like that road surface characteristics show poor correlation with *Fuel.Consumption*. Therefore, analysis of other statistics should be performed to identify the possibly existing trends.

	Fuel.Consumption	Altitude.AVG	Traveled.Time	Gross.Weight_kg	Speed.AVG	Acceleration	Torque.AVG	Revs.AVG	Geom.Abs_Radius	Geom.Gradient	Geom.Crossfall	TRACS.LPV03m	TRACS.LPV10m	TRACS.LPV30m	TRACS.Pav_Texture	Temperature_deg_C	Humidity_.	Atm_Pressure_hPa	WindSpeedDir_kn
Fuel.Consumption	1	0.02	-0.02	0.31	0	-0.16	0.69	-0.03	-0.01	0.77	-0.05	-0.01	-0.01	-0.01	0	-0.06	0.04	-0.01	0.03
Altitude.AVG	0.02	1	-0.08	0	0.07	-0.01	0	0.03	-0.06	0.02	-0.09	-0.06	-0.01	-0.03	-0.28	0.04	-0.06	0	0
Traveled.Time	-0.02	-0.08	1	-0.17	-0.26	-0.17	0.03	0.05	0.03	0.03	0.04	0.03	0.03	-0.01	0.01	-0.03	0.04	0	0
Gross.Weight_kg	0.31	0	-0.17	1	0.05	0.04	0.17	-0.3	0.01	0	0.01	0.01	0	-0.01	-0.03	0.01	0.04	-0.01	0.02
Speed.AVG	0	0.07	-0.26	0.05	1	0.06	0.01	0.25	0.19	-0.02	-0.01	-0.09	-0.12	-0.17	-0.01	0	-0.03	0.01	0
Acceleration	-0.16	-0.01	-0.17	0.04	0.06	1	-0.1	-0.06	-0.02	-0.05	0	0	0	0.02	0	0	0	0	0
Torque.AVG	0.69	0	0.03	0.17	0.01	-0.1	1	0.09	-0.02	0.56	-0.04	-0.02	-0.01	0	-0.02	-0.07	0.05	-0.01	0.03
Revs.AVG	-0.03	0.03	0.05	-0.3	0.25	-0.06	0.09	1	0.05	0.01	-0.01	-0.03	-0.04	-0.05	0	-0.01	-0.02	0	0
Geom.Abs_Radius	-0.01	-0.06	0.03	0.01	0.19	-0.02	-0.02	0.05	1	0.06	0.17	-0.05	-0.16	-0.39	0.06	0.03	-0.06	0	-0.01
Geom.Gradient	0.77	0.02	0.03	0	-0.02	-0.05	0.56	0.01	0.06	1	-0.03	-0.01	-0.03	-0.07	-0.01	0.01	-0.01	0	-0.02
Geom.Crossfall	-0.05	-0.09	0.04	0.01	-0.01	0	-0.04	-0.01	0.17	-0.03	1	0.01	0	-0.03	0.22	0.01	-0.02	0.01	-0.01
TRACS.LPV03m	-0.01	-0.06	0.03	0.01	-0.09	0	-0.02	-0.03	-0.05	-0.01	0.01	1	0.89	0.64	0.02	0	0	0	0
TRACS.LPV10m	-0.01	-0.01	0.03	0	-0.12	0	-0.01	-0.04	-0.16	-0.03	0	0.89	1	0.81	0.01	-0.01	0.01	0	0
TRACS.LPV30m	-0.01	-0.03	-0.01	-0.01	-0.17	0.02	0	-0.05	-0.39	-0.07	-0.03	0.64	0.81	1	0.06	-0.02	0.03	-0.01	0
TRACS.Pav_Texture	0	-0.28	0.01	-0.03	-0.01	0	-0.02	0	0.06	-0.01	0.22	0.02	0.01	0.06	1	0.01	-0.04	0	0.02
Temperature_deg_C	-0.06	0.04	-0.03	0.01	0	0	-0.07	-0.01	0.03	0.01	0.01	0	-0.01	-0.02	0.01	1	-0.5	-0.16	-0.08
Humidity_.	0.04	-0.06	0.04	0.04	-0.03	0	0.05	-0.02	-0.06	-0.01	-0.02	0	0.01	0.03	-0.04	-0.5	1	-0.02	0.02
Atm_Pressure_hPa	-0.01	0	0	-0.01	0.01	0	-0.01	0	0	0	0.01	0	0	-0.01	0	-0.16	-0.02	1	0.12
WindSpeedDir_kn	0.03	0	0	0.02	0	0	0.03	0	-0.01	-0.02	-0.01	0	0	0	0.02	-0.08	0.02	0.12	1

Figure 72. Correlation matrix for the data from the medium trucks App 2, from the Dataset 2.

Table 45 reports the results of the analysis of the p-values for the medium trucks App 2.

Table 45 - Summary of the computed p-values of the correlated variables for the medium trucks App 2, from the Dataset 2.

Variable	p-value	Significance level
Gross.Weight_kg	< 2.00 e^{-16}	99.9 %
Geom.Gradient	< 2.00 e^{-16}	99.9 %
Torque.AVG	< 2.00 e^{-16}	99.9 %
Revs.AVG	4.75 e^{-13}	99.9 %
Speed.AVG	4.63 e^{-09}	99.9 %
Acceleration	< 2.00 e^{-16}	99.9 %
Traveled.Time	1.04 e^{-10}	99.9 %
Geom.Abs_Radius	1.97 e^{-06}	99.9 %
Geom.Crossfall	1.02 e^{-12}	99.9 %
TRACS.LPV03m	1.74 e^{-12}	99.9 %
TRACS.LPV10m	0.013	95 %
TRACS.LPV30m	< 2.00 e^{-16}	99.9 %
TRACS.Pav_Texture	< 2.00 e^{-16}	99.9 %
Altitude.AVG	< 2.00 e^{-16}	99.9 %
Atm_Pressure_hPa	< 2.00 e^{-16}	99.9 %
Temperature_deg_C	< 2.00 e^{-16}	99.9 %
Humidity_.	0.131	0 %
Delta.Angle_deg	0.145	0 %
WindSpeedDir_kn	< 2.00 e^{-16}	99.9 %
WindSpeed.Cro_kn	0.278	0 %

From the table, it is possible to see that most of variables considered can be classified significant at the 99.9% significance level, with *TRACS.LPV10m*, which is

identified to be significant at 95%. Variables such as *Humidity_.*, *Delta.Angle_deg* and *WindSpeed.Cro_kn* are classified instead to be insignificant for estimating the fuel consumption of the considered fleet of trucks, which partially confirms the findings of the analysis conducted on the other vehicle types.

Analysis of the adjusted-R^2 shows that, also in this case, the algorithm identifies 30 out of 77 variables to play an important role in the modelling of the fuel consumption of the considered fleet of trucks using LR (see Figure 73). However, the figure also shows that after approximately 20 variables the model does not improve its accuracy significantly. This matches the results of the analysis that the present study conducted on different vehicle types (see Appendix E.2.1.2 and E.2.1.3).

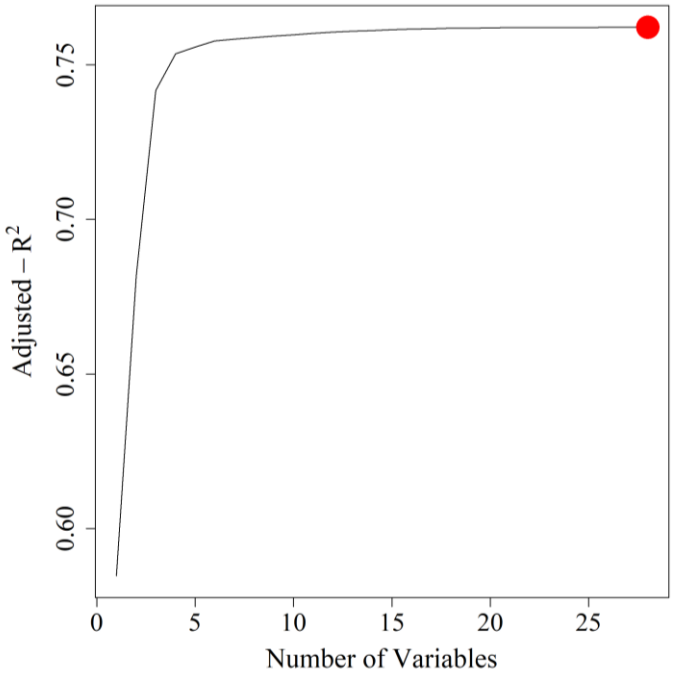


Figure 73. Analysis of the adjusted-R^2 for the medium trucks App 2, from the Dataset 2.

Table 46 reports results of the Lasso analysis. As in the case of light trucks driven along M and A roads (light trucks App 2), also in this case (medium trucks App

2), the Lasso reduces the coefficient assigned to the *TRACS.LPV03m* to zero, meaning that this should not play a significant role in the estimation of the fuel consumption of the considered fleet of trucks.

Table 46 - Summary of the Lasso coefficients for the medium trucks App2, from the Dataset 2.

Variable	Lasso Coefficient
Gross.Weight_kg	0.432
Geom.Gradient	0.594
Torque.AVG	0.267
Revs.AVG	0.117
Speed.AVG	-0.023
Acceleration	-0.053
Traveled.Time	-0.0047
Geom.Abs_Radius	-0.0091
Geom.Crossfall	-0.0169
TRACS.LPV03m	0
TRACS.LPV10m	-0.0066
TRACS.LPV30m	0.0188
TRACS.Pav_Texture	0.0154
Altitude.AVG	-0.0055
Atm_Pressure_hPa	-0.0164
Temperature_deg_C	-0.0341
Humidity_.	-0.052
Delta.Angle_deg	-0.0097
WindSpeedDir_kn	0.040
WindSpeed.Cro_kn	-0.0097

In Figure 74, rank of the variables for the medium trucks App 2 is shown.

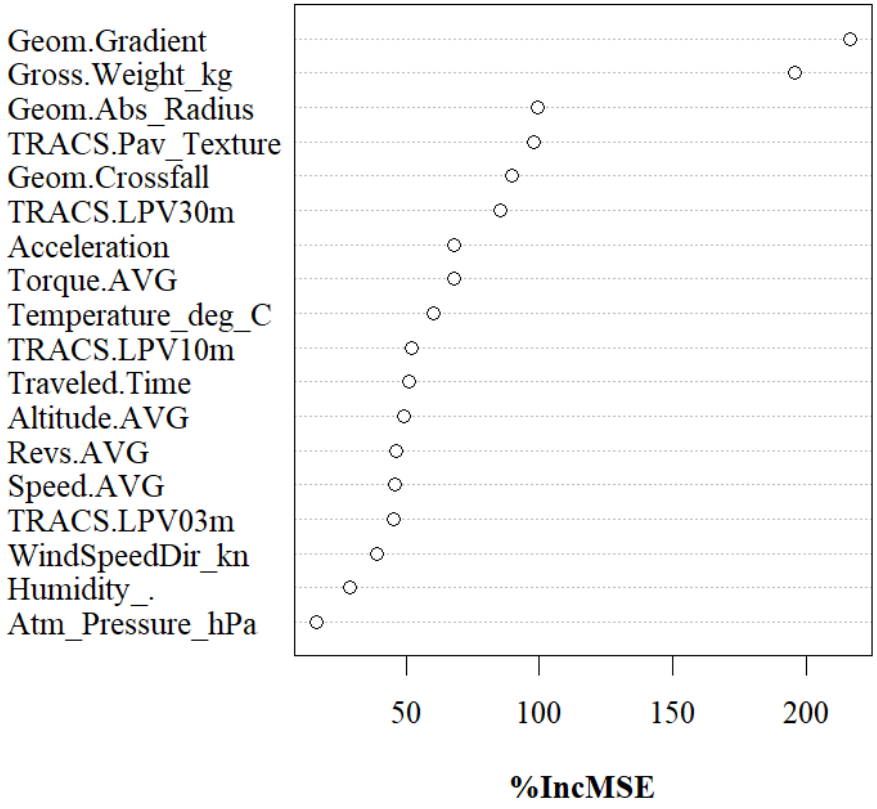


Figure 74. Rank of variable importance as increase in MSE for the medium trucks App 2, from the Dataset 2, as computed by the RF algorithm.

It is important to notice that (see Figure 74), for this truck type and conditions considered, *Torque.AVG* and *Revs.AVG* are ranked lower than some road characteristics. In fact, in this case, immediately after *Geom.Gradient* and *Gross.Weight_kg*, *Geom.Abs_Radius* and *TRACS.Pav_Texture* followed by *Geom.Crossfall* and *TRACS.LPV30m* are ranked. This means that, these variables have a high influence on the resulting MSE of a RF model developed using the available data. Also, TRACS.LPV03m is ranked quite low but still higher than other factors such as *WindSpeedDir_kn*, which should have much more considerable effect on *Fuel.Consumption*. However, this might be the case as the available data for wind

speed are measured every hour at the weather station location and not exactly where and when the truck drives, which can introduce some bias.

E.2.1.5 Heavy trucks on A roads (App 1)

From an analysis of the correlation matrix (see Figure 75), it is possible to see that, also for the heavy trucks App 1, the highest linearly correlated variables with *Fuel.Consumption* are *Geom.Gradient*, *Torque.AVG* and *Gross.Weight_kg*. Following these it is possible to see that analysis of the Pearson's correlation coefficients shows that *Revs.AVG* and *Acceleration* seem to have a quite significant role.

	Fuel.Consumption	Altitude.AVG	Traveled.Time	Gross.Weight_kg	Speed.AVG	Acceleration	Torque.AVG	Revs.AVG	Geom.Abs_Radius	Geom.Gradient	Geom.Crossfall	TRACS.LPV03m	TRACS.LPV10m	TRACS.LPV30m	TRACS.Pav_Texture	Temperature_deg_C	Humidity_.	Atm_Pressure_hPa	WindSpeedDir_kn
Fuel.Consumption	1	0.04	-0.02	0.33	-0.01	-0.21	0.69	0.17	-0.02	0.69	-0.08	-0.01	-0.01	-0.02	-0.05	-0.07	0.06	-0.01	0.02
Altitude.AVG	0.04	1	-0.06	0.02	0.07	0.01	0.03	0.01	-0.11	0.04	-0.32	-0.01	0.01	-0.01	-0.63	-0.09	0.12	0.01	0.01
Traveled.Time	-0.02	-0.06	1	-0.09	-0.23	-0.17	0.02	-0.05	0.09	0.01	0.03	0	0	-0.01	0.04	-0.01	0.01	0	0.01
Gross.Weight_kg	0.33	0.02	-0.09	1	-0.01	0.01	0.3	0.01	0.03	0.03	-0.02	-0.01	-0.01	-0.01	-0.03	0	0.03	-0.01	-0.01
Speed.AVG	-0.01	0.07	-0.23	-0.01	1	0.05	0	0.42	0.15	-0.02	-0.02	-0.11	-0.12	-0.15	0	-0.03	0.05	0	0
Acceleration	-0.21	0.01	-0.17	0.01	0.05	1	-0.13	-0.1	-0.03	-0.08	0	0.01	0.01	0.01	-0.01	0.01	-0.01	0	0
Torque.AVG	0.69	0.03	0.02	0.3	0	-0.13	1	0.15	0	0.49	-0.05	-0.02	-0.02	-0.03	-0.04	-0.07	0.05	-0.01	0.02
Revs.AVG	0.17	0.01	-0.05	0.01	0.42	-0.1	0.15	1	0.06	0.09	0	-0.03	-0.03	-0.05	0.04	-0.02	0.02	-0.01	0
Geom.Abs_Radius	-0.02	-0.11	0.09	0.03	0.15	-0.03	0	0.06	1	0.04	0.16	-0.04	-0.06	-0.16	0.08	0.02	0	-0.01	-0.01
Geom.Gradient	0.69	0.04	0.01	0.03	-0.02	-0.08	0.49	0.09	0.04	1	-0.07	0	-0.01	-0.03	-0.12	0.02	-0.01	0	-0.02
Geom.Crossfall	-0.08	-0.32	0.03	-0.02	-0.02	0	-0.05	0	0.16	-0.07	1	0.01	0	0.01	0.37	0.03	-0.02	-0.01	0
TRACS.LPV03m	-0.01	-0.01	0	-0.01	-0.11	0.01	-0.02	-0.03	-0.04	0	0.01	1	0.98	0.85	0	0	-0.01	0	0
TRACS.LPV10m	-0.01	0.01	0	-0.01	-0.12	0.01	-0.02	-0.03	-0.06	-0.01	0	0.98	1	0.93	-0.01	0	-0.01	0	0
TRACS.LPV30m	-0.02	-0.01	-0.01	-0.01	-0.15	0.01	-0.03	-0.05	-0.16	-0.03	0.01	0.85	0.93	1	0.02	0.01	-0.02	0	0
TRACS.Pav_Texture	-0.05	-0.63	0.04	-0.03	0	-0.01	-0.04	0.04	0.08	-0.12	0.37	0	-0.01	0.02	1	0.05	-0.06	0	0
Temperature_deg_C	-0.07	-0.09	-0.01	0	-0.03	0.01	-0.07	-0.02	0.02	0.02	0.03	0	0	0.01	0.05	1	-0.48	-0.17	-0.15
Humidity_.	0.06	0.12	0.01	0.03	0.05	-0.01	0.05	0.02	0	-0.01	-0.02	-0.01	-0.01	-0.02	-0.06	-0.48	1	-0.05	0.04
Atm_Pressure_hPa	-0.01	0.01	0	-0.01	0	0	-0.01	-0.01	-0.01	0	-0.01	0	0	0	0	-0.17	-0.05	1	0.21
WindSpeedDir_kn	0.02	0.01	0.01	-0.01	0	0	0.02	0	-0.01	-0.02	0	0	0	0	0	-0.15	0.04	0.21	1

Figure 75. Correlation matrix for the data from the heavy trucks App 1, from the Dataset 2.

Table 47 - Summary of the computed p-values of the correlated variables for the heavy trucks App 1, from the Dataset 2.

Variable	p-value	Significance level
Gross.Weight_kg	< 2.00 e^{-16}	99.9 %
Geom.Gradient	< 2.00 e^{-16}	99.9 %
Torque.AVG	< 2.00 e^{-16}	99.9 %
Revs.AVG	< 2.00 e^{-16}	99.9 %
Speed.AVG	< 2.00 e^{-16}	99.9 %
Acceleration	< 2.00 e^{-16}	99.9 %
Traveled.Time	< 2.00 e^{-16}	99.9 %
Geom.Abs_Radius	< 2.00 e^{-16}	99.9 %
Geom.Crossfall	< 2.00 e^{-16}	99.9 %
TRACS.LPV03m	< 2.00 e^{-16}	99.9 %
TRACS.LPV10m	< 2.00 e^{-16}	99.9 %
TRACS.LPV30m	< 2.00 e^{-16}	99.9 %
TRACS.Pav_Texture	< 2.00 e^{-16}	99.9 %
Altitude.AVG	< 2.00 e^{-16}	99.9 %
Atm_Pressure_hPa	< 2.00 e^{-16}	99.9 %
Temperature_deg_C	< 2.00 e^{-16}	99.9 %
Humidity_.	1.63 e^{-13}	99.9 %
Delta.Angle_deg	0.648	0 %
WindSpeedDir_kn	< 2.00 e^{-16}	99.9 %
WindSpeed.Cro_kn	0.342	0 %

Correlation is very low instead with LPV and SMTD. Again, from Figure 75, it is possible to notice that correlation of LPV and *Speed.AVG* is quite significant (negative), and that *Altitude.AVG* correlates quite well with *TRACS.Pav_Texture* (SMTD).

In Table 47, results of the analysis of *p-values* for the heavy trucks App 1 are reported. As it is possible to see, *Delta.Angle_deg* and *WindSpeed.Cro_kn* are the only variables which coefficients are considered to be insignificant for a LR of the variables to estimate *Fuel.Consumption*. All the others, are confirmed to play a significant role at 99.9% significance level, including LPV at 3, 10 and 30 metres wavelength and TRACS.Pav_Texture measurements. Again, this is a confirmation of the findings of the other case studies analysed in this study, which confirm the importance of considering these variables in modelling the fuel consumption of the considered vehicles.

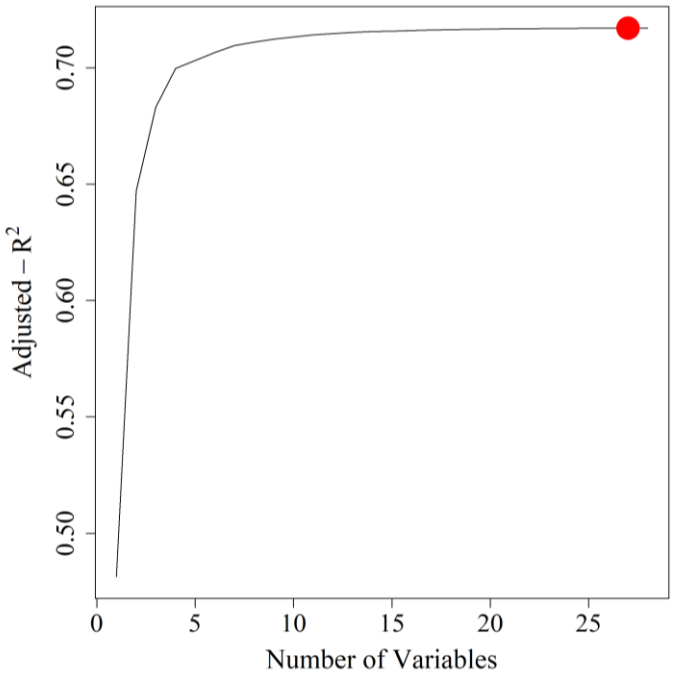


Figure 76. Analysis of the adjusted-R^2 for the heavy trucks App 1, from the Dataset 2.

Figure 76 reports results of the analysis of the adjusted-R^2 for the heavy trucks App 1. In the figure, it is possible to see that 28 out of 77 variables available are identified to contribute at improving the accuracy of an LR, including LPV and SMTD

measurements. Also in this case, it is possible to notice that, after approximately 20 variables added to the model, no significant improvement of the adjusted-R^2 is obtained, with small improvements.

Table 48 - Summary of the Lasso coefficients for the heavy trucks App 1, from the Dataset 2.

Variable	Lasso Coefficient
Gross.Weight_kg	0.201
Geom.Gradient	0.508
Torque.AVG	0.361
Revs.AVG	0.116
Speed.AVG	-0.039
Acceleration	-0.130
Traveled.Time	-0.031
Geom.Abs_Radius	-0.043
Geom.Crossfall	-0.026
TRACS.LPV03m	0
TRACS.LPV10m	-0.0019
TRACS.LPV30m	0
TRACS.Pav_Texture	0.057
Altitude.AVG	-0.0015
Atm_Pressure_hPa	-0.0012
Temperature_deg_C	-0.053
Humidity_.	-0.061
Delta.Angle_deg	0.0047
WindSpeedDir_kn	0.020
WindSpeed.Cro_kn	0.0069

Table 48 reports results of the Lasso analysis performed for the heavy trucks App 1. Results show that *TRACS.LPV03m* and *TRACS.LPV30m* are identified to be insignificant by the algorithm, to model *Fuel.Consumption*. For this case study, all the other variables seem to be considered important by the Lasso.

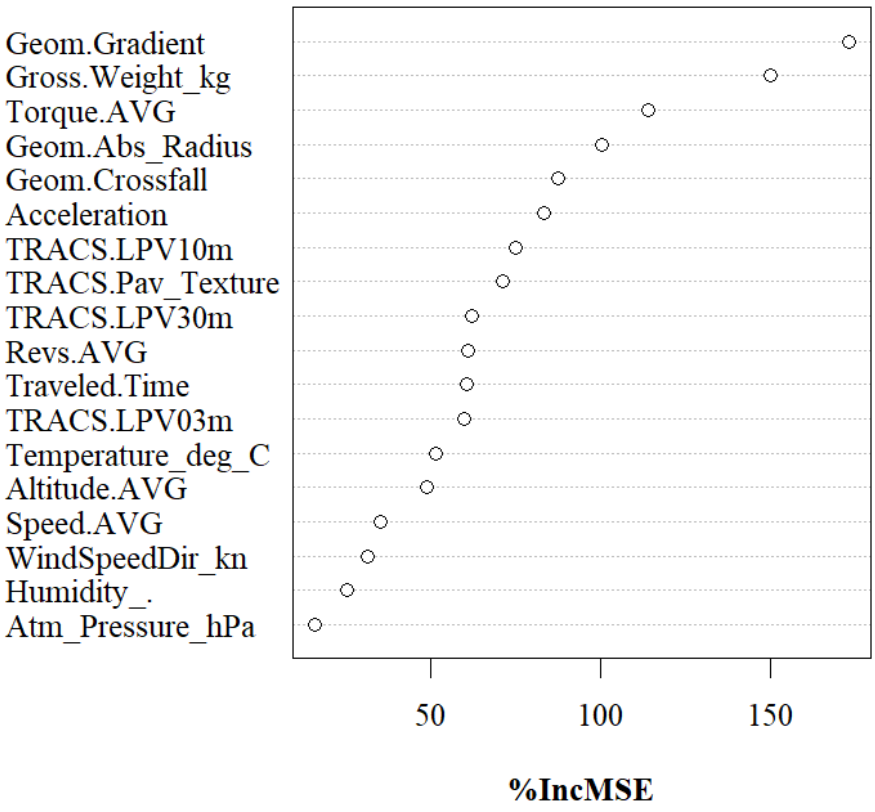


Figure 77. Rank of variable importance as increase in MSE the heavy trucks App 1, from the Dataset 2, as computed by the RF algorithm.

Finally, Figure 77 show the results of the rank of variables made by the RF algorithm. Again, while some variables such as *Geom.Gradient*, *Gross.Weight_kg* and *Torque.AVG* are placed at the very top of the rank (as it was the case for the other case study considered), *Geom.Abs_Radius*, *Geom.Crossfall*, *TRACS.LPV10m*, *TRACS.Pav_Texture* and TRACS.LPV30m are ranked just after them. Also in this

case, *TRACS.LPV03m* seems to be less important than other variables, but yet more important than *Speed.AVG* and weather conditions data.

E.2.1.6 Heavy trucks on Motorways and A roads (App 2)

Finally, results of the variable selection analysis conducted on the heavy trucks App 2 are reported. Figure 78 shows results of the analysis of Pearson's correlation coefficients using correlation matrices.

	Fuel.Consumption	Altitude.AVG	Traveled.Time	Gross.Weight_kg	Speed.AVG	Acceleration	Torque.AVG	Revs.AVG	Geom.Abs_Radius	Geom.Gradient	Geom.Crossfall	TRACS.LPV03m	TRACS.LPV10m	TRACS.LPV30m	TRACS.Pav_Texture	Temperature_deg_C	Humidity_.	Atm_Pressure_hPa	WindSpeedDir_kn
Fuel.Consumption	1	0.03	-0.02	0.26	-0.03	-0.21	0.74	0.19	-0.03	0.76	-0.05	-0.02	-0.03	-0.04	0.03	-0.05	0.04	-0.01	0.03
Altitude.AVG	0.03	1	-0.15	0.02	0.16	0.02	0.03	0.08	-0.15	0.01	-0.14	-0.08	-0.03	-0.01	-0.29	-0.03	0.03	0.01	0.03
Traveled.Time	-0.02	-0.15	1	-0.11	-0.28	-0.2	0.01	-0.05	0.03	-0.01	0.07	0.06	0.06	0.02	0.06	0	0.01	0	-0.01
Gross.Weight_kg	0.26	0.02	-0.11	1	-0.01	0.04	0.24	0.01	0.02	0	-0.02	-0.01	-0.02	-0.02	-0.03	0	0.03	-0.01	0.01
Speed.AVG	-0.03	0.16	-0.28	-0.01	1	0.07	-0.02	0.38	0.15	-0.02	-0.03	-0.14	-0.16	-0.21	-0.03	-0.04	0.04	0.01	0.01
Acceleration	-0.21	0.02	-0.2	0.04	0.07	1	-0.13	-0.12	-0.01	-0.06	0	0	0	0.02	-0.02	0.01	0	0	-0.01
Torque.AVG	0.74	0.03	0.01	0.24	-0.02	-0.13	1	0.16	-0.03	0.58	-0.04	-0.04	-0.04	-0.04	-0.01	-0.05	0.04	-0.01	0.03
Revs.AVG	0.19	0.08	-0.05	0.01	0.38	-0.12	0.16	1	0.06	0.12	-0.01	-0.07	-0.08	-0.09	0.03	-0.02	0.01	0	0.01
Geom.Abs_Radius	-0.03	-0.15	0.03	0.02	0.15	-0.01	-0.03	0.06	1	0.07	0.17	-0.04	-0.14	-0.37	0.06	0.03	-0.05	0	0
Geom.Gradient	0.76	0.01	-0.01	0	-0.02	-0.06	0.58	0.12	0.07	1	-0.03	0	-0.04	-0.1	0.01	0.02	-0.02	0	-0.01
Geom.Crossfall	-0.05	-0.14	0.07	-0.02	-0.03	0	-0.04	-0.01	0.17	-0.03	1	0.01	0	-0.02	0.25	0.01	-0.01	0.01	-0.01
TRACS.LPV03m	-0.02	-0.08	0.06	-0.01	-0.14	0	-0.04	-0.07	-0.04	0	0.01	1	0.94	0.72	0	0.02	-0.03	0	-0.01
TRACS.LPV10m	-0.03	-0.03	0.06	-0.02	-0.16	0	-0.04	-0.08	-0.14	-0.04	0	0.94	1	0.84	0.01	0.01	-0.01	0	-0.01
TRACS.LPV30m	-0.04	-0.01	0.02	-0.02	-0.21	0.02	-0.04	-0.09	-0.37	-0.1	-0.02	0.72	0.84	1	0.05	0	0	-0.01	-0.02
TRACS.Pav_Texture	0.03	-0.29	0.06	-0.03	-0.03	-0.02	-0.01	0.03	0.06	0.01	0.25	0	0.01	0.05	1	0.02	-0.03	0	0.01
Temperature_deg_C	-0.05	-0.03	0	0	-0.04	0.01	-0.05	-0.02	0.03	0.02	0.01	0.02	0.01	0	0.02	1	-0.51	-0.16	-0.1
Humidity_.	0.04	0.03	0.01	0.03	0.04	0	0.04	0.01	-0.05	-0.02	-0.01	-0.03	-0.01	0	-0.03	-0.51	1	-0.02	0.04
Atm_Pressure_hPa	-0.01	0.01	0	-0.01	0.01	0	-0.01	0	0	0	0.01	0	0	-0.01	0	-0.16	-0.02	1	0.15
WindSpeedDir_kn	0.03	0.03	-0.01	0.01	0.01	-0.01	0.03	0.01	0	-0.01	-0.01	-0.01	-0.01	-0.02	0.01	-0.1	0.04	0.15	1

Figure 78. Correlation matrix for the data from heavy trucks App 2, from the Dataset 2.

Table 49 - Summary of the computed *p-values* of the correlated variables for the heavy trucks App 2, from the Dataset 2.

Variable	p-value	Significance level
Gross.Weight_kg	< 2.00 e^{-16}	99.9 %
Geom.Gradient	< 2.00 e^{-16}	99.9 %
Torque.AVG	< 2.00 e^{-16}	99.9 %
Revs.AVG	< 2.00 e^{-16}	99.9 %
Speed.AVG	< 2.00 e^{-16}	99.9 %
Acceleration	< 2.00 e^{-16}	99.9 %
Traveled.Time	< 2.00 e^{-16}	99.9 %
Geom.Abs_Radius	< 2.00 e^{-16}	99.9 %
Geom.Crossfall	< 2.00 e^{-16}	99.9 %
TRACS.LPV03m	< 2.00 e^{-16}	99.9 %
TRACS.LPV10m	0.114	95 %
TRACS.LPV30m	< 2.00 e^{-16}	99.9 %
TRACS.Pav_Texture	< 2.00 e^{-16}	99.9 %
Altitude.AVG	< 2.00 e^{-16}	99.9 %
Atm_Pressure_hPa	< 2.00 e^{-16}	99.9 %
Temperature_deg_C	< 2.00 e^{-16}	99.9 %
Humidity_.	1.53 e^{-07}	99.9 %
Delta.Angle_deg	0.381	0 %
WindSpeedDir_kn	< 2.00 e^{-16}	99.9 %
WindSpeed.Cro_kn	0.472	0 %

Again, from an analysis of the figures it is possible to see that the variables which show the highest linear correlation with *Fuel.Consumption* are *Geom.Gradient*, *Torque.AVG*, and *Gross.Weight_kg*, while road surface condition measurements show poor correlation with that. Additionally, it is possible to see that LPV measurements show some correlation with *Speed.AVG*, meaning that lower speeds are associated to

higher (worse) road profile variance. This is shown by all the cases considered in the study, and it might be an important consideration to make when drawing conclusions for the study.

Table 49 shows results of the analysis of p-values for the heavy trucks App 2. The table shows that all measurements, apart from *Delta.Angle_deg* and *WindSpeed.Cro_kn*, show significant regression coefficients which should help in improving the developed LR model. From Table 49, it is also possible to see that from the analysis of the p-values, it is not possible to exclude road surface condition measurements from the LR model for the considered truck type and conditions. This matches the observations made on the same vehicle type driving along A roads only (see Appendix E.2.1.5), and those of the other cases considered (from Appendix E.2.1.1 to E.2.1.4).

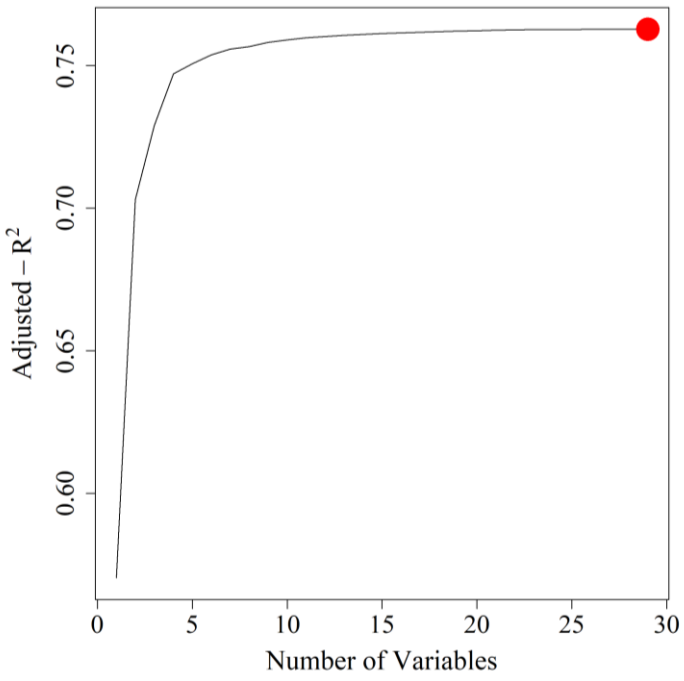


Figure 79. Analysis of the adjusted-R^2 for the heavy trucks App 2, from the Dataset 2.

In Figure 79, analysis of the adjusted-R^2 is represented. From the figure it is possible to see that the algorithm finds at 30 out of 77 variables available for the study the optimal amount to be included in the regression analysis. Again, it is possible to notice also that after 20 variables the increase in adjusted-R^2 is small.

Table 50 - Summary of the Lasso coefficients for the heavy trucks App 2, from the Dataset 2.

Variable	Lasso Coefficient
Gross.Weight_kg	0.177
Geom.Gradient	0.549
Torque.AVG	0.350
Revs.AVG	0.118
Speed.AVG	-0.049
Acceleration	-0.132
Traveled.Time	-0.025
Geom.Abs_Radius	-0.034
Geom.Crossfall	-0.017
TRACS.LPV03m	-0.017
TRACS.LPV10m	0
TRACS.LPV30m	0.029
TRACS.Pav_Texture	0.038
Altitude.AVG	0.0072
Atm_Pressure_hPa	-0.012
Temperature_deg_C	-0.043
Humidity_.	-0.0021
Delta.Angle_deg	-0.0006
WindSpeedDir_kn	0.025
WindSpeed.Cro_kn	0.0056

Table 50 reports results of the Lasso analysis performed on the heavy trucks App 2. Results show that the only variable which is recognised by the algorithm to play an insignificant role in the modelling of fuel consumption for the considered fleet of trucks is the *TRACS.LPV10m.* This means that inclusion of variables such as *Delta.Angle_deg* or *WindSpeed.Cro_kn* should help the developed model to improve accuracy of its final estimates for the considered truck type driving along M and A roads (App 2).

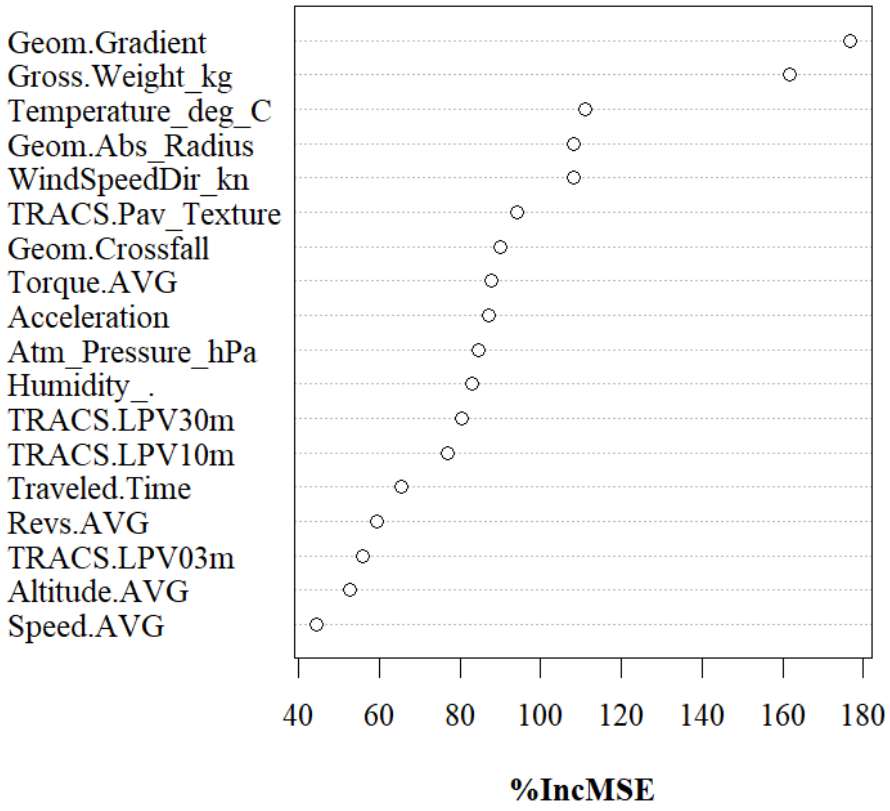


Figure 80. Rank of variable importance as increase in MSE for the heavy trucks App 2, from the Dataset 2, as computed by the RF algorithm.

Finally, results from the rank of variables performed by the RF is reported (see Figure 80) and analysed. Again, from an analysis of Figure 80 it is possible to see that *Geom.Gradient* and *Gross.Weight_kg* are the variables, which are ranked to

influence the fuel consumption of the considered fleet of trucks the most. After Geom.Gradient and Gross.Weight_kg, for the heavy trucks App 2 the algorithm ranks *Temperature_deg_C*, *Geom.Abs_Radius*, *WindSpeedDir_kn*, and *TRACS.Pav_Texture*. From the figure, it is possible to see also that, for this vehicle type and situation considered, road conditions seem to have a lower impact than shown for other vehicles in similar situation (from Appendix E.2.1.1 to E.2.1.4).

Appendix E.2.2 Modelling 2

This section of the thesis is to summarise results of the modelling phase of the study. Again, also in this section, results are discussed separately for different vehicle types, separate datasets and applications of the developed method.

E.2.2.1 Light trucks on A roads (App 1)

Fitting an LR to the selection of variables listed in Appendix E.2.1.1, the Equation (31), LR for light trucks App 1, is obtained.

$$\begin{aligned} Fuel.Consumption &= 14.38 + 6.7\times10^{-4}\times Gross.Weight + 2.81\times Geom.Gradient - 0.075\times Speed.AVG + \\ &+0.11\times\ Torque.AVG + 0.008\times Revs.AVG - 13.23\times Acceleration - 3.36\times10^{-4}\times Traveled.Time + \\ &-8.65\times10^{-4}\times Geom.Abs_Radius - 0.077\times Geom.Crossfall + 0.012\times TRACS.LPV03m + 0.007\times TRACS.LPV10m + \\ &-7.77\times10^{-4}\times TRACS.LPV30m + 0.57\times TRACS.Pav_Texture - 0.005\times Altitude.AVG - 0.013\times Atm_Pressure_hPa + \\ &-0.068\times Temperature_deg_C - 0.025\times Humidity_. + 0.0029\times WindSpeed.Dir_kn \end{aligned} \quad (31)$$

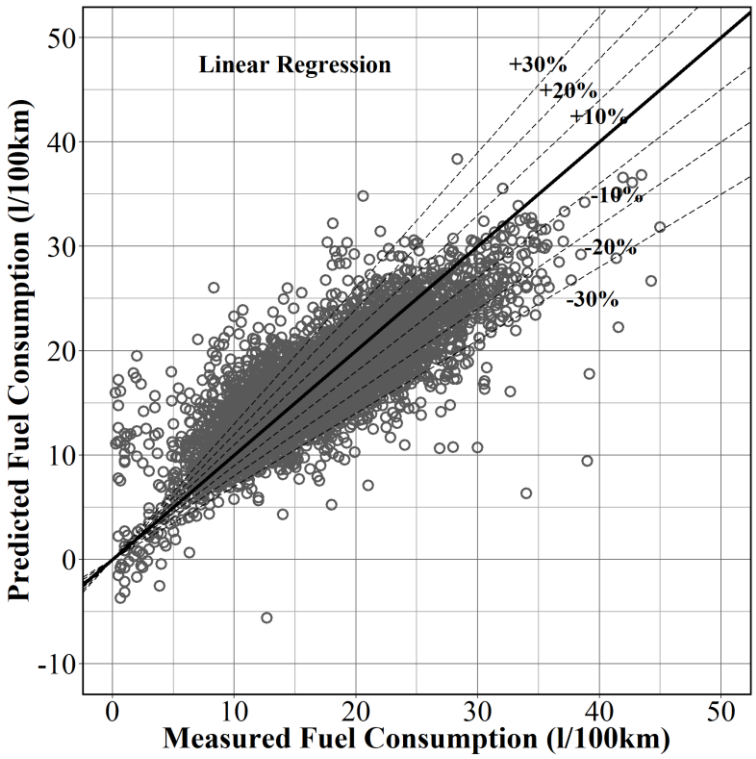


Figure 81. Fit of the developed LR for the light trucks App 1, from the Dataset 2.

Using the Equation (31), it is then possible to fit the developed model to the available dataset for the light trucks App 1 (see Figure 81). From the figure, it is possible to see that, apart from some outliers, the model tends to fit quite well low and average fuel consumption, tending to underestimate high measurements of fuel consumption. This might be due to the high impact of non-linearity in these other cases.

For this set, a total of 12,844 records are available. Of these, 9,633 were used to train SVR, RF and ANN, and 3,211 to test them. For this truck type, because of the reduced amount of data available, no validation set is used and performance are evaluated on the test set, averaged across the ten-fold cross-validation process (see sect. 3.4.2).

For this case study, the developed SVR model was a regression model, with RBF kernel function and tolerance set to 0.001. Figure 82 shows the fit of the test set for the ten-fold cross-validation process applied to the SVR model for the light trucks App 1 dataset. From the figure it is possible to see that the SVR tends to be quite robust in the estimates made on the test set.

Figure 83 shows the fit of the developed SVR model to the light trucks App 1 dataset. The figure shows that there is some improvement from the fit of the LR model (see Figure 82), comparable to the performance of RF (see Figure 85) and ANN (see Figure 87) (see Table 51). In fact, although some outliers and the low amount of data available, the model tends to estimate quite well the whole range of the observed fuel consumption.

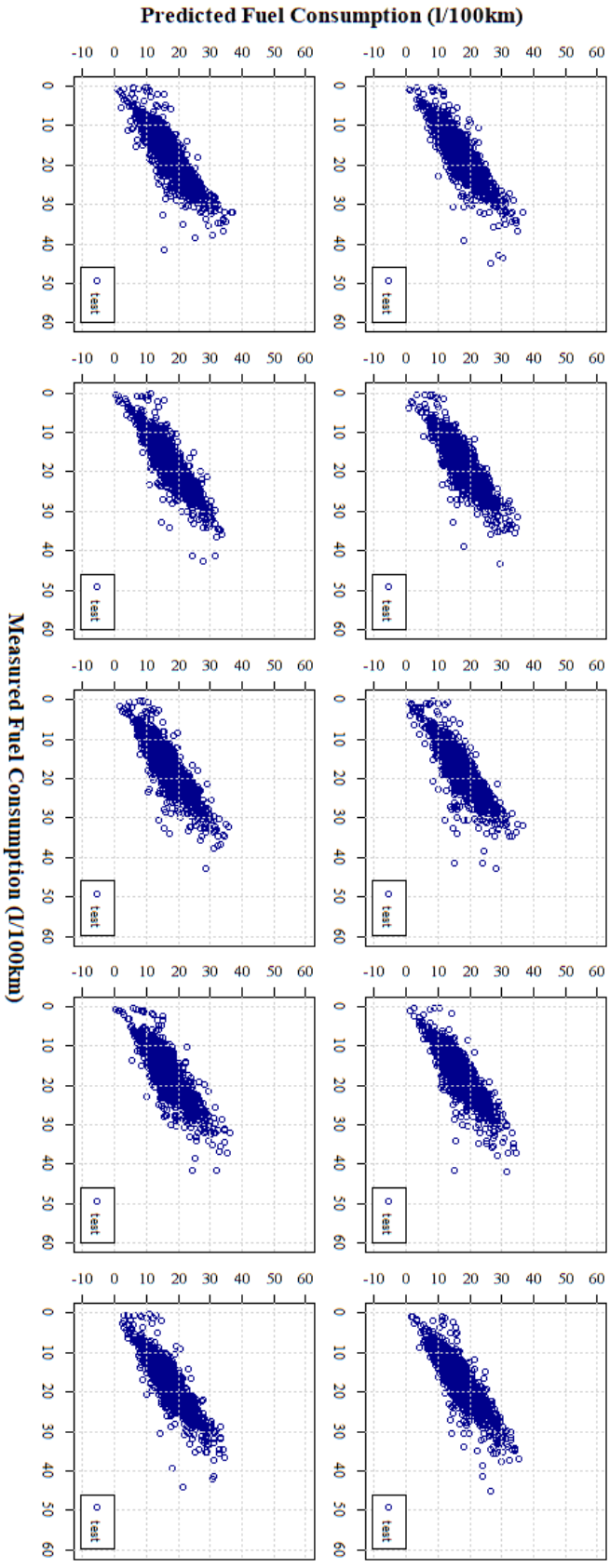


Figure 82. Ten-fold cross-validation of the SVR model for the light trucks App 1, from the Dataset 2.

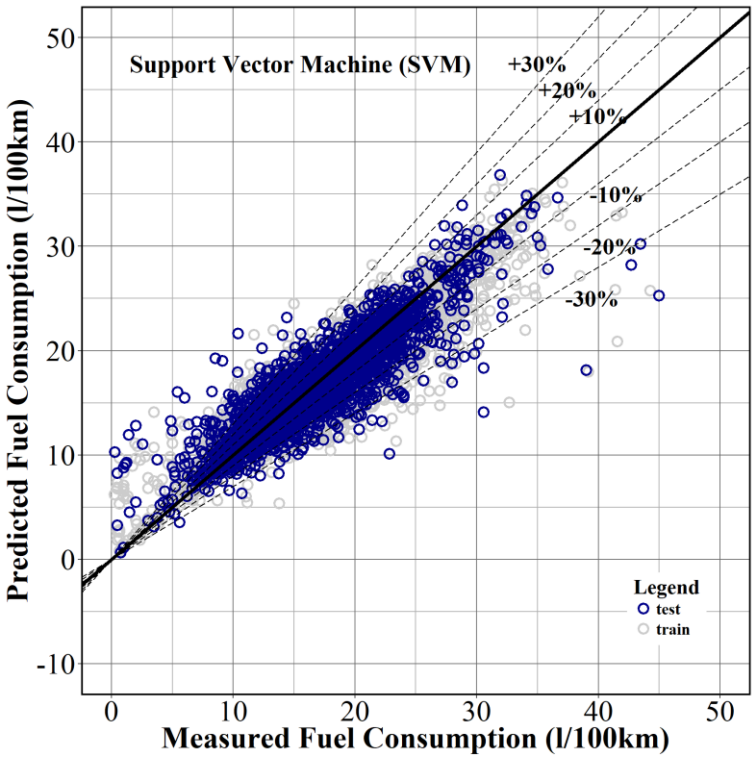


Figure 83. Fit of the developed SVR model for the light trucks App 1, from the Dataset 2.

The developed RF is composed of 250 trees and in each of these it uses 6 randomly sampled variables to make predictions. Figure 84 show the results of the cross-validation process applied to the RF model for the light trucks App 1. Also in this case, the model demonstrates to be quite robust at making estimates on different test sets.

Figure 85 shows the fit of the RF model to the data from the light trucks App 1. The model seems the best so far with very small variance, but some overfitting at the extremes. This especially happens on the training set and is due to characteristics of RF models which, when the number of data samples is low, generally tend to overfit the training set (Breiman, 2001).

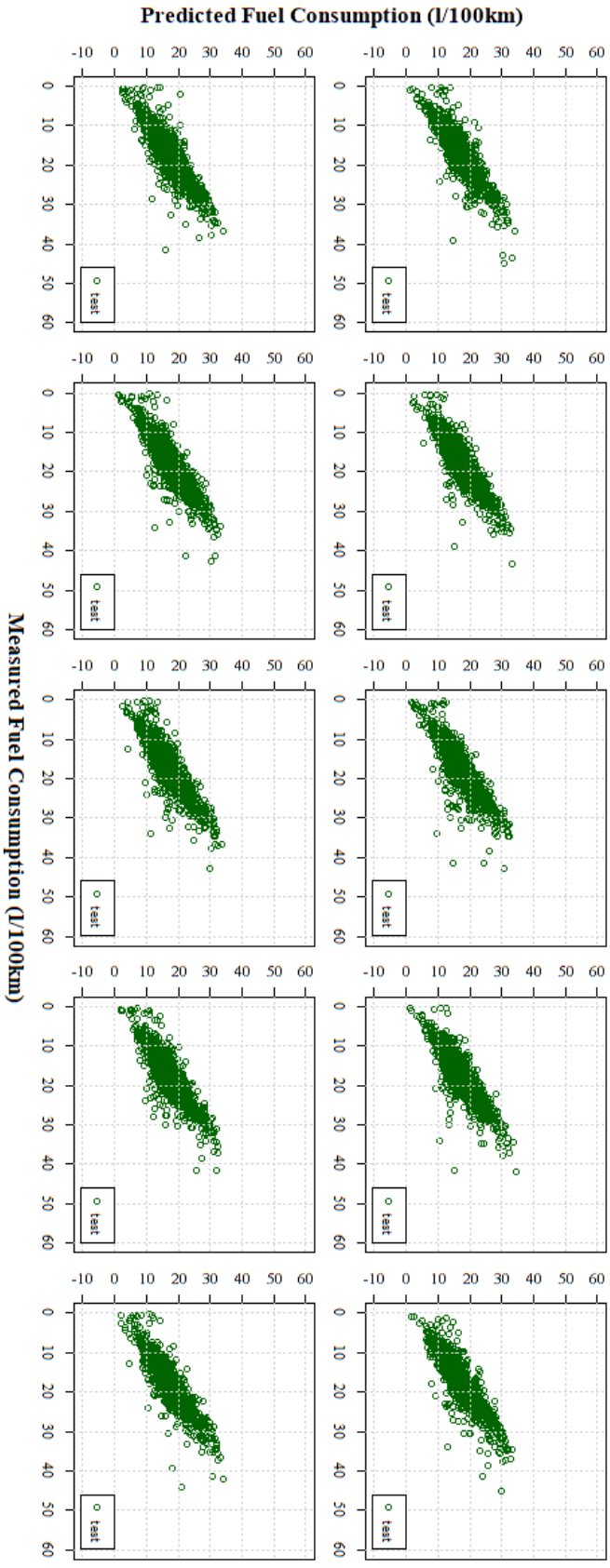


Figure 84. Ten-fold cross-validation of the RF model for light trucks App 1, from the Dataset 2.

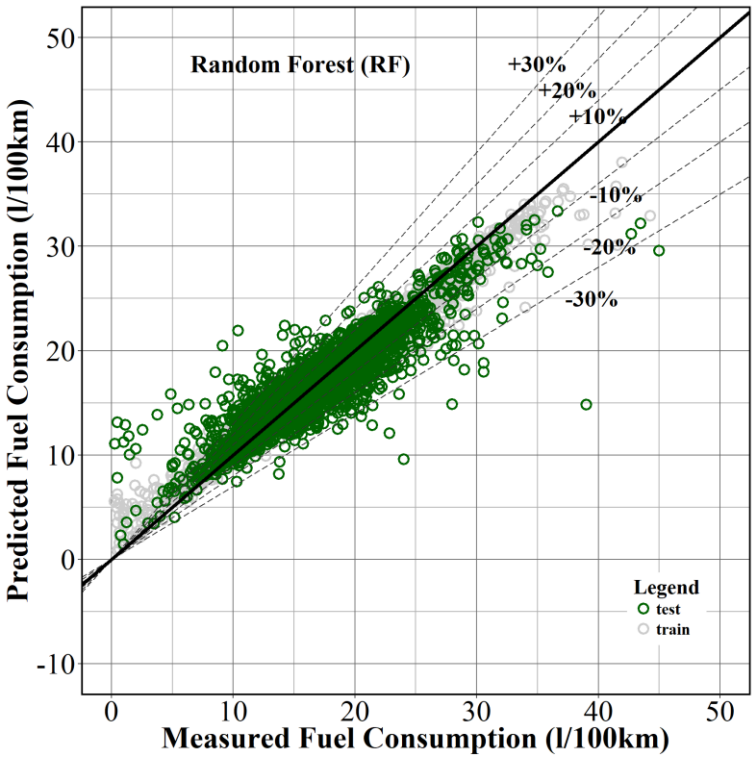


Figure 85. Fit of the developed RF model for the light trucks App 1, from the Dataset 2.

This also causes the model to not be able to estimate some extreme values of fuel consumption (low and high) (see Figure 85). Again, from the figure, it is possible to say that the RF outperforms the LR model (see Figure 82) and, apart from some outliers, which could not be fitted by any of the investigated techniques, it is fair to say that the RF tends to estimate quite well the whole range of the observed fuel consumption, better than the SVR (see Figure 83) and the ANN (see Figure 87) (see Table 51).

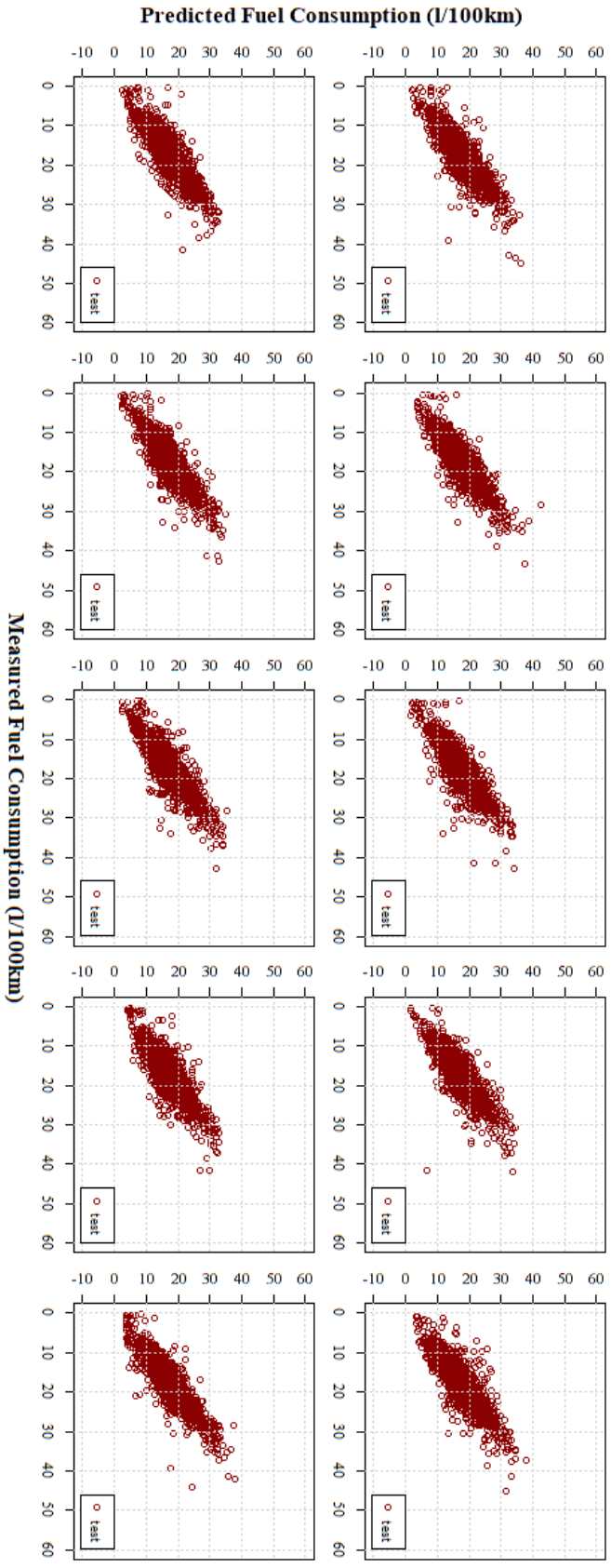


Figure 86. Ten-fold cross-validation of the ANN model for the light trucks App 1, from the Dataset 2.

Then, the developed ANN and its performance are considered. The ANN is structured on two layers of 14 and 4 neurons respectively. The model was trained using the rprop- algorithm (Riedmiller *et al.*, 1993), and it used the logistic activation function to identify patterns through the data. Figure 86 shows how the ANN model performs on the test set under the cross-validation process. From the figure it is possible to say that the ANN is quite robust and that split of the available data does not influence the performance of the developed model significantly.

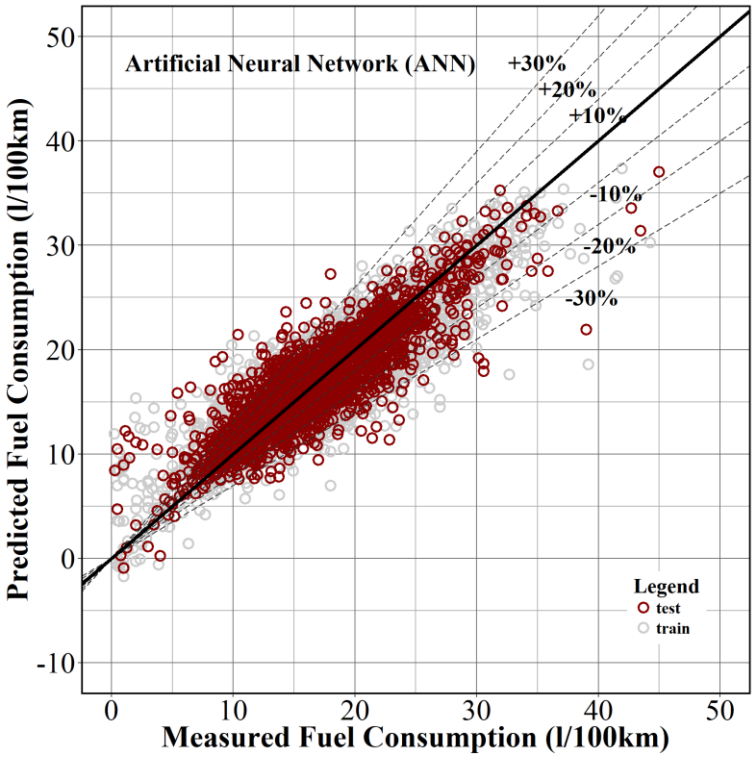


Figure 87. Fit of the developed ANN model to the data for the light trucks App 1, from the Dataset 2.

Figure 87 shows the fit of the developed ANN model to the data for the light trucks App 1. The model shows similar performance to those of the SVR model (see Figure 83), higher than that of the LR model (see Figure 82), but slightly lower than

the RF (see Figure 85) (see Table 51). It is possible to see that, like in the case of SVR and RF, the ANN cannot fit some outliers, but generally performs well overall. This is common also to the SVR and the RF models and is probably due to the lack of examples for the ML algorithms to detect any trend for these records.

Finally, for the light trucks App 1, summary of the performance of the developed models is reported (see Table 51). From the table, it is possible to see that the model that performs best at estimating the overall fuel consumption of light trucks App1 is the RF. The model, in fact, shows the highest R^2, and lowest RMSE and MAE, both in the cross-validation process, and on the validation set, in line with performance of the models recently calibrated and currently existing in the literature (e.g. Chatti *et al.*, 2012). The only negative aspect of the RF is the cross-validation time which, although still reasonable, for this set of data is over two times higher than that of SVR and ANN.

Table 51 - Summary of performance of the regression models developed for the light trucks App 1.

Model	R^2_{cv}	$RMSE_{cv}$	MAE_{cv}	$Time_{cv}$	R^2_{test}	$RMSE_{test}$	MAE_{test}
	(-)	*(l/100 km)*	*(l/100 km)*	*(mins)*	*(-)*	*(l/100 km)*	*(l/100 km)*
LR	0.69	2.71	1.93	-	-	-	-
SVR	0.80	2.23	1.53	8.63	0.79	2.26	1.56
RF	0.82	2.13	1.46	20.15	0.81	2.15	1.44
ANN	0.79	2.34	1.67	7.51	0.78	2.30	1.64
N.B.: The subscript 'test' stays for test set (one of the ten of cross-validation), and 'cv' for cross-validation.							

E.2.2.2 Light trucks on Motorways and A roads (App 2)

Fitting an LR to the selection of variables listed in Appendix E.2.1.2, the Equation (32), LR for light trucks App 2, is obtained.

$$
\begin{aligned}
&Fuel.Consumption = 13.92 + 6.7\times10^{-4}\times Gross.Weight + 3.13\times Geom.Gradient - 0.035\times Speed.AVG + \\
&+0.099\times\ Torque.AVG + 0.007\times Revs.AVG - 15.83\times Acceleration - 0.0045\times Traveled.Time + \\
&-3.22\times10^{-4}\times Geom.Abs_Radius - 0.043\times Geom.Crossfall + 0.010\times TRACS.LPV03m - 0.11\times TRACS.LPV10m + \\
&-0.034\times TRACS.LPV30m + 0.57\times TRACS.Pav_Texture - 0.002\times Altitude.AVG - 0.014\times Atm_Pressure_hPa + \\
&-0.080\times Temperature_deg_C - 0.025\times Humidity_. + 0.003\times WindSpeed.Dir_kn
\end{aligned}
\qquad (32)
$$

Using the Equation (32) it is then possible to fit the developed model to the available dataset for the light trucks App 2 (see Figure 88). From the figure, it is possible to see that, also in this case, apart from some outliers, while the model tends to fit well low and average fuel consumption, this generally tends also to underestimate the high fuel consumption. Again, this might be due to the high impact of non-linearity in these cases.

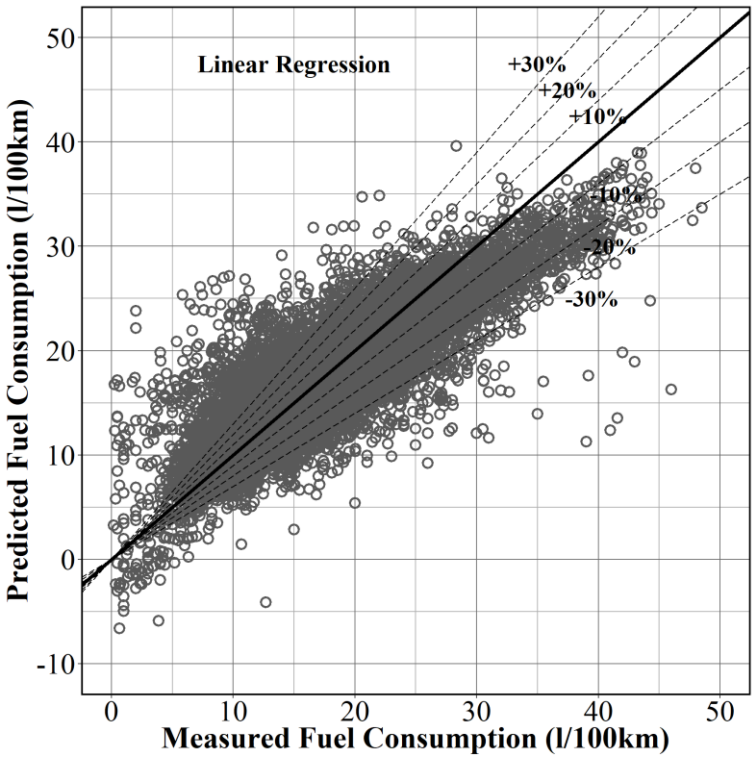


Figure 88. Fit of the developed LR for the light trucks App 2, from the Dataset 2.

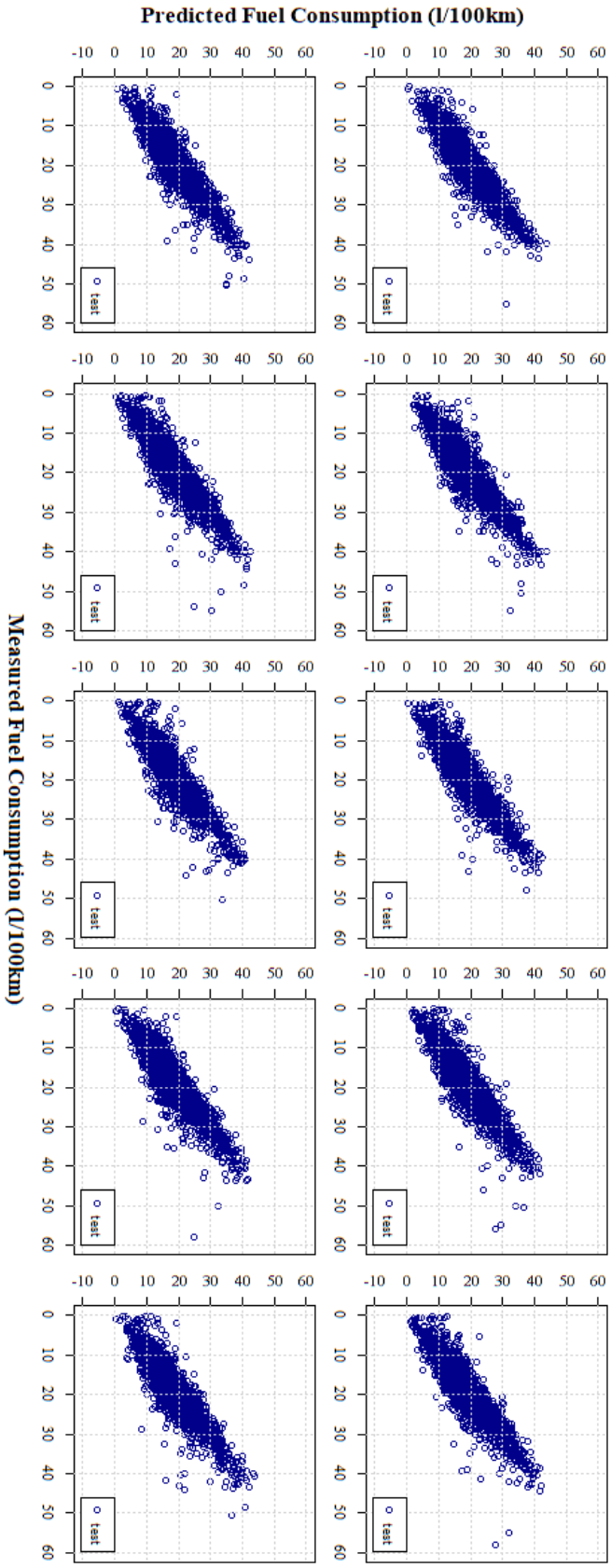


Figure 89. Ten-fold cross-validation of the SVR model for the light trucks App 2, from the Dataset 2.

For this set, 119,661 records are available. Of these, 85,199 were used to train SVR, RF and ANN, 28,400 to test, and 6,062 to validate the developed models on data from trucks not considered in the other sets of data.

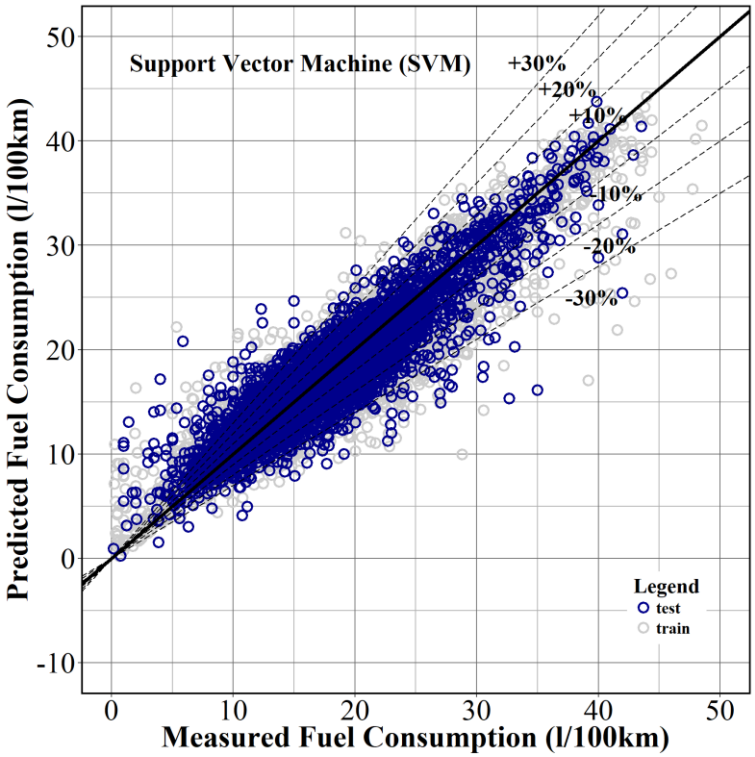


Figure 90. Fit of the developed SVR for the light trucks App 2, from the Dataset 2.

For this case study, the SVR model was a regression model, with RBF kernel function and tolerance set to 0.001. Figure 89 shows the fit of the test set for the ten-fold cross-validation process applied to the SVR model for the light trucks App 2 dataset. From the figure it is possible to see that the SVR tends to be quite robust and consistent in the estimates made on the test set.

Figure 90 show the fit of the developed SVR model to the light trucks App 2 dataset. The figure shows that there is quite visible improvement from the fit of the

LR model (see Figure 88), comparable to the performance of RF (see Figure 91) and ANN (see Figure 93) (see Table 52). In this case, although some outliers and the amount of data available, the model performs quite well over the whole range of fuel consumption.

The developed RF is composed of 250 trees and in each of these it uses 6 randomly sampled variables to make predictions. Figure 92 show the results of the cross-validation process applied to the RF model for the light trucks App 2. Also in this case, the model demonstrates to be quite robust at making estimates on different test sets.

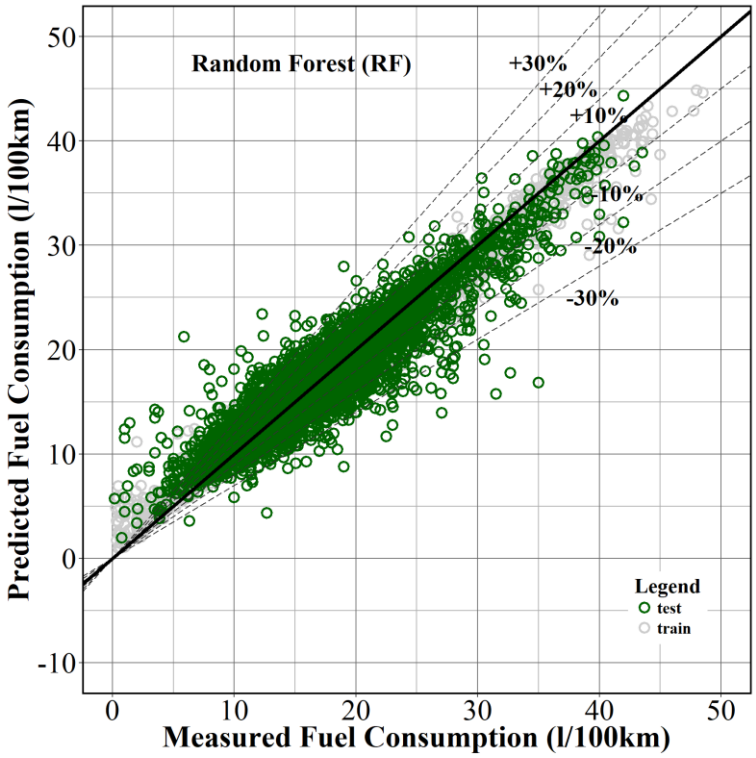


Figure 91. Fit of the developed RF to light trucks for the light trucks App 2, from the Dataset 2.

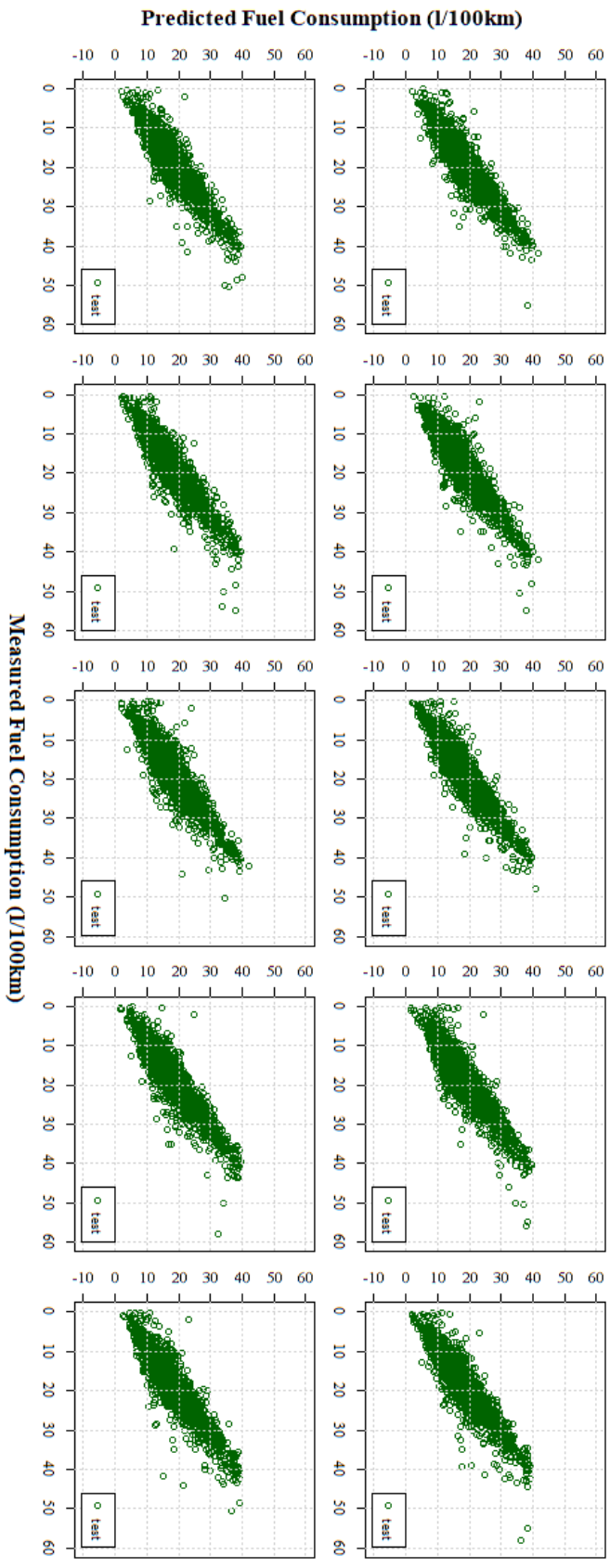


Figure 92. Ten-fold cross-validation of the RF model for the light trucks App 2, from the Dataset 2.

Figure 91 shows the fit of the RF model to the data from the light trucks App 1. The model seems to be the best so far, with very small variance, and not much overfitting at the extremes (which was the case for the light trucks App 1, see Appendix E.2.2.1). Again the model cannot fit some outliers, but it generally performs better than any of the other developed models (LR, SVR and ANN) (see Table 52).

Then, the developed ANN and its performance are considered. The ANN is structured on two layers of 14 and 4 neurons respectively. The model was trained using the rprop- algorithm (Riedmiller *et al.*, 1993), and it used the logistic activation function to identify patterns through the data.

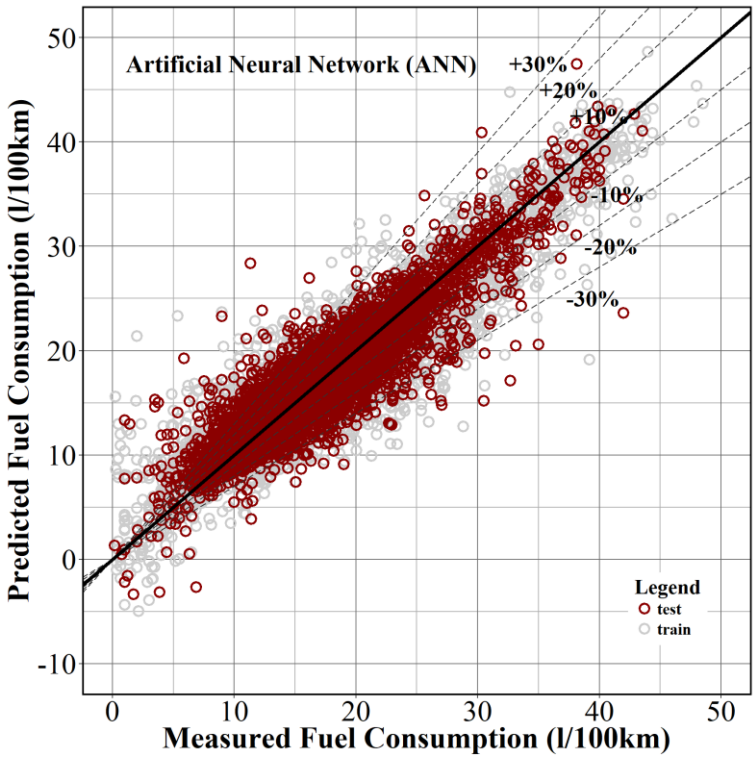


Figure 93. Fit of the developed ANN for the light trucks App 2, from the Dataset 2.

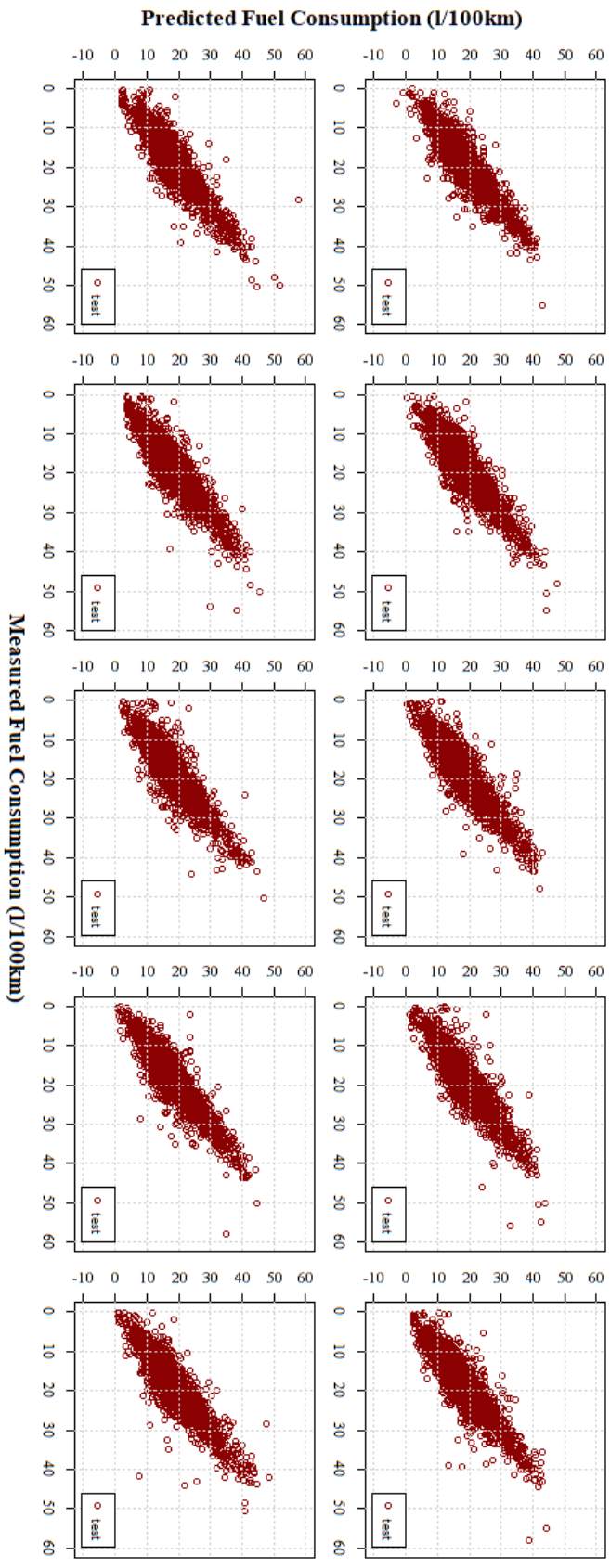


Figure 94. Ten-fold cross-validation of the ANN model for the light trucks App 2, from the Dataset 2.

Figure 94 shows how the ANN model performs on the test set under the cross-validation process. From the figure it is possible to say that the ANN is quite robust and that split of the available data does not influence the performance of the developed model significantly.

Figure 93 shows the fit of the developed ANN model to the data for the light trucks App 2. The model shows similar performance to those of the SVR model (see Figure 90), higher than that of the LR model (see Figure 88), but lower than the RF (see Figure 91) (see Table 52).

Table 52 - Summary of performance of the regression models developed for the light trucks App 2.

Model	R^2_{cv}	$RMSE_{cv}$	MAE_{cv}	$Time_{cv}$	R^2_{test}	$RMSE_{test}$	MAE_{test}
	(-)	*(l/100 km)*	*(l/100 km)*	*(mins)*	*(-)*	*(l/100 km)*	*(l/100 km)*
LR	0.75	2.93	2.16	-	-	-	-
SVR	0.85	1.95	1.47	745	0.86	1.93	1.45
RF	0.87	1.87	1.37	833	0.88	1.86	1.38
ANN	0.84	1.99	1.51	802	0.86	1.97	1.49
N.B.: The subscript 'test' stays for test set (one of the ten of cross-validation), and 'cv' for cross-validation.							

Finally, for the light trucks App 2, summary of the performance of the developed models is reported (see Table 52). From the table, it is possible to see that the model that performs best at estimating the overall fuel consumption of light trucks App 2 is the RF. The model, in fact, shows the highest R^2, and lowest RMSE and MAE, both in the cross-validation process, and on the validation set. Again, also in this case, performance seem to be in line with those achieved by calibration of the physical/mechanistic models existing in the literature (e.g. Chatti *et al.*, 2012). Although, also in this case, the RF shows a higher cross-validation time compared to SVR and ANN, these are comparable for the quantity of the data considered.

E.2.2.3 Medium trucks on A roads (App 1)

Fitting an LR to the selection of variables listed in Appendix E.2.1.3, the Equation (33), LR for medium trucks App 1, is obtained.

$$\begin{aligned} Fuel.Consumption &= 24.52 + 3.2\times10^{-4}\times Gross.Weight + 6.83\times Geom.Gradient - 0.052\times Speed.AVG + \\ &+0.196\times\ Torque.AVG + 6.93\times10^{-4}\times Revs.AVG - 41.40\times Acceleration - 0.017\times Traveled.Time + \\ &-0.0022\times Geom.Abs_Radius - 0.102\times Geom.Crossfall + 0.027\times TRACS.LPV03m - 0.044\times TRACS.LPV10m + \\ &+0.016\times TRACS.LPV30m + 1.51\times TRACS.Pav_Texture - 0.005\times Altitude.AVG - 0.020\times Atm_Pressure_hPa + \\ &-0.112\times Temperature_deg_C - 0.0036\times Humidity_. + 0.003\times WindSpeed.Dir_kn \end{aligned} \quad (33)$$

Using the Equation (33) it is then possible to fit the developed model to the available dataset for the medium trucks App 1 (see Figure 95).

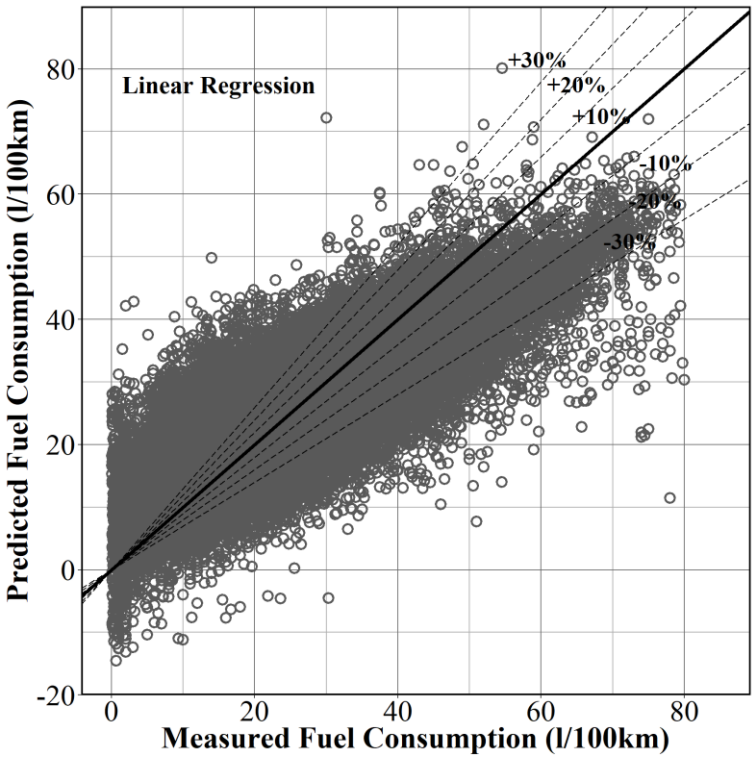


Figure 95. Fit of the developed LR for the medium trucks App 1, from the Dataset 2.

From the figure, it is possible to see that the model tends to fit well only average fuel consumption, while it tends to underestimate the high measurements of fuel consumption. In this case, also estimation of low fuel consumption seems to create

some problems to the model, with the LR that makes also a few negative predictions, which are not physically explainable. Again, this might be due to the high impact of non-linearity in these cases.

For this set, a total of 167,111 records are available. Of these, 118,990 were used to train SVR, RF and ANN, 39,663 to test, and to 8,458 validate the developed models on data from trucks not considered in the other sets of data.

For this case study, the SVR model was a regression model, with RBF kernel function and tolerance set to 0.001. Figure 96 shows the performance of the develop SVR model to fit the test set over the ten-fold cross-validation process applied to the medium trucks App 1.

From the figure, it is possible to say that the model seems quite consistent, with similar results that are obtained for different splits of the considered dataset. That confirms results for smaller datasets seen earlier.

Figure 97 shows how the developed SVR model fits the dataset of the medium trucks App 1. From the figure it is possible to say that, thanks to the higher amount of data available, performance are improved, with extreme values that are fitted quite well also for the validation set. In comparison to the performance of the other ML models developed for medium trucks App 1, it is possible to say that, by looking just at the figures, performance of the SVR (see Figure 97) seems to be in line with those of the RF model (see Figure 99) and the ANN (see Figure 101) (see Table 53).

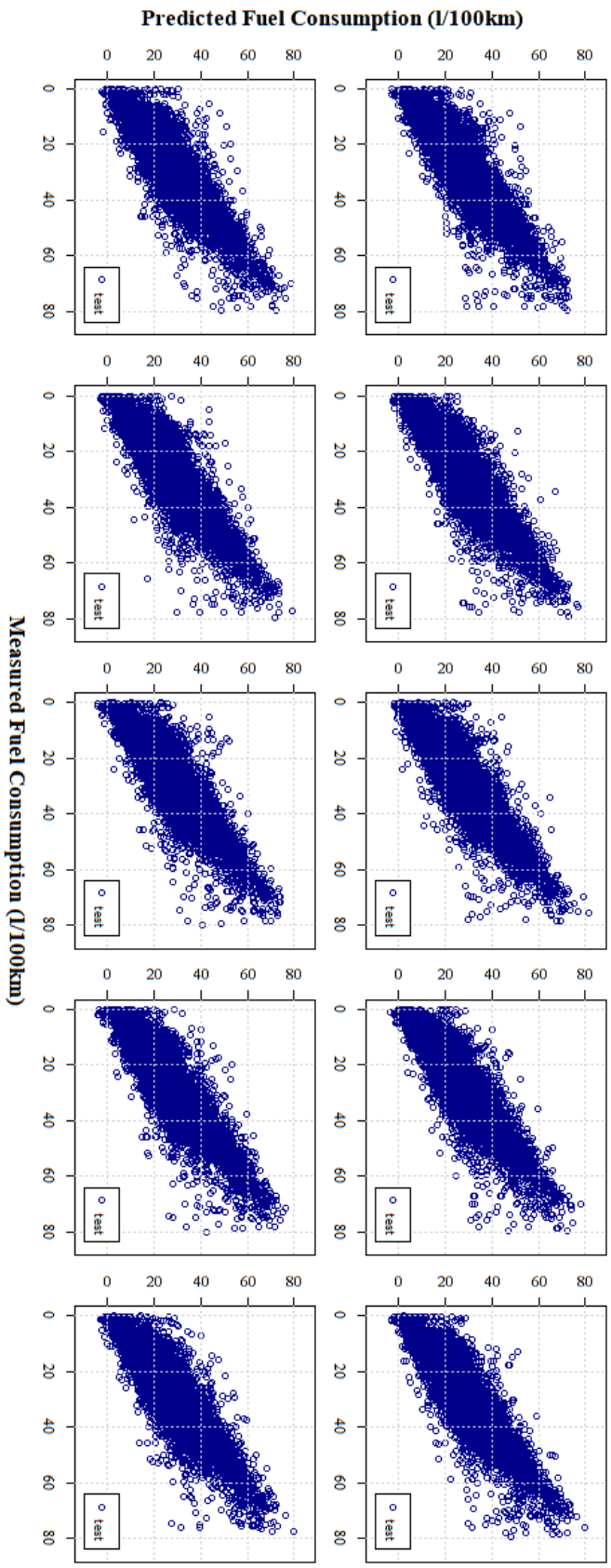


Figure 96. Ten-fold cross-validation of the SVR model for the medium trucks App 1, from the Dataset 2.

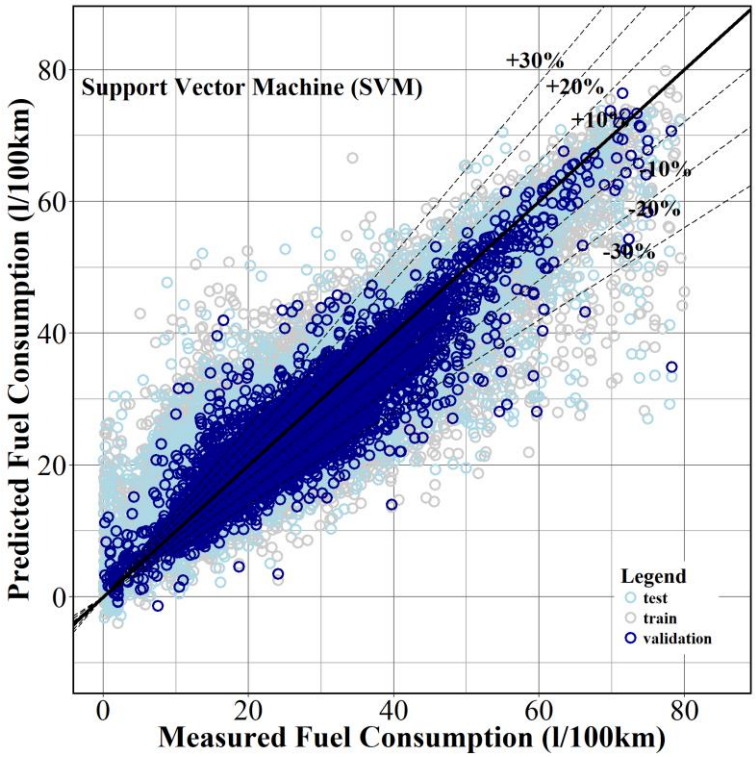


Figure 97. Fit of the developed SVR for the medium trucks App 1, from the Dataset 2.

The developed RF is composed of 250 trees and in each of these it uses 6 randomly sampled variables to make predictions. Figure 98 represents the fit of the RF model on the test set over the ten-fold cross-validation process, for the medium trucks App 1. From the figure it is possible to see that performance of the RF over the cross-validation are quite consistent, which indicate robustness of the model at estimating fuel consumption for different test sets.

Figure 99 reports instead results of the fit of the developed RF model on the medium trucks App 1 dataset. The figure shows that there are some outliers and that the model tends to underestimate high fuel consumption. However, this is the model which performs best (see Table 53).

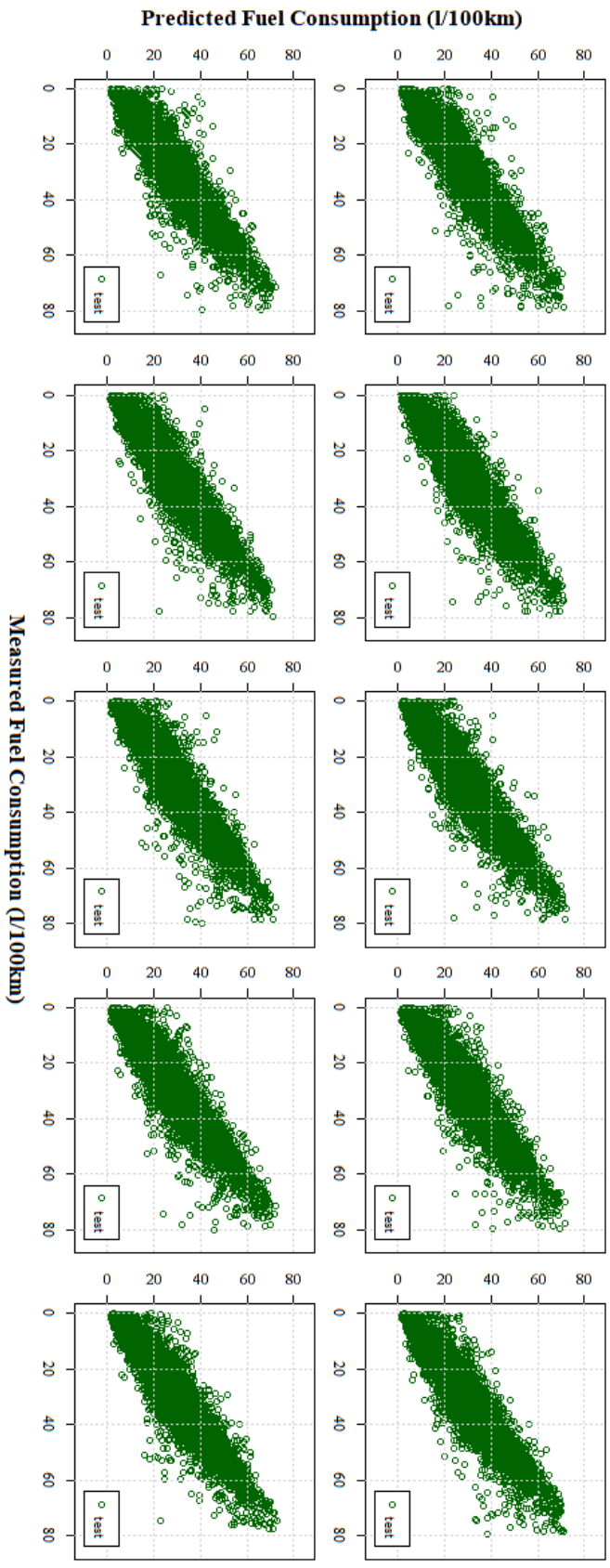


Figure 98. Ten-fold cross-validation of the RF model for the medium trucks App 1, from the Dataset 2.

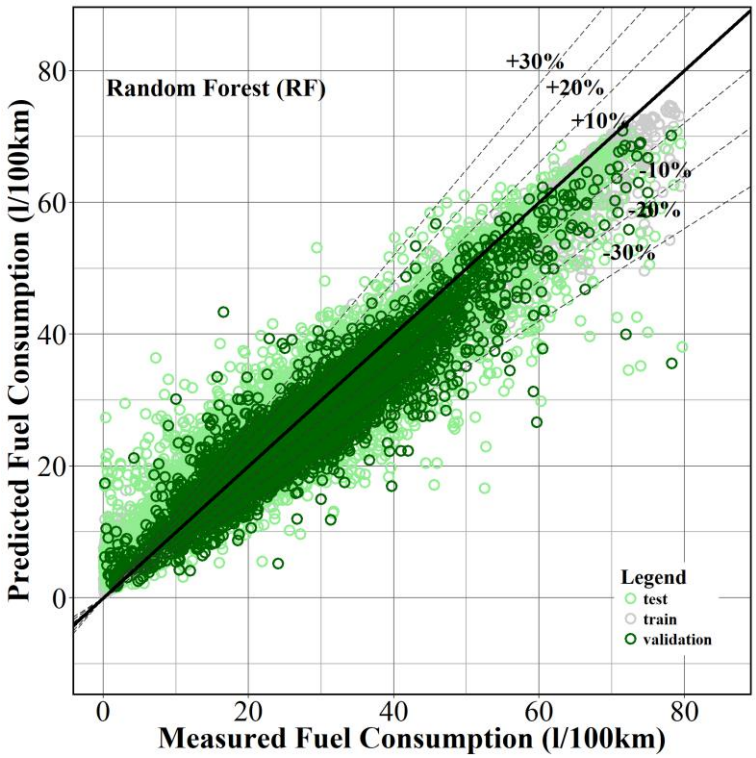


Figure 99. Fit of the developed RF for the medium trucks App 1, from the Dataset 2.

Finally, the developed ANN and its performance are considered. The ANN is structured on two layers of 14 and 4 neurons respectively. The model was trained using the rprop- algorithm (Riedmiller *et al.*, 1993), and it used the logistic activation function to identify patterns through the data.

In Figure 100, representation of the fit of the test set over the ten-fold cross-validation process is shown for the develop ANN for the medium trucks App 1 dataset. The figure shows that also this model seems to have quite constant performance on every test set over the cross-validation.

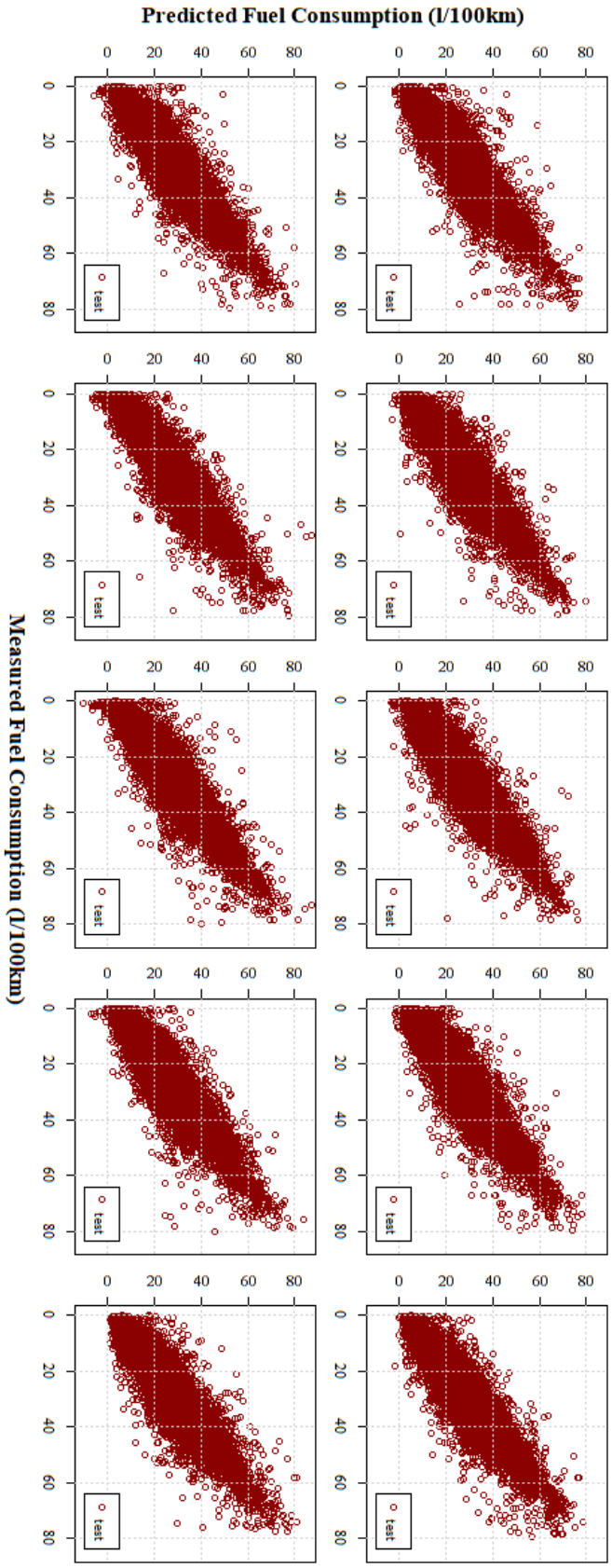


Figure 100. Ten-fold cross-validation of the ANN model for the medium trucks App 1, from the Dataset 2.

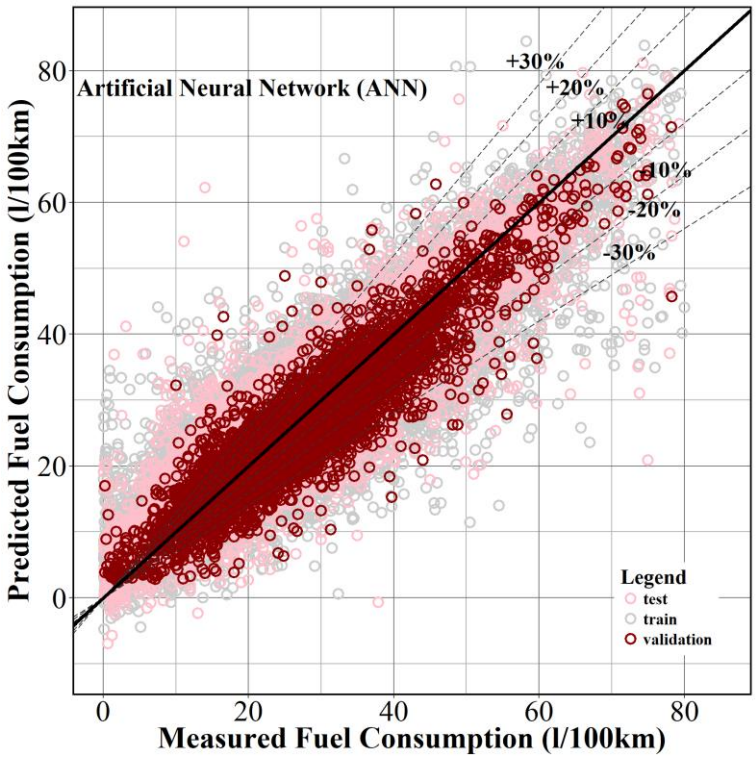


Figure 101. Fit of the developed ANN for the medium trucks App 1, from the Dataset 2.

Figure 101 shows instead fit of the ANN model to the medium trucks App 1 set. As already mentioned, the model seems to have performance close to those of the SVR (see Figure 97), and therefore superior to that of the LR (see Figure 95) but lower than the RF (see Figure 99) (see Table 53).

In Table 53, a comparison of the numerical performance between the different regression models developed in the study is made, for the considered case study. The table shows that the RF outperforms all the other models achieving (for the very first time in the study) an R^2 of over 0.90, and RMSE and MAE between 40% and 45% less than the LR model. This represents a quite good level of accuracy, similar (or superior in some cases) to that of existing models (e.g. Chatti *et al.*, 2012).

Table 53 - Summary of performance of the regression models developed for the medium trucks App 1.

Model	R^2_{cv}	$RMSE_{cv}$	MAE_{cv}	$Time_{cv}$	R^2_{val}	$RMSE_{val}$	MAE_{val}
	(-)	*(l/100 km)*	*(l/100 km)*	*(mins)*	*(-)*	*(l/100 km)*	*(l/100 km)*
LR	0.69	5.35	3.83	-	-	-	-
SVR	0.85	3.78	2.54	1,963	0.84	3.82	2.57
RF	0.91	3.12	2.05	1,646	0.89	3.16	2.08
ANN	0.84	3.74	2.63	1,878	0.84	3.76	2.64
N.B.: The subscript 'val' stays for validation set, and 'cv' for cross-validation.							

E.2.2.4 Medium trucks on Motorways and A roads (App 2)

Fitting an LR to the selection of variables listed in Appendix E.2.1.4, the Equation (34), LR for medium trucks App 2, is obtained.

$$\begin{aligned} Fuel.Consumption &= 28.40 + 3.04\times10^{-4}\times Gross.Weight + 6.83\times Geom.Gradient - 0.019\times Speed.AVG + \\ &+0.198\times Torque.AVG + 5.29\times10^{-4}\times Revs.AVG - 33.02\times Acceleration - 0.010\times Traveled.Time + \\ &-0.0018\times Geom.Abs_Radius - 0.127\times Geom.Crossfall - 0.125\times TRACS.LPV03m - 0.025\times TRACS.LPV10m + \\ &+0.035\times TRACS.LPV30m + 1.27\times TRACS.Pav_Texture + 0.003\times Altitude.AVG - 0.021\times Atm_Pressure_hPa + \\ &-0.103\times Temperature_deg_C - 0.0017\times Humidity_. + 0.004\times WindSpeed.Dir_kn \end{aligned} \quad (34)$$

From Figure 102, it is possible to see the fit of the developed LR model (see Equation (34)) to the medium trucks App 2 dataset. The figure shows that, as in the case of medium trucks App 1, the model tends to fit well only average fuel consumption while it struggles with extreme values. This seems to be a characteristic of the LR models, and it might be due to the effect of non-linearity, which for extreme values of fuel consumption might be significant.

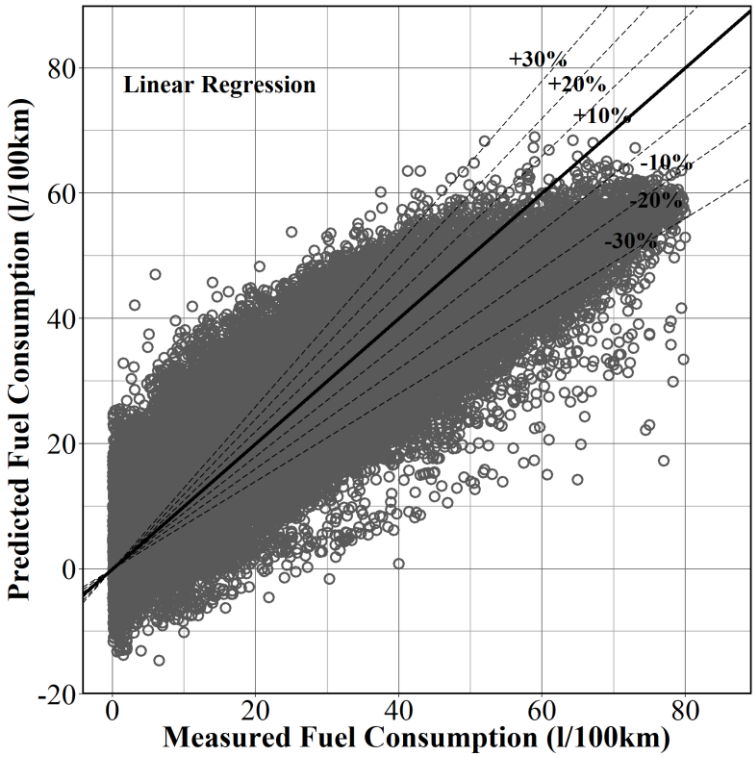


Figure 102. Fit of the developed LR for the medium trucks App 2, from the Dataset.

For this set, a total of 193,753 records are available. Of these, 137,845 were used to train SVR, RF and ANN, 45,948 to test, and 9,960 to validate the developed models on data from trucks not considered in the other sets of data.

For this case study, the SVR model was a regression model, with RBF kernel function and tolerance set to 0.001. Figure 103 represents the fit of the SVR model on the test set over the ten-fold cross-validation process, for the medium trucks App 2.

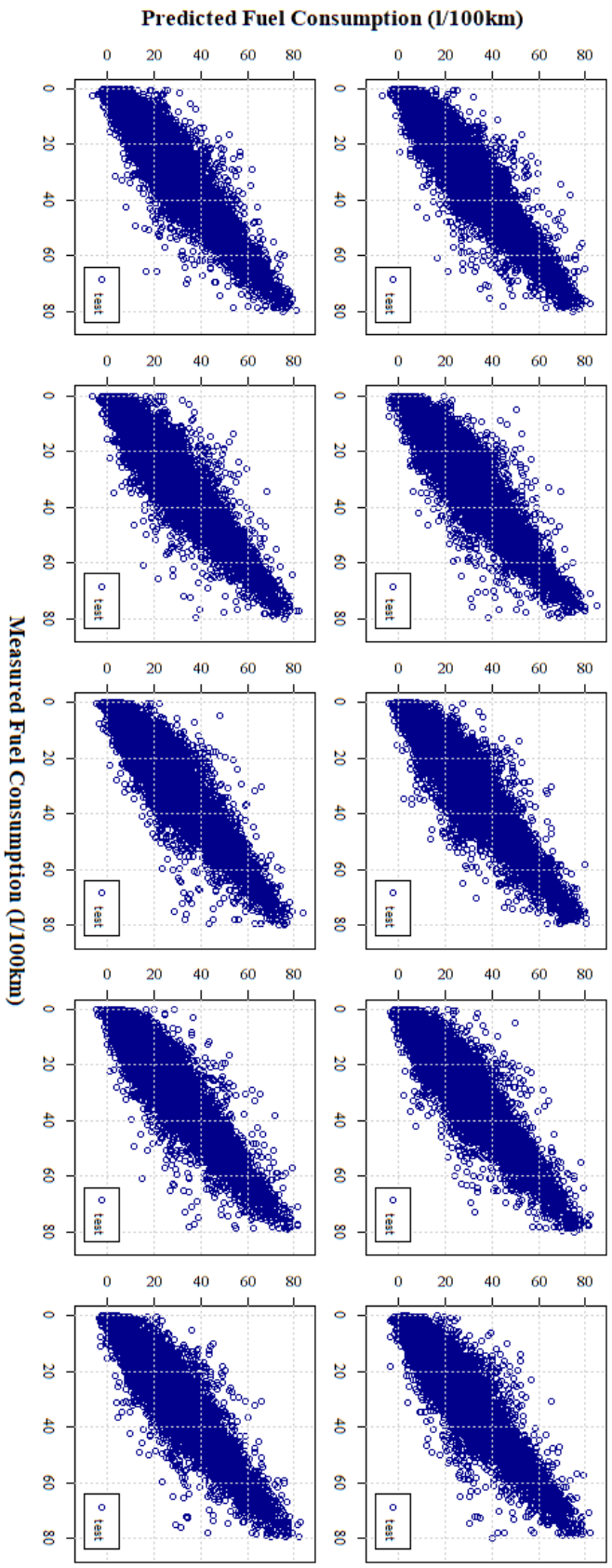


Figure 103. Ten-fold cross-validation of the SVR model for the medium trucks App 2, from the Dataset 2.

From the figure it is possible to see that performance of the SVR over the cross-validation can be considered to be quite constant, which indicate robustness of the model at estimating fuel consumption on different datasets.

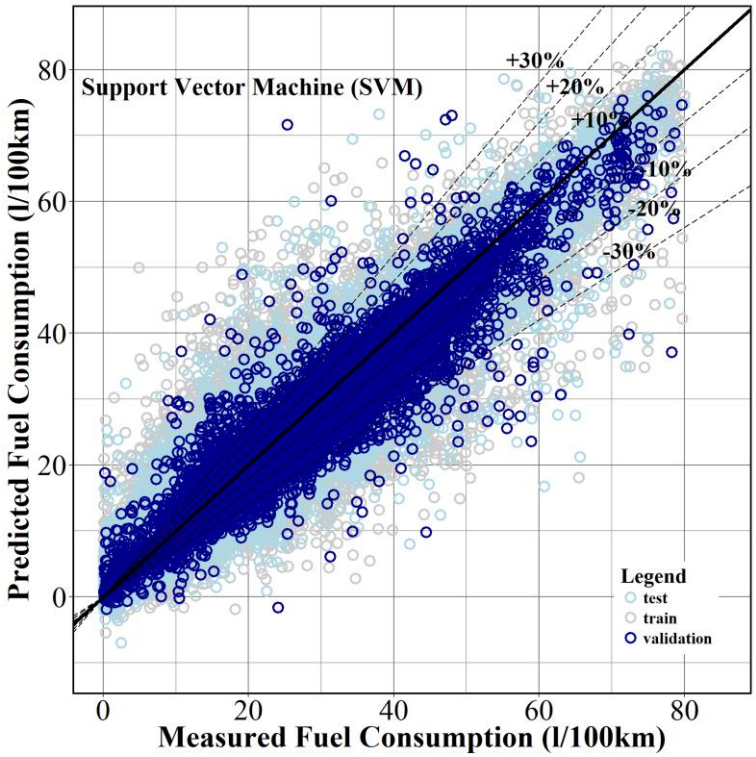


Figure 104. Fit of the developed SVR for the medium trucks App 2, from the Dataset 2.

Figure 104 reports instead results of the fit of the developed SVR model on the medium trucks App 2 dataset. The figure shows that although the model generally tends to underestimate low fuel consumption for the considered set of data. Important is to highlight the fact that for the quantity of data analysed in this case study, the SVR, RF and ANN seem to fit extreme values of fuel consumption more sensibly. In comparison to RF (see Figure 106) and ANN (see Figure 108), performance of the SVR seems to be slightly lower than that of the RF, in line with that of the ANN (see Table 54).

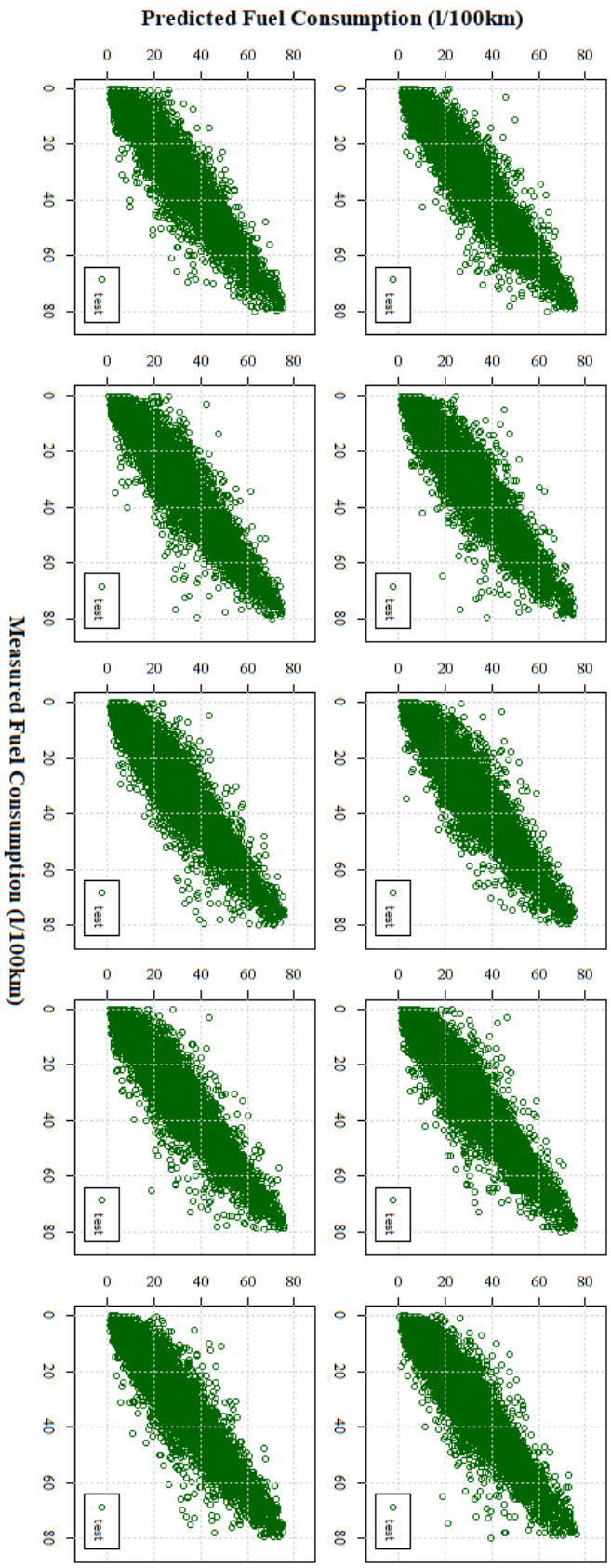


Figure 105. Ten-fold cross-validation of the RF model for the medium trucks App 2, from the Dataset 2.

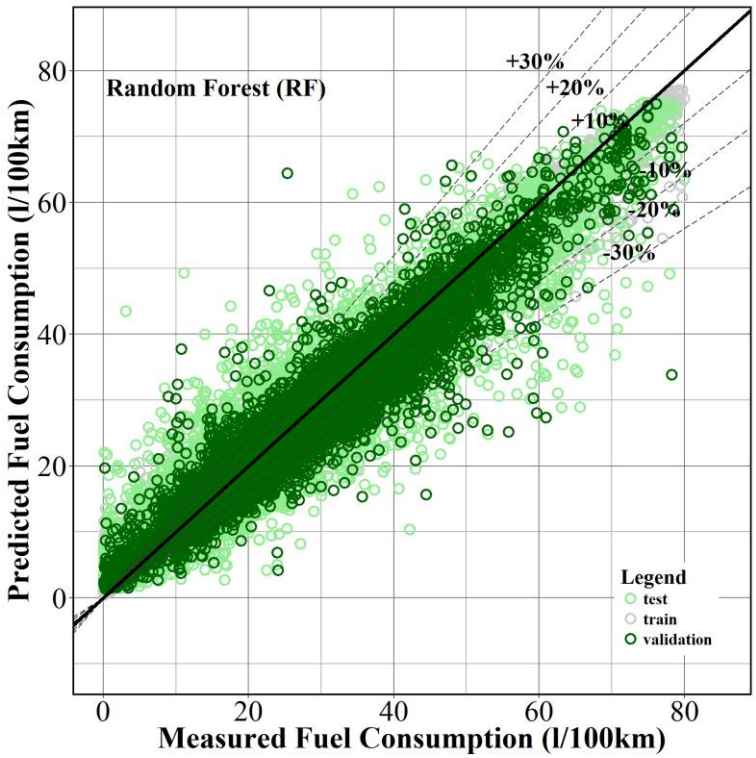


Figure 106. Fit of the developed RF for the medium trucks App 2, from the Dataset 2.

The developed RF is composed of 250 trees and in each of these it uses 6 randomly sampled variables to make predictions. Figure 105 represents the fit of the RF model on the test set over the ten-fold cross-validation process, for the medium trucks App 2. From the figure it is possible to see that performance of the RF over the cross-validation are quite constant, which indicate robustness of the model at estimating fuel consumption on different test sets.

Figure 106 reports instead results of the fit of the developed SVR model on the medium trucks App 2 dataset. The figure shows that although the model generally tends to underestimate low fuel consumption for the considered set of data. In comparison to SVR (see Figure 104) and ANN (see Figure 108), performance of the RF model seems to be slightly superior (see Table 54).

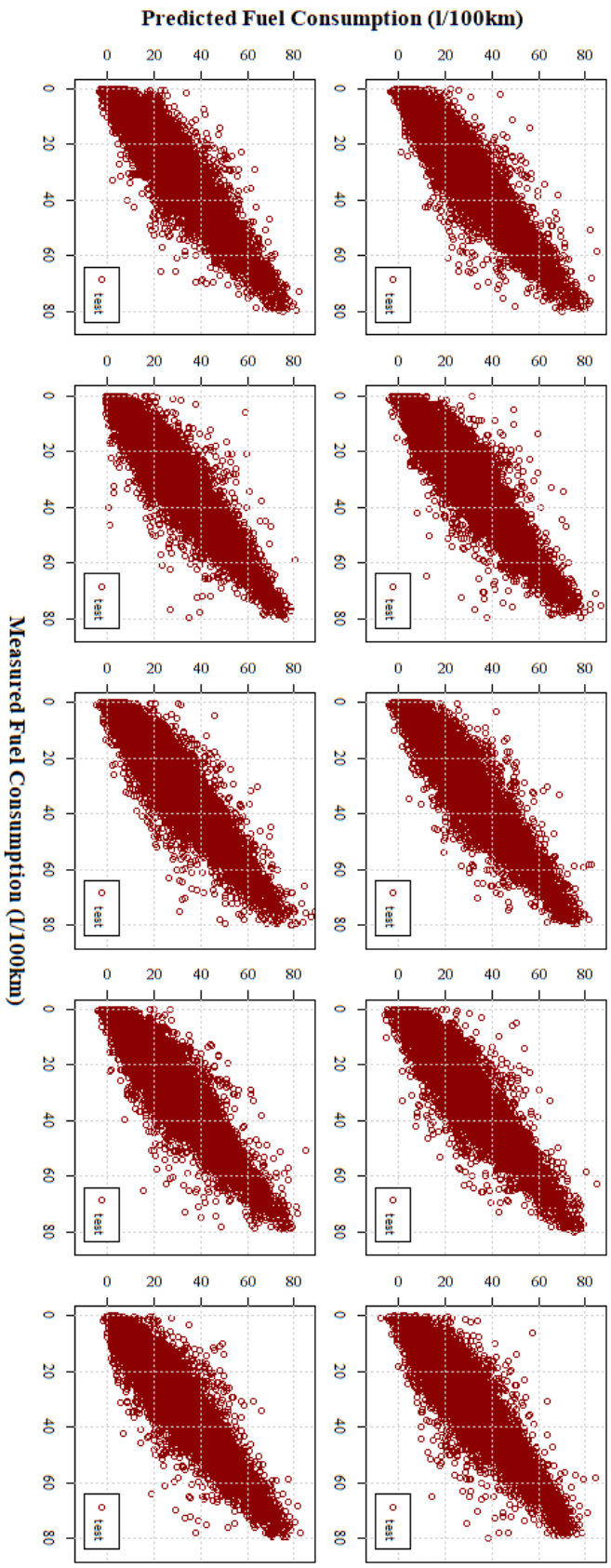


Figure 107. Ten-fold cross-validation of the ANN model for the medium trucks App 2, from the Dataset 2.

Finally, the developed ANN and its performance are considered. The ANN is structured on two layers of 14 and 4 neurons respectively. The model was trained using the rprop- algorithm (Riedmiller *et al.*, 1993), and it used the logistic activation function to identify patterns through the data. Figure 107 represents the fit of the developed ANN model on the test set over the ten-fold cross-validation process, for the medium trucks App 2. From the figure it is possible to see that performance of the ANN over the cross-validation are quite consistent, which indicate robustness of the model at estimating fuel consumption on different test sets.

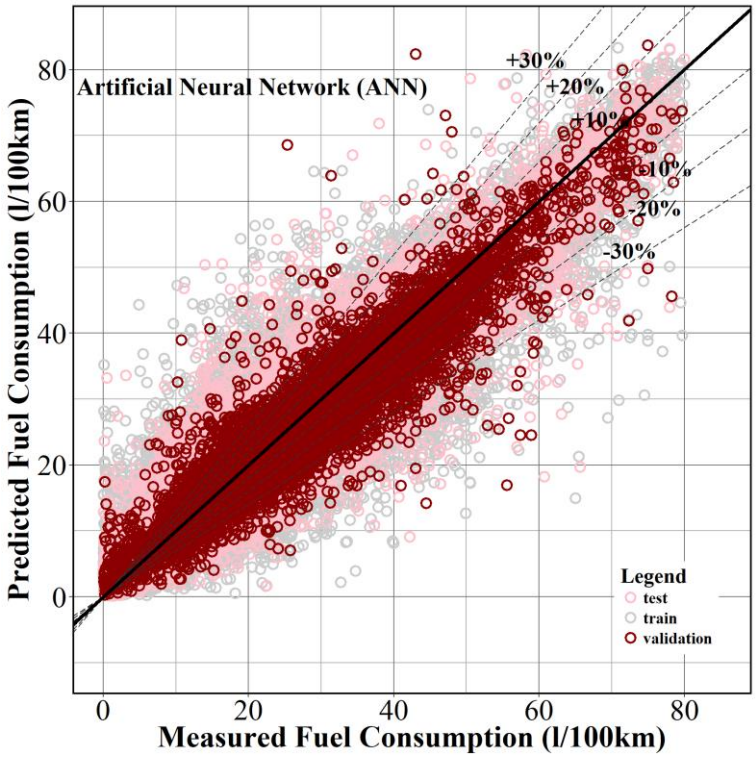


Figure 108. Fit of the developed ANN for the medium trucks App 2, from the Dataset 2.

Figure 108 reports instead results of the fit of the developed ANN model on the medium trucks App 2 dataset. The figure shows that although the model generally tends to underestimate low fuel consumption for the considered set of data. The model

shows similar performance to the SVR (see Figure 104), while it looks to be outperformed by the RF (see Figure 106) (see Table 54).

Table 54 summarises performance of the developed models for the cross-validation and the validation sets. From the table it is possible to see that the RF is the model that (again) performs best, with an R^2 of 0.91, and RMSE and MAE, which reduce of approximately 32%, in comparison to the LR. This makes the developed RF superior for accuracy to the most recently calibrated models currently in the literature (e.g. Chatti *et al.*, 2012).

Table 54 - Summary of performance of the regression models developed for the medium trucks App 2.

Model	R^2_{cv}	$RMSE_{cv}$	MAE_{cv}	$Time_{cv}$	R^2_{val}	$RMSE_{val}$	MAE_{val}
	(-)	*(l/100 km)*	*(l/100 km)*	*(mins)*	*(-)*	*(l/100 km)*	*(l/100 km)*
LR	0.76	5.74	4.14	-	-	-	-
SVR	0.88	4.50	3.12	4,180	0.88	4.51	3.12
RF	0.91	3.95	2.78	4,358	0.91	3.95	2.77
ANN	0.88	4.62	3.25	4,747	0.87	4.61	3.26
N.B.: The subscript 'val' stays for validation set, and 'cv' for cross-validation.							

E.2.2.5 Heavy trucks on A roads (App 1)

Fitting an LR to the selection of variables listed in Appendix E.2.1.5, the Equation (35), LR for heavy trucks App 1, is obtained.

$$\begin{aligned} Fuel.Consumption = {} & 26.45 + 2.03\times10^{-4}\times Gross.Weight + 7.83\times Geom.Gradient - 0.068\times Speed.AVG + \\ & +0.266\times Torque.AVG + 0.008\times Revs.AVG - 39.64\times Acceleration - 0.018\times Traveled.Time + \\ & -0.0023\times Geom.Abs_Radius - 0.147\times Geom.Crossfall + 0.067\times TRACS.LPV03m - 0.061\times TRACS.LPV10m + \\ & +0.011\times TRACS.LPV30m + 1.55\times TRACS.Pav_Texture - 0.0055\times Altitude.AVG - 0.018\times Atm_Pressure_hPa + \\ & -0.11\times Temperature_deg_C - 0.0075\times Humidity_. + 0.003\times WindSpeed.Dir_kn \end{aligned} \quad (35)$$

Figure 109 shows how the developed LR model (Equation (35)) fits the available dataset for heavy trucks App 1. Again, also in this case, the LR model seems to being not able to estimate the overall fuel consumption of the considered fleet of trucks sensibly. The model seems subject to overfitting, which causes overestimate of

low fuel consumption and underestimate high consumption for the considered heavy trucks driving along A roads.

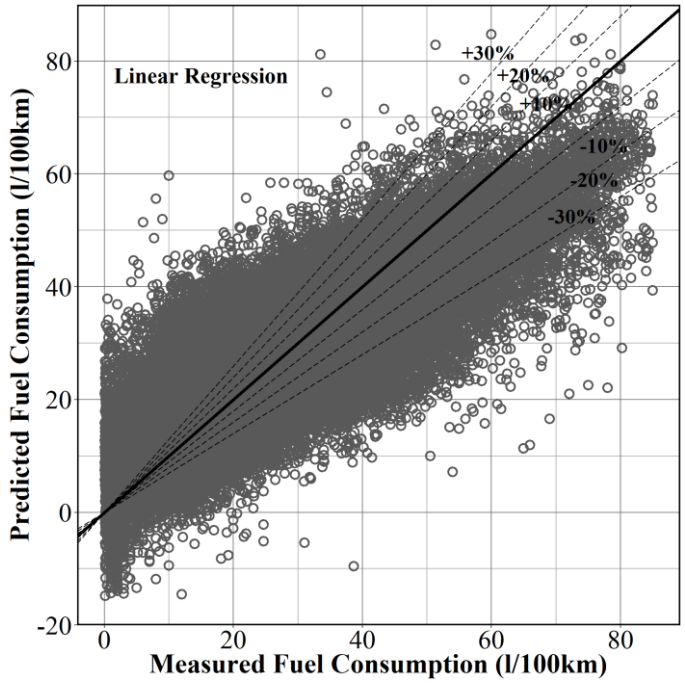


Figure 109. Fit of the developed LR for the heavy trucks App 1, from the Dataset 2.

For this set, a total of 278,607 records are available. Of these, 198,653 were used to train SVR, RF and ANN, 66,217 to test, and 13,737 to validate the developed models on data from trucks not considered in the other sets of data.

For this case study, the SVR model was a regression model, with RBF kernel function and tolerance set to 0.001. Figure 110 shows how the SVR model, developed for heavy trucks App 1, performs on the test set over the ten-fold cross-validation process.

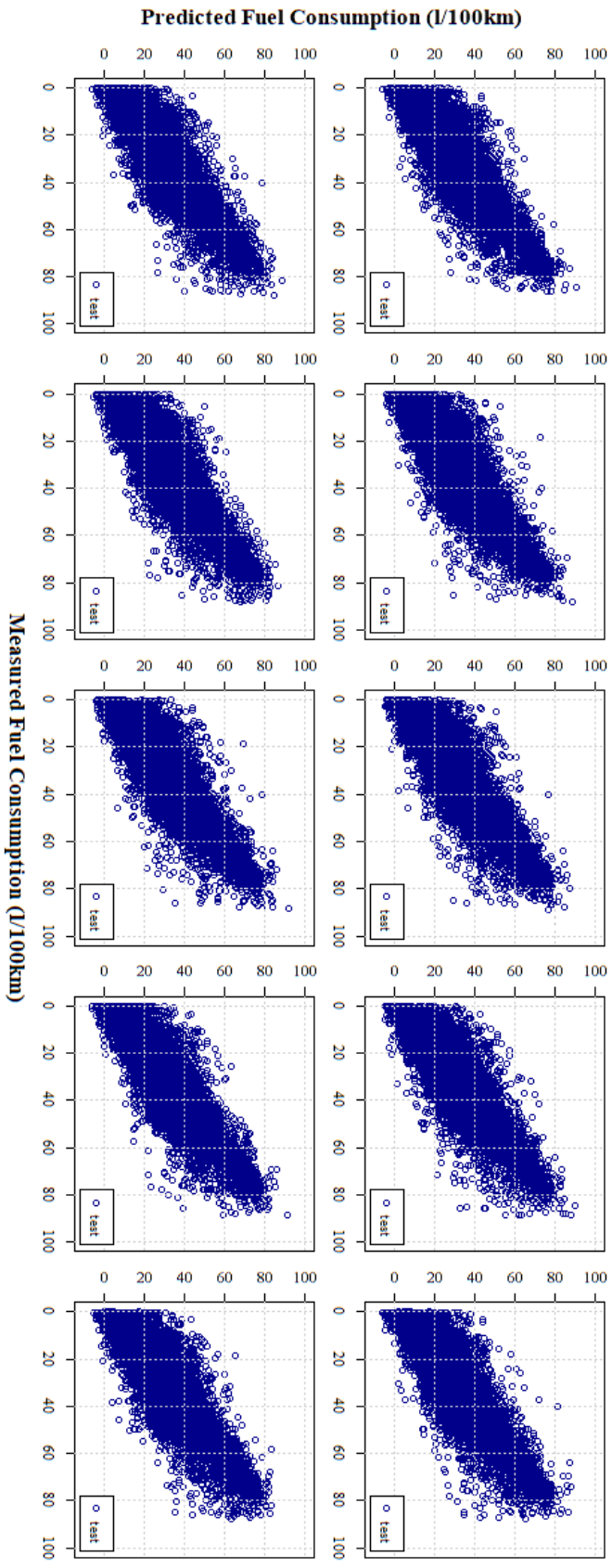


Figure 110. Ten-fold cross-validation of the SVR model for the heavy trucks App 1, from the Dataset 2.

From the figure, it is possible to say that this performs consistently over the ten-fold cross-validation, which indicate robustness of the model at modelling the fuel consumption of the considered fleet of trucks for different test sets.

Figure 111 shows how the developed SVR model fits the available dataset for the heavy trucks App 1. From the figure, it is possible to say that, thanks to the large amount of data available for this case study, the developed SVR performs quite well, also on extreme values. From a comparison of performance of the SVR with the developed RF (see Figure 113) and ANN (see Figure 115), it is possible to say that the SVR shows again performance similar to those of the ANN, which are outperformed by the RF (see Table 55).

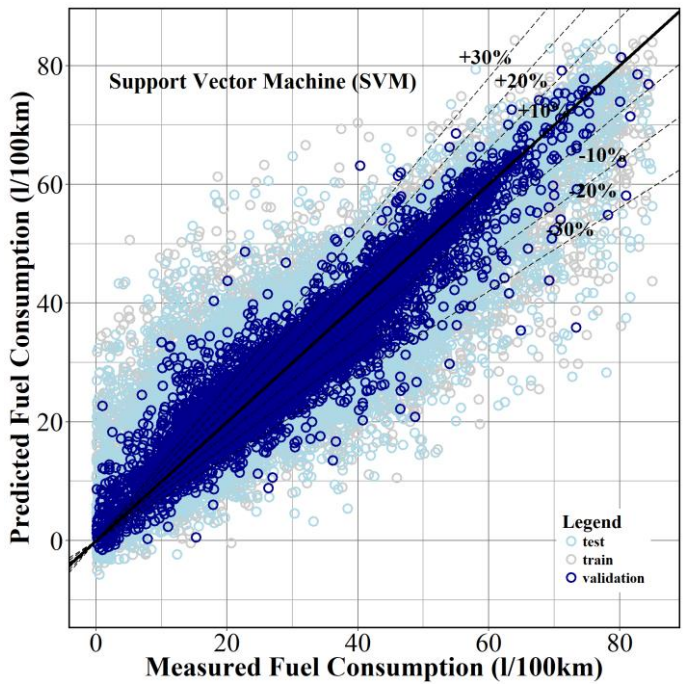


Figure 111. Fit of the developed SVR for the heavy trucks App 1, from the Dataset 2.

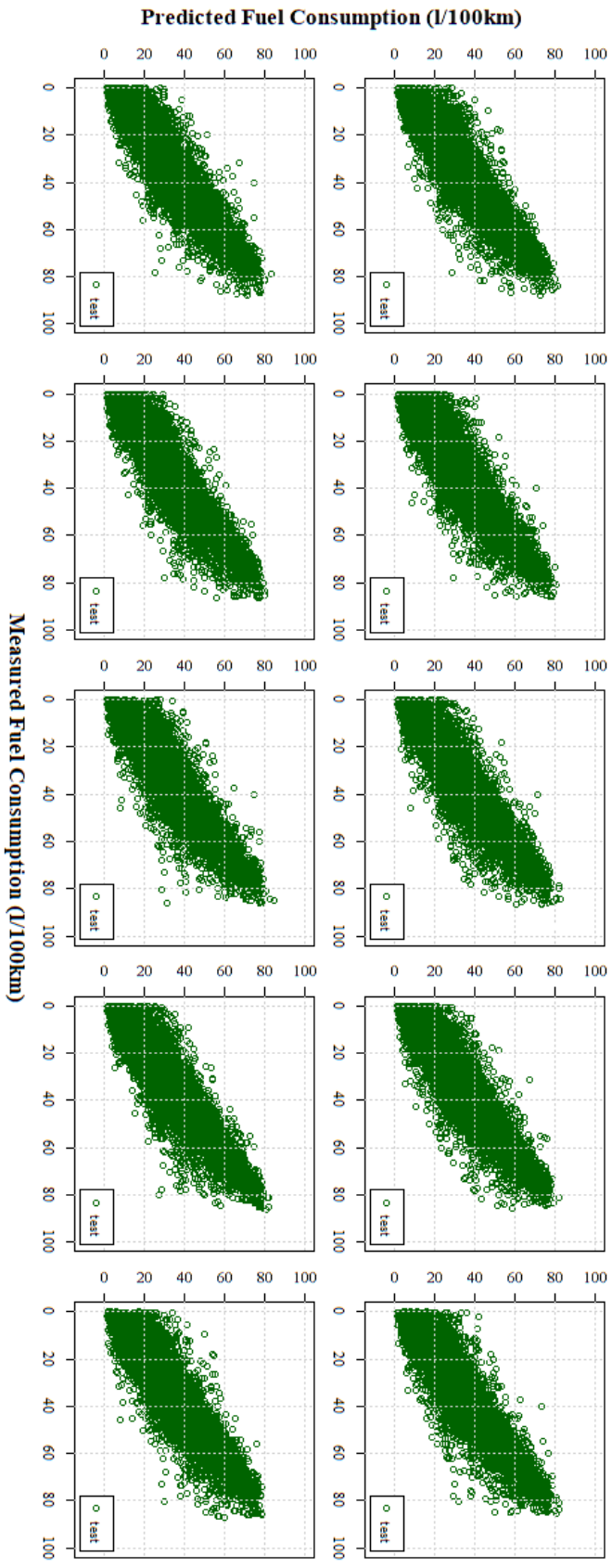


Figure 112. Ten-fold cross-validation of the RF model for the heavy trucks App 1, from the Dataset 2.

The developed RF is composed of 250 trees and in each of these it uses 6 randomly sampled variables to make predictions. In Figure 112 results of the cross-validation process of the RF model applied to the heavy trucks App 1 dataset are reported. The figure shows that the model performs quite consistently all throughout the ten-fold cross-validation, which indicates robustness of the model at estimating the fuel consumption of considered fleet of trucks.

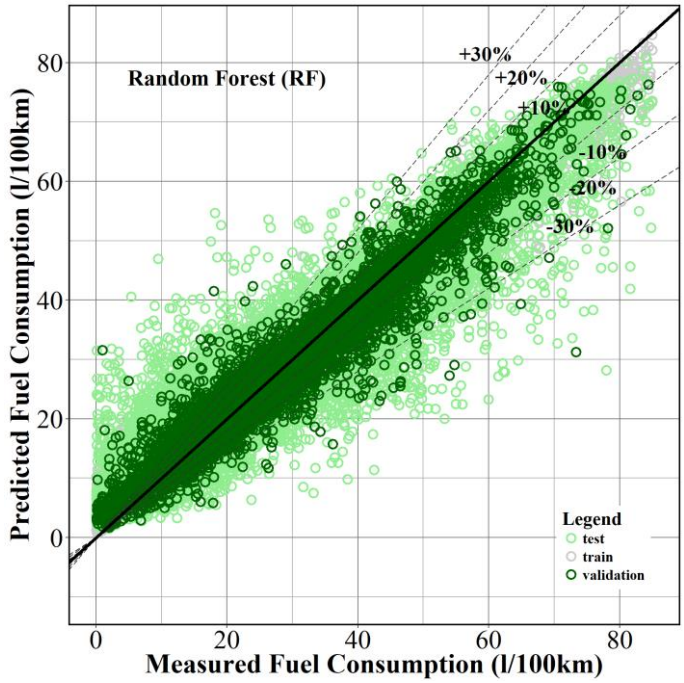


Figure 113. Fit of the developed RF for the heavy trucks App 1, from the Dataset 2.

Figure 113 shows how the RF model, developed for the heavy trucks App 1, fits the available dataset. From a comparison of the fits of the RF with those of the SVR (see Figure 111) and the ANN (see Figure 115), it is possible to say that the RF seems to outperform both the other two models, reducing variance to a minimum (see Table 55).

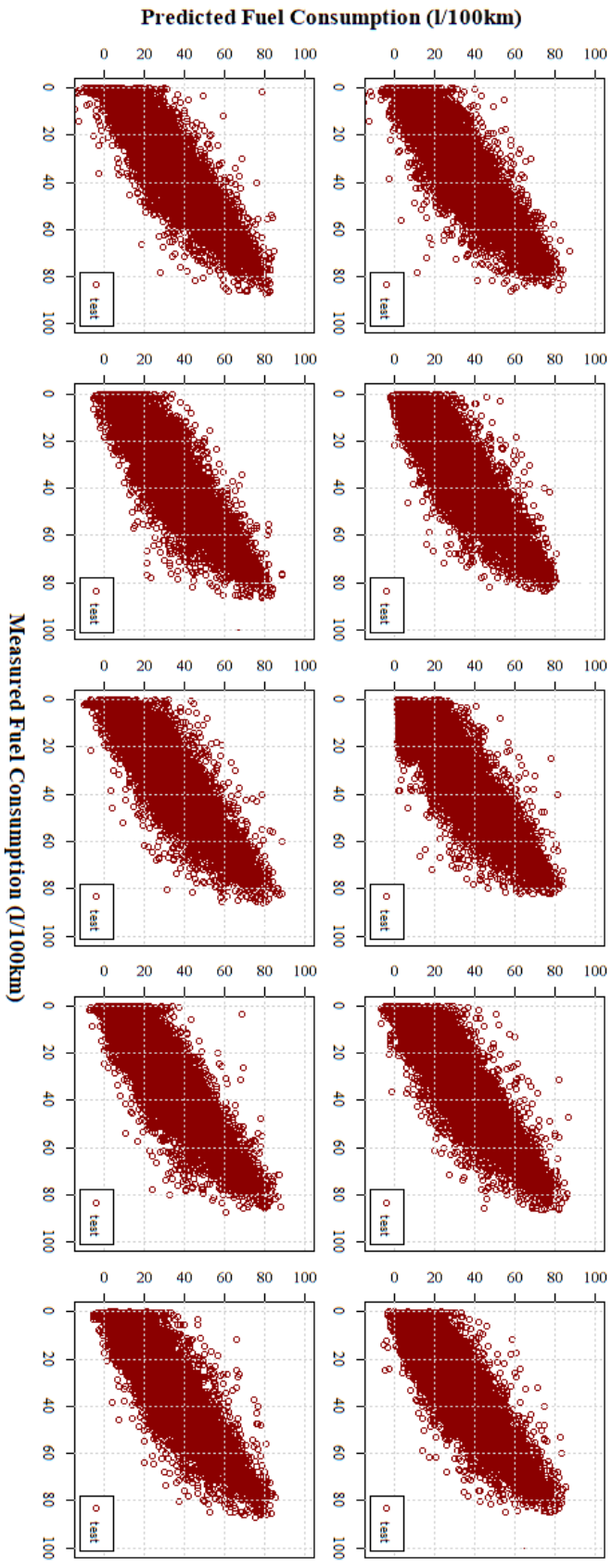


Figure 114. Ten-fold cross-validation of the ANN model for the heavy trucks App 1, from the Dataset 2.

Finally, the developed ANN and its performance are considered. The ANN is structured on two layers of 14 and 4 neurons respectively. The model was trained using the rprop- algorithm (Riedmiller *et al.*, 1993), and it used the logistic activation function to identify patterns through the data. In Figure 114 results of the cross-validation process of the ANN model applied to the heavy trucks App 1 dataset are reported. The figure shows that the model performs quite consistently all throughout the ten-fold cross-validation, which indicates robustness of the model at estimating the fuel consumption of considered fleet of trucks.

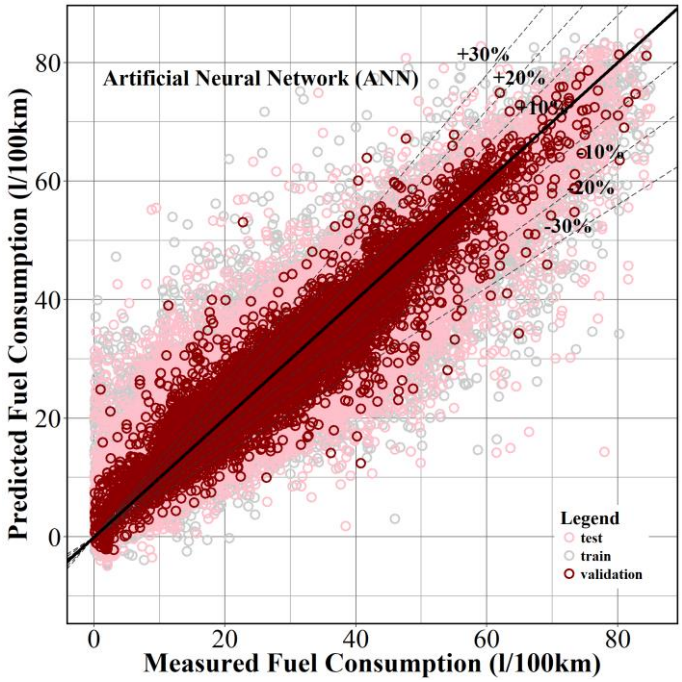


Figure 115. Fit of the developed ANN for the heavy trucks App 1, from the Dataset 2.

Figure 115 shows how the ANN model, developed for the heavy trucks App 1, fits the available dataset. From a comparison of the fits of the ANN with those of the SVR (see Figure 111) and the RF (see Figure 113), it is possible to say that the

ANN seems to perform well on the available data, similarly to the SVR, and slightly worse than the RF (see Table 55).

Table 55 - Summary of performance of the regression models developed for the heavy trucks App 1.

Model	R^2_{cv}	$RMSE_{cv}$	MAE_{cv}	$Time_{cv}$	R^2_{val}	$RMSE_{val}$	MAE_{val}
	(-)	*(l/100 km)*	*(l/100 km)*	*(mins)*	*(-)*	*(l/100 km)*	*(l/100 km)*
LR	0.72	5.87	4.17	-	-	-	-
SVR	0.90	3.63	2.52	5,128	0.90	3.63	2.52
RF	0.94	3.00	1.98	6,576	0.93	3.02	1.99
ANN	0.89	3.72	2.66	5,289	0.90	3.74	2.66
N.B.: The subscript 'val' stays for validation set, and 'cv' for cross-validation.							

In Table 55, comparison of the performance of all the regression models developed for the heavy trucks App 1 is reported. Analysis of the values showed in the table, reveal that the RF remains the model which performs best, with its performance that increase to 0.93 R^2, for the validation set, and RMSE and MAE which become as low as 49% less than the LR. Thanks to the larger quantity of data fed to the ML models, also performance of the SVR and ANN improve with their R^2s, which reach 0.90 on the validation set. This makes any of SVR, RF and ANN superior for performance to the HDM-4 recently calibrated for similar vehicles in the US (Chatti *et al.*, 2012).

E.2.2.6 Heavy trucks on Motorways and A roads (App 2)

Fitting an LR to the selection of variables listed in Appendix E.2.1.6, the Equation (36), LR for heavy trucks App 2, is obtained.

$$\begin{aligned} Fuel.Consumption &= 31.96 + 2.2\times10^{-4}\times Gross.Weight + 7.60\times Geom.Gradient - 0.104\times Speed.AVG + \\ &+0.264\times Torque.AVG + 0.008\times Revs.AVG - 34.78\times Acceleration - 0.014\times Traveled.Time + \\ &-0.0024\times Geom.Abs_Radius - 0.114\times Geom.Crossfall - 0.115\times TRACS.LPV03m + 0.012\times TRACS.LPV10m + \\ &-0.021\times TRACS.LPV30m + 1.61\times TRACS.Pav_Texture - 0.0038\times Altitude.AVG - 0.021\times Atm_Pressure_hPa + \\ &-0.104\times Temperature_deg_C - 0.0056\times Humidity_. + 0.005\times WindSpeed.Dir_kn \end{aligned} \quad (36)$$

Figure 116 shows how the LR model developed for the heavy trucks App 2 (Equation (36)) fits the available data. Again, it is possible to see that the LR model provides estimates with large variance and tend to overestimate low fuel consumption, while it underestimates high values. Again, this is probably due to the high effect of non-linearity in the problem.

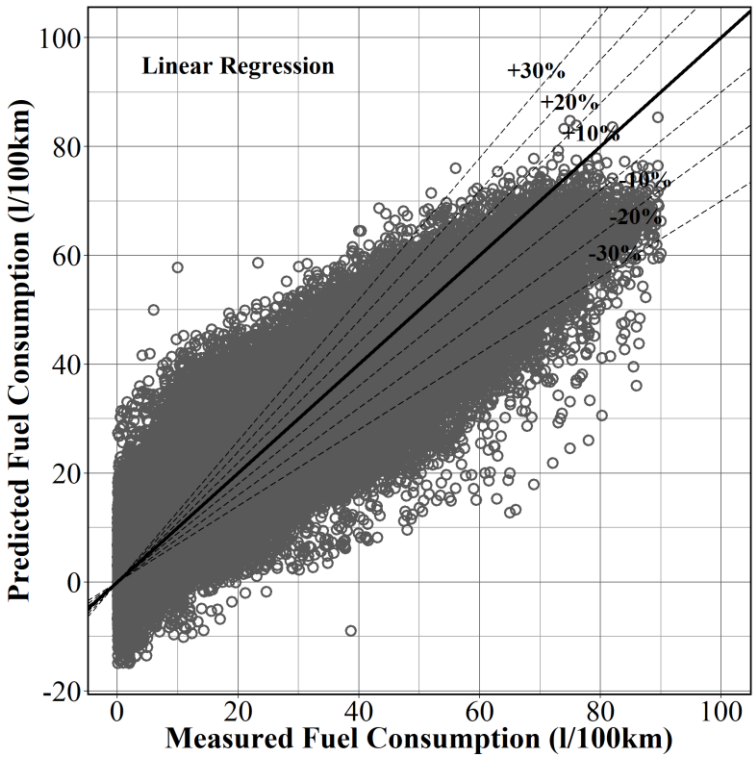


Figure 116. Fit of the developed LR for the heavy trucks App 2, from the Dataset 2.

For this set, a total of 291,073 records are available. Of these, 208,082 were used to train SVR, RF and ANN, 69,630 to test, and 13,631 to validate the developed models on data from trucks not considered in the other sets of data.

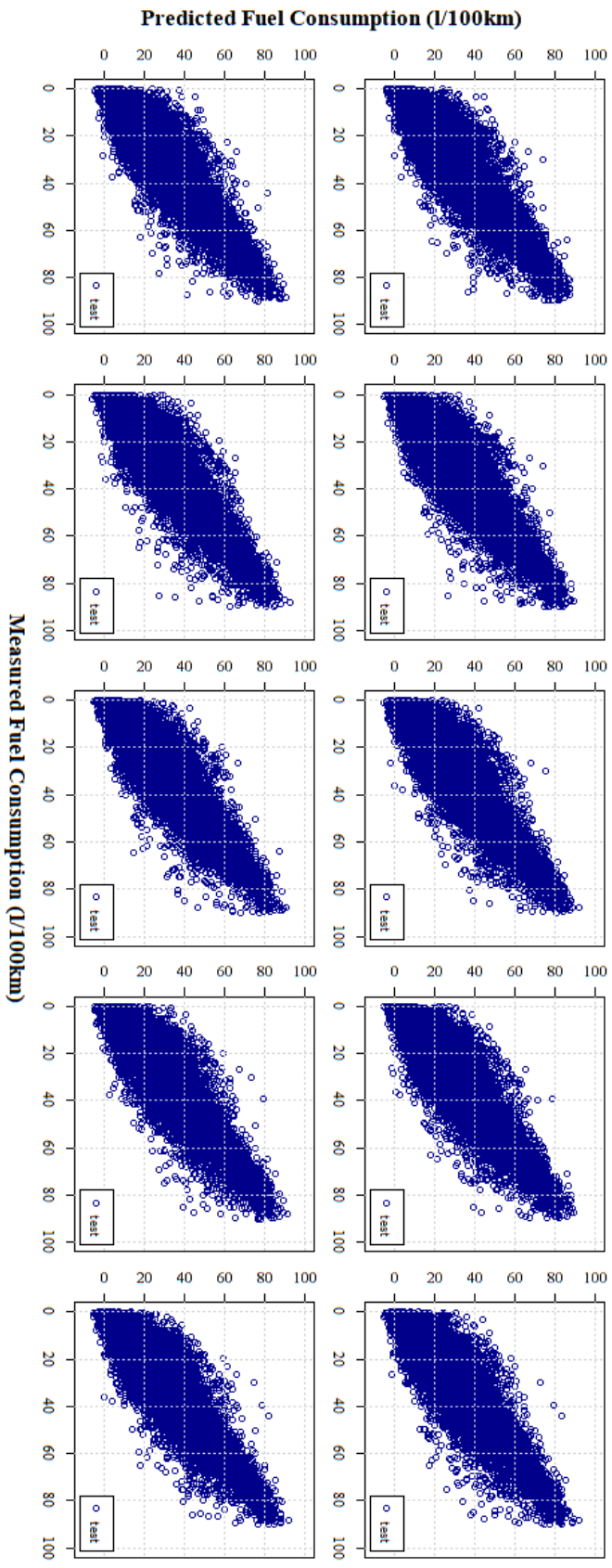


Figure 117. Ten-fold cross-validation of the SVR model for the heavy trucks App 2, from the Dataset 2.

Also in this case, the SVR model was a regression model, with RBF kernel function and tolerance set to 0.001. Figure 117 shows the fit of the SVR model developed for the heavy trucks App 2 on the test set over the ten-fold cross-validation process. The figure shows how the model keeps quite consistent performance with quite low variance of the estimates made on each test set analysed.

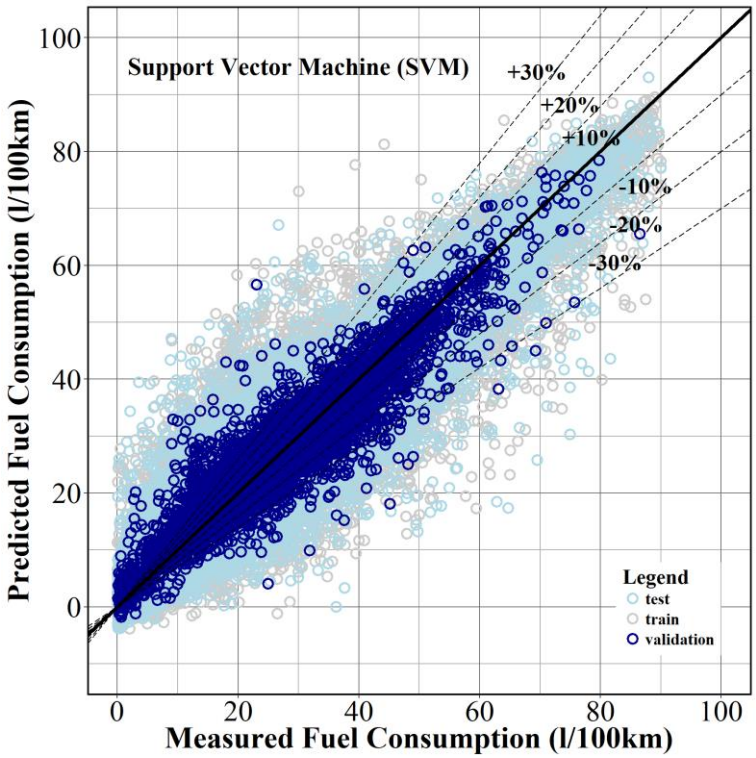


Figure 118. Fit of the developed SVR for the heavy trucks App 2, from the Dataset 2.

Figure 118 shows the fit of the developed SVR model for the available dataset from heavy trucks App 2. In particular, the figure shows how the model, developed for heavy trucks App 2, has good performance also on the validation set. Performance of the SVR model seem good and, although lower than those of the RF (see Figure 120), similar to those of the developed ANN (see Figure 122) (see Table 56).

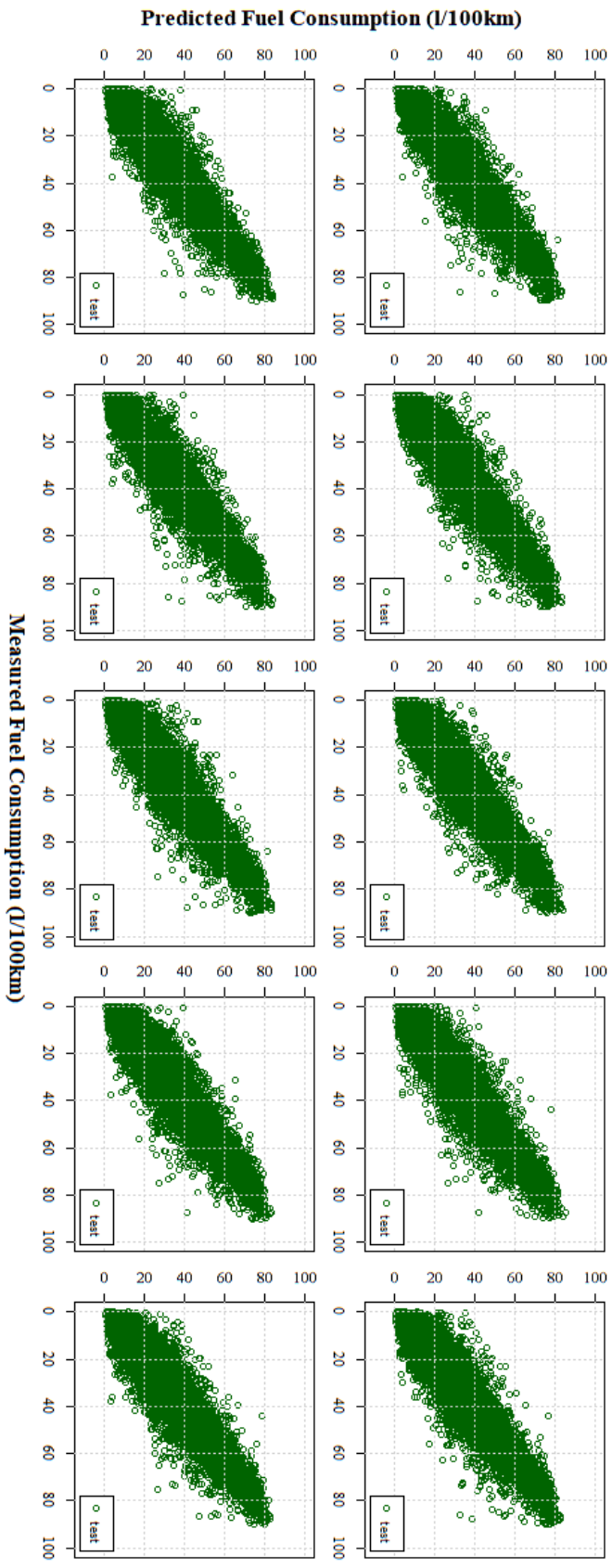


Figure 119. Ten-fold cross-validation of the RF model for the heavy trucks App 2, from the Dataset 2.

The developed RF is composed of 250 trees and in each of these it uses 6 randomly sampled variables to make predictions. Figure 120 shows the fit of the RF model developed for the heavy trucks App 2 on the test set over the ten-fold cross-validation process. The figure shows how the model keeps quite consistent performance with quite low variance of the estimates made on each test set analysed.

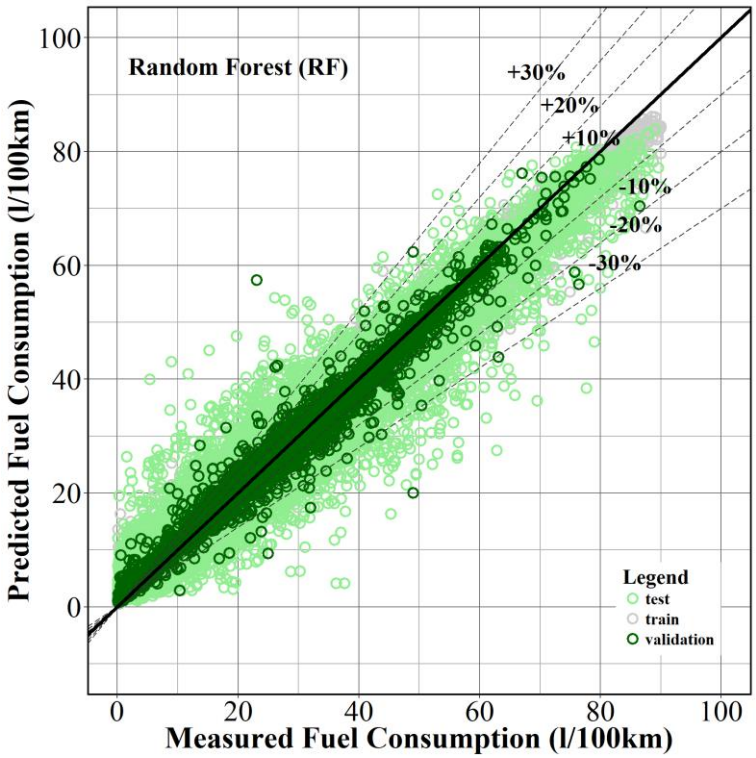


Figure 120. Fit of the developed RF for the heavy trucks App 2, from the Dataset 2.

Figure 119 shows the fit of the developed RF model for the available dataset from heavy trucks App 2. In particular, the figure shows how the model, developed for heavy trucks App 2, has good performance also on the validation set.

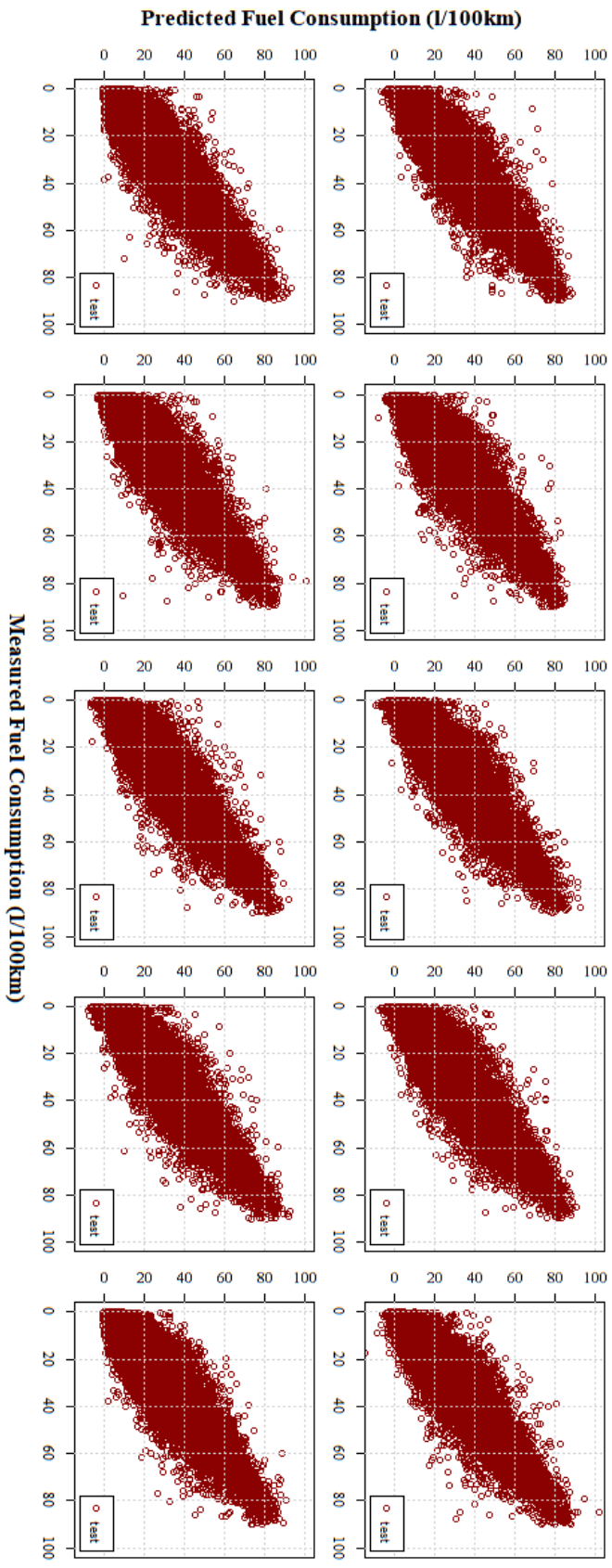


Figure 121. Ten-fold cross-validation of the ANN model for the heavy trucks App 2, from the Dataset 2.

Performance of the RF model seem much better than those of the other models (see Figure 118 and 122) (see Table 56). However, such higher performance might be due to specific characteristics of the considered validation set, which might have been slightly overfitted.

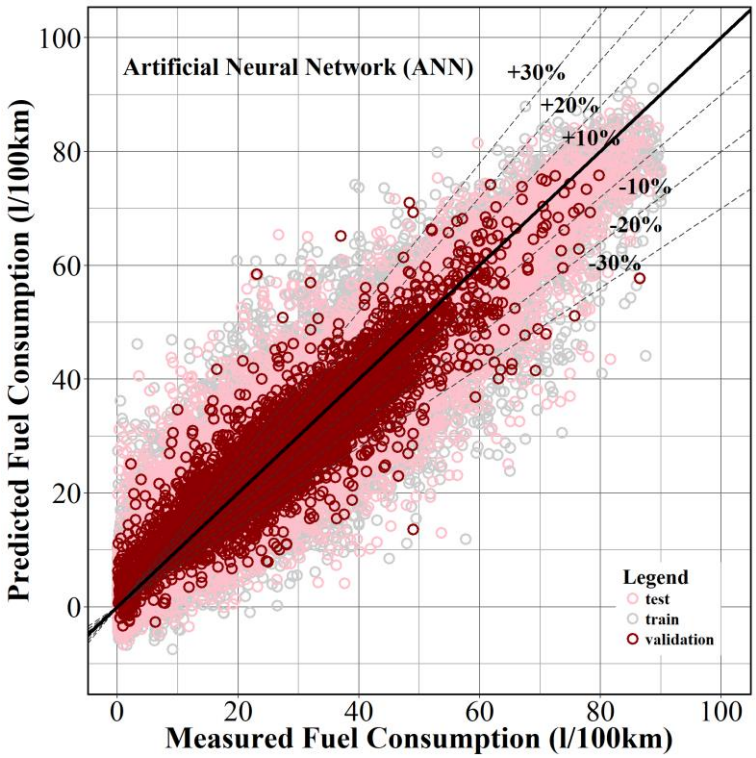


Figure 122. Fit of the developed ANN for the heavy trucks App 2, from the Dataset 2.

The developed ANN is structured on two layers of 14 and 4 neurons respectively. The model was trained using the rprop- algorithm (Riedmiller *et al.*, 1993), and it used the logistic activation function to identify patterns through the data. Figure 121 shows the fit of the ANN model developed for the heavy trucks App 2 on the test set over the ten-fold cross-validation process. The figure shows how the model

keeps quite consistent performance with quite low variance of the estimates made on each test set analysed.

Figure 122 shows the fit of the developed ANN model for the available dataset from heavy trucks App 2. In particular, the figure shows how the model, developed for heavy trucks App 2, has good performance also on the validation set. Performance of the ANN model seem good, similar to those of the developed SVR (see Figure 118), but lower than those of the RF (see Figure 120) (see Table 56).

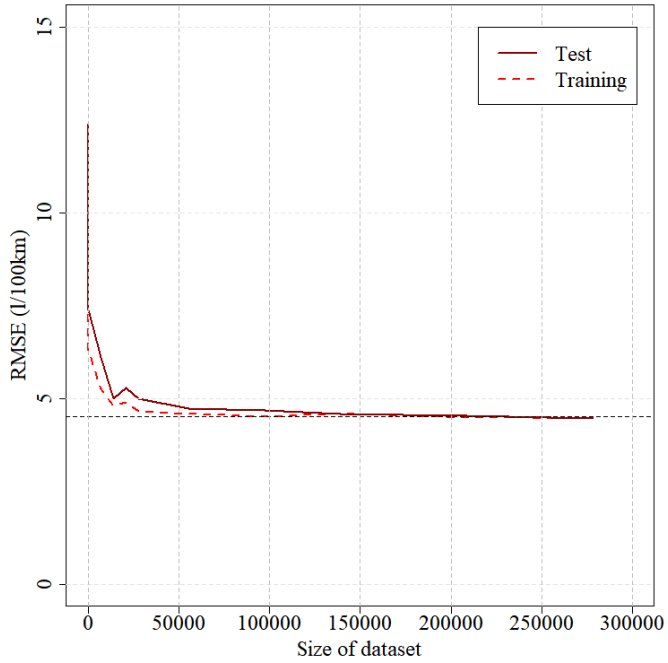


Figure 123. RMSE learning curve for the ANN fitted for the heavy trucks App 2, from the Dataset 2.

Figure 123 and 124 report the learning curves for RMSE and MAE respectively, calculated on the heavy trucks App 2 dataset. The curves were build using the ANN model only. From an analysis of these curves, it is possible to see that both RMSE and MAE seem to converge for the quantity of data analysed in this case study. In particular, both curves seem to converge for an amount of data between

150,000 and 200,000. This means that probably feeding more data to the models developed for heavy trucks would not improve accuracy of the models. Probably, that might help instead with light and medium trucks.

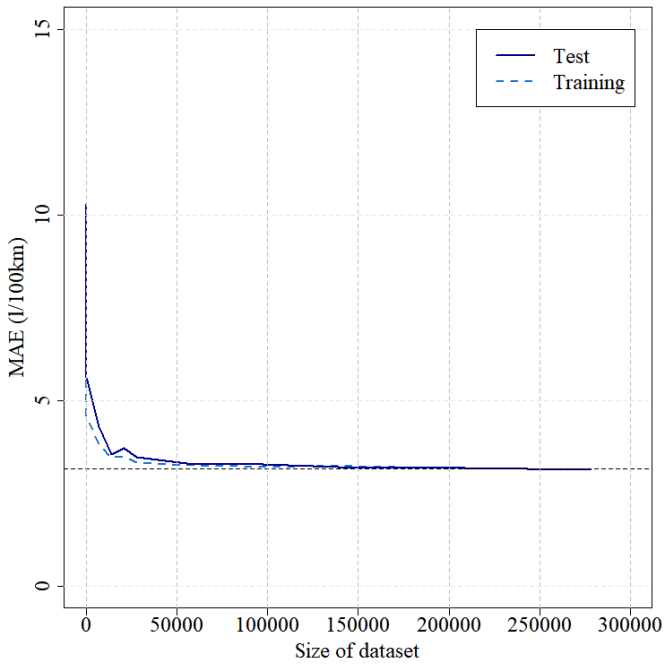


Figure 124. MAE learning curve for the ANN fitted for the heavy trucks App 2, from the Dataset 2.

Finally, the developed ANN1 and its performance are considered. The ANN1 is structured on three layers of 22, 30 and 22 neurons respectively. The model was trained using the rprop- algorithm (Riedmiller *et al.*, 1993), and it used the logistic activation function to identify patterns through the data. Figure 125 shows the fit of the ANN1 model developed for the heavy trucks App 2 on the test set over the ten-fold cross-validation process.

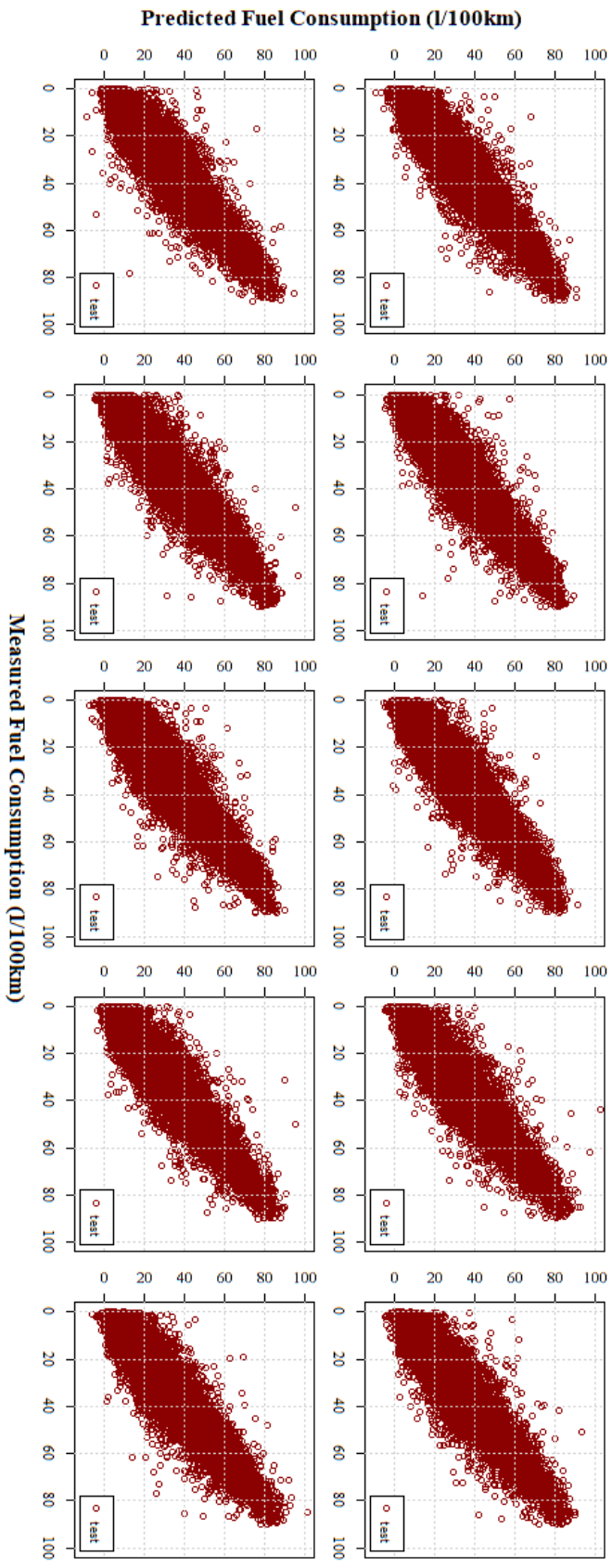


Figure 125. Ten-fold cross-validation of the ANN1 model for the heavy trucks App 2, from the Dataset 2.

The figure shows how the model keeps quite consistent performance with quite low variance of the estimates made on each test set analysed. In particular, variance of the estimates made by the ANN1 seem to be lower than that of the ANN (see Figure 121).

Figure 126 shows the fit of the developed ANN1 model for the available dataset from heavy trucks App 2.

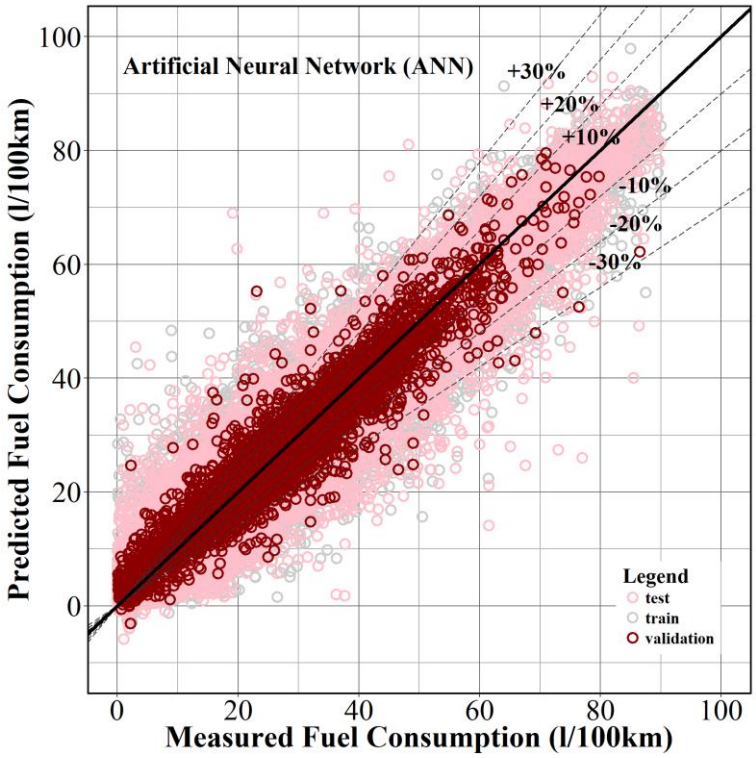


Figure 126. Fit of the developed ANN1 for the heavy trucks App 2, from the Dataset 2.

In particular, the figure shows how the model, developed for heavy trucks App 2, has good performance also on the validation set. Performance of this model seem to be comparable to those of the RF (see Figure 120), higher than those of the SVR (see Figure 118) and ANN (see Figure 122) (see Table 56).

Table 56 - Summary of performance of the regression models developed for the heavy trucks App 2.

Model	R^2_{cv}	$RMSE_{cv}$	MAE_{cv}	$Time_{cv}$	R^2_{val}	$RMSE_{val}$	MAE_{val}
	(-)	*(l/100 km)*	*(l/100 km)*	*(mins)*	*(-)*	*(l/100 km)*	*(l/100 km)*
LR	0.76	6.50	4.67	-	-	-	-
SVR	0.89	4.36	2.95	5,394	0.88	3.96	2.77
RF	0.94	3.31	2.17	7,963	0.97	1.63	0.91
ANN	0.88	4.54	3.18	5,953	0.86	4.39	3.12
ANN1	0.92	3.76	2.64	10,833	0.92	3.77	2.62
N.B.: The subscript 'val' stays for validation set, and 'cv' for cross-validation.							

Finally, a comparison of the performance of all the developed models for heavy trucks App 2 is reported in Table 56. This shows that RF and ANN1 are the models, which perform best at estimating the overall fuel consumption of the considered fleet of trucks. In particular, it is possible to see that the more complex structure of the ANN1 is able to improve performance of the algorithm at a level comparable to that of the RF model. That was not happening for a reduced amount of data such as that of the Dataset 1 (see Appendix E.1.2). Additionally, it is possible to see that, although performance of the RF is high also in the cross-validation, these are lower than those resulted on the validation set, which confirms a slight overfitting on the latter. Overall, it is possible to say that ML models perform better than the LR with R^2 that improves of up to 22% (for the RF model), and RMSE and MAE which improve of up to 80% (for the RF model). This, thanks to the application of ML, and secondary use of large quantities of data from existing databases owned by truck fleet managers, road agencies and the meteorological institutes, allows to achieve a level of accuracy higher to that of the HDM-4 recently calibrated for similar vehicles in the US (Chatti *et al.*, 2012).

Appendix E.2.3 Parametric Analysis 2

Finally, a parametric analysis was performed to see how each of the developed models addresses the impacts of each of the considered variables on the fuel consumption of each of the truck fleets part of the Dataset 2. Again, separate discussions are made for different sets part of the Dataset 2.

E.2.3.1 Light trucks on A roads (App 1)

Following the same schema used in the previous sections, results of the parametric analysis performed on the light trucks App 1, are reported first.

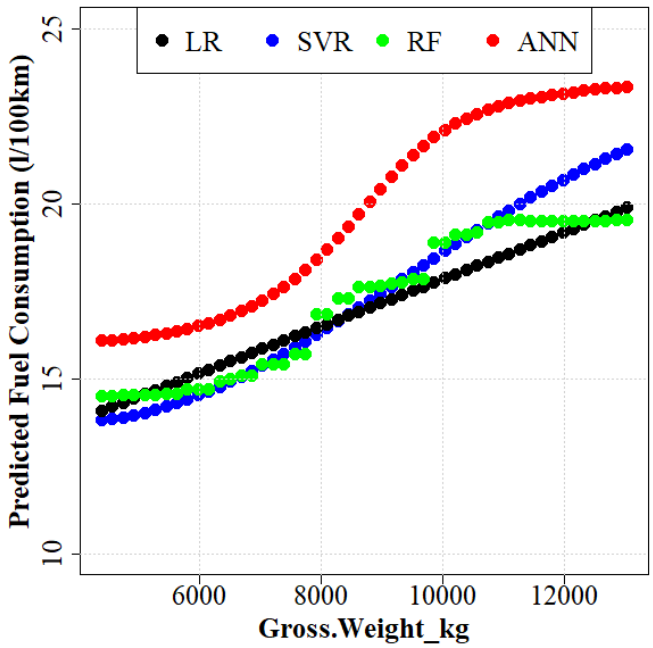


Figure 127. Parametric analysis of *Gross.Weight_kg* (kg) for the light trucks App 1.

Figure 127 shows variation of fuel consumption for the light trucks App 1 due to *Gross.Weight_kg*. From the figure, it is possible to see that all the developed models identify higher fuel consumption associated with higher vehicle weight (including payload). While LR, SVR and RF seems to show similar patterns, results

of the ANN seem to be influenced by other variables and the amount of data considered in the study. However, all models seem to attribute to *Gross.Weight_kg* a variation of about 8 l/100 km (from 14 l/100 km to 22 l/100 km) over a range of about 8000 kg (from 4000 kg to 12,000 kg).

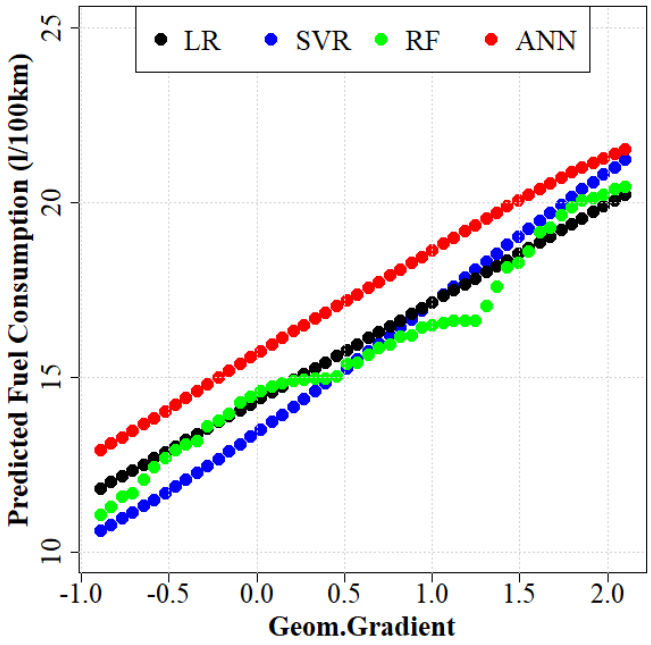


Figure 128. Parametric analysis of *Geom.Gradient* (%) for the light trucks App 1.

Figure 128 shows variation of fuel consumption for the considered fleet of trucks due to changes of the *Geom.Gradient*. All models seem to agree on the identified pattern, which sees fuel consumption to increase due to *Geom.Gradient* from about 10 l/100 km to approximately 22 l/100 km over a range of road gradient from -1% to 2%.

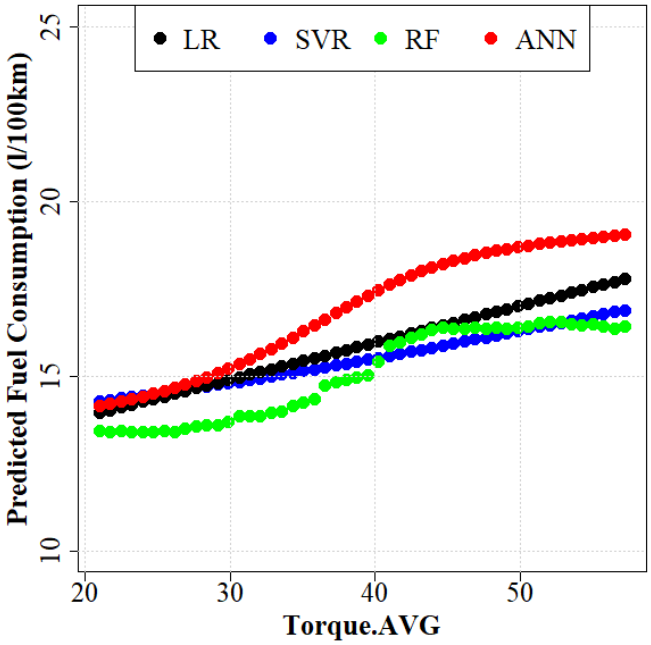


Figure 129. Parametric analysis of *Torque.AVG* (%) for the light trucks App 1.

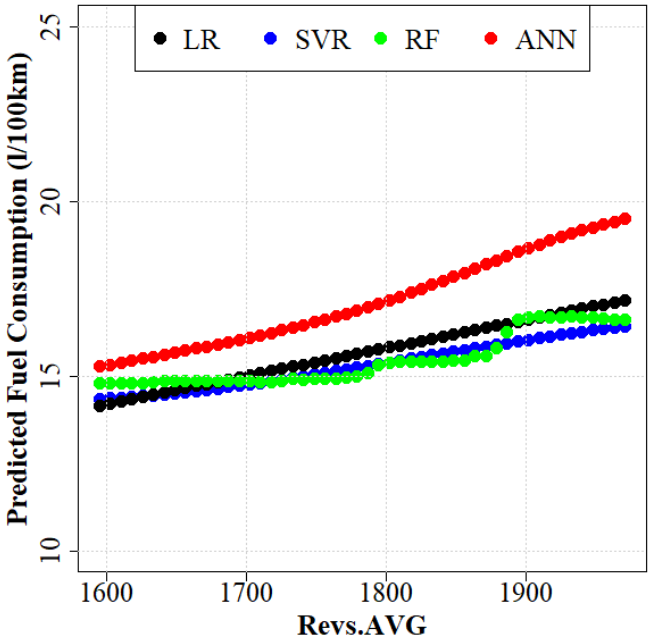


Figure 130. Parametric analysis of *Revs.AVG* (rpm) for the light trucks App 1.

Figure 129 and 130 show variation of fuel consumption for the considered fleet of truck for changes in the used *Torque.AVG* and *Revs.AVG* respectively. For

these similar trends are shown by each of the developed models with fuel consumption that increases for higher *Torque.AVG* and *Revs.AVG* of up to 5 l/100 km (from 14 l/100 km to 19 l/100 km) over a range of 20% - 60% for *Torque.AVG* and 1600 rpm - 2000 rpm for *Revs.AVG*.

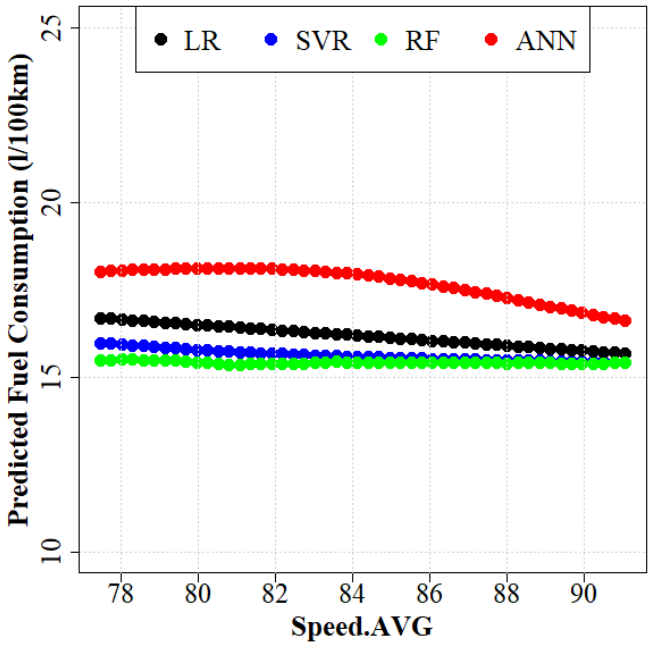


Figure 131. Parametric analysis of *Speed.AVG* (km/h) for the light trucks App 1.

Figure 131 shows variation of fuel consumption for the considered fleet of trucks for changes of *Speed.AVG*. From the figure, it is possible to see that the considered range of speeds is quite small, from 78 km/h to 90 km/h. In this range all models agree on the fact that higher speed seem to produce lower fuel consumption. Only the RF model seems to detect a very small increase which would be physically explainable. However, trends identified by the LR, SVR and ANN are like those detected by the same techniques on the Dataset 1 (see Appendix E.1.3). As explained, this might be due to the predominant effect of *Geom.Gradient*, which may cause vehicles driving downhill to drive slightly faster without consuming any additional

fuel, and vehicles driving uphill consuming more to drive slightly slower. That, due to the reduce amount of data considered, affects the results.

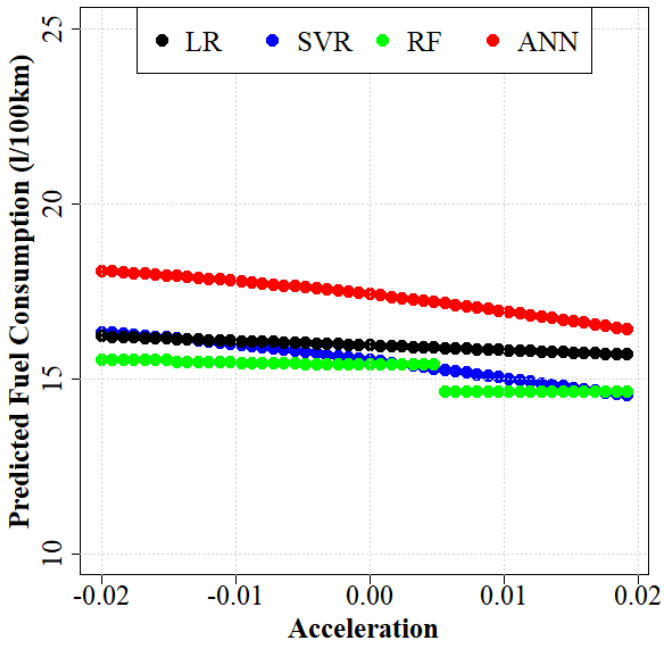


Figure 132. Parametric analysis of *Acceleration* (m/s^2) for the light trucks App 1.

Confirmation of what said regarding *Speed.AVG* can be found in Figure 132. This shows variation of fuel consumption for the considered fleet of trucks for changes of *Acceleration*. In fact, although data are filtered to include only records from trucks driving at relatively constant speed some variation of speed is still detected. However, trends identified by the developed models show that lower fuel consumption is associated to positive acceleration, which, again, it might highlight again an effect of *Geom.Gradient* on the results.

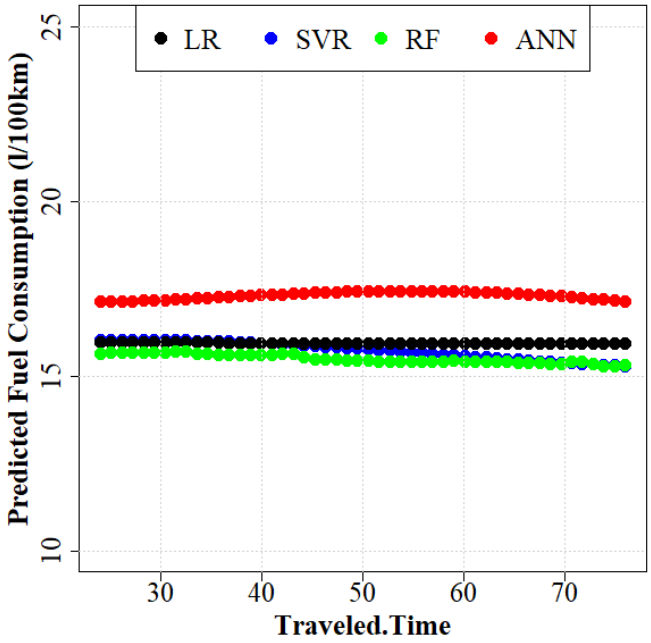


Figure 133. Parametric analysis of *Traveled.Time* (s) for the light trucks App 1.

Figure 133 shows variation of fuel consumption for the considered fleet of trucks for different periods of travel (*Traveled.Time*). From the figure, it is possible to see that RF and SVR seem to agree on the fact that lower fuel consumption is associated to longer time between one and the following record of data. LR seems to show no significant variation, while the ANN seems to indicate that fuel consumption might increase due to higher *Traveled.Time*. Variations are small, but probably the amount of data analysed is not enough to allow all the models to identify a unique pattern.

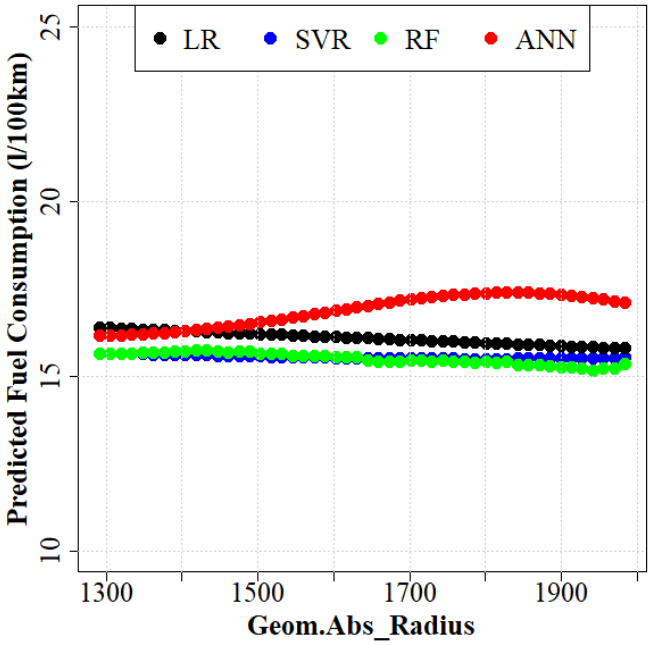


Figure 134. Parametric analysis of *Geom.Abs_Radius* (m) for the light trucks App 1.

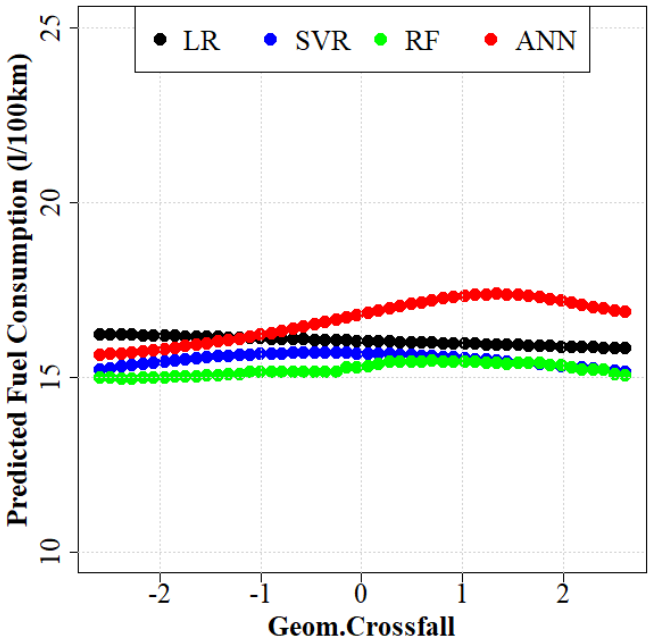


Figure 135. Parametric analysis of *Geom.Crossfall* (%) for the light trucks App 1.

Figure 134 and 135 show variation of fuel consumption for the considered fleet of trucks for changes of *Geom.Abs_Radius* and *Geom.Crossfall* respectively. These are characteristics of the road infrastructure which are established in the design phase of the road and that are correlated one to each other. In fact, in order to guarantee an adequate disposal of rainwater and balance of vehicles *Geom.Crossfall* varies with *Geom.Abs_Radius* (Highways Agency, 2006a). From an analysis of Figure 134 and 135 it is possible to see that, the ANN shows trends somewhat different from those of LR, SVR and RF. However, it is possible to notice that all models tend to show similar trends between *Geom.Crossfall* and *Geom.Abs_Radius*, with Figure 134 and 135 that result quite similar. This is due to the correlation of these two variables, and highlights stability of the applied algorithms.

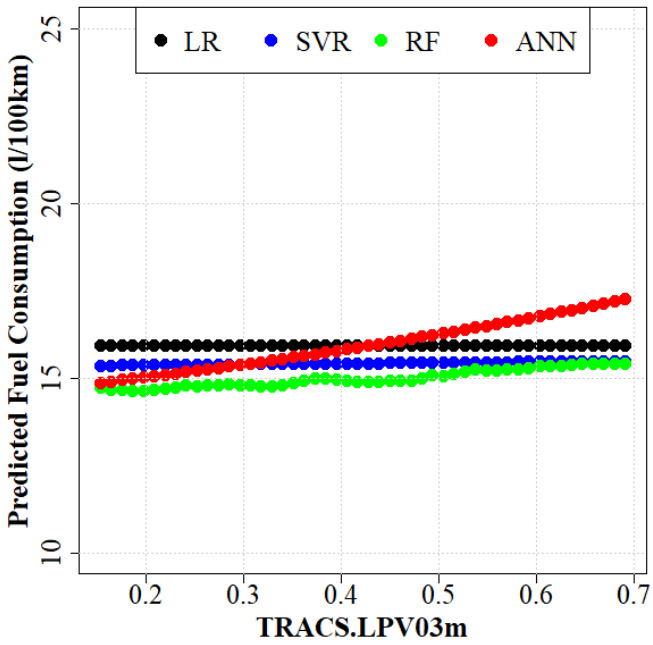


Figure 136. Parametric analysis of *TRACS.LPV03m* (mm^2) for the light trucks App 1.

Figure 136 shows variation of fuel consumption due to changes of *TRACS.LPV03m*. This is a characteristic of the road surface and, although the considered range of *TRACS.LPV03m* is very small, it is possible to see that while LR and SVR seem to identify no significant increase or decrease in fuel consumption due to variation of LPV at 3 metres wavelength, RF and ANN detect an increase. That would be variation of about 1 – 1.5 l/100 km over the considered range of *TRACS.LPV03m*.

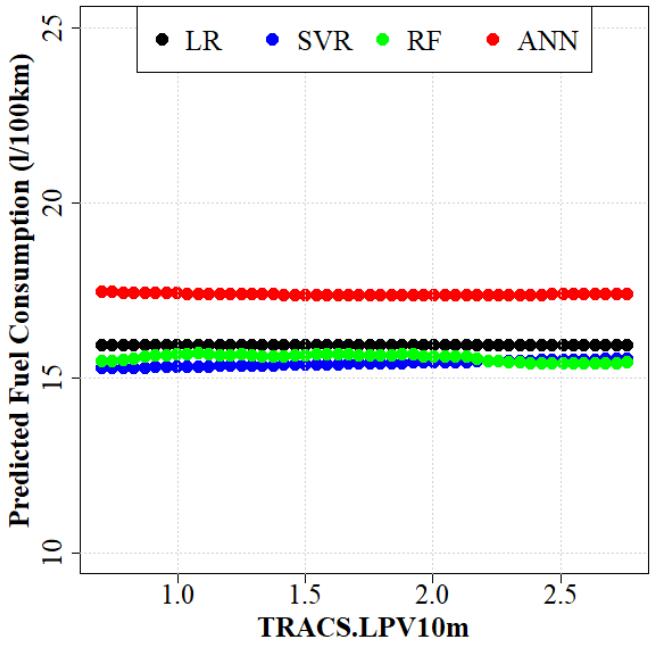


Figure 137. Parametric analysis of *TRACS.LPV10m* (mm^2) for the light trucks App 1.

Figure 137 shows variation of fuel consumption due to changes of *TRACS.LPV10m*. This is another characteristic of the road surface, and it is possible to see that all models seem to identify a very little effect of this on the fuel consumption of the considered fleet of trucks. Although, the considered range of *TRACS.LPV10m* is very small (0.8 mm^2 - 2.7 mm^2), the identified trend seems to be insignificant.

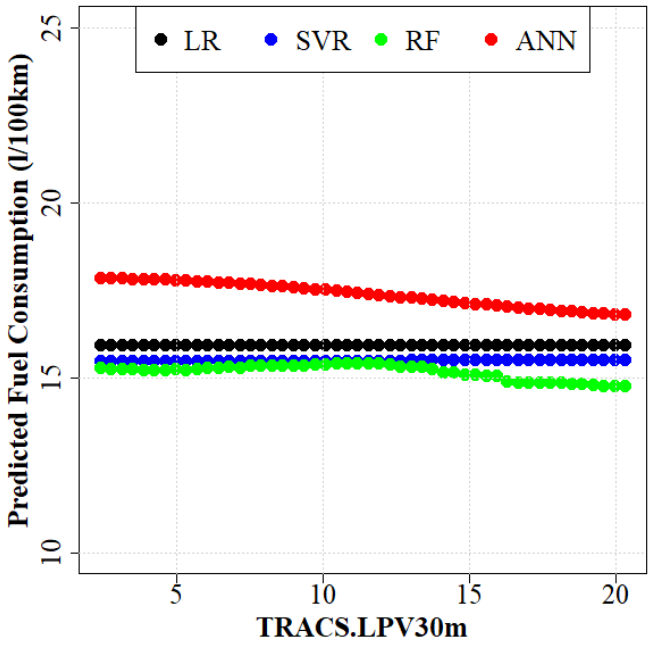


Figure 138. Parametric analysis of *TRACS.LPV30m* (mm^2) for the light trucks App 1.

Figure 138 shows variation of fuel consumption due to changes of *TRACS.LPV30m*. This is the third wavelength of LPV investigated. From an analysis of Figure 138, it is possible to see that LR and SVR identify a small increase of fuel consumption due to *TRACS.LPV30m*. Instead, RF and ANN, which show similar patterns, seem to identify a decrease of fuel consumption over the same range of road characteristics (between 1 mm^2 and 20 mm^2). Although this result might be affected by the correlation of *TRACS.LPV30m* with *Speed.AVG*, it must be considered that, the small variation of LPV at 30 metres considered, together with the small amount of data analysed might not help the used algorithm at identifying a unique pattern through the data.

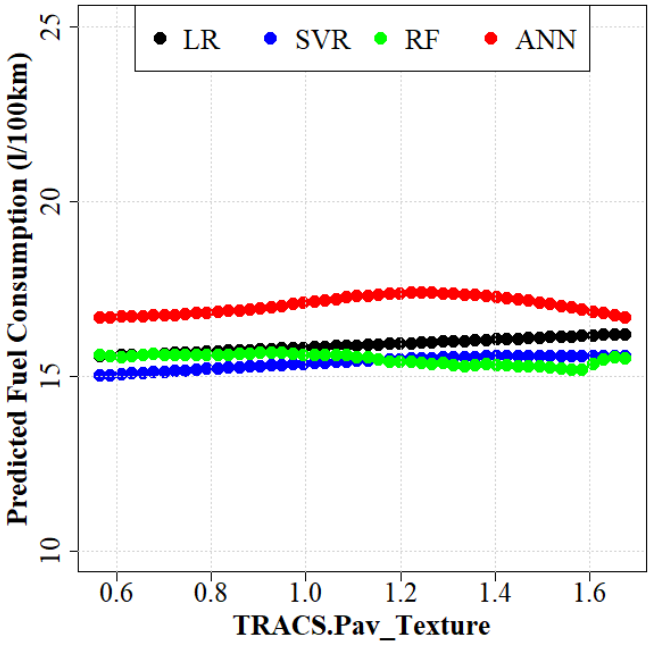


Figure 139. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the light trucks App 1.

Figure 139 shows variation of fuel consumption due to changes of *TRACS.PAV_Texture*. From the figure, it is possible to see that, while the ANN identifies a slightly different pattern, perhaps due to overfitting of few data, all other models seems to agree on the fact that a higher SMTD produces higher fuel consumption. In particular, over a range of *TRACS.PAV_Texture* between 0.6 mm and 1.65 mm, fuel consumption increases by approximately 0.5 l/100 km, for the considered fleet of trucks.

Figure 140 and 141 show variation of fuel consumption due to *Altitude.AVG* and *Atm_Pressure_hPa* respectively. From the figures, it looks like the amount of data analysed does not allow the applied algorithms to identify any clear trend. In fact, all models define different trends, probably due to overfitting of different data.

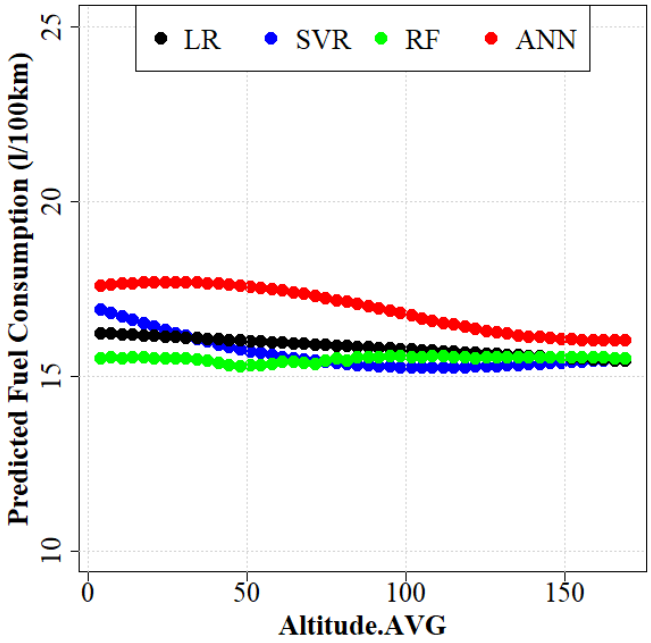


Figure 140. Parametric analysis of *Altitude.AVG* (mamls) for the light trucks App 1.

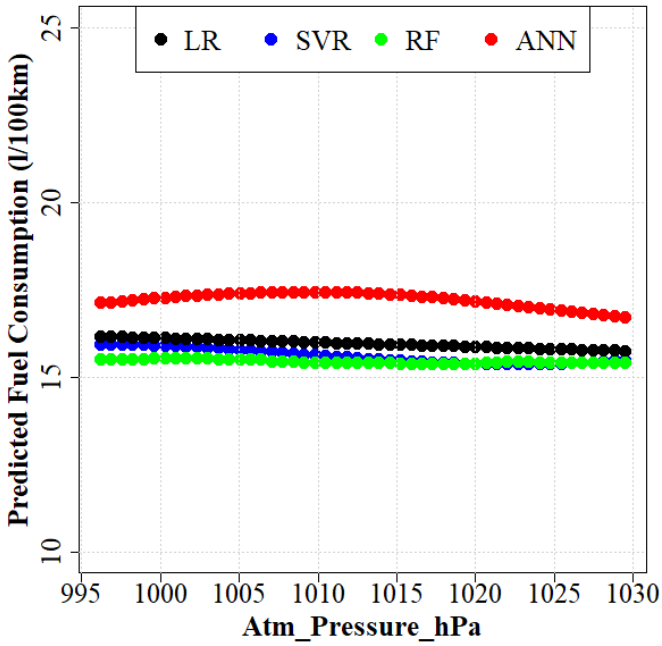


Figure 141. Parametric analysis of *Atm_Pressure_hPa* (hPa) for the light trucks App 1.

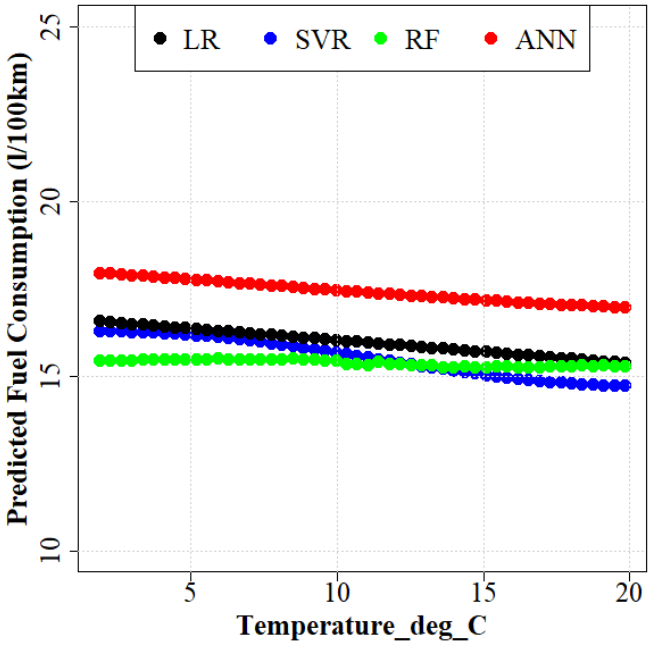


Figure 142. Parametric analysis of *Temperature_deg_C* (°C) for the light trucks App 1.

Figure 142 and 143 show variation of fuel consumption due to changes of *Temperature_deg_C* and *Humidity_*. Due to correlation of these two variables, it is possible to see that the two figures result in similar trends identified by the applied algorithms for both variables. In fact, all the developed models identify a decrease in fuel consumption due to higher temperature. Similarly, the developed models define variations of fuel consumption due to *Humidity_*. Although measurements of these variables happen at miles from the location where trucks are actually driving, results show that these can help at improving accuracy and interpretability of the developed models.

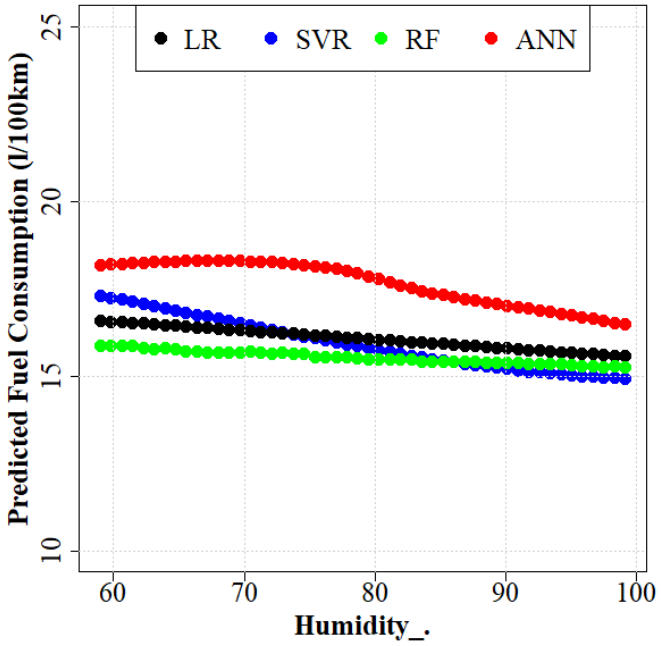


Figure 143. Parametric analysis of *Humidity_*. (%) for the light trucks App 1.

Figure 144 shows variation of fuel consumption due to changes of *WindSpeed.Dri_kn*. As specified at sect. 4.2.2, positive *WindSpeed.Dri_kn* identify wind which blows in opposite direction to the direction of travel of the considered truck. Again, from the figure, it is possible to say that the adopted algorithms are able to identify an increase in fuel consumption due to stronger wind. In fact, all the developed models identify the increasing trend. However, due to the fact that measurements are not measured exactly at the position where trucks are driving, impacts are estimated low, between 1.0 and 2.0 l/100 km over a range of wind speeds between ± 100 kn. Because, from a physical perspective, an increase of vehicle speed should produce an increase of air drag proportional to the square of the difference in vehicle speed (Guzzella *et al.*, 2015), considering the effect of wind blowing in opposite direction to that of travel of a vehicle to have similar effects to a change of the vehicle speed itself (which in reference to the air is reasonable), these estimates might be considered unrealistic (Saltzman *et al.*, 1999; Hammache *et al.*, 2001;

Karlsson *et al.*, 2015; Ekman *et al.*, 2016). However, it is interesting to see that on average trucks driving with windy weather conditions tend to consume more than others, and that SVR, RF and ANN are able to detect that, also for a relatively small amount of data as that considered in the light trucks App 1.

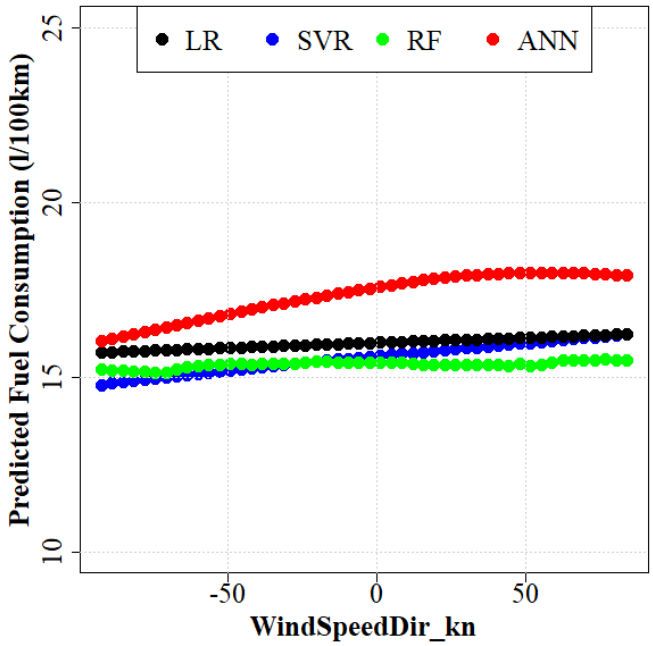


Figure 144. Parametric analysis of *WindSpeedDir_kn* (kn) for the light trucks App 1.

Table 57 and 58 report impacts on fuel consumption calculated between the 25th - 75th and 5th - 95th percentiles of the considered variables, respectively. This provides a numerical estimation of how each of the developed models estimates the impact of each of the considered variables on fuel consumption, it helps in the interpretation of the parametric analysis, and provides with values to compare with results from the studies conducted in the past, draw conclusions and make recommendations. As mentioned, two calculations were made. One for the 25th - 75th percentiles (see Table 57) and another for the 5th - 95th percentiles (see Table 58). This

is because although there might be more confidence in the estimates made within the first range, the latter might be of better interest for the considered study, as this will include the extreme conditions of the road surface, and their impact on the fuel consumption of the considered fleet of trucks.

Table 57 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the light trucks App 1.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	2.96	18.31	4.65	28.80	4.52	27.98	5.99	37.10
Geom.Gradient	4.29	26.57	5.64	34.91	3.26	20.18	4.56	28.21
Torque.AVG	1.95	12.09	1.36	8.39	2.75	17.06	3.43	21.25
Revs.AVG	1.54	9.51	1.20	7.44	0.71	4.40	2.28	14.15
Speed.AVG	-0.52	-3.21	-0.26	-1.59	0.03	0.16	-0.74	-4.56
Acceleration	-0.26	-1.64	-0.94	-5.82	-0.82	-5.07	-0.84	-5.21
Traveled.Time	-0.01	-0.06	-0.46	-2.87	-0.19	-1.20	0.15	0.90
Geom.Abs_Radius	-0.31	-1.89	-0.09	-0.57	-0.30	-1.87	0.99	6.14
Geom.Crossfall	-0.21	-1.28	-0.09	-0.58	0.34	2.13	1.36	8.40
TRACS.LPV03m	0.00	0.02	0.06	0.40	0.41	2.56	1.21	7.46
TRACS.LPV10m	0.01	0.05	0.13	0.83	-0.17	-1.05	-0.03	-0.18
TRACS.LPV30m	-0.01	-0.04	0.01	0.08	-0.24	-1.46	-0.65	-4.02
TRACS.Pav_ Texture	0.32	2.00	0.34	2.08	-0.29	-1.78	0.46	2.86
Altitude.AVG	-0.41	-2.52	-0.63	-3.92	0.15	0.94	-1.34	-8.32
Atm_Pressure_hPa	-0.22	-1.33	-0.45	-2.78	-0.12	-0.72	-0.23	-1.44
Temperature_deg_C	-0.61	-3.78	-1.08	-6.70	-0.25	-1.54	-0.54	-3.35
Humidity_.	-0.52	-3.23	-1.36	-8.40	-0.28	-1.74	-1.19	-7.36
WindSpeedDir_kn	0.26	1.64	0.71	4.39	0.00	0.02	1.19	7.35

Table 58 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the light trucks App 1.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	5.32	32.97	7.29	45.14	5.00	30.95	7.14	44.23
Geom.Gradient	7.72	47.83	9.80	60.70	8.83	54.66	8.03	49.75
Torque.AVG	3.51	21.76	2.39	14.77	3.07	19.01	4.74	29.36
Revs.AVG	2.76	17.12	1.95	12.06	1.88	11.63	3.93	24.32
Speed.AVG	-0.93	-5.77	-0.51	-3.16	-0.08	-0.49	-1.25	-7.75
Acceleration	-0.48	-2.95	-1.66	-10.26	-0.93	-5.73	-1.51	-9.33
Traveled.Time	-0.02	-0.10	-0.69	-4.28	-0.32	-1.97	0.08	0.50
Geom.Abs_Radius	-0.55	-3.41	-0.13	-0.83	-0.48	-2.98	1.08	6.66
Geom.Crossfall	-0.37	-2.30	-0.02	-0.14	0.26	1.58	1.35	8.37
TRACS.LPV03m	0.01	0.04	0.12	0.72	0.76	4.72	2.16	13.38
TRACS.LPV10m	0.01	0.09	0.24	1.50	-0.07	-0.41	-0.06	-0.34
TRACS.LPV30m	-0.01	-0.08	0.02	0.14	-0.45	-2.80	-0.98	-6.07
TRACS.Pav_ Texture	0.58	3.61	0.55	3.41	-0.21	-1.29	0.16	1.02
Altitude.AVG	-0.73	-4.54	-1.37	-8.48	-0.02	-0.10	-1.58	-9.77
Atm_Pressure_hPa	-0.39	-2.40	-0.52	-3.19	-0.11	-0.65	-0.35	-2.18
Temperature_deg_C	-1.10	-6.81	-1.52	-9.42	-0.15	-0.90	-0.92	-5.70
Humidity_.	-0.94	-5.81	-2.26	-14.02	-0.60	-3.70	-1.57	-9.72
WindSpeedDir_kn	0.48	2.95	1.33	8.22	0.26	1.62	1.86	11.52

E.2.3.2 Light trucks on Motorways and A roads (App 2)

This section is dedicated to the results of the parametric analysis performed on the light trucks App 2, from the Dataset 2. Figure 145 shows variation of fuel consumption due to changes of *Gross.Weight_kg*. From the figure, it is possible to see that a non-linear behaviour is detected by all ML models which is approximated instead with a line by the LR. The estimated impact of *Gross.Weight_kg* for this type of truck seem to be quite high, with a variation of about 10 l/100 km (from 13 l/100

km up to 21 l/100 km) between 4,000 kg and 16,000 kg. On a note, it is possible to observe that, for this vehicle type, the range of *Gross.Weight_kg* investigated is larger than that analysed for the light trucks App 1.

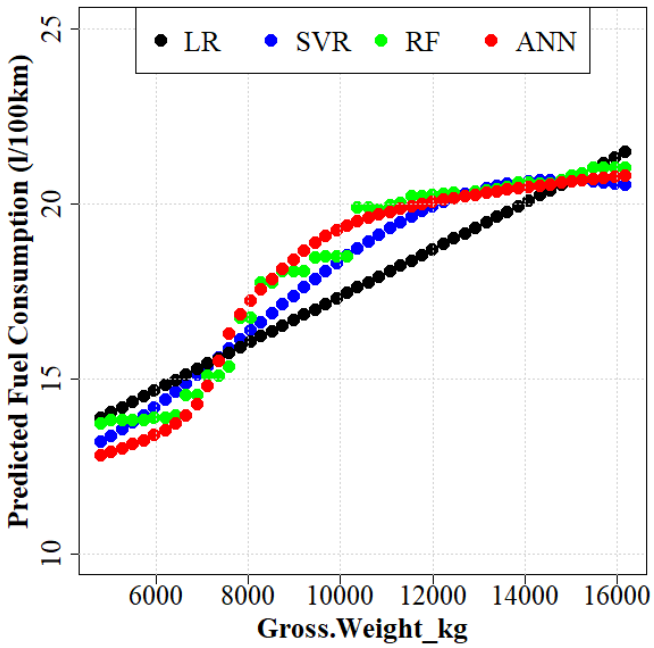


Figure 145. Parametric analysis of *Gross.Weight_kg* (kg) for the light trucks App 2.

While, it is unrealistic that trucks are more heavily loaded on motorways than what they are on A roads, it must be considered that the *Gross.Weight_kg* is estimated by Microlise based on engine performance parameters (see sect. 3.1.2), and therefore this estimate might be affected by the different distribution of vehicle speeds and engine performance typical of the two road types considered in the study.

Figure 146 shows instead variation of fuel consumption due to changes of *Geom.Gradient* for the considered fleet of trucks. The figure shows that all the models make similar estimates with fuel consumption which can vary of approximately 15 l/100 km (from 10 l/100 km up to 25 l/100 km), in a range of *Geom.Gradient* between -1.5 % and 2 %.

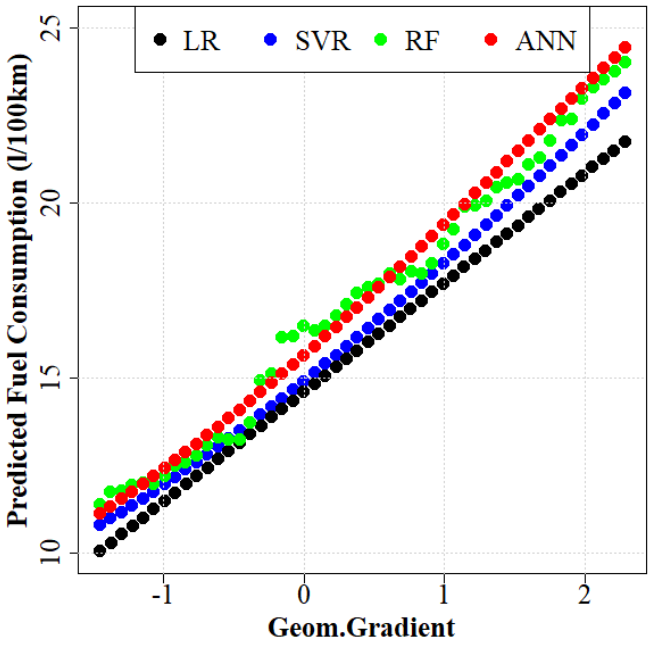


Figure 146. Parametric analysis of *Geom.Gradient* (%) for the light trucks App 2.

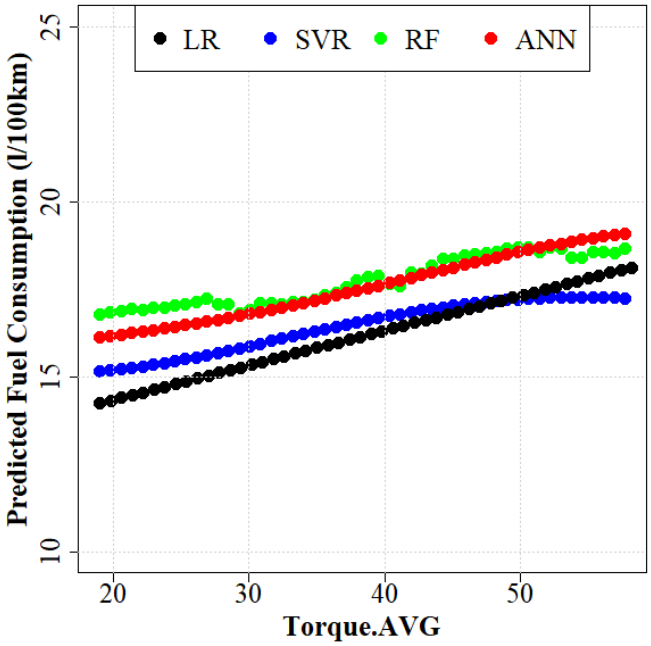


Figure 147. Parametric analysis of *Torque.AVG* (%) for the light trucks App 2.

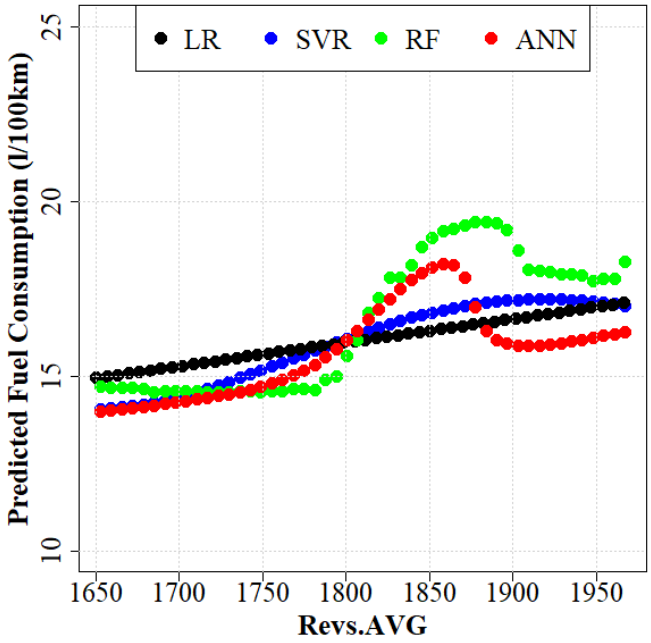


Figure 148. Parametric analysis of *Revs.AVG* (rpm) for the light trucks App 2.

Figure 147 and 148 show variation of fuel consumption for changes of *Torque.AVG* and *Revs.AVG* respectively. From the figures, it is possible to see that all models tend to agree on the identified patterns for both type of measurements considered. In particular, the models show that fuel consumption increases with higher *Torque.AVG* or higher *Revs.AVG* of approximately 3 or 4 l/100 km over the considered range. In regard to the *Revs.AVG*, it is possible to see that RF and ANN seem to generate some overfitting on the fuel consumption estimates for high values of *Revs.AVG*. However, because RF and ANN behave very similarly, this might be due to a considerable amount of specific records, which make the models behave this way. This requires further investigation.

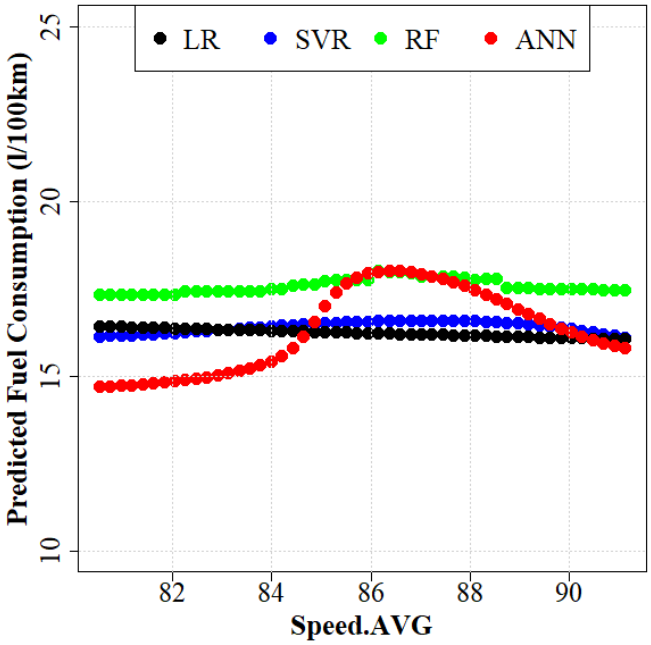


Figure 149. Parametric analysis of *Speed.AVG* (km/h) for the light trucks App 2.

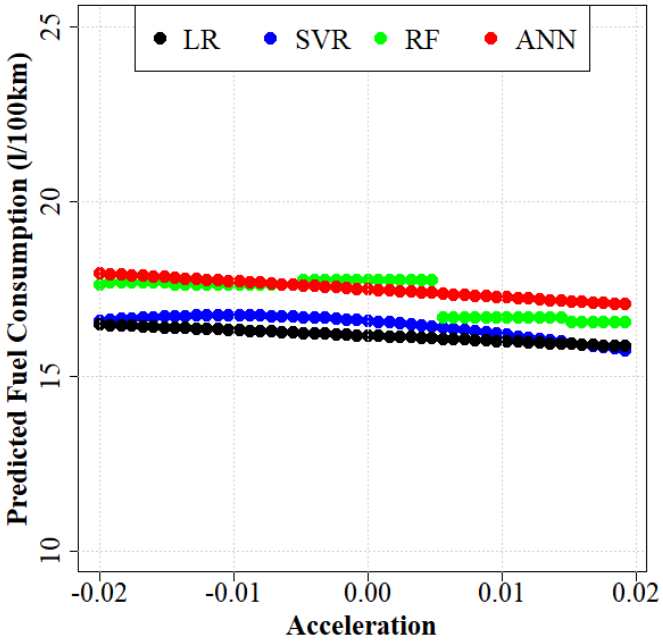


Figure 150. Parametric analysis of *Acceleration* (m/s^2) for the light trucks App 2.

Figure 149 and 150 show variation of fuel consumption for changes of *Speed.AVG* and *Acceleration*. Thanks to the larger quantity of data fed to the developed models, it

is possible to see that the applied algorithms are able to make a difference between the fuel consumption due to *Speed.AVG* and that due to *Acceleration*. This might be due to the higher amount of data considered, which allows the ML algorithms (i.e. SVR, RF and ANN) to highlight the hidden patterns. The fact that, for similar (or higher) amount of the data considered, similar patterns are detected for the medium and heavy trucks, in both the App 1 and the App 2 (from Appendix E.2.3.3 to sect. E.2.3.6), gives more confidence in the statement above. Regarding *Speed.AVG*, Figure 149 shows that, although some overfitting of the ANN model is visible, the ML models seem to agree that fuel consumption increase with higher speeds, also in consideration of a small range of speeds. This can be estimated to be approximately 1.5 l/100 km, over a range of speeds from 80 km/h to 95 km/h. In regard to *Acceleration* (see Figure 150), instead, all models seem to identify a reduction of fuel consumption of approximately 1 or 2 l/100 km over a range of ± 0.02 m/s^2. Although this might not be intuitive as a finding, this might be due to predominance of the effect of road gradient. This is also confirmed by the other case studies (from Appendix E.2.3.3 to E.2.3.6).

Figure 151 shows variation of fuel consumption due to changes of *Traveled.Time*. From the figure, it is possible to see that the developed models identify similar trends, with this variable, which do not seem to affect the fuel consumption of the considered fleet of trucks significantly with maximum variation of less than 0.5 l/100 km over the investigated range.

Figure 152 shows variation of fuel consumption due to change of *Geom.Abs_Radius*. The figure shows that all models seem to estimate different trends, which might be due to overfitting of different set of data. The LR model for example shows no impact, SVR and ANN a reduction of fuel consumption while the RF identifies an increase. It is therefore difficult to evaluate reliability of these estimates.

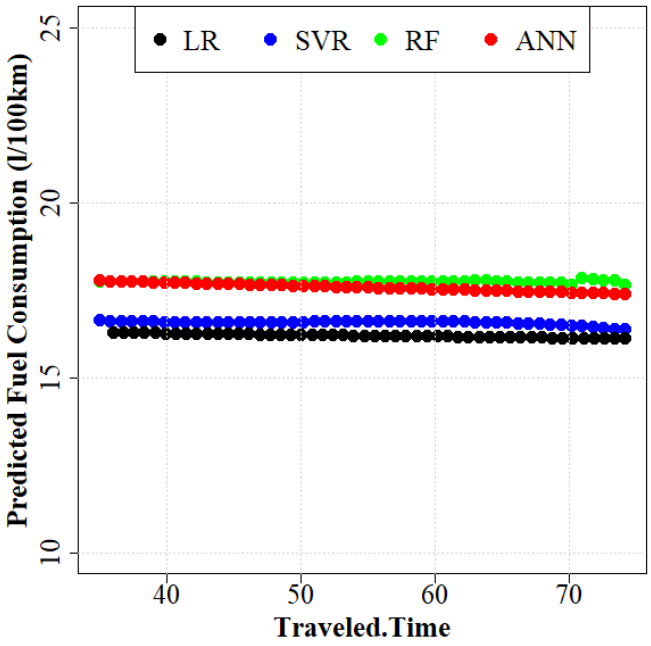


Figure 151. Parametric analysis of *Traveled.Time* (s) for the light trucks App 2.

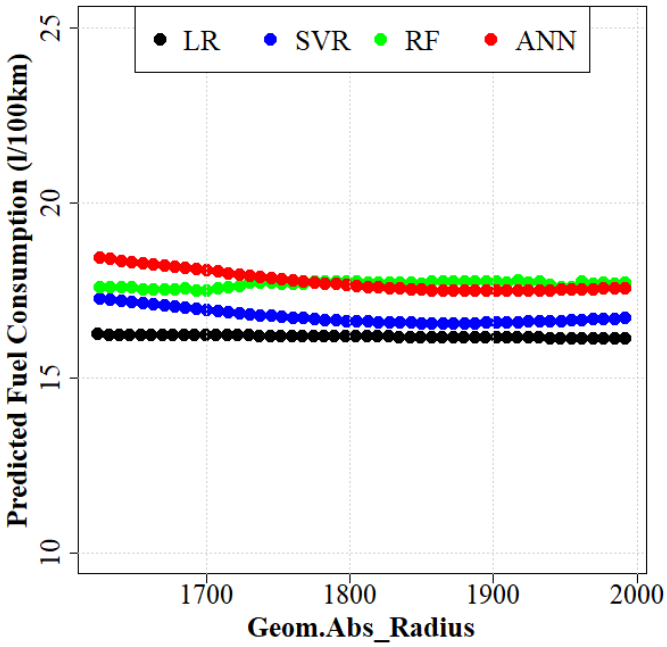


Figure 152. Parametric analysis of *Geom.Abs_Radius* (m) for the light trucks App 2.

For example, we saw earlier, that *Geom.Crossfall* is correlated to *Geom.Abs_Radius* due to the geometrical design of road pavements. However, in this case study, the developed models seem to find different patterns of fuel consumption variation for *Geom.Crossfall* and *Geom.Abs_Radius*. Figure 153 shows variation of fuel consumption due to changes of *Geom.Crossfall*. From the figure, it is possible to see that while the LR model seems to identify no significant impact of this variable on fuel consumption, SVR, RF and ANN seem to agree on the fact that higher fuel consumption is associated to positive *Geom.Crossfall*. This means that between a road at a turn and a straight road fuel consumption can increase of about 1 l/100 km, on motorways and primary roads in the UK, and it might be due to the higher speed characteristic of straight road segments.

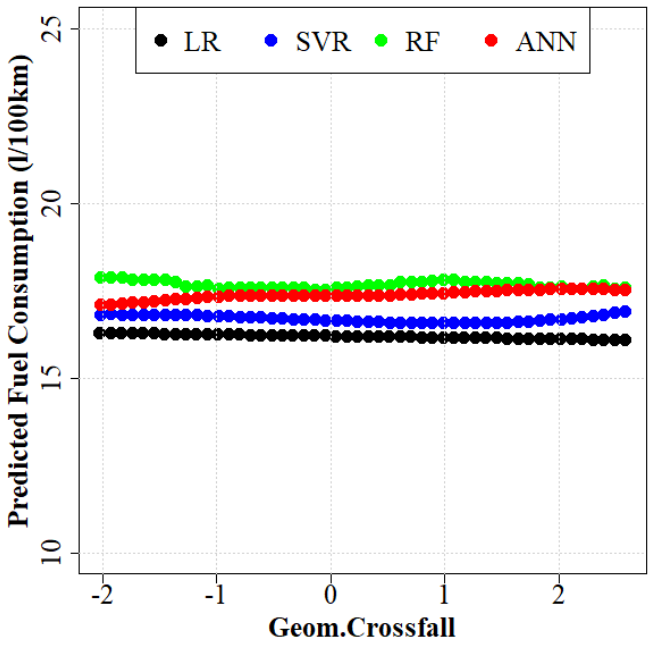


Figure 153. Parametric analysis of *Geom.Crossfall* (%) for the light trucks App 2.

Figure 154 shows variation of fuel consumption for changes of *TRACS.LPV03m.* From the figure, it is possible to see that the considered range is smaller in this case than for the case in which A roads only were considered. This is true also for the other road condition measurements(see Figure 155, 156, and 157) and truck types considered in the study (from Appendix E.2.3.3 to E.2.3.6), and is probably due to the fact that majority of road segments on motorways are in very good conditions, which reduces the range of 5^{th} - 95^{th} percentiles considered in the parametric analysis. The figure shows that for the considered range of road surface characteristics, and for the considered truck type, no significant effect seems to be detected by any of the developed models.

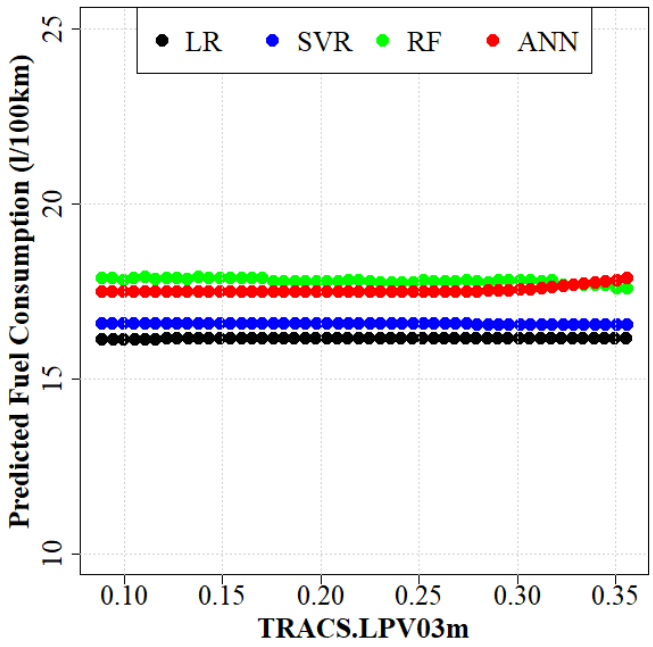


Figure 154. Parametric analysis of *TRACS.LPV03m* (mm^2) for the light trucks App 2.

Only the ANN seems to show a slight increase in fuel consumption due to higher *TRACS.LPV03m* but, since none of the other models detects that, this might be

due to some overfitting from a few records. However, it must be noticed that, in comparison to the previous case study, a much smaller range of road surface conditions is investigate, which can affect validity and applicability of the obtained results.

Figure 155 shows, instead, variation of fuel consumption due to changes of *TRACS.LPV10m.* The figure shows that while LR and SVR seem to do not detect any impact from this variable on fuel consumption, RF and ANN seem to agree that higher fuel consumption can be associated to higher values of *TRACS.LPV10m.* However, also in this case, the considered range of road characteristics is small and these results might be representative of these conditions only.

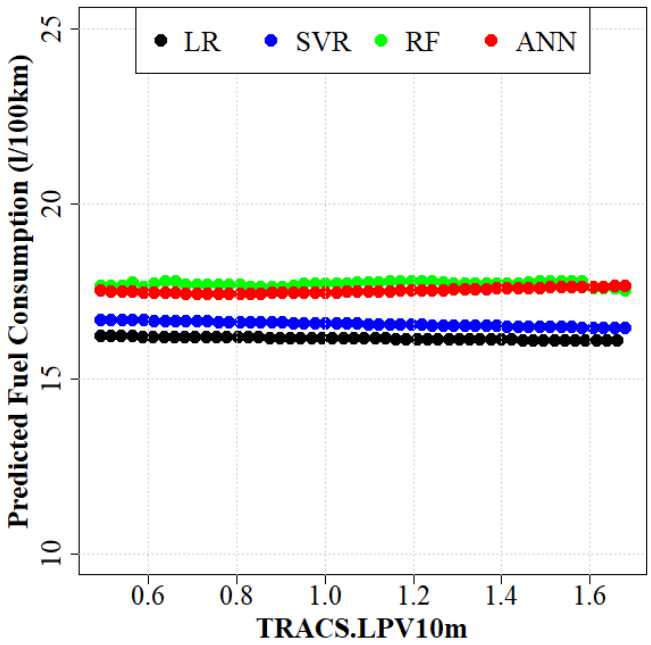


Figure 155. Parametric analysis of *TRACS.LPV10m* (mm^2) for the light trucks App 2.

Figure 156 shows variation of fuel consumption for changes of *TRACS.LPV30m.* From the figure, it is possible to see that all the developed models

seem to estimate an increase of fuel consumption (of approximately 1 l/100 km) over the considered range of road surface characteristics (between 3.5 and 12 mm^2). The fact that all models detect similar patterns, gives more confidence about the identified trend, which only partially matches the findings for the light trucks App 1, which showed a potential decrease of fuel consumption for *TRACS.LPV30m* higher than 12 mm^2.

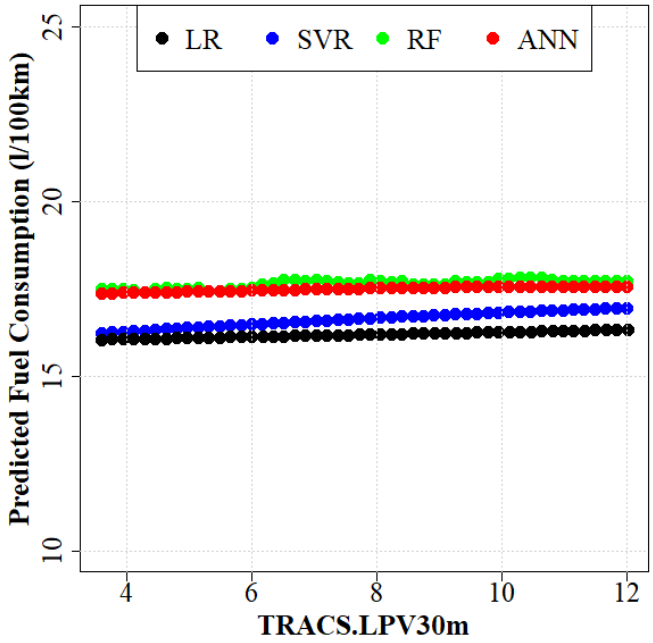


Figure 156. Parametric analysis of *TRACS.LPV30m* (mm^2) for the light trucks App 2.

That could be due to possible lower speeds adopted by drivers, who, due to the high vibrations induced by the deformations of the road surface on the vehicle, decelerate producing lower fuel consumption (Hammarström *et al.*, 2012). However, this requires further investigation.

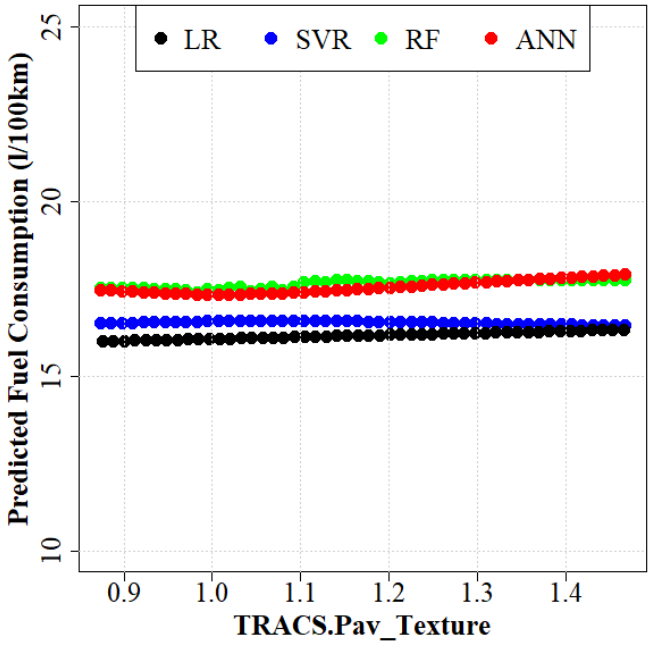


Figure 157. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the light trucks App 2.

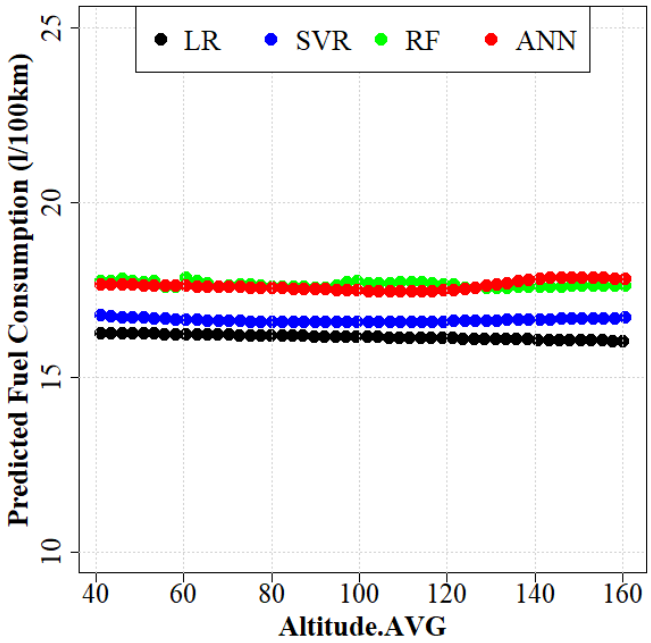


Figure 158. Parametric analysis of *Altitude.AVG* (mamls) for the light trucks App 2.

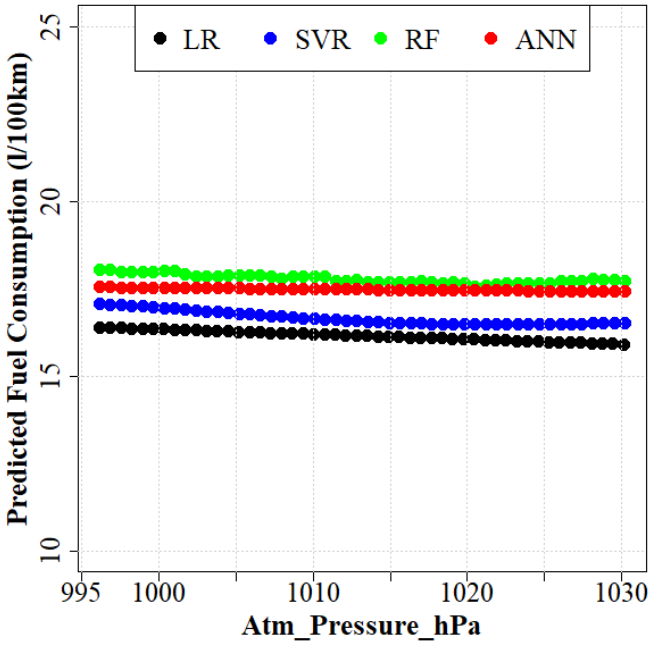


Figure 159. Parametric analysis of *Atm_Pressure_hPa* (hPa) for the light trucks App 2.

Figure 157 shows variation of fuel consumption for changes of *TRACS.Pav_Texture* (SMTD). The figure shows that all models seem to agree on the fact that, in the considered range of road characteristics, higher *TRACS.Pav_Texture* produces an increase of fuel consumption of about 1 l/100 km in a range between 0.9 and 1.5 mm of SMTD.

Figure 158 and 159 show variation of fuel consumption for changes of *Altitude.AVG* and *Atm_Pressure_hPa* respectively. For these two variables, and for the considered truck types, all the developed models seem to not find any strong correlation between these and fuel consumption. Trends are hardly visible, but it is possible to see that models are quite robust and avoid overfitting.

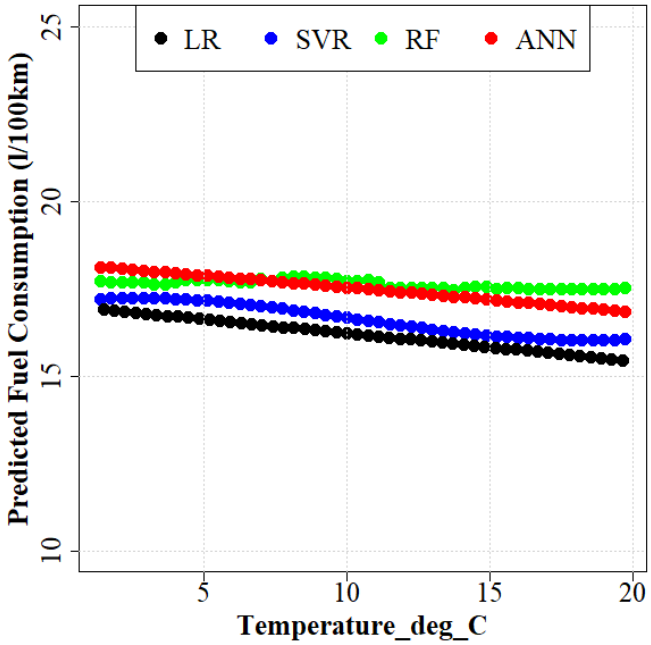


Figure 160. Parametric analysis of *Temperature_deg_C* (°C) for the light trucks App 2.

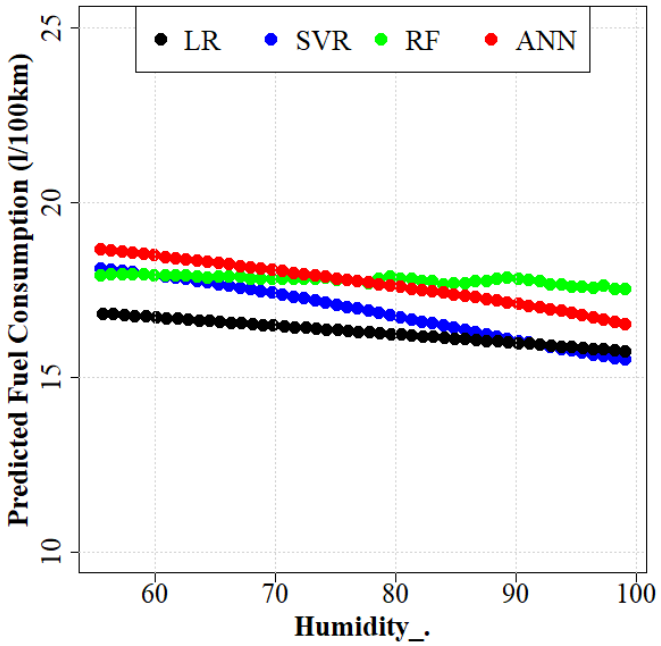


Figure 161. Parametric analysis of *Humidity_*. (%) for the light trucks App 2.

Figure 160 and 161 show variation of fuel consumption for changes of *Temperature_deg_C* and *Humidity_*. respectively. Both figures show that all models

agree on the fact that higher temperature and humidity produces lower fuel consumption. Similar trends are also shown in the figure, highlighting the correlation existing between the two variables. From an analysis of the two figures, it is possible to say that the models seem to attribute up to approximately 2.5 l/100 km to changes in *Temperature_deg_C* and *Humidity_*. For these two variables, ranges investigated are quite large and characteristic of weather conditions in the UK.

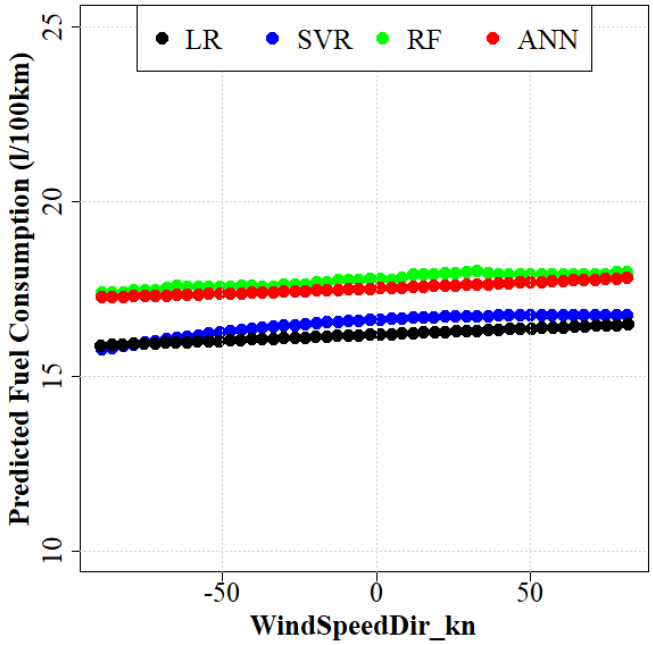


Figure 162. Parametric analysis of *WindSpeedDir_kn* (kn) for the light trucks App 2.

Figure 162 shows variation of fuel consumption due to changes in *WindSpeedDir_kn*. The figure shows that the models detect an increase of fuel consumption due to stronger wind blowing in opposite direction to the direction of travel of the considered trucks. This is physically reasonable and confirmed by all the other case study investigated in the study.

Table 59 and 60 report impacts on fuel consumption calculated between the 25^{th} - 75^{th} and 5^{th} - 95^{th} percentiles of the considered variables, respectively.

Table 59 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the light trucks App 2.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	3.87	24.14	4.83	30.08	5.27	32.84	4.81	29.95
Geom.Gradient	5.95	37.09	6.32	39.39	6.75	42.08	6.98	43.46
Torque.AVG	1.98	12.36	1.45	9.02	1.46	9.08	1.71	10.68
Revs.AVG	1.09	6.79	2.36	14.70	4.88	30.39	1.86	11.56
Speed.AVG	-0.19	-1.18	0.24	1.50	0.37	2.31	2.32	14.45
Acceleration	-0.32	-1.97	-0.49	-3.05	-0.95	-5.92	-0.46	-2.84
Traveled.Time	-0.09	-0.55	0.00	0.02	0.06	0.37	-0.19	-1.18
Geom.Abs_Radius	-0.06	-0.38	-0.34	-2.10	0.18	1.12	-0.55	-3.43
Geom.Crossfall	-0.10	-0.63	-0.20	-1.23	0.20	1.26	0.16	1.02
TRACS.LPV03m	0.01	0.09	-0.02	-0.10	-0.12	-0.76	0.01	0.06
TRACS.LPV10m	-0.07	-0.42	-0.12	-0.75	0.05	0.33	0.13	0.82
TRACS.LPV30m	0.14	0.90	0.38	2.35	0.25	1.58	0.12	0.76
TRACS.Pav_ Texture	0.17	1.06	-0.07	-0.41	0.29	1.84	0.36	2.22
Altitude.AVG	-0.11	-0.71	0.01	0.04	-0.04	-0.25	0.02	0.10
Atm_Pressure_hPa	-0.25	-1.55	-0.36	-2.24	-0.29	-1.79	-0.06	-0.39
Temperature_deg_C	-0.74	-4.59	-0.96	-5.98	-0.18	-1.14	-0.65	-4.03
Humidity_.	-0.55	-3.40	-1.44	-8.96	-0.11	-0.68	-1.03	-6.40
WindSpeedDir_kn	0.30	1.89	0.48	2.96	0.38	2.38	0.28	1.72

This provides a numerical estimation of how each of the developed models estimates the impact of each of the considered variables on fuel consumption, it helps in the interpretation of the parametric analysis, and provides with values to compare

with results from the studies conducted in the past, draw conclusions and make recommendations.

Table 60 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the light trucks App 2.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	6.97	43.45	7.26	45.24	7.21	44.90	7.80	48.57
Geom.Gradient	10.71	66.76	11.25	70.10	11.53	71.86	12.23	76.24
Torque.AVG	3.57	22.25	2.08	13.00	1.72	10.72	2.79	17.40
Revs.AVG	1.96	12.22	3.05	19.01	3.04	18.99	2.07	12.93
Speed.AVG	-0.34	-2.13	0.11	0.68	0.17	1.07	1.32	8.27
Acceleration	-0.57	-3.55	-0.74	-4.64	-1.12	-6.99	-0.82	-5.10
Traveled.Time	-0.16	-0.98	-0.18	-1.11	0.08	0.48	-0.34	-2.11
Geom.Abs_Radius	-0.11	-0.68	-0.57	-3.55	0.10	0.65	-0.85	-5.27
Geom.Crossfall	-0.18	-1.14	-0.04	-0.26	-0.25	-1.57	0.43	2.69
TRACS.LPV03m	0.03	0.16	-0.03	-0.18	-0.20	-1.23	0.24	1.52
TRACS.LPV10m	-0.12	-0.76	-0.22	-1.35	-0.05	-0.33	0.13	0.83
TRACS.LPV30m	0.26	1.61	0.68	4.21	0.23	1.46	0.18	1.12
TRACS.Pav_ Texture	0.31	1.91	-0.05	-0.31	0.23	1.45	0.40	2.49
Altitude.AVG	-0.21	-1.28	-0.06	-0.39	-0.14	-0.87	0.20	1.23
Atm_Pressure_hPa	-0.45	-2.79	-0.55	-3.41	-0.26	-1.60	-0.12	-0.72
Temperature_deg_C	-1.32	-8.25	-1.19	-7.42	-0.23	-1.41	-1.17	-7.28
Humidity_.	-0.98	-6.12	-2.41	-15.03	-0.40	-2.47	-1.89	-11.79
WindSpeedDir_kn	0.55	3.40	0.93	5.77	0.52	3.22	0.50	3.10

As mentioned, two calculations were made. One for the $25^{th} – 75^{th}$ percentiles (see Table 59) and another for the $5^{th} – 95^{th}$ percentiles (see Table 60). This is because although there might be more confidence in the estimates made within the first range, the latter might be of better interest for the considered study, as this might include the

extreme conditions of the road surface, and their impact on the fuel consumption of the considered fleet of trucks.

E.2.3.3 Medium trucks on A roads (App 1)

This section is dedicated to the results of the parametric analysis performed on the medium trucks App 1, part of the Dataset 2. Figure 163 shows variation of fuel consumption for changes of *Gross.Weight_kg*. From the figure it is possible to see that all models identify similar patterns, with an increment of fuel consumption of about 10 - 12 l/100 km associated to an increment of *Gross.Weight_kg* of about 30,000 kg (from 10,000 to 44,000 kg).

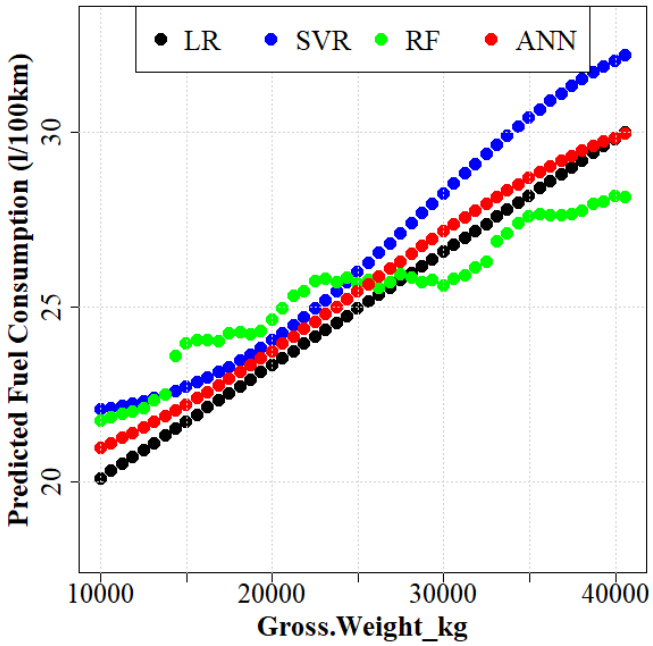


Figure 163. Parametric analysis of *Gross.Weight_kg* (kg) for the medium trucks App 1.

It is possible to see (Figure 163) that the RF model seems to have a kind of 'steps' behaviour. This is due to the nature of the model, which is based on the theory of decision trees, but it can be associated also to the uncertainties in the data, given, in

this case, by calculation of the *Gross.Weight_kg* from the proprietary software of Microlise, which does not allow the algorithm to determine smooth trend lines.

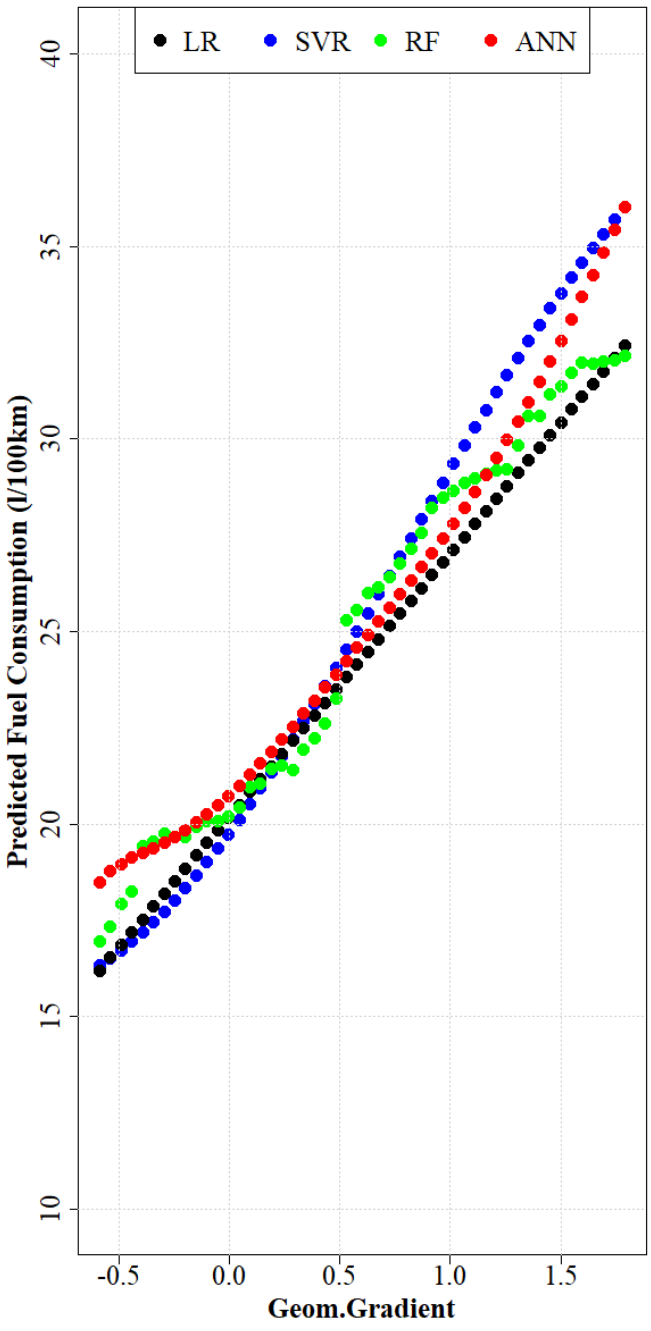


Figure 164. Parametric analysis of *Geom.Gradient* (%) for the medium trucks App 1.

Figure 164 show variation of fuel consumption due to changes of *Geom.Gradient*. Again, the figure shows that trends are quite defined by all the developed models, which tend to agree with each other. That produces up to 20 l/100 km difference in fuel consumption within a range of 2.5% between -0.5% and 2% road gradient. Although the considered range of *Geom.Gradient* cannot be considered exhaustive of the whole range of gradients allowed by the DMRB along primary roads (Highways Agency, 2006b), this represents a quite good range characteristic of large part of the network of reference. Notice that the considered range of *Geom.Gradient* is not symmetrically distributed from 0%. This is probably because for drivers it is easier to drive at a constant speed uphill than downhill.

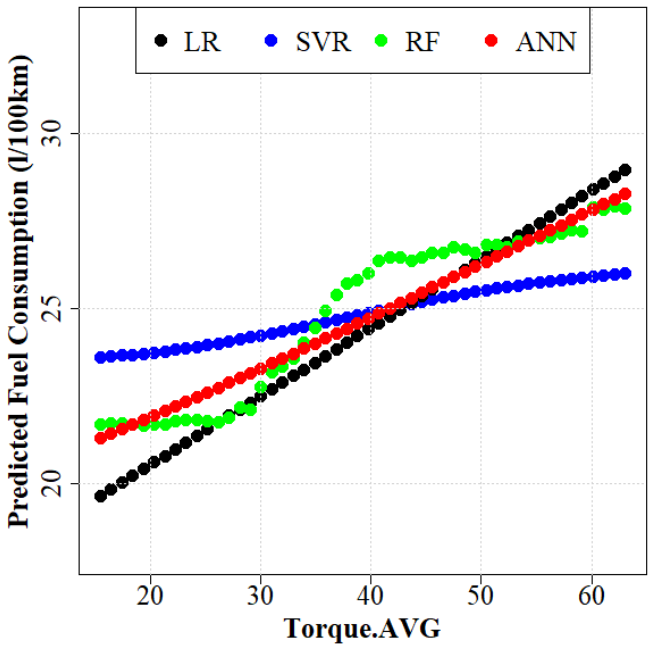


Figure 165. Parametric analysis of *Torque.AVG* (%) for the medium trucks App 1.

Figure 165 and 166 show variation of fuel consumption due to *Torque.AVG* and *Revs.AVG* respectively. In regard to the first (see Figure 165), it is possible to say

that all models seem to agree that higher *Torque.AVG* produces higher fuel consumption. The SVR seems to underfit the data in this case, as all the other models developed show similar trends instead. Maximum variation, in this case is shown by the LR model, which estimates that about 8 l/100 km can be caused by an increment of *Torque.AVG* from 15% to 60%. Regarding the *Revs.AVG*, the figure (see Figure 166) shows that while LR, SVR and ANN seem to agree on the fact that higher revolutions cause higher fuel consumption, the RF seem to overfit the data showing a poorly interpretable pattern. Overall, by analysing the figure, it is possible to say that 2 l/100 km might be affected by higher Revs.AVG in the range from 1,100 to 1,700 rpm.

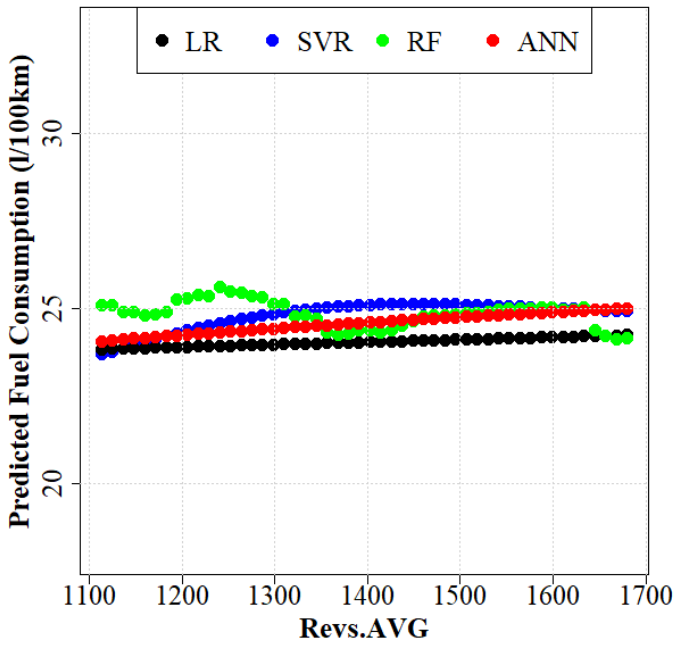


Figure 166. Parametric analysis of *Revs.AVG* (rpm) for the medium trucks App 1.

Figure 167 and 168 show variation of fuel consumption due to changes of *Speed.AVG* and *Acceleration* respectively. As it was seen in the previous section about the light

trucks App 2 (see Appendix E.2.3.2), for the quantity of data analysed the applied algorithms are able to detect the different trends of fuel consumption due to *Speed.AVG* and *Acceleration*. In fact, all the used algorithms seem to detect an increment of fuel consumption due to higher vehicle speed and lower fuel consumption for positive acceleration (due to the effect of road gradient).

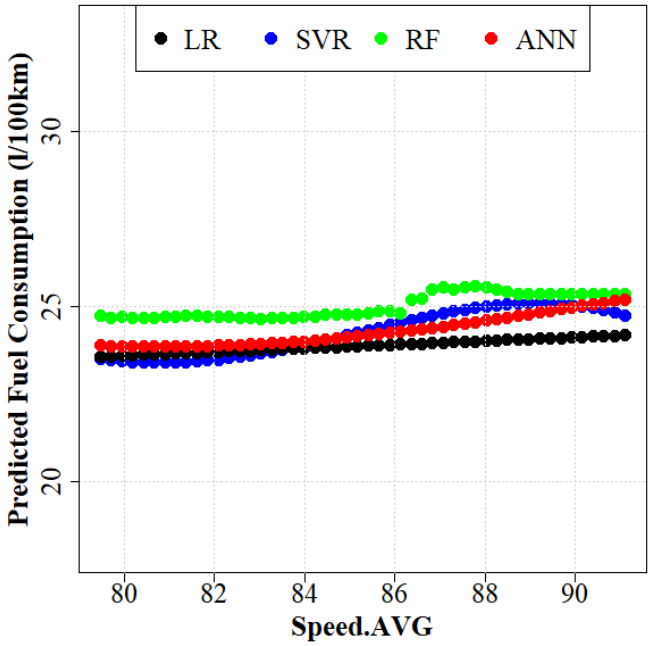


Figure 167. Parametric analysis of *Speed.AVG* (km/h) for the medium trucks App 1.

In regard to *Speed.AVG* (see Figure 167), it can be seen that fuel consumption can increase of about 1.5 l/100 km for a change of *Speed.AVG* between 80 and 95 km/h. For *Acceleration* (see Figure 168), instead, excluding the estimates made by the SVR which seems somewhat overestimated, decrement of fuel consumption can be estimated of approximately -2 l/100 km over a range of *Acceleration* from -0.02 to +0.03 m/s^2.

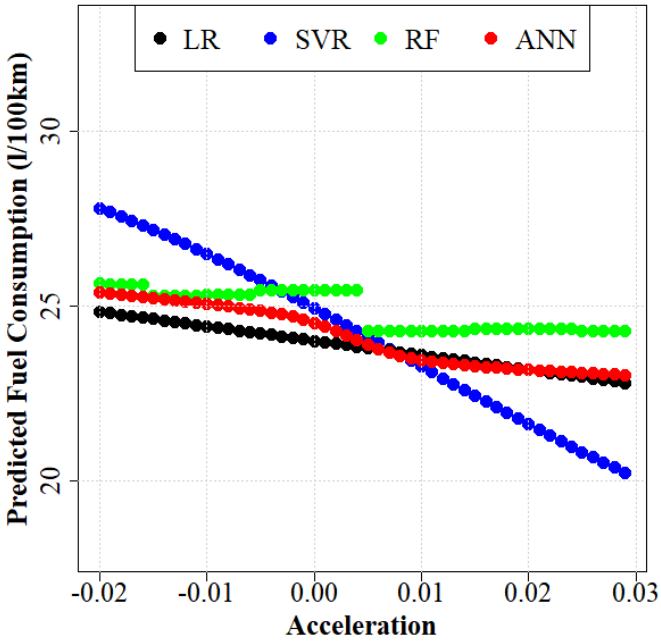


Figure 168. Parametric analysis of *Acceleration* (m/s^2) for the medium trucks App 1.

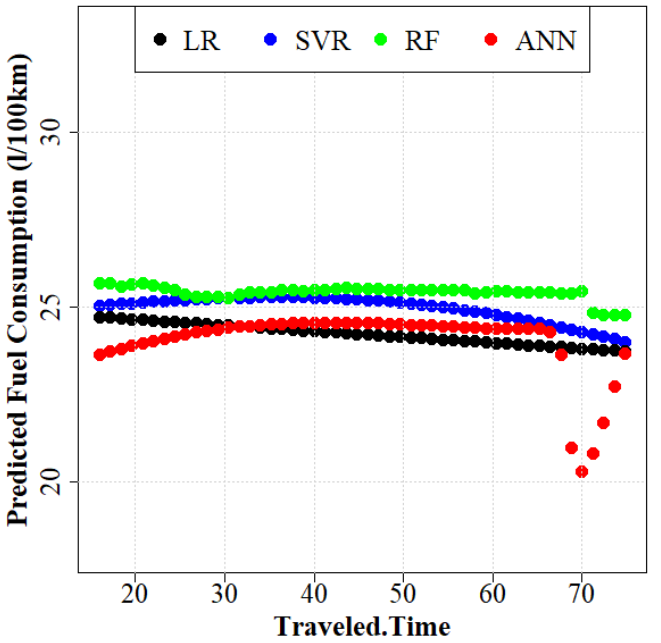


Figure 169. Parametric analysis of *Traveled.Time* (s) for the medium trucks App 1.

Figure 169 shows variation of fuel consumption due to changes of *Traveled.Time*. For this variable, all models detect different patterns, with some overfitting given by the ANN at the extremes. Therefore, although it looks like LR, SVR and RF detect a decrease of fuel consumption due to higher *Traveled.Time*, it is difficult to make definite conclusions, and in this sense further investigation is needed.

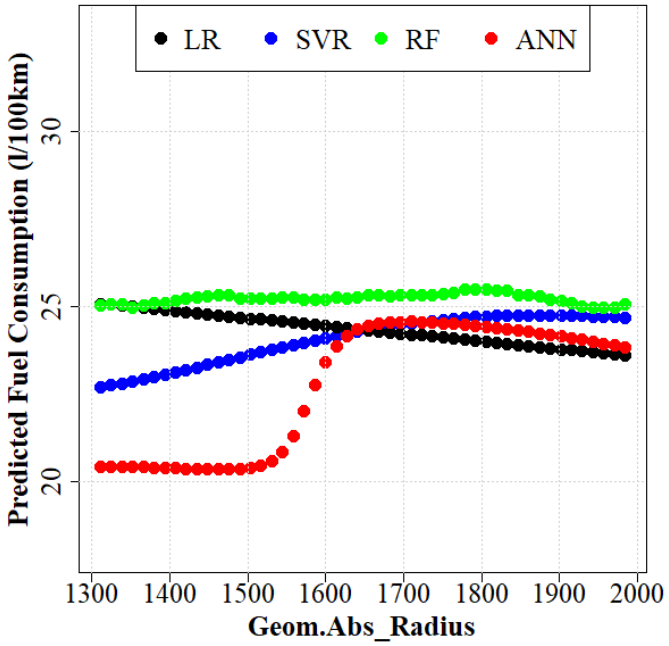


Figure 170. Parametric analysis of *Geom.Abs_Radius* (m) for the medium trucks App 1.

Figure 170 and 171 show variation of fuel consumption due to changes of *Geom.Abs_Radius* and *Geom.Crossfall* respectively. Although it is expected that these two variables should be correlated, it is possible to see that all models apart from the LR identify different trends for the two. In particular, it looks like the ANN is overfitting the data regarding the *Geom.Abs_Radius*, while the SVR is probably overfitting those regarding the *Geom.Crossfall*. The only model which show consistent responses are the LR and the RF, which show though opposite trends.

Therefore, in regard to *Geom.Abs_Radius* and *Geom.Crossfall* is difficult to make conclusions and further investigation is probably required.

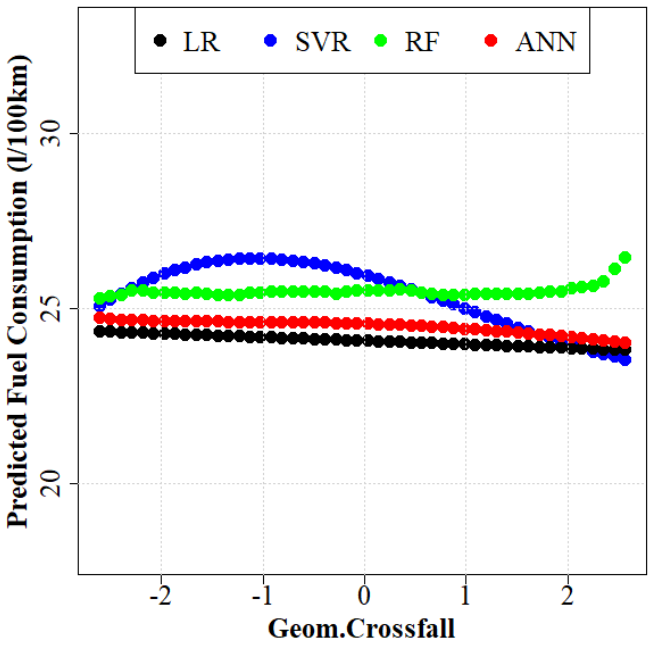


Figure 171. Parametric analysis of *Geom.Crossfall* (%) for the medium trucks App 1.

Figure 172 shows variation of fuel consumption due to changes of *TRACS.LPV03m*. The figure shows that while the LR and SVR seem to identify the variable to be insignificant, and RF and ANN seem to agree on the fact that higher *TRACS.LPV03m* produces an increment in fuel consumption of about 1 l/100 km over a range of 0.30 mm^2 for *TRACS.LPV03m* (between 0.15 and 0.45 mm^2).

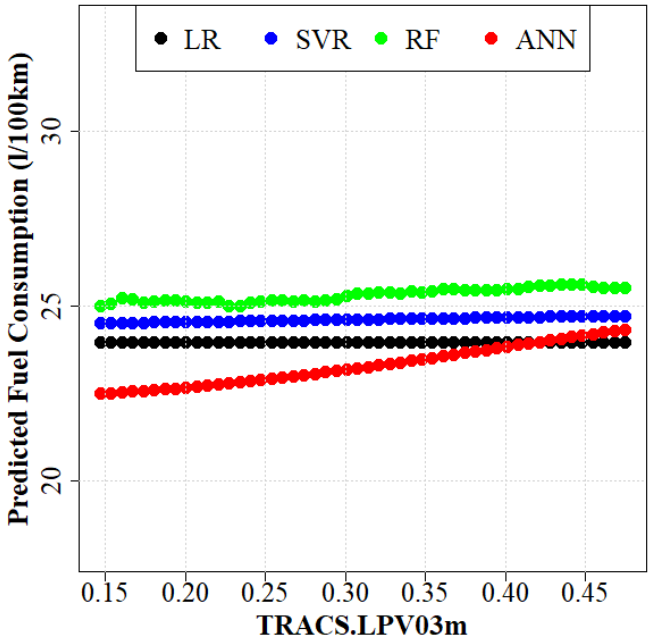


Figure 172. Parametric analysis of *TRACS.LPV03m* (mm^2) for the medium trucks App 1.

Figure 173 shows variation of fuel consumption due to changes of *TRACS.LPV10m*. The figure shows that while the LR and the RF seems to identify the variable to be substantially insignificant, the ANN produces some overfitting, and the SVR seems to address higher fuel consumption of about 1 l/100 km produced by an increment of *TRACS.LPV10m* of 2.5 mm^2 between 0 and 2.5 mm^2.

Figure 174 shows variation of fuel consumption due to changes of *TRACS.LPV30m*. The figure shows that while SVR and ANN seem to identify a decrease in fuel consumption due to higher *TRACS.LPV30m*, which might be affected by a reduction of vehicle speed, the LR and RF seem to agree that on increment of this variable produces and increment in fuel consumption over the considered range.

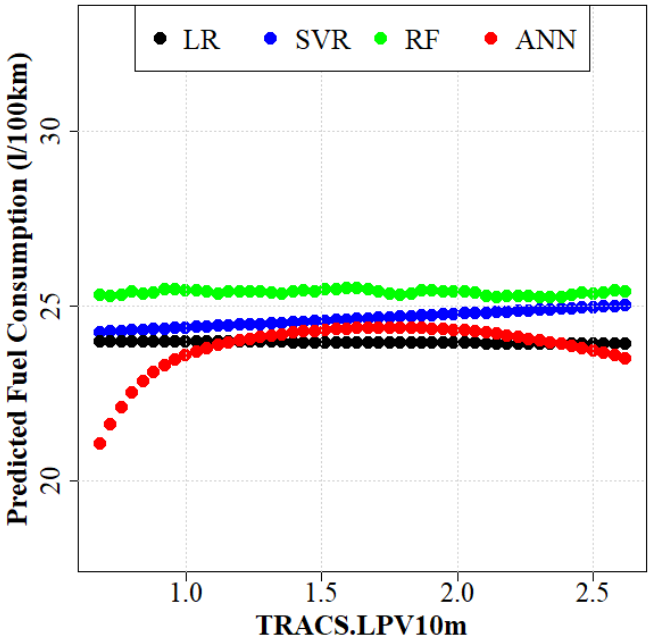


Figure 173. Parametric analysis of *TRACS.LPV10m* (mm^2) for the medium trucks App 1.

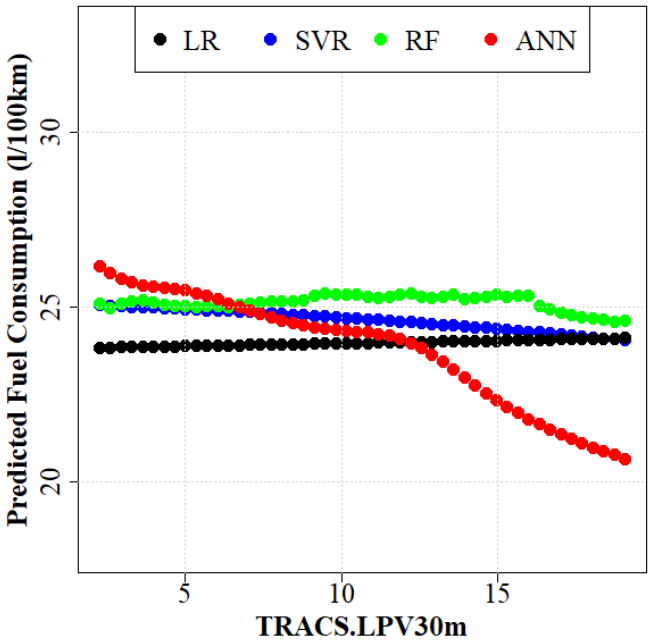


Figure 174. Parametric analysis of *TRACS.LPV30m* (mm^2) for the medium trucks App 1.

In particular, apart from the ANN, which seems to heavily overfit the data, the SVR shows that *TRACS.LPV30m* can reduce fuel consumption by 1 l/100 km, while LR and RF that this can increase fuel consumption by about the same quantity. That is within a considered range of 20 mm^2, which still identifies a road in fairly good conditions (Highways Agency, 2008a).

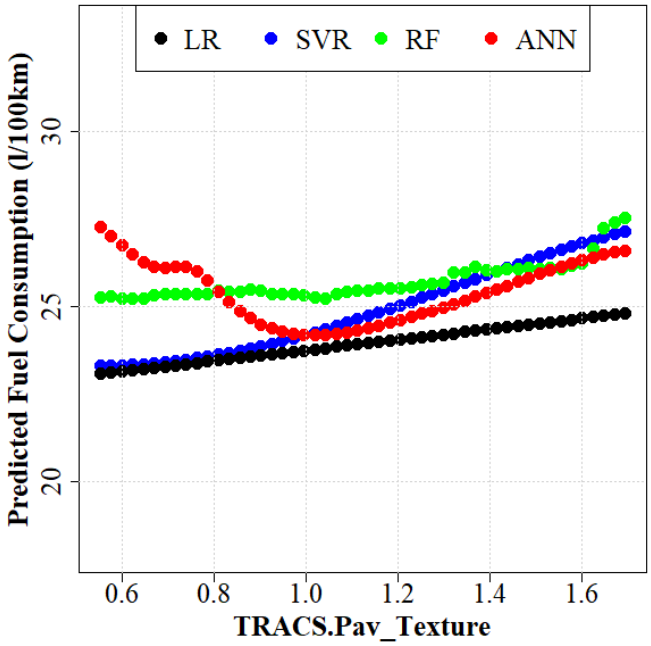


Figure 175. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the medium trucks App 1.

Figure 175 shows variation of fuel consumption for changes of *TRACS.Pav_Texture* (SMTD). Although some overfitting of the ANN for low SMTD, it is possible to see that all developed models identify an increment of fuel consumption due to higher values of *TRACS.Pav_Texture*. In particular, also in a reduced range of SMTD investigated, such as that of the present case study, variations identified by the developed models can be as high as 3 - 3.5 l/100 km, over a range between 0.6 and 1.8 mm.

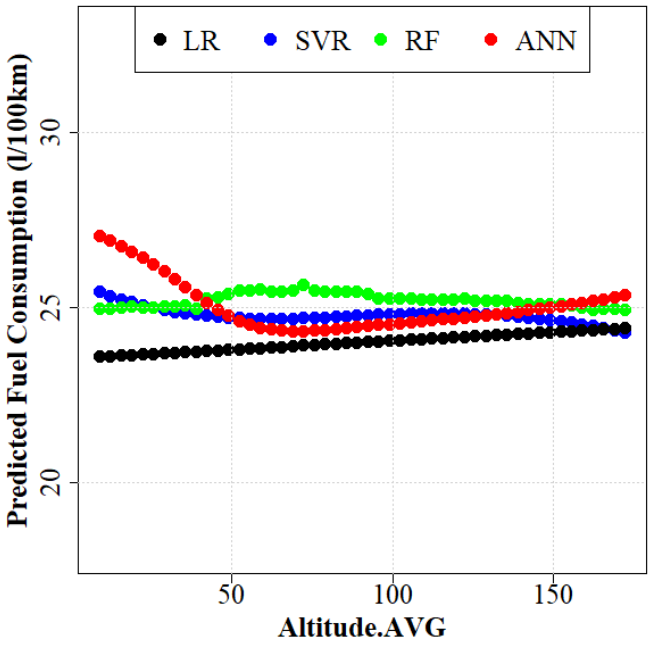


Figure 176. Parametric analysis of *Altitude.AVG* (mamls) for the medium trucks App 1.

Figure 176 shows variation of fuel consumption due to changes of Altitude.AVG. Although the detected pattern seem confusing, with different trends identified by different models, it is possible to see that the ANN detect a trend similar to that identified for *TRACS.Pav_Texture*, which would be meaningful of the correlation existing between the two variables, but less meaningful from a physical perspective.

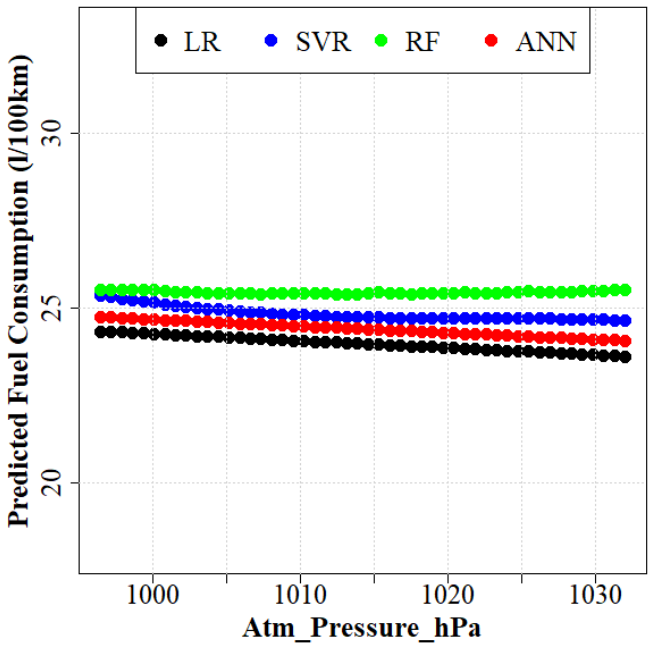


Figure 177. Parametric analysis of *Atm_Pressure_hPa* (hPa) for medium trucks App 1.

Figure 177 shows variation of fuel consumption for changes of *Atm_Pressure_hPa*. Although the RF model seems to show no significant impact from this variable, LR, SVR and ANN seem to agree on the fact that fuel consumption reduces for higher *Atm_Pressure_hPa*. Important is to notice that, while in previous case studies due to the low quantity of data fed to the models, these were not able to agree on the detected trends, this happens now. Variation of fuel consumption can be estimated at about -1 l/100 km over the considered range of *Atm_Pressure_hPa*.

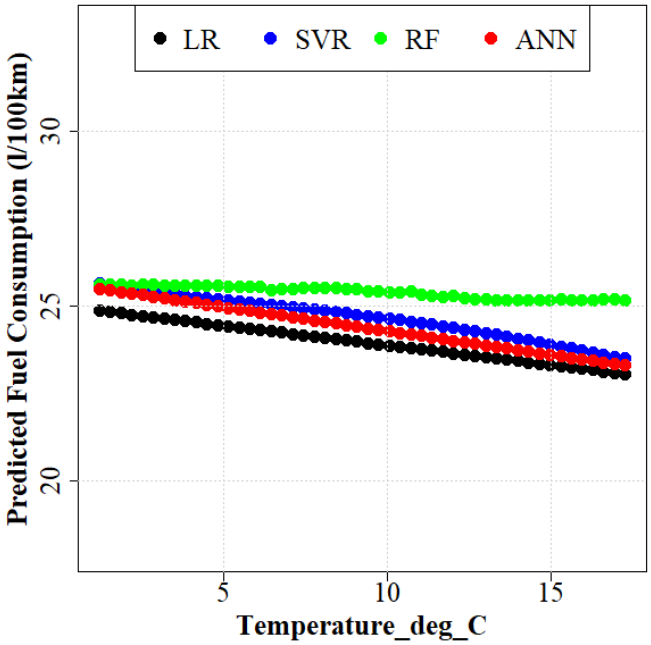


Figure 178. Parametric analysis of *Temperature_deg_C* (°C) for the medium trucks App 1.

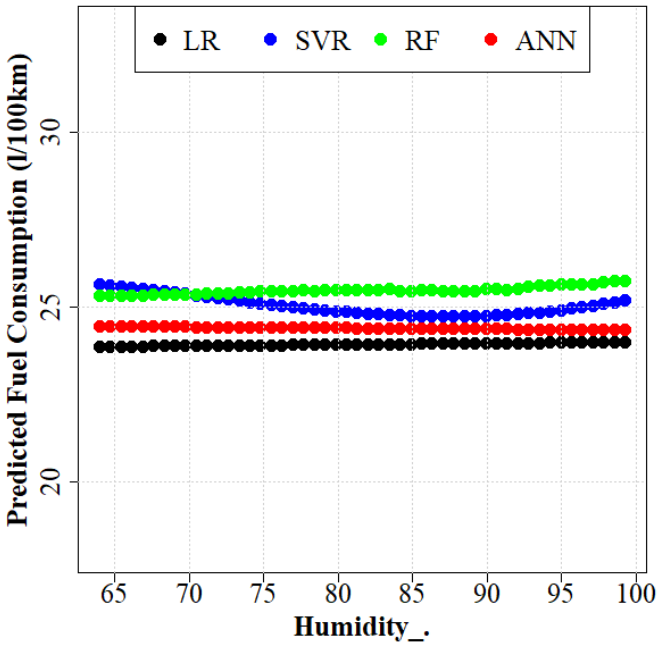


Figure 179. Parametric analysis of *Humidity_*. (%) for the medium trucks App 1.

Figure 178 and 179 show variation of fuel consumption due to changes of *Temperature_deg_C* and *Humidity_*. Again, it is possible to see that for *Temperature_deg_C* all models agree on the fact that higher temperature causes lower fuel consumption to the considered fleet of trucks. That can be estimated at about 2 l/100 km over a range of *Temperature_deg_C* between 0 and 20 °C. In regard to Humidity_., apart from the SVR, all the other models seem to agree on the fact that impact of this variables is very small, perhaps negligible, for this type of truck.

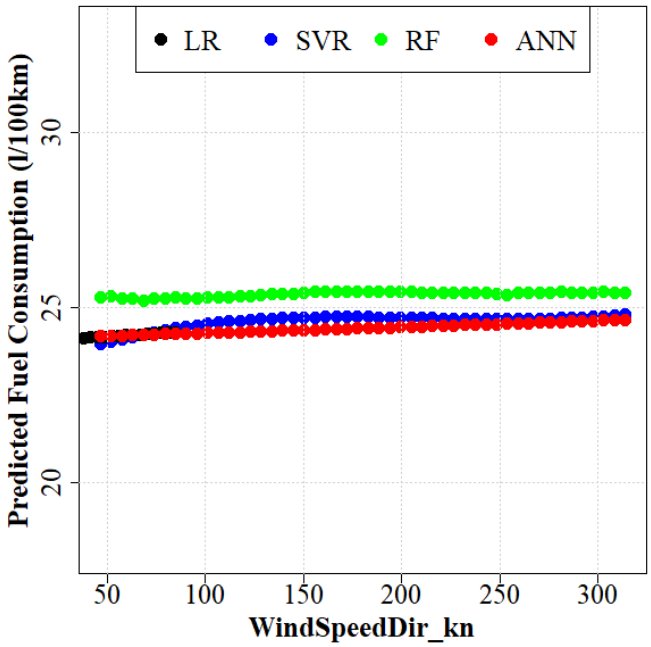


Figure 180. Parametric analysis of *WindSpeedDir_kn* (kn) for the medium trucks App 1.

Figure 180 shows variation of fuel consumption due to *WindSpeedDir_kn*. Again, also in this case study, all models seem to agree on the fact that higher fuel consumption (up to 1 l/100 km) is associated to stronger wind speed (measured at the weather station).

Table 61 and 62 report impacts on fuel consumption calculated between the 25^{th} - 75^{th} and 5^{th} - 95^{th} percentiles of the considered variables, respectively. This provides a numerical estimation of how each of the developed models estimates the impact of each of the considered variables on fuel consumption, it helps in the interpretation of the parametric analysis, and provides with values to compare with results from the studies conducted in the past, draw conclusions and make recommendations.

Table 61 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the medium trucks App 1.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	5.05	20.92	6.24	25.87	2.27	9.42	5.19	21.51
Geom.Gradient	8.28	34.29	11.40	47.21	9.02	37.37	8.58	35.53
Torque.AVG	4.75	19.68	1.51	6.27	5.06	20.96	3.60	14.90
Revs.AVG	0.20	0.83	0.52	2.14	-0.72	-2.99	0.48	1.99
Speed.AVG	0.31	1.29	1.51	6.26	0.85	3.53	0.71	2.96
Acceleration	-1.04	-4.29	-4.07	-16.88	-0.96	-3.98	-1.76	-7.31
Traveled.Time	-0.50	-2.08	-0.43	-1.77	0.11	0.45	0.03	0.13
Geom.Abs_Radius	-0.75	-3.13	1.31	5.42	0.17	0.69	4.08	16.89
Geom.Crossfall	-0.27	-1.11	-1.59	-6.57	0.01	0.05	-0.24	-0.99
TRACS.LPV03m	0.00	0.02	0.11	0.45	0.35	1.43	0.99	4.10
TRACS.LPV10m	-0.04	-0.18	0.38	1.59	-0.08	-0.31	0.36	1.49
TRACS.LPV30m	0.14	0.57	-0.50	-2.09	0.27	1.13	-2.67	-11.06
TRACS.Pav_ Texture	0.88	3.64	2.27	9.42	0.60	2.47	-0.05	-0.21
Altitude.AVG	0.42	1.73	0.08	0.32	-0.09	-0.36	-0.16	-0.65
Atm_Pressure_hPa	-0.36	-1.51	-0.24	-1.00	0.01	0.04	-0.35	-1.43
Temperature_deg_C	-0.92	-3.81	-0.98	-4.04	-0.40	-1.67	-1.12	-4.65
Humidity_.	0.07	0.27	-0.51	-2.12	0.11	0.44	-0.05	-0.23
WindSpeedDir_kn	0.31	1.28	0.11	0.45	0.13	0.54	0.24	0.97

Table 62 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the medium trucks App 1.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	9.09	37.65	9.60	39.76	6.10	25.27	8.48	35.13
Geom.Gradient	14.90	61.72	18.44	76.41	14.63	60.60	15.50	64.21
Torque.AVG	8.55	35.42	2.28	9.44	6.19	25.64	6.39	26.48
Revs.AVG	0.36	1.49	1.18	4.88	-0.71	-2.94	0.87	3.60
Speed.AVG	0.56	2.32	1.50	6.23	0.67	2.79	1.19	4.94
Acceleration	-1.86	-7.72	-7.00	-29.02	-1.34	-5.56	-2.28	-9.46
Traveled.Time	-0.90	-3.74	-0.85	-3.51	-0.86	-3.55	-2.91	-12.05
Geom.Abs_Radius	-1.36	-5.63	1.97	8.17	-0.09	-0.39	3.57	14.78
Geom.Crossfall	-0.48	-2.00	-1.47	-6.09	0.29	1.22	-0.58	-2.39
TRACS.LPV03m	0.01	0.03	0.20	0.81	0.51	2.10	1.69	6.99
TRACS.LPV10m	-0.08	-0.33	0.69	2.87	0.07	0.31	2.11	8.73
TRACS.LPV30m	0.25	1.03	-0.90	-3.75	-0.29	-1.21	-5.00	-20.71
TRACS.Pav_ Texture	1.58	6.56	3.59	14.87	1.36	5.65	-0.62	-2.57
Altitude.AVG	0.75	3.12	-0.87	-3.59	-0.04	-0.16	-1.72	-7.11
Atm_Pressure_hPa	-0.65	-2.71	-0.64	-2.66	-0.04	-0.17	-0.62	-2.58
Temperature_deg_C	-1.66	-6.86	-1.92	-7.97	-0.47	-1.93	-2.02	-8.37
Humidity_.	0.12	0.49	-0.59	-2.43	0.31	1.30	-0.10	-0.41
WindSpeedDir_kn	0.56	2.31	0.70	2.91	0.12	0.49	0.42	1.75

As mentioned, two calculations were made. One for the 25^{th} - 75^{th} percentiles (see Table 61) and another for the 5^{th} - 95^{th} percentiles (see Table 62). This is because although there might be more confidence in the estimates made within the first range, the latter might be of better interest for the considered study, as this might include the extreme conditions of the road surface, and their impact on the fuel consumption of the considered fleet of trucks.

E.2.3.4 Medium trucks on Motorways and A roads (App 2)

This section is dedicated to the results of the parametric analysis performed on the medium trucks App 2, from the Dataset 2.

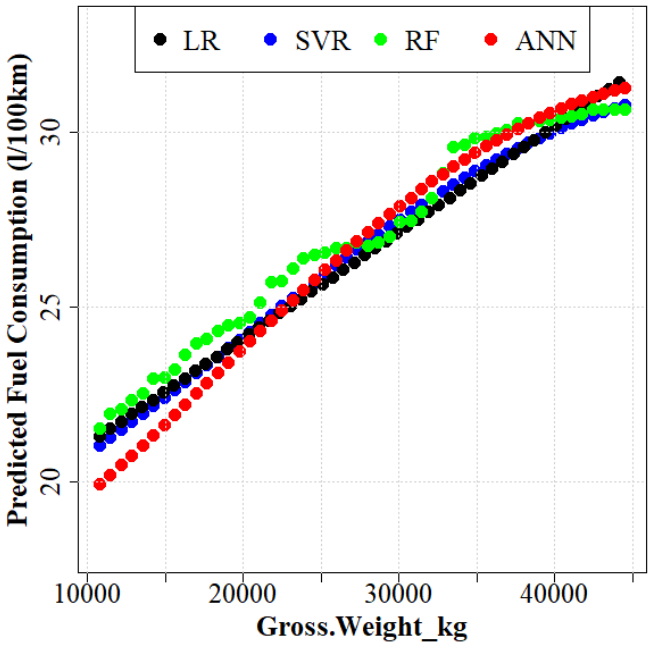


Figure 181. Parametric analysis of *Gross.Weight_kg* (kg) for the medium trucks App 2.

Figure 181 shows variation of fuel consumption due to changes of *Gross.Weight_kg*. The figure shows that, as seen for the other case studies considered, all models tend to identify similar patterns with fuel consumption that increases of about 11 l/100 km over a range of vehicle weight from 10,000 to 50,000 kg.

Figure 182 shows variation of fuel consumption due to changes of *Geom.Gradient*. From the figure, it is possible to see that all models seem to agree on the defined trends with fuel consumption that increases of about 20 to 30 l/100 km over a range of road gradient between -1 % and 2.5 %.

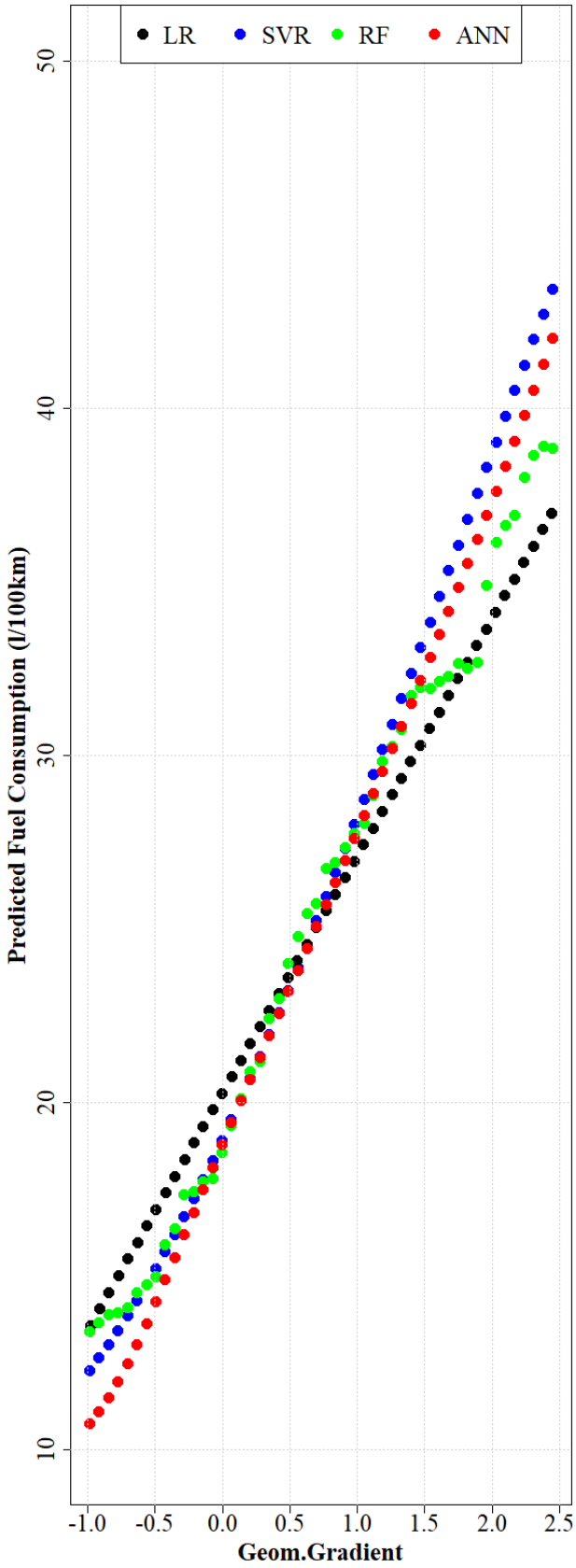


Figure 182. Parametric analysis of *Geom.Gradient* (%) for the medium trucks App 2.

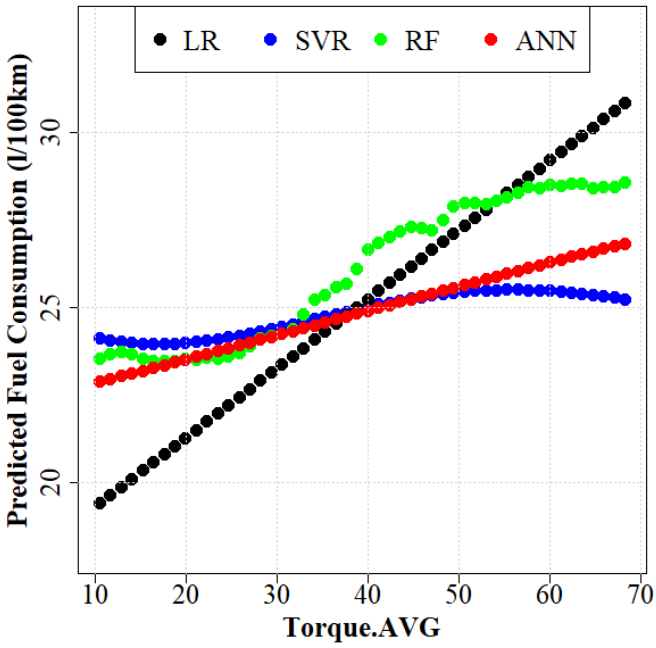


Figure 183. Parametric analysis of *Torque.AVG* (%) for the medium trucks App 2.

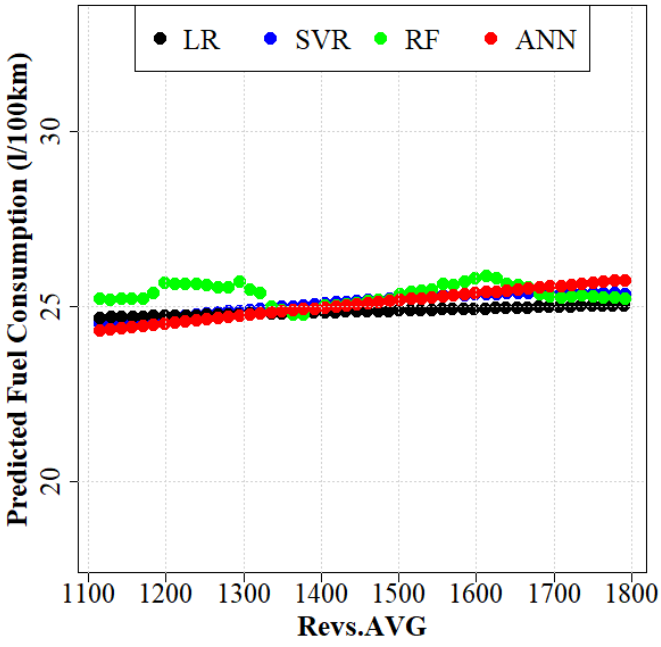


Figure 184. Parametric analysis of *Revs.AVG* (rpm) for the medium trucks App 2.

Figure 183 and 184 show variation of fuel consumption due to changes of *Torque.AVG* and *Revs.AVG*, respectively. In regard to the *Torque.AVG* (see Figure 183), it is possible to see that while the ML models seem to agree on the identified patterns, the LR detects a different behaviour probably due to the effect of non-linearity that cannot be accounted for using the latter. From the figure, for SVR, RF and ANN, variation of fuel consumption due to *Torque.AVG* can be estimated at approximately 4 l/100 km over a range of Torque.AVG from 10 % to 70 %. In regard to Revs.AVG, instead, it looks like that for this truck type all models agree on the fact that this variables is less important, with maximum 1 or 1.5 l/100 km affected between 1,100 and 1,800 rpm.

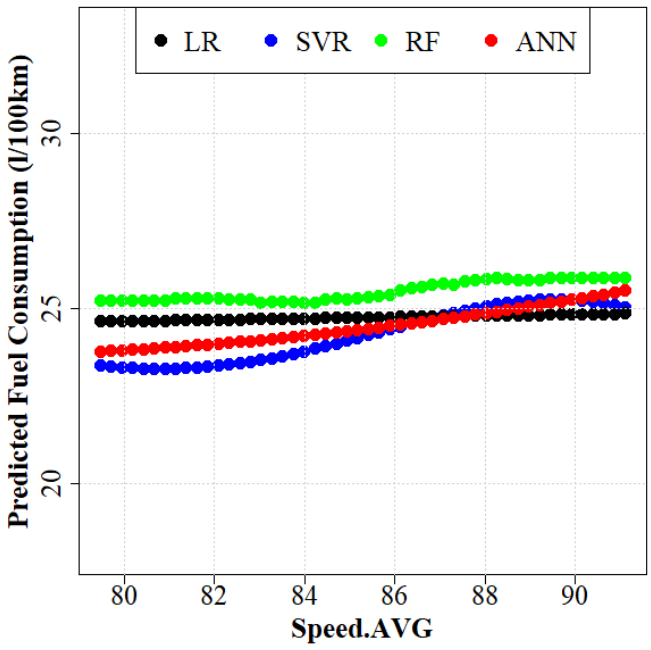


Figure 185. Parametric analysis of *Speed.AVG* (km/h) for the medium trucks App 2.

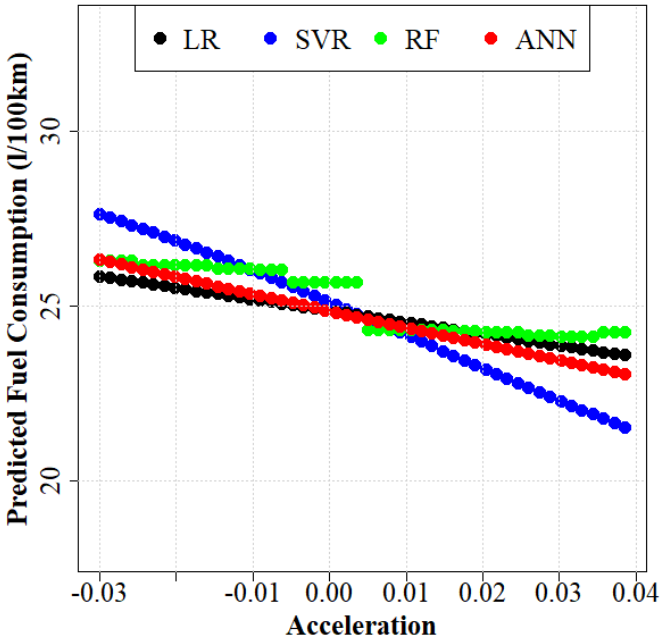


Figure 186. Parametric analysis of *Acceleration* (m/s^2) for the medium trucks App 2.

Figure 185 and 186 show variation of fuel consumption due to changes of *Speed.AVG* and *Acceleration*, respectively. In regard to the first (see Figure 185), it is possible to see that, as seen for other case studies earlier, ML models detect an increase of fuel consumption due to higher *Speed.AVG*. That can be estimated at approximately 2 l/100 km, over a range between 80 and 95 km/h. Only the LR model seems to be not able to detect the same trend for *Speed.AVG*, again, probably because of its inability to account for the non-linearity in the problem. For the Acceleration (see Figure 186), as seen for other case studies. All models detect a reduction of fuel consumption due to positive *Acceleration*. However, as explained earlier, this is probably due to the combined effect of *Geom.Gradient*. This effect can be estimated of about -2.5 l/100 km over a range of Acceleration between -0.03 and +0.04 m/s^2.

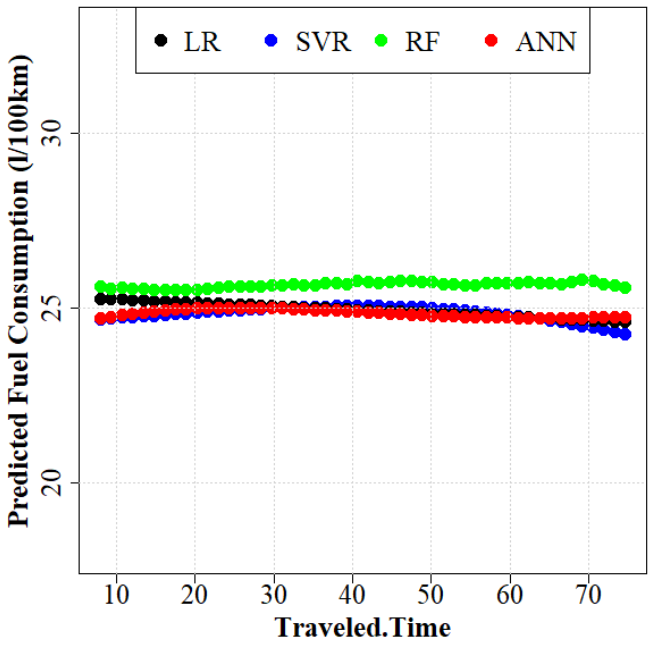


Figure 187. Parametric analysis of *Traveled.Time* (s) for the medium trucks App 2.

Figure 187 shows variation of fuel consumption due to changes of Traveled.Time. Again, also in this case study, the effect of this variable on the fuel consumption of the considered fleet of trucks seems to be minimal, with all models that show poorly visible trends.

Figure 188 and 189 show variation of fuel consumption for changes of *Geom.Abs_Radius* and *Geom.Crossfall*. In regard to the first (see Figure 188), it is possible to see that LR, RF and ANN show similar trends while it looks that the SVR seems to suffer of some overfitting on this variables, and probably tends to fit some outliers in the data, which makes the model to diverge from the trends identified by the others. Similarly, for *Geom.Crossfall* (see Figure 189), the SVR identify a pattern differently from what LR, RF and ANN do, which might confirm the overfitting of the SVR on these two variables.

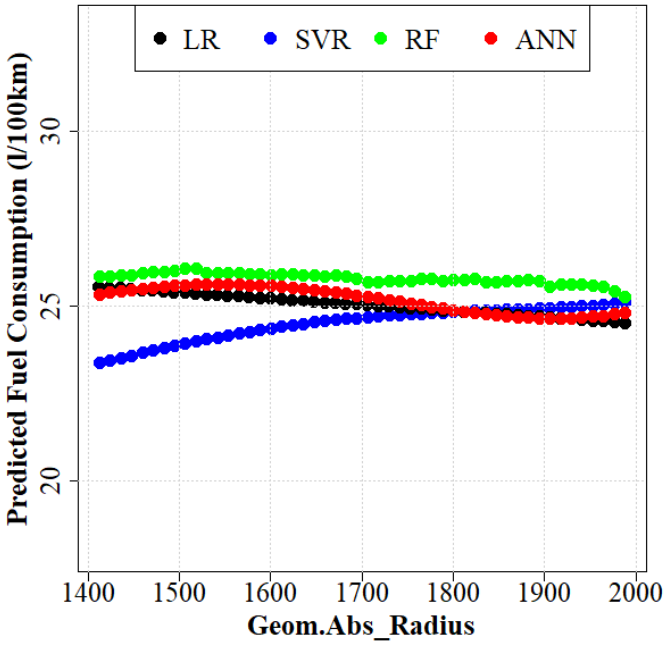


Figure 188. Parametric analysis of *Geom.Abs_Radius* (m) for the medium trucks App 2.

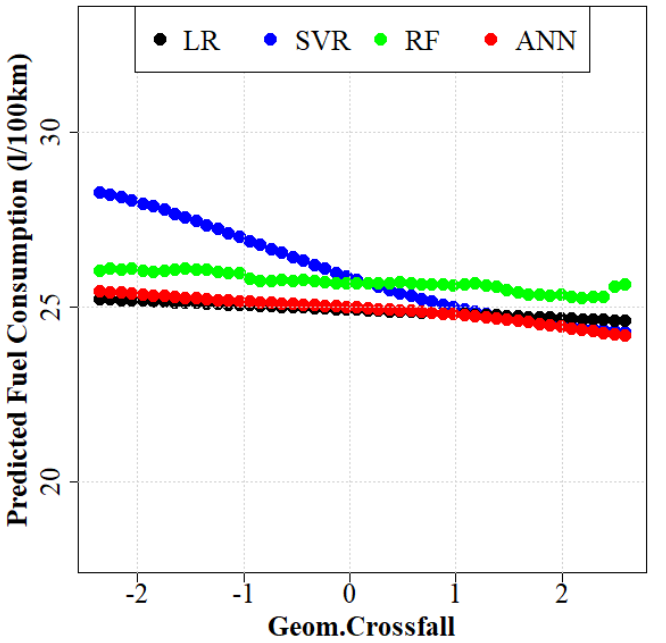


Figure 189. Parametric analysis of *Geom.Crossfall* (%) for the medium trucks App 2.

Figure 190 shows variation of fuel consumption for changes of *TRACS.LPV03m*. From the figure it is possible to see that, within the considered range of road characteristics (from 0.1 to 0.5 mm^2), the LR does not detect any particular change, SVR and RF detect a small increase (of approximately 1 l/100 km), and the ANN detects a small reduction of fuel consumption. Interestingly, results from SVR and RF agree with the results from previous case studies (see Appendix E.2.3.1, E.2.3.2 and E.2.3.3).

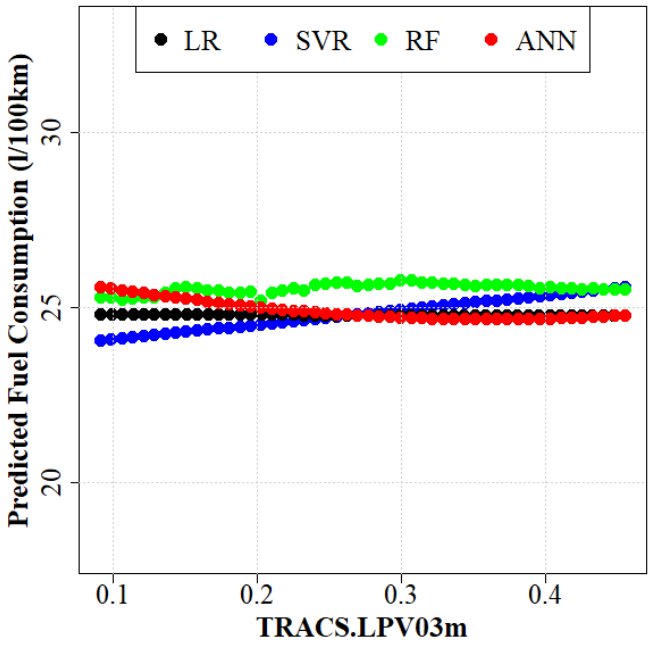


Figure 190. Parametric analysis of *TRACS.LPV03m* (mm^2) for the medium trucks App 2.

Figure 191 shows variation of fuel consumption due to changes of *TRACS.LPV10m*. From the figure it is possible to see that LR, RF and ANN detects a very small increase while the opposite behaviour is identified by the SVR. Assuming the trend identified by LR, RF and ANN correct, that would cause an increase of about 0.5 l/100 km over a range of road characteristics between 0.5 and 2.5 mm^2.

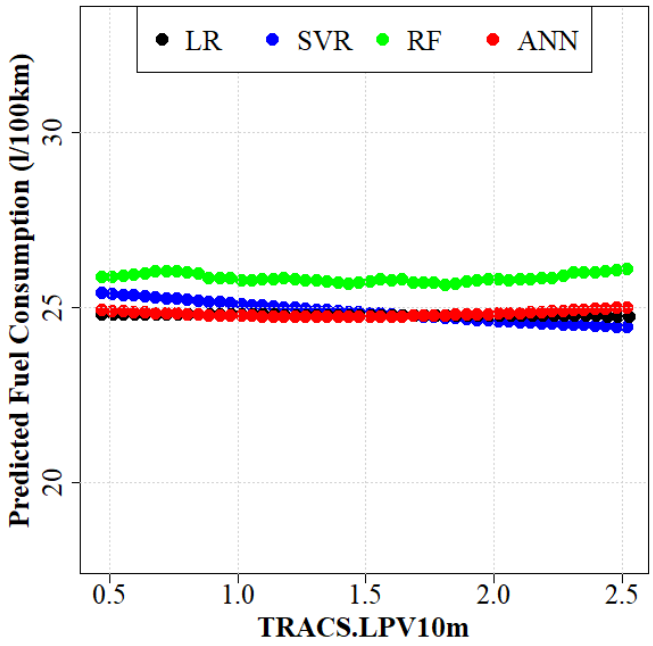


Figure 191. Parametric analysis of *TRACS.LPV10m* (mm^2) for the medium trucks App 2.

Figure 192 shows variation of fuel consumption due to changes of *TRACS.LPV30m*. From the figure, it is possible to see that, LR, SVR and RF seem to detect and overall small increase of fuel consumption, in the same order of magnitude of that estimated for *TRACS.LPV10m* (see Figure 191). However, the ANN detects a decrease, which might be associated to changes in vehicle speed (similar trends were identified earlier).

Figure 193 shows variation of fuel consumption due to changes of *TRACS.Pav_Texture* (SMTD). The figure shows that while LR, RF and ANN seem to agree on an increase of fuel consumption associated to higher SMTD (of approximately 1 l/100 km, over a range of SMTD between 0.6 and 1.6 mm^2), a different trend is initially identified by the SVR, which ends up to converge in the trend identified by the LR model.

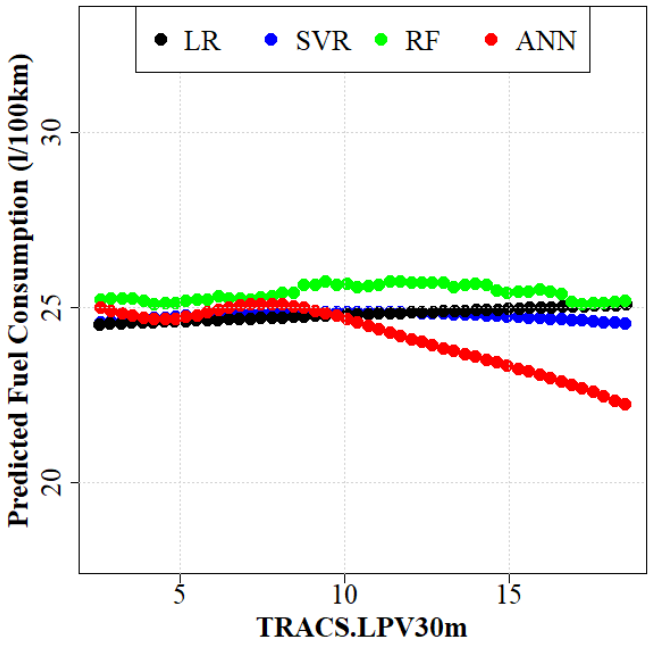


Figure 192. Parametric analysis of *TRACS.LPV30m* (mm^2) for the medium trucks App 2.

Figure 194 shows variation of fuel consumption due to changes of Altitude.AVG. The figure shows that LR, SVR and RF seem to agree on the fact that an increase of fuel consumption is associated to higher altitudes. That might be due to the lower density of air (and less oxygen for combustion in the engine), but, that reminds also of the correlation between this variable and the *TRACS.Pav_Texture*, which shows also similar trends. Therefore, also for this, further investigation is required. Opposite trend is shown by the ANN, which could be related to the lower density of air and less aerodynamic friction. However, that may be also overfitting.

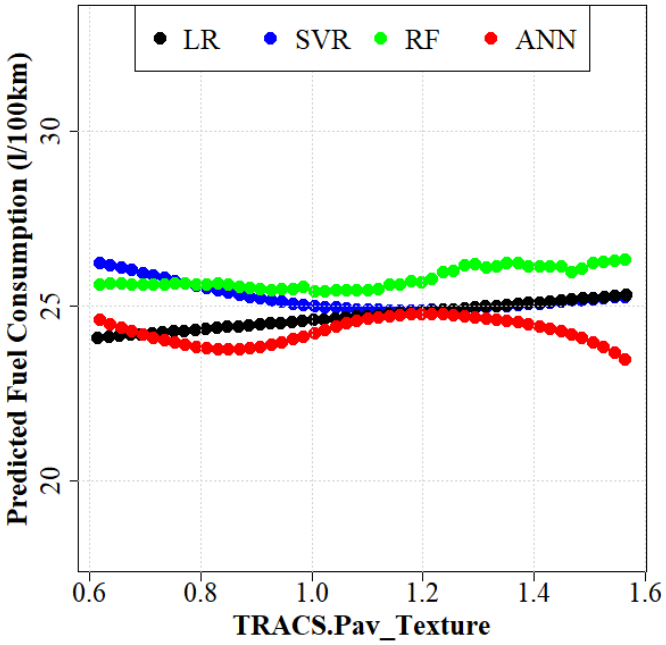


Figure 193. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the medium trucks App 2.

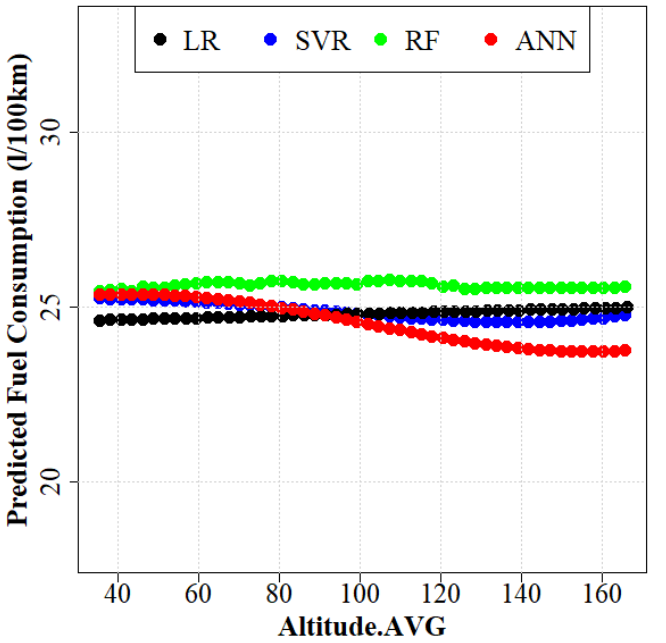


Figure 194. Parametric analysis of *Altitude.AVG* (mamls) for the medium trucks App 2.

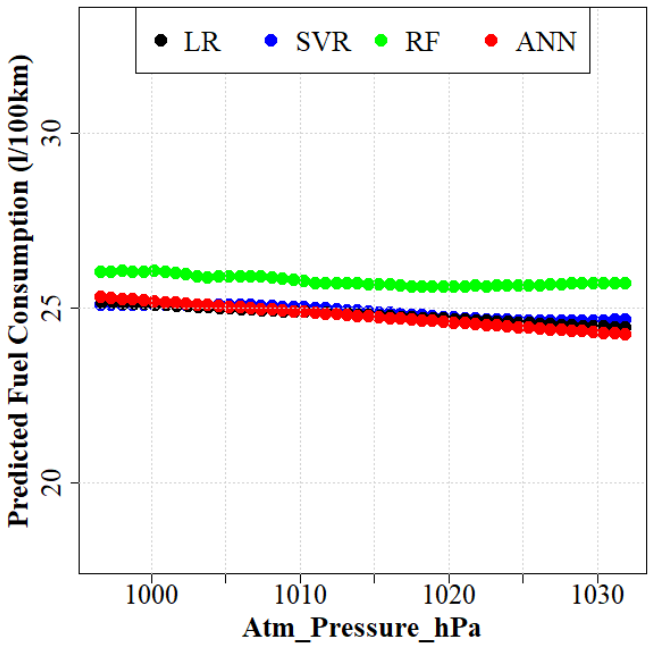


Figure 195. Parametric analysis of *Atm_Pressure_hPa* (hPa) for medium trucks App 2.

Figure 195 shows variation of fuel consumption due to changes of *Atm_Pressure_hPa*. The figure shows how LR, SVR and ANN detect very similar pattern, almost identical, which highlights reduction of fuel consumption for higher *Atm_Pressure_hPa*. A similar, but less visible, pattern is shown by the RF model, which might indicate that the engine of the considered truck type might work more efficiently when *Atm_Pressure_hPa* is high (more oxygen for combustion). Additionally, high *Atm_Pressure_hPa* is usually associated to good weather and thus, on average, higher air temperature.

Figure 196 shows variation of fuel consumption due to changes of *Temperature_deg_C*. For this, all models seem to highlight an increase of fuel consumption due to higher air temperature. In particular, the figure shows that, a reduction of approximately 2.5 l/100 km is expected for temperature from 0 ℃ to 20

°C. That, because of the similarities, would explain the trends highlighted by the developed models for *Atm_Pressure_hPa* (see Figure 195).

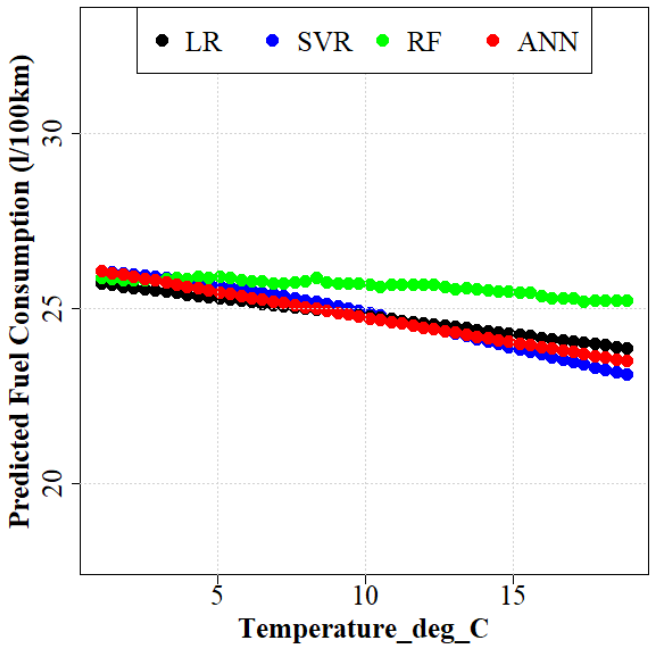


Figure 196. Parametric analysis of *Temperature_deg_C* (°C) for the medium trucks App 2.

Figure 197 shows variation of fuel consumption due to changes of *Humidity_*. The figure shows that LR, RF and ANN seem to agree on the fact that this variable does not produce any significant effect. However, a significant reduction of fuel consumption is detected by the SVR, in the range of conditions considered in the study. That could be correlated to *Temperature_deg_C* and its effects (see Figure 196).

Figure 198 shows variation of fuel consumption due to changes of *WindSpeedDir_kn*. Again, as seen in other case studies earlier (see Appendix E.2.3.1, E.2.3.2 and E.2.3.3), the figure shows that all the developed models detect an increase in fuel consumption due to stronger wind. Which is reasonable.

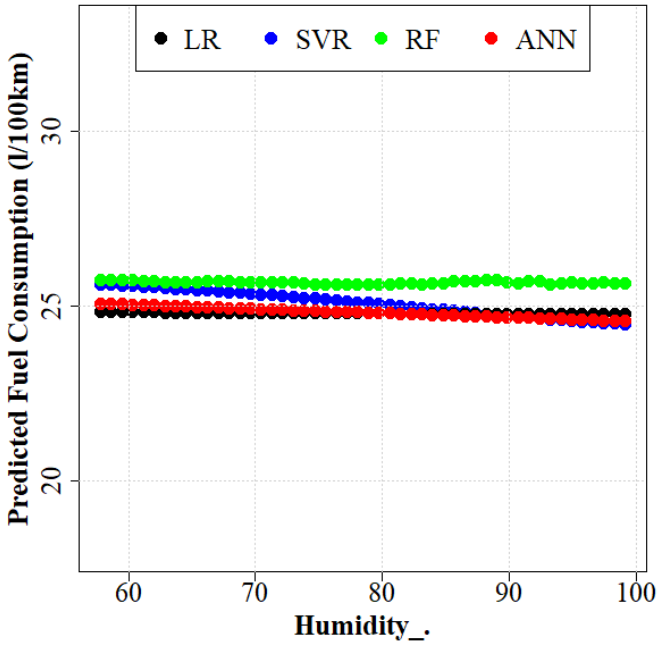


Figure 197. Parametric analysis of *Humidity_*. (%) for the medium trucks App 2.

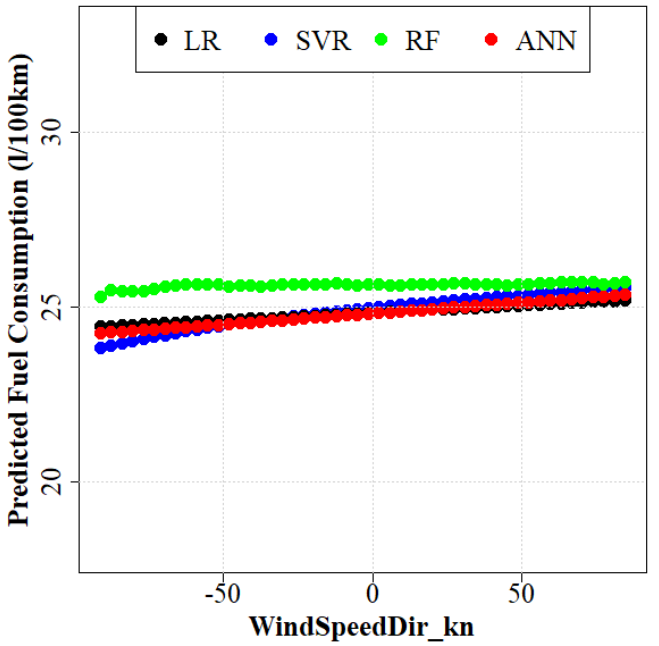


Figure 198. Parametric analysis of *WindSpeedDir_kn* (kn) for the medium trucks App 2.

Table 63 and 64 report impacts on fuel consumption calculated between the 25^{th} - 75^{th} and 5^{th} - 95^{th} percentiles of the considered variables, respectively.

Table 63 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the medium trucks App 2.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	5.16	20.82	5.47	22.06	5.55	22.37	6.48	26.13
Geom.Gradient	11.94	48.16	16.60	66.94	14.49	58.46	15.97	64.44
Torque.AVG	5.83	23.52	1.39	5.61	4.41	17.80	2.05	8.27
Revs.AVG	0.18	0.73	0.53	2.15	0.31	1.26	0.73	2.94
Speed.AVG	0.12	0.46	1.71	6.88	0.57	2.29	0.87	3.52
Acceleration	-1.16	-4.66	-3.23	-13.02	-1.80	-7.25	-1.67	-6.75
Traveled.Time	-0.34	-1.37	-0.03	-0.11	0.13	0.52	-0.27	-1.08
Geom.Abs_Radius	-0.52	-2.10	0.77	3.09	-0.26	-1.03	-0.85	-3.43
Geom.Crossfall	-0.32	-1.29	-2.42	-9.74	-0.39	-1.58	-0.50	-2.02
TRACS.LPV03m	-0.02	-0.09	0.78	3.14	0.16	0.64	-0.48	-1.93
TRACS.LPV10m	-0.03	-0.11	-0.51	-2.05	-0.05	-0.19	0.04	0.16
TRACS.LPV30m	0.28	1.15	-0.02	-0.09	0.33	1.34	-1.42	-5.75
TRACS.Pav_ Texture	0.61	2.48	-0.49	-1.96	0.47	1.91	0.88	3.55
Altitude.AVG	0.19	0.75	-0.54	-2.17	-0.18	-0.73	-1.30	-5.24
Atm_Pressure_hPa	-0.38	-1.52	-0.42	-1.68	-0.28	-1.13	-0.54	-2.18
Temperature_deg_C	-0.93	-3.77	-1.60	-6.47	-0.39	-1.57	-1.32	-5.31
Humidity_.	-0.04	-0.15	-0.64	-2.59	0.04	0.18	-0.25	-1.03
WindSpeedDir_kn	0.38	1.54	0.81	3.27	0.00	0.00	0.56	2.25

This provides a numerical estimation of how each of the developed models estimates the impact of each of the considered variables on fuel consumption, it helps in the interpretation of the parametric analysis, and provides with values to compare

with results from the studies conducted in the past, draw conclusions and make recommendations.

Table 64 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the medium trucks App 2.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	9.29	37.48	9.21	37.14	8.71	35.11	10.81	43.60
Geom.Gradient	21.49	86.69	28.59	115.34	24.33	98.14	28.69	115.72
Torque.AVG	10.49	42.33	1.29	5.22	4.74	19.13	3.64	14.68
Revs.AVG	0.33	1.32	0.85	3.42	0.09	0.36	1.32	5.32
Speed.AVG	0.21	0.84	1.86	7.49	0.64	2.57	1.55	6.26
Acceleration	-2.08	-8.39	-5.62	-22.68	-2.15	-8.68	-3.01	-12.14
Traveled.Time	-0.61	-2.47	-0.27	-1.09	0.21	0.86	-0.02	-0.07
Geom.Abs_Radius	-0.94	-3.78	1.57	6.35	-0.25	-1.01	-0.69	-2.80
Geom.Crossfall	-0.58	-2.33	-3.83	-15.45	-0.81	-3.25	-1.13	-4.54
TRACS.LPV03m	-0.04	-0.17	1.40	5.65	0.26	1.05	-0.81	-3.27
TRACS.LPV10m	-0.05	-0.19	-0.91	-3.67	0.13	0.52	0.03	0.14
TRACS.LPV30m	0.51	2.07	0.01	0.04	-0.14	-0.58	-2.33	-9.40
TRACS.Pav_ Texture	1.11	4.46	-0.97	-3.91	0.59	2.38	-0.52	-2.09
Altitude.AVG	0.33	1.35	-0.58	-2.33	0.06	0.25	-1.64	-6.60
Atm_Pressure_hPa	-0.68	-2.74	-0.46	-1.85	-0.33	-1.32	-0.97	-3.93
Temperature_deg_C	-1.68	-6.79	-2.71	-10.92	-0.61	-2.47	-2.37	-9.54
Humidity_.	-0.07	-0.26	-1.07	-4.32	-0.08	-0.32	-0.46	-1.85
WindSpeedDir_kn	0.69	2.77	1.59	6.41	0.24	0.97	1.01	4.06

As mentioned, two calculations were made. One for the $25^{th} – 75^{th}$ percentiles (see Table 63) and another for the $5^{th} – 95^{th}$ percentiles (see Table 64). This is because although there might be more confidence in the estimates made within the first range, the latter might be of better interest for the considered study, as this might include the

extreme conditions of the road surface, and their impact on the fuel consumption of the considered fleet of trucks.

E.2.3.5 Heavy trucks on A roads (App 1)

A parametric analysis for the heavy trucks App 1, from the Dataset 2 was performed. Figure 199 shows variation of fuel consumption for changes of *Gross.Weight_kg*. From the figure, it is possible to see that all models agree on the detected pattern, with fuel consumption that increases with *Gross.Weight_kg* of about 7 l/100 km between 15,000 and 50,000 kg.

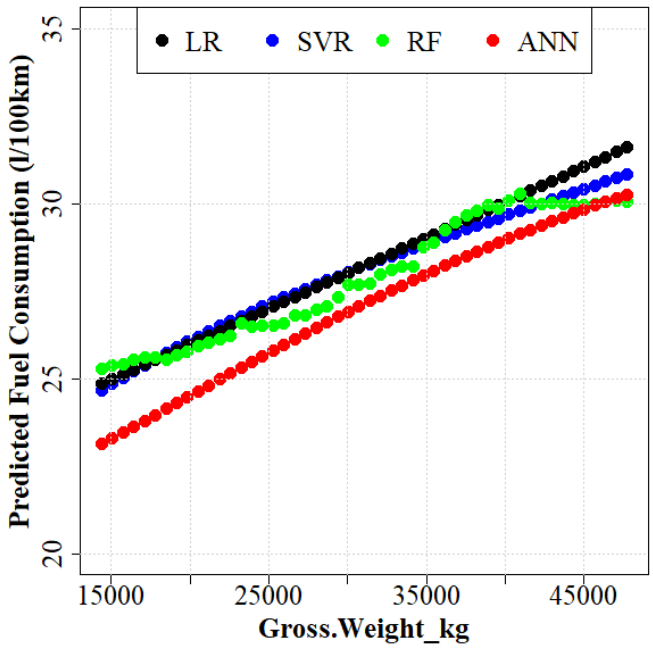


Figure 199. Parametric analysis of *Gross.Weight_kg* (kg) for the heavy trucks App 1.

Figure 200 shows variation of fuel consumption for changes of *Geom.Gradient*. The figure shows that, also in this case, all models tend to agree on the detected trends.

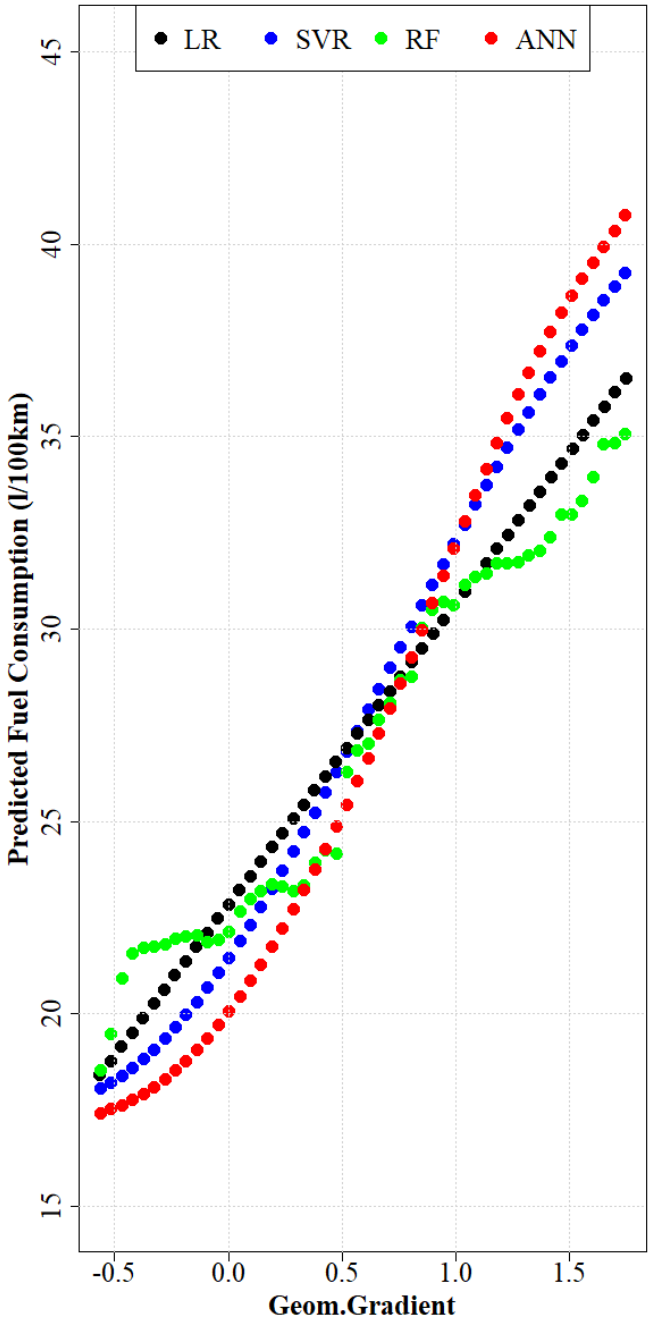


Figure 200. Parametric analysis of *Geom.Gradient* (%) for the heavy trucks App 1.

With this variable, fuel consumption is estimated to vary of approximately 24 l/100 km between –0.5 % and +1.8 %. This represents the highest variation of fuel consumption seen so far, probably due to the higher empty vehicle weight and bigger engines installed on this truck type. Additionally, it is possible to see that, thanks to

the large quantity of records available, SVR, RF and ANN detect a change in fuel consumption which seems asymptotic for flat or downhill roads, which, in consideration of the fact that any vehicle has to consume some fuel also going downhill due to idling, is reasonable.

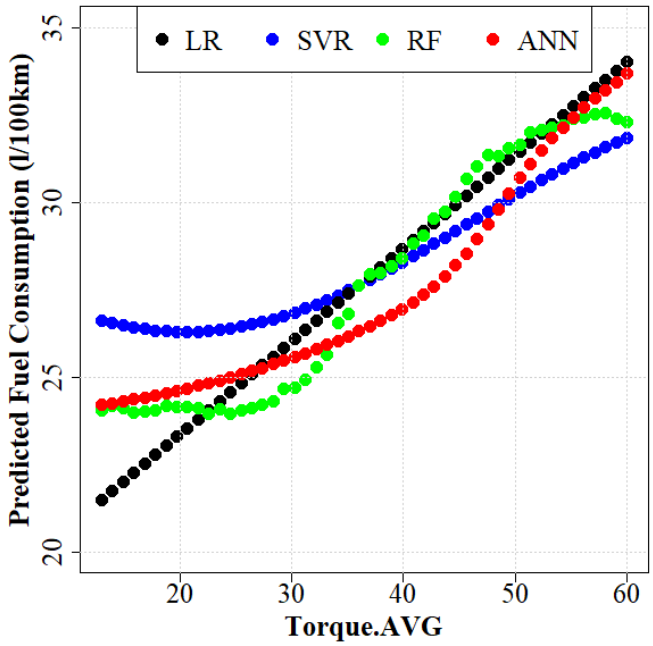


Figure 201. Parametric analysis of *Torque.AVG* (%) for the heavy trucks App 1.

Figure 201 and 202 show variation of fuel consumption due to changes of *Torque.AVG* and *Revs.AVG*, respectively. In regard to the first, while SVR, RF and ANN detect similar non-linear patterns, this is approximated by a straight line by the LR. That might produce bias in the final estimates made by the LR model. Variation of fuel consumption with *Torque.AVG* (see Figure 201) can be estimated at about 8.5 l/100 km, which is comparable to the impact detected for the *Gross.Weight_kg* (see Figure 199). For the *Revs.AVG* (see Figure 202), all models seem to identify an increase of fuel consumption due to higher rpm of the engine. Variation, for this

variable, can be estimated as up to approximately 2 - 2.5 l/100 km between 1,100 and 1,300 rpm.

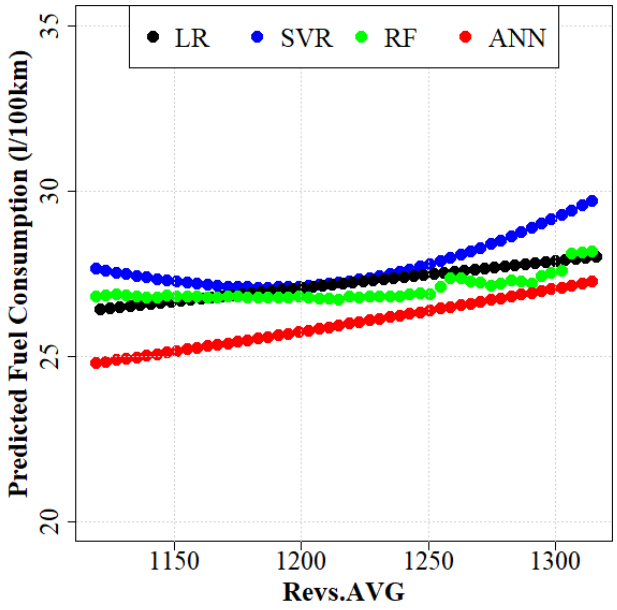


Figure 202. Parametric analysis of *Revs.AVG* (rpm) for the heavy trucks App 1.

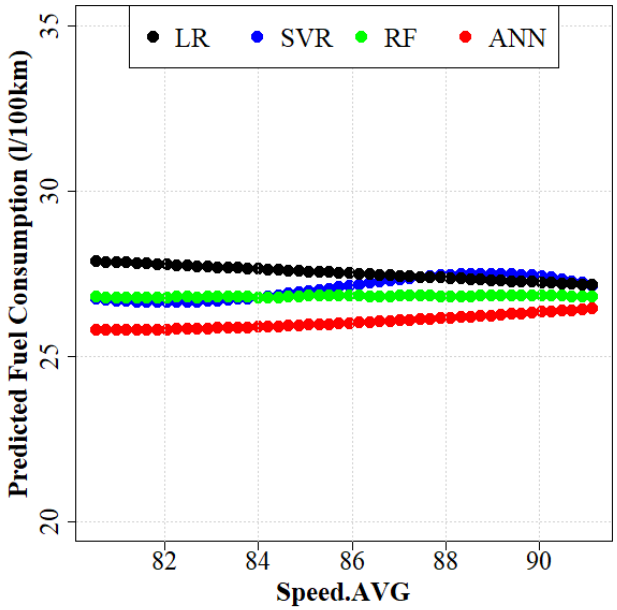


Figure 203. Parametric analysis of *Speed.AVG* (km/h) for the heavy trucks App 1.

Figure 203 shows variation of fuel consumption due to *Speed.AVG*. From the figure, it is possible to see that, apart from the LR, all the others seem to detect an increase in fuel consumption (of about 1 l/100 km), over the considered range of vehicle speeds (between 80 and 95 km/h).

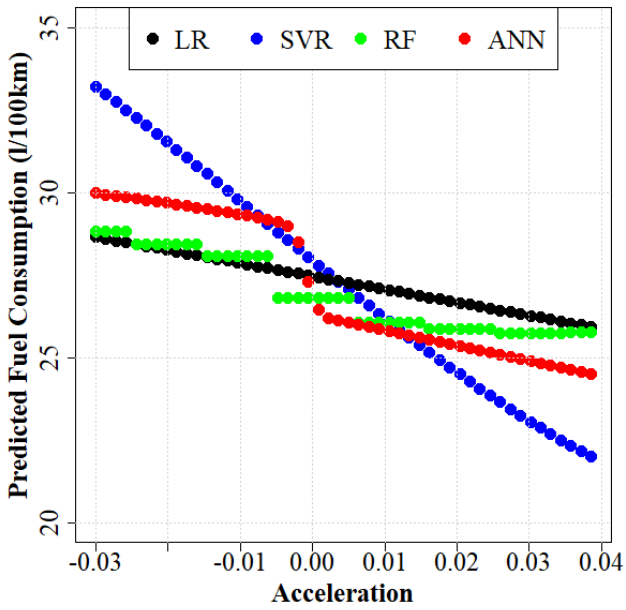


Figure 204. Parametric analysis of *Acceleration* (m/s^2) for the heavy trucks App 1.

Figure 204 shows variation of fuel consumption due to changes of *Acceleration*. Again, also in this case, the identified pattern seems to be affected by the effect of *Geom.Gradient*, with higher positive acceleration associated to lower fuel consumption. Similar patterns are identified also for other vehicle types (see Appendix E.2.3.1, E.2.3.2, E.2.3.3 and E.2.3.4). For the Acceleration, the estimated reduction of fuel consumption can be estimated to be of approximately 5 l/100 km between -0.03 and +0.04 m/s^2.

Figure 205 shows variation of fuel consumption due to changes of *Traveled.Time*. It is possible to see that, while the LR identifies a reduction of fuel

consumption associated with higher *Traveled.Time*, opposite trend is identified by the ML algorithms. For SVR, RF and ANN the estimated increase of fuel consumption is of approximately 1.5 - 2 l/100 km over a range of times between 10 and 70 s.

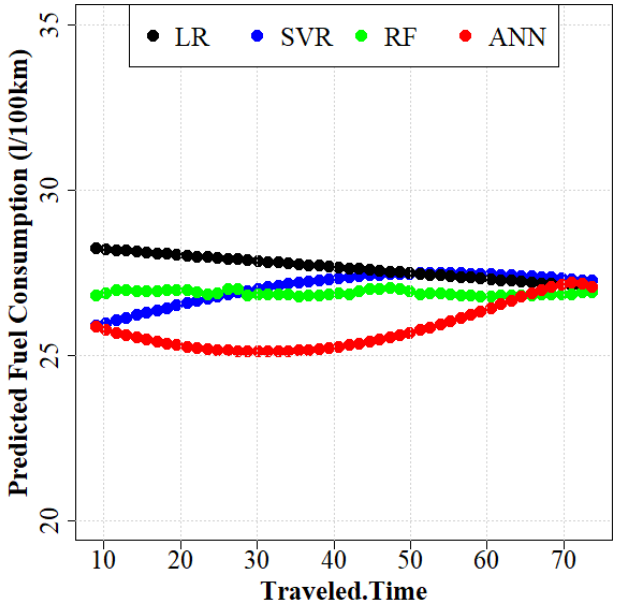


Figure 205. Parametric analysis of *Traveled.Time* (s) for the heavy trucks App 1.

Figure 206 and 207 show variation of fuel consumption due to changes of *Geom.Abs_Radius* and *Geom.Crossfall*, respectively. Although it was seen how these two variables should be somewhat correlated, the developed SVR and ANN models seem to identify different trends for the two. LR and RF seem instead to detect same trends for the two variables but in opposite direction one to the other. It is therefore impossible to give an adequate explanation of these results. Interestingly, for the *Geom.Abs_Radius*, while the LR model estimates a reduction of fuel consumption associated to straight roads, SVR, RF and ANN detect an increment, probably due to higher *Speed.AVG*.

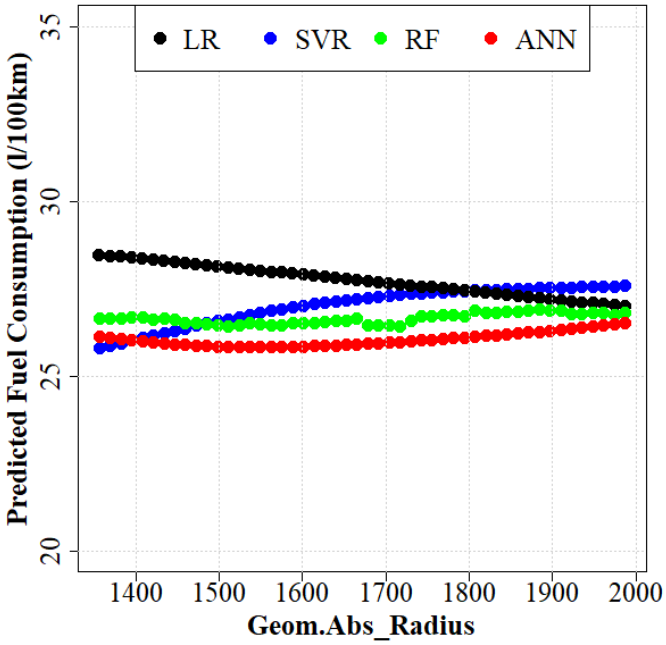


Figure 206. Parametric analysis of *Geom.Abs_Radius* (m) for the heavy trucks App 1.

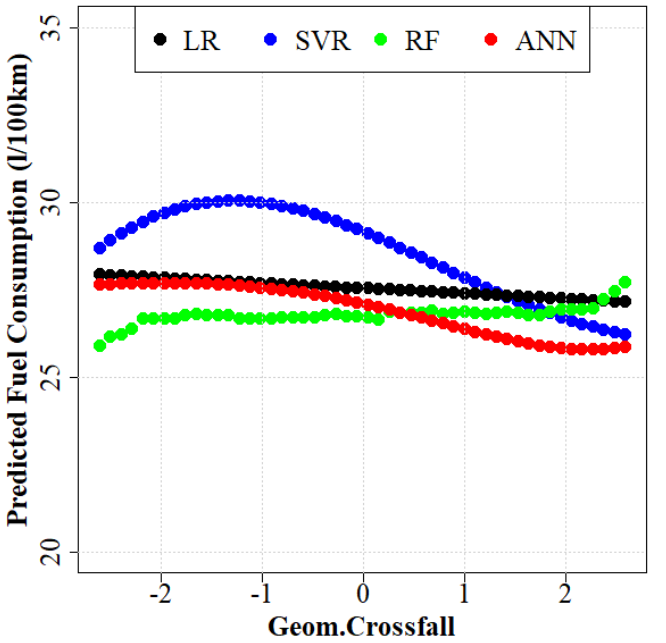


Figure 207. Parametric analysis of *Geom.Crossfall* (%) for the heavy trucks App 1.

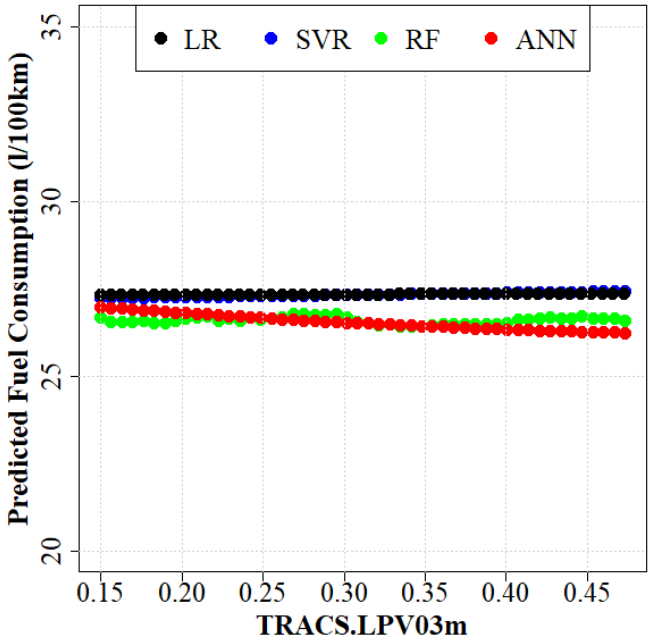


Figure 208. Parametric analysis of *TRACS.LPV03m* (mm^2) for the heavy trucks App 1.

Figure 208 shows variation of fuel consumption due to changes of *TRACS.LPV03m*. Although trends identified by the developed models seem quite close one to the other, it is possible to see that while LR and SVR detect a small increment of fuel consumption, RF and ANN detect a reduction. For this reason, it is not possible to make any claim on this result, and further investigation is required before drawing conclusions.

Figure 209 shows variation of fuel consumption due to changes of *TRACS.LPV10m*. Also in this case, although trends identified by the developed models seem quite close one to the other, it is possible to see that while LR and RF detect a substantially null variation of fuel consumption, the SVR detects and increment, and the ANN a reduction. Therefore, it is difficult to make any claim on this result, and further investigation is required before any conclusion can be made.

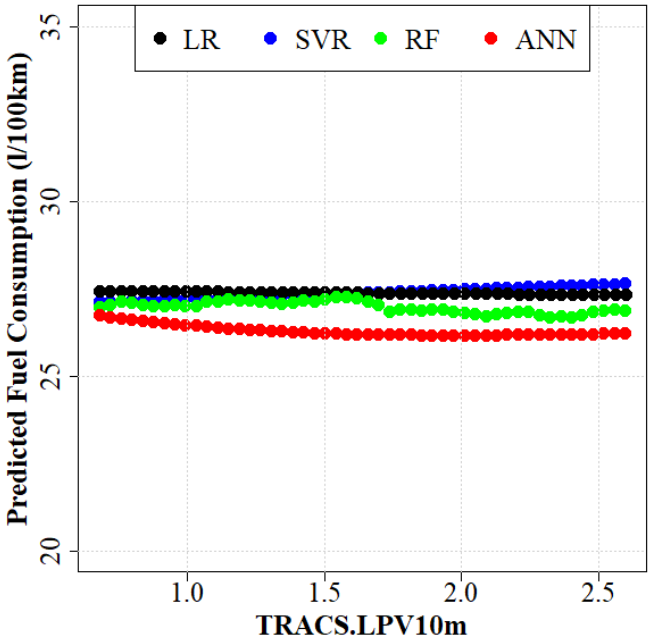


Figure 209. Parametric analysis of *TRACS.LPV10m* (mm^2) for the heavy trucks App 1.

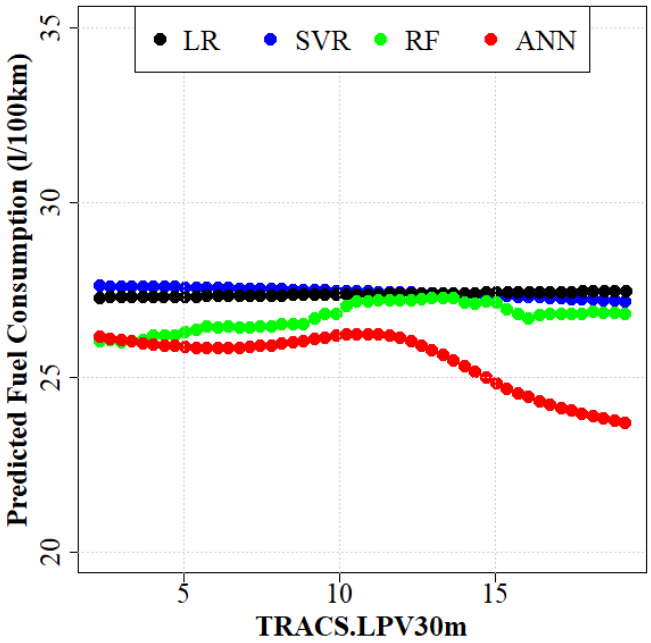


Figure 210. Parametric analysis of *TRACS.LPV30m* (mm^2) for the heavy trucks App 1.

Figure 210 shows variation of fuel consumption due to changes of *TRACS.LPV30m*. From the figure it is possible to see that while LR and RF seem to agree on a increment of fuel consumption due to *TRACS.LPV30m*. Although the ANN initially follows the same trend identified by the RF it then drops, showing an overall reduction of fuel consumption due to the effect of *TRACS.LPV30m*. Regarding the SVR, the identified trend seems detecting a reduction of fuel consumption with *TRACS.LPV30m*. However, this should be verified numerically.

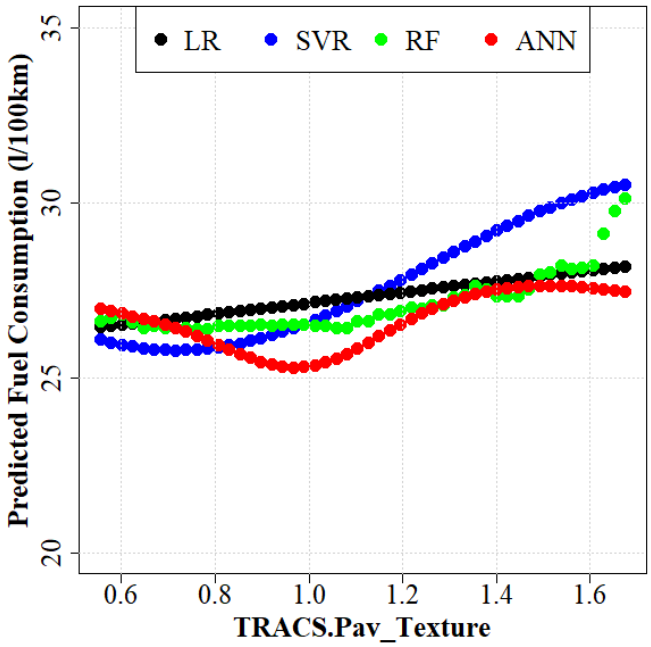


Figure 211. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the heavy trucks App 1.

Figure 211 shows variation of fuel consumption due to *TRACS.Pav_Texture* (SMTD). The figure shows that all models detect an increase of fuel consumption due to this variable. While this is linearly approximated by the LR, it looks like SVR, RF and ANN detect a non-linear behaviour. The ANN seems to show some overfitting. Nevertheless, impacts detected by the developed models can be estimated between 2

and 4 l/100 km difference between 0.6 and 1.6 mm of SMTD, which represent a significant amount of fuel.

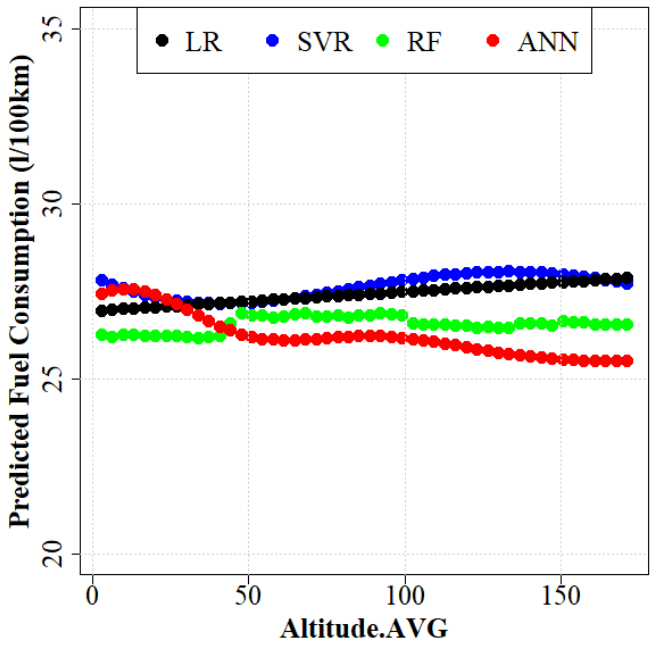


Figure 212. Parametric analysis of *Altitude.AVG* (mamls) for the heavy trucks App1.

Figure 212 shows variation of fuel consumption due to changes of *Altitude.AVG*. From the figure it is possible to see that, apart from some overfitting given by the ANN, all the other models seem to agree on the fact that higher fuel consumption is produced at higher altitudes. This might be due to the lower air density, but, as seen earlier for medium trucks App 2 (see Appendix E.2.3.4), that could be due also to the correlation of this variable with SMTD, which could introduce some bias in the final estimates.

Figure 213 shows variation of fuel consumption due to changes of *Atm_Pressure_hPa*. From the figure, it is possible to see that, apart from the RF which shows no significant impact of this variable on fuel consumption, all the other

developed models seem to agree that a reduction of fuel consumption is produced by higher *Atm_Pressure_hPa*. That, as higher pressure is associated to higher air density, would agree with the results seen for *Altitude.AVG* (see Figure 212), but, since high atmospheric pressure is also commonly associated to good weather, that could be also related to variations of air temperature (see Figure 214). Excluding results of the RF model, reduction of fuel consumption due to *Atm_Pressure_hPa* is estimated to be approximately 1.5 l/100 km between 995 and 1035 hPa.

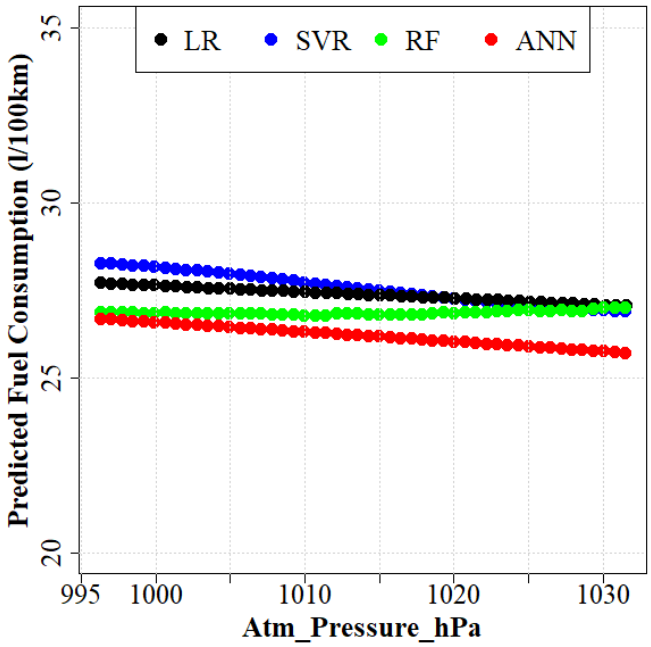


Figure 213. Parametric analysis of *Atm_Pressure_hPa* (hPa) for the heavy trucks App 1.

Figure 214 shows variation of fuel consumption due to changes of *Temperature_deg_C*. From the figure, it is possible to see that all the developed models agree on the fact that higher temperature reduces fuel consumption. That could be associated to higher efficiency of the engines of the considered fleet of trucks. From

a numerical perspective, impact of this variable on fuel consumption can be estimated to be approximately -2.5 l/100 km between 0 °C and 20 °C.

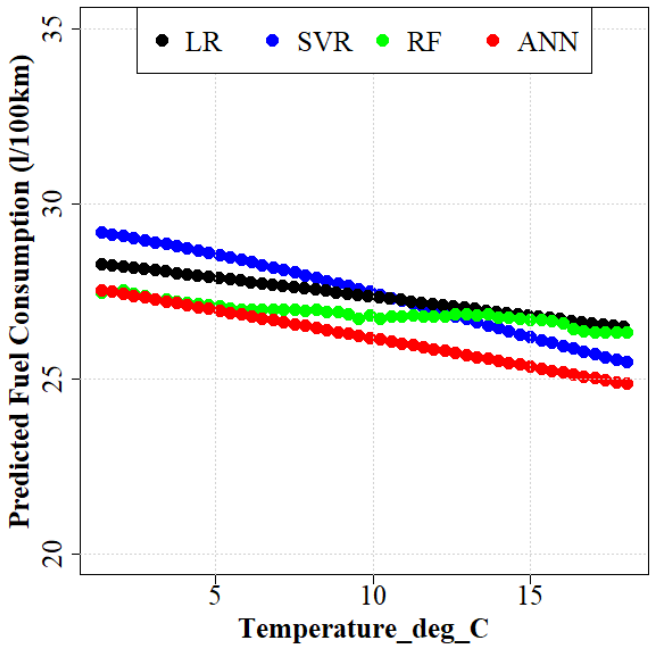


Figure 214. Parametric analysis of *Temperature_deg_C* (°C) for the heavy trucks App 1.

Figure 215 shows variation of fuel consumption due to changes of *Humidity_*. From the figure, it is possible to see that trends detected by all models are similar and slightly increasing for higher *Humidity_*. However, due to the entity of this variation, it is not possible to estimate that visually.

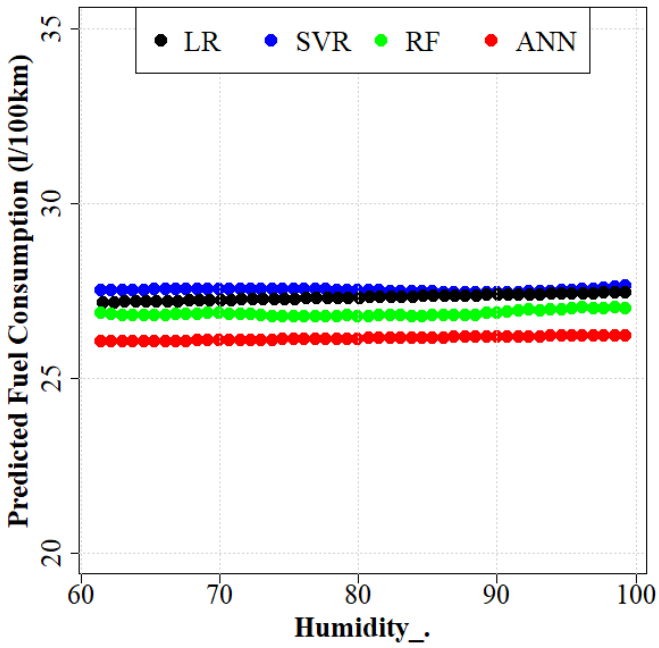


Figure 215. Parametric analysis of *Humidity_*. (%) for the heavy trucks App 1.

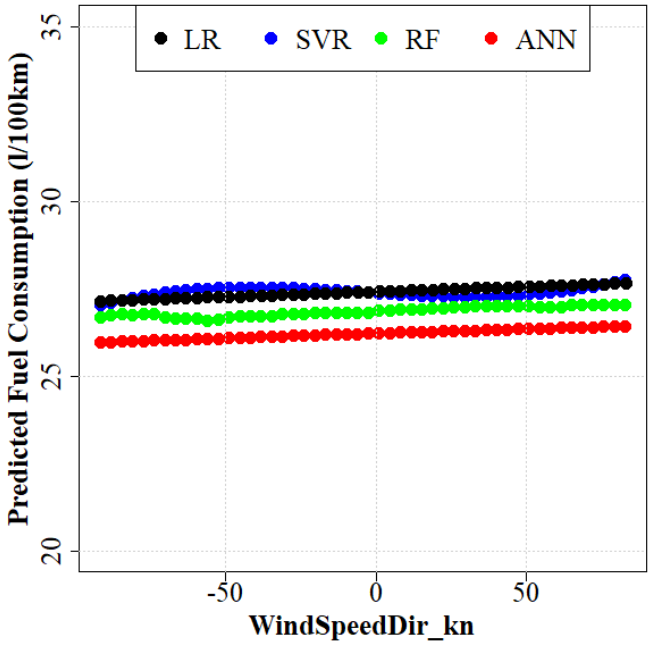


Figure 216. Parametric analysis of *WindSpeedDir_kn* (kn) for the heavy trucks App 1.

Figure 216 shows variation of fuel consumption due to changes of *WindSpeedDir_kn*. As seen for other case studies, the figure confirms that truck fleet

fuel consumption tends to increase with stronger wind blowing in opposite direction to the direction of travel. That is physically reasonable but, due to the distance between where the truck drives and where the measurement is taken, the identified variation results smaller than expected (approximately 1 l/100 km over a range of 200 kn). More accurate measurements of wind speed at the truck location might help at obtaining improved estimates.

Table 65 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the heavy trucks App 1.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	3.44	12.59	2.97	10.87	3.82	13.97	3.78	13.80
Geom.Gradient	9.24	33.77	12.67	46.31	9.50	34.73	14.46	52.85
Torque.AVG	6.39	23.34	3.38	12.36	7.25	26.51	4.45	16.28
Revs.AVG	0.81	2.96	0.92	3.35	0.57	2.06	1.24	4.54
Speed.AVG	-0.36	-1.33	0.82	2.98	0.00	0.00	0.34	1.25
Acceleration	-1.39	-5.07	-6.07	-22.17	-2.20	-8.04	-4.14	-15.13
Traveled.Time	-0.60	-2.19	0.76	2.79	-0.02	-0.08	0.93	3.41
Geom.Abs_Radius	-0.76	-2.76	0.90	3.27	0.35	1.26	0.30	1.11
Geom.Crossfall	-0.39	-1.43	-2.47	-9.02	0.04	0.13	-1.44	-5.25
TRACS.LPV03m	0.01	0.04	0.10	0.36	-0.10	-0.35	-0.38	-1.40
TRACS.LPV10m	-0.06	-0.22	0.26	0.97	-0.41	-1.51	-0.21	-0.78
TRACS.LPV30m	0.09	0.34	-0.21	-0.78	0.75	2.75	-0.86	-3.13
TRACS.Pav_ Texture	0.89	3.24	3.18	11.61	1.03	3.76	1.51	5.52
Altitude.AVG	0.47	1.72	0.90	3.30	0.28	1.03	-0.71	-2.59
Atm_Pressure_hPa	-0.33	-1.20	-0.84	-3.05	0.03	0.10	-0.49	-1.81
Temperature_deg_C	-0.92	-3.36	-1.99	-7.28	-0.23	-0.85	-1.37	-5.01
Humidity_.	0.14	0.53	-0.10	-0.36	0.01	0.05	0.10	0.36
WindSpeedDir_kn	0.26	0.95	-0.24	-0.87	0.37	1.34	0.24	0.87

Table 66 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the heavy trucks App 1.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	6.20	22.65	5.67	20.72	4.59	16.77	6.65	24.29
Geom.Gradient	16.64	60.79	19.94	72.88	14.47	52.88	22.01	80.42
Torque.AVG	11.49	42.00	4.90	17.90	8.34	30.48	8.71	31.83
Revs.AVG	1.46	5.34	1.69	6.18	0.77	2.80	2.25	8.21
Speed.AVG	-0.66	-2.40	0.62	2.26	0.04	0.14	0.59	2.15
Acceleration	-2.50	-9.13	-10.46	-38.22	-3.09	-11.28	-5.24	-19.15
Traveled.Time	-1.08	-3.94	1.35	4.94	-0.05	-0.20	1.38	5.03
Geom.Abs_Radius	-1.36	-4.97	1.68	6.12	0.19	0.71	0.31	1.14
Geom.Crossfall	-0.70	-2.57	-2.47	-9.02	0.83	3.03	-1.86	-6.80
TRACS.LPV03m	0.02	0.07	0.18	0.65	0.09	0.34	-0.69	-2.53
TRACS.LPV10m	-0.11	-0.40	0.48	1.74	-0.20	-0.73	-0.49	-1.77
TRACS.LPV30m	0.17	0.61	-0.38	-1.40	0.80	2.91	-2.21	-8.06
TRACS.Pav_ Texture	1.60	5.84	4.28	15.64	1.49	5.43	0.67	2.46
Altitude.AVG	0.85	3.10	0.19	0.69	0.36	1.31	-2.01	-7.33
Atm_Pressure_hPa	-0.59	-2.16	-1.31	-4.77	0.12	0.42	-0.89	-3.25
Temperature_deg_C	-1.66	-6.06	-3.43	-12.54	-1.16	-4.24	-2.46	-9.01
Humidity_.	0.26	0.95	0.04	0.16	0.16	0.60	0.17	0.64
WindSpeedDir_kn	0.47	1.71	0.47	1.71	0.30	1.08	0.43	1.58

Table 65 and 66 report impacts on fuel consumption calculated between the 25^{th} - 75^{th} and 5^{th} - 95^{th} percentiles of the considered variables, respectively. This provides a numerical estimation of how each of the developed models estimates the impact of each of the considered variables on fuel consumption, it helps in the interpretation of the parametric analysis, and provides with values to compare with results from the studies conducted in the past, draw conclusions and make

recommendations. As mentioned, two calculations were made. One for the 25th – 75th percentiles (see Table 65) and another for the 5th – 95th percentiles (see Table 66). This is because although there might be more confidence in the estimates made within the first range, the latter might be of better interest for the considered study, as this might include the extreme conditions of the road surface, and their impact on the fuel consumption of the considered fleet of trucks.

E.2.3.6 Heavy trucks on Motorways and A roads (App 2)

In this section, result from the parametric analysis of the heavy trucks App 2, from the Dataset 2, are reported.

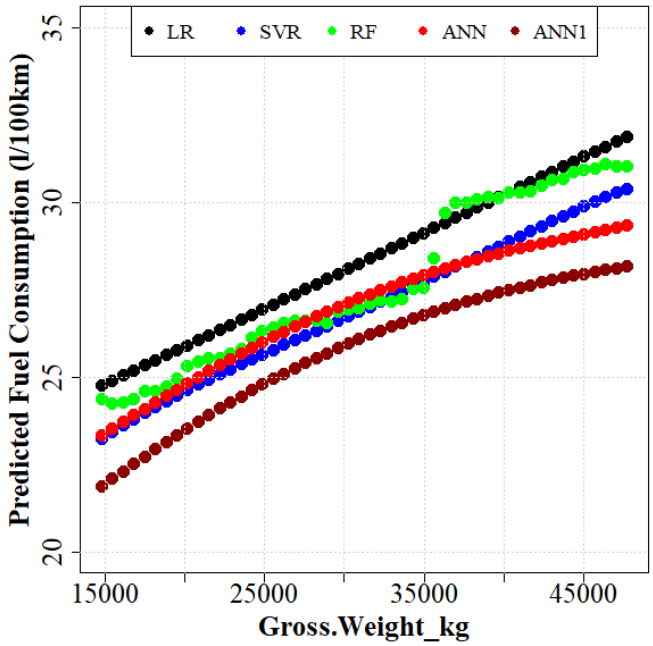


Figure 217. Parametric analysis of *Gross.Weight_kg* (kg) for the heavy trucks App 2.

Because this is about trucks similar to those in the Dataset 1, and from roads of similar characteristics, comparison of results is possible to a certain extent. In this last application of the study, similarly to the RF, the ANN1 demonstrated to

outperform the performance of the SVR and ANN at estimating the overall fuel consumption of the considered fleet of trucks. For this reason, parametric analysis of this model is also performed.

Figure 217 shows variation of fuel consumption due to changes of *Gross.Weight_kg*. All the developed models seem to agree on the fact that over a range between 15,000 and 50,000 kg, fuel consumption increases of about 7 l/100 km.

Figure 218 shows variation of fuel consumption due to changes of *Geom.Gradient*. The figure shows that, again all models agree on the fact that fuel consumption increases of up to about 36 l/100 km between -1 % and +2 %, which represents the highest impact detected so far.

Figure 219 and 220 show variation of fuel consumption due to changes of *Torque.AVG* and *Revs.AVG*, respectively. In regard to *Torque.AVG*, as it was seen in previous case studies, the LR provides a linear approximation of a non-linear behaviour that SVR, RF, ANN and ANN1 detect. Apart from the LR model, all the other detect an increment of fuel consumption of approximately 5 l/100 km between 10 % and 70 % of *Torque.AVG*. Regarding *Revs.AVG*, the models seem to agree all on the fact that higher values of this variable produce an increment in fuel consumption of approximately 2 l/100 km between 1,100 and 1,300 rpm.

Figure 221 shows variation of fuel consumption due to changes of *Speed.AVG*. Although the LR and ANN seem to underfit the data, showing a reduction of fuel consumption of the restricted range of vehicle speed considered in the study. SVR, RF and the ANN1 show a small increment, of about 1 l/100 km between 80 and 95 km/h.

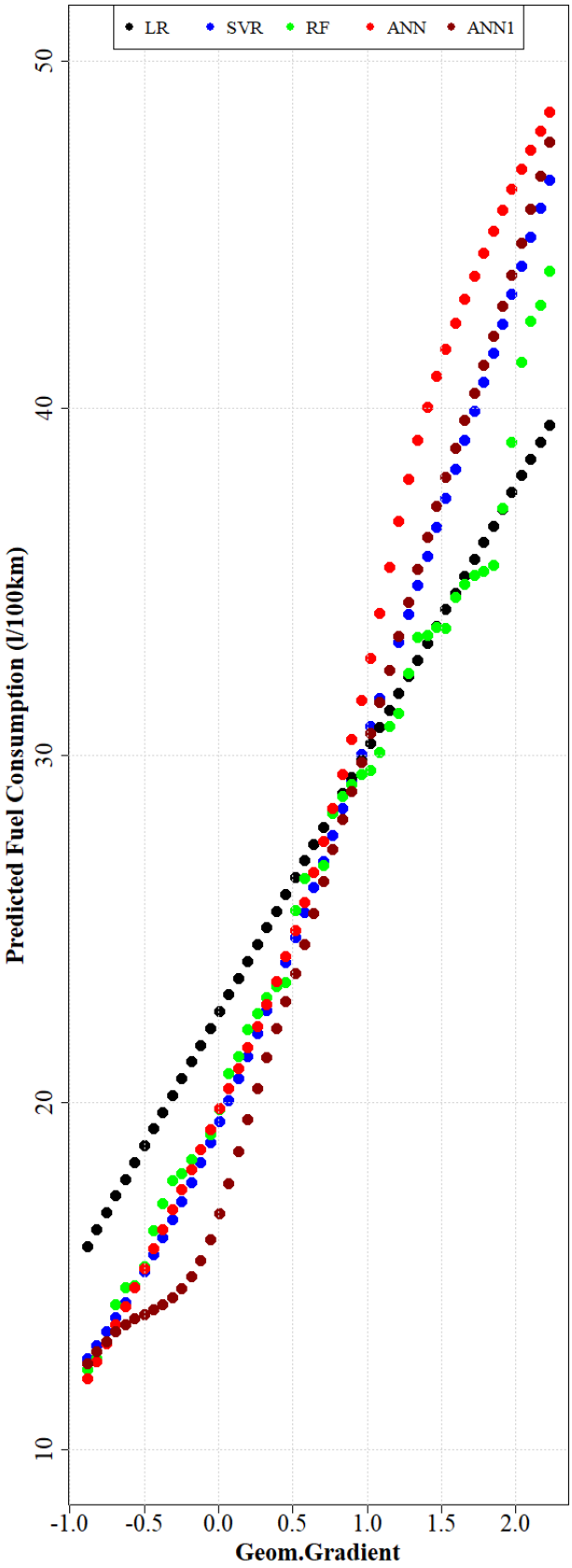


Figure 218. Parametric analysis of *Geom.Gradient* (%) for the heavy trucks App 2.

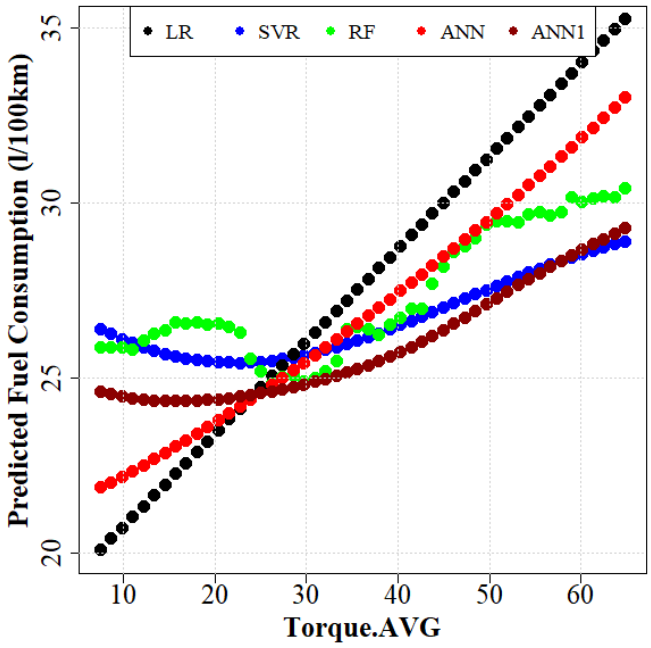


Figure 219. Parametric analysis of *Torque.AVG* (%) for the heavy trucks App 2.

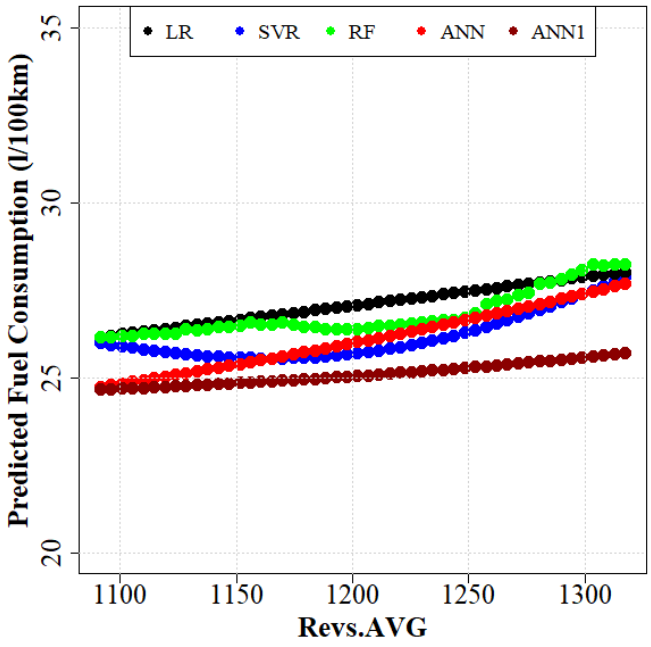


Figure 220. Parametric analysis of *Revs.AVG* (rpm) for the heavy trucks App 2.

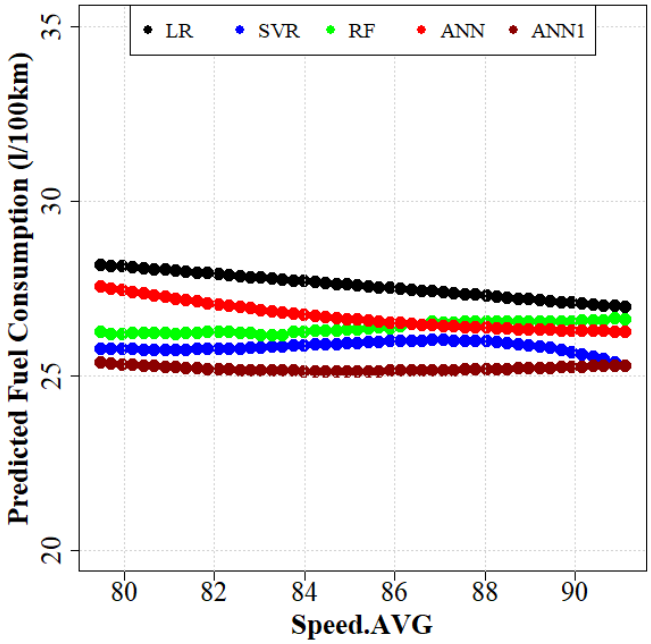


Figure 221. Parametric analysis of *Speed.AVG* (km/h) for the heavy trucks App 2.

Figure 222 shows variation of fuel consumption due to changes of *Acceleration*. Again, the figure shows that, the restricted range of *Acceleration* considered in the study is probably due to the effect of *Geom.Gradient* on *Speed.AVG*. In fact, all models show a quite clear trend, which sees fuel consumption reducing because of higher positive acceleration. Apart of the SVR, which seems to overfit a bit the available data, effect of *Acceleration* can be estimated to vary between 3 and 5 l/100 km, between -0.04 and +0.06 m/s^2.

Figure 223 shows variation of fuel consumption due to changes of *Traveled.Time*. In this regard, models provide different estimations with substantially different trends observed by the applied algorithms. Due to the large quantity of data considered in this case study, and in lieu of the results seen in previous case studies (see Appendix E.2.3.1, E.2.3.2, E.2.3.3, E.2.3.4 and E.2.3.5), that might be due more likely to overfitting of the applied algorithms, than to actual trends.

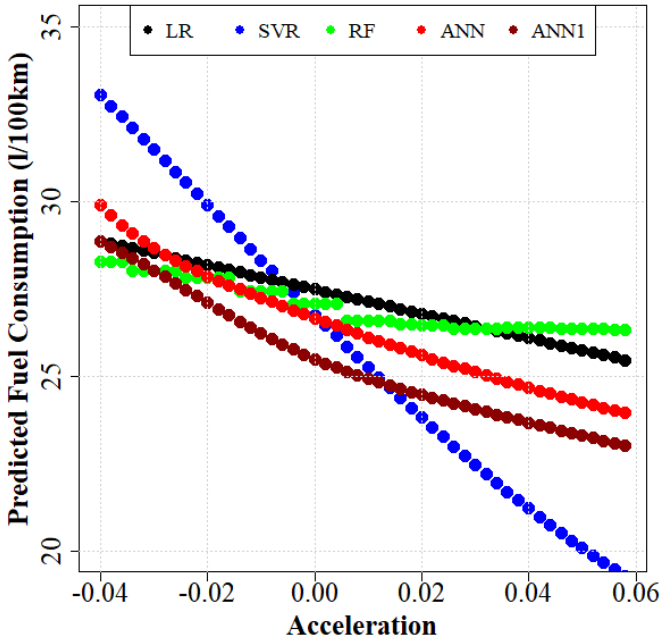


Figure 222. Parametric analysis of *Acceleration* (m/s^2) for the heavy trucks App 2.

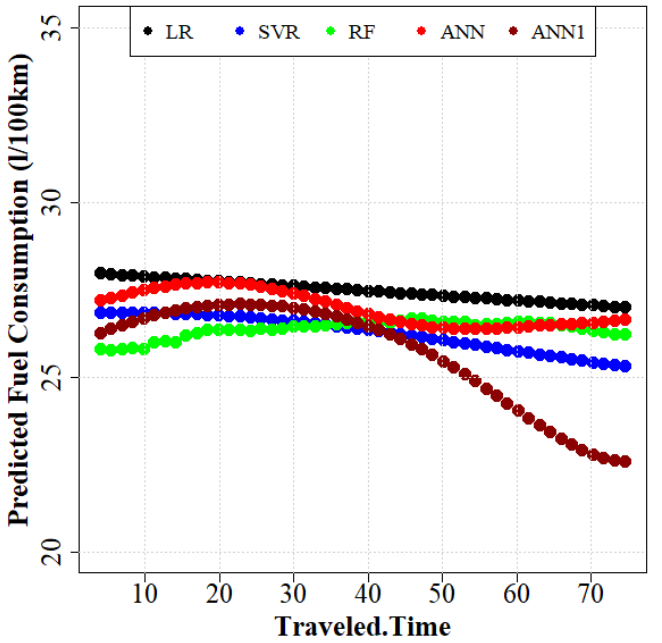


Figure 223. Parametric analysis of *Traveled.Time* (s) for the heavy trucks App 2.

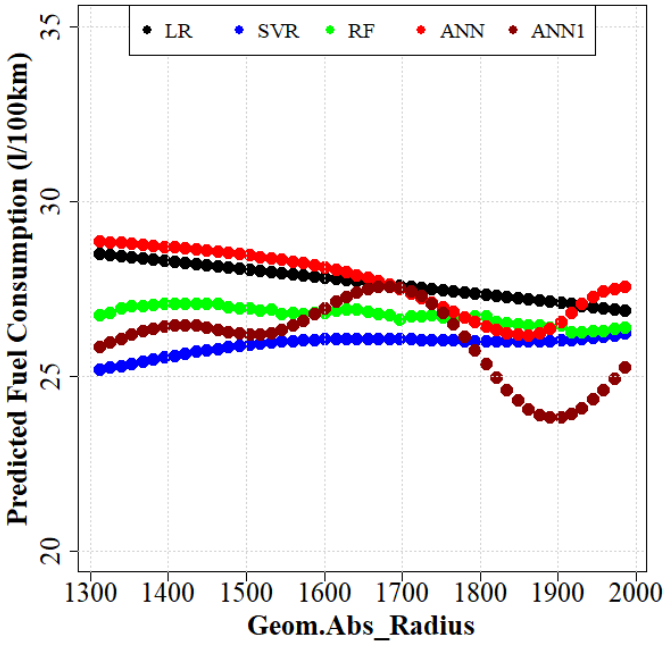


Figure 224. Parametric analysis of *Geom.Abs_Radius* (m) for the heavy trucks App 2.

Figure 224 shows variation of fuel consumption due to changes of *Geom.Abs_Radius*. The figure shows that, while the ANN1 seems to absolutely overfit these data, the other models identify two trends. In particular, while LR and ANN seems to estimate a reduction (of about 1.5 l/100 km) of fuel consumption due to higher *Geom.Abs_Radius*, between 1,300 and ≥ 2,000 m radius, RF and SVR identify an increment (of about 1.5 l/100 km), over the same range. Increase of fuel consumption due to larger *Geom.Abs_Radius* might be due to higher vehicle speed associated with it.

Figure 225 shows variation of fuel consumption due to changes of *Geom.Crossfall*. Interestingly, while LR and RF show substantially insignificant trends, SVR, ANN and ANN1 show that a reduction (of about 3 to 5 l/100 km) of fuel consumption is associated to positive crossfall. That would be in line with results

estimated by LR and ANN for the *Geom.Abs_Radius* (see Figure 224), which should be somewhat correlated to *Geom.Crossfall*.

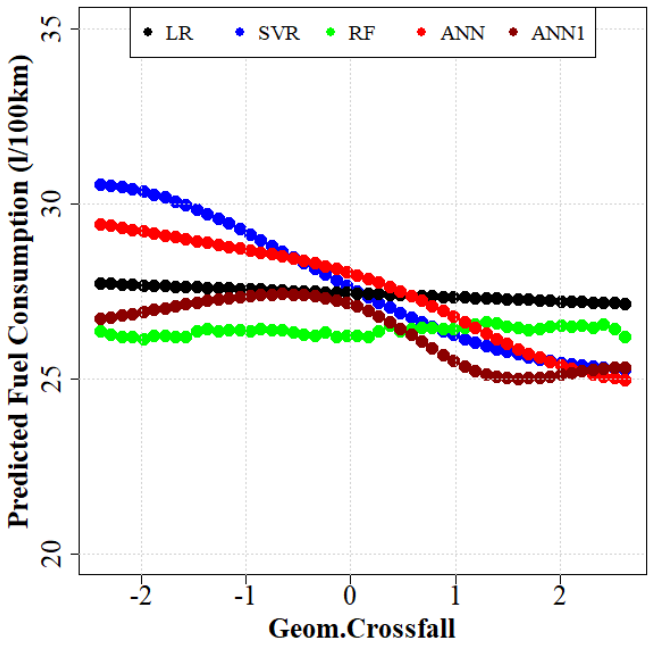


Figure 225. Parametric analysis of *Geom.Crossfall* (%) for the heavy trucks App 2.

Figure 226 shows variation of fuel consumption due to changes of *TRACS.LPV03m*. From the figure, it is possible to see that, while the LR model estimates insignificant changes of fuel consumption due to these variables, SVR, RF and ANN identify a clear trend through the data, with the ANN1 that, overfitting apart, shows a similar behaviour identified. Increment of fuel consumption associated with *TRACS.LPV03m* can be estimated to be up to approximately 1.5 l/100 km, between 0.1 and 0.55 mm^2.

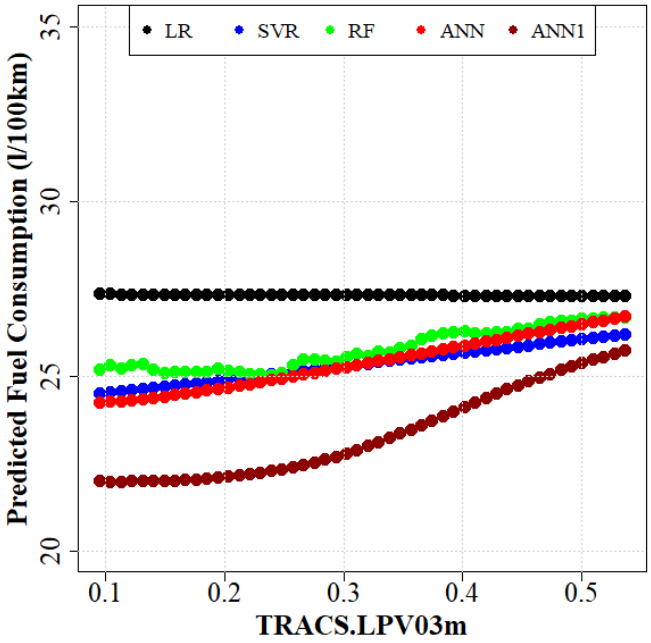


Figure 226. Parametric analysis of *TRACS.LPV03m* (mm^2) for the heavy trucks App 2.

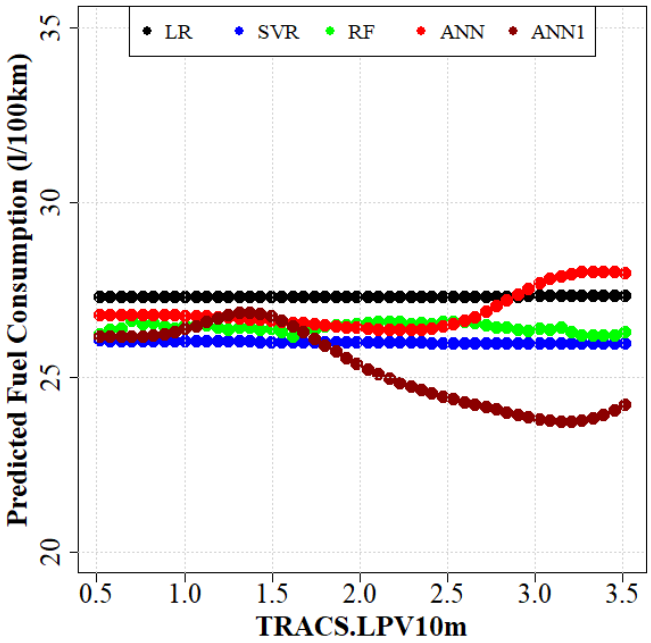


Figure 227. Parametric analysis of *TRACS.LPV10m* (mm^2) for the heavy trucks App 2.

Figure 227 shows variation of fuel consumption due to changes of *TRACS.LPV10m*. The figure shows that the ANN and ANN1 models seems to detect opposite patterns, probably due to some overfitting of the data available. In fact, LR, RF and SVR seem to agree that the effect of this variable is nearly insignificant over the considered range of road surface characteristics (from 0.5 to 3.5 mm^2).

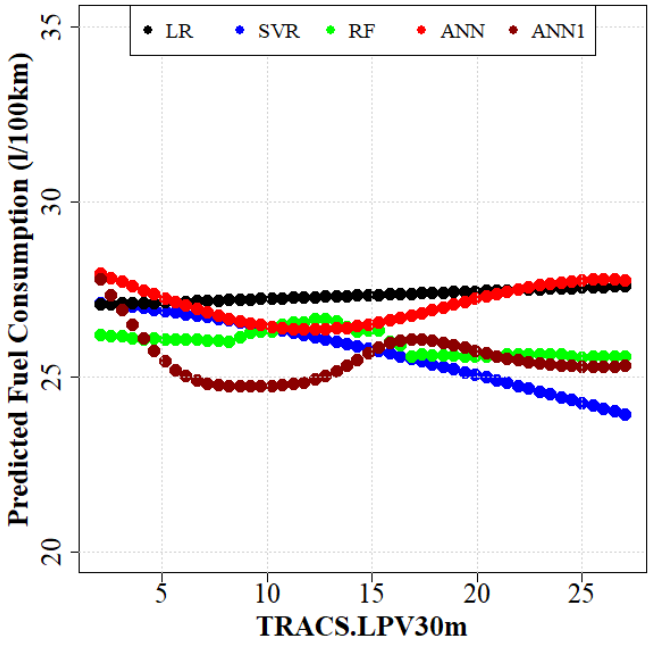


Figure 228. Parametric analysis of *TRACS.LPV30m* (mm^2) for the heavy trucks App 2.

Figure 228 shows variation of fuel consumption due to changes of *TRACS.LPV30m*. While the ANN detects an increment, of which the LR gives a linear approximation, all the other models developed seem disagree with the patterns identified by the firsts. In fact, the SVR estimates a reduction, the RF very small changes, while the ANN1 seems to overfit the data again.

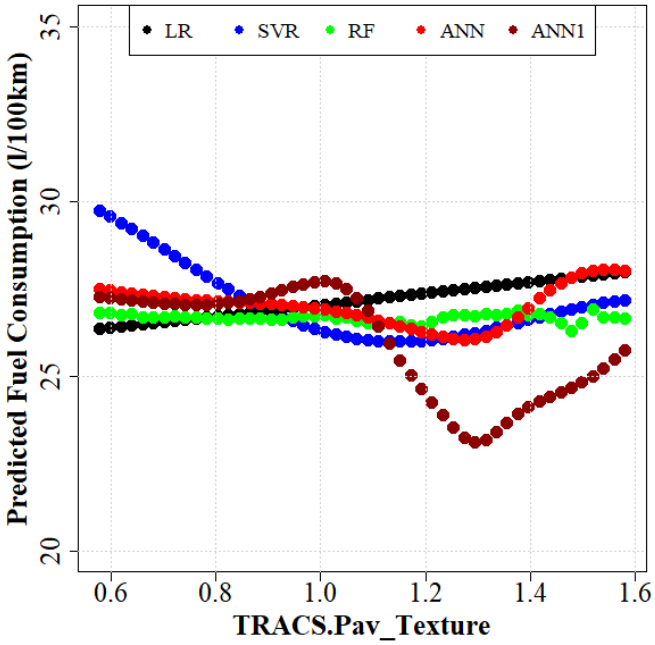


Figure 229. Parametric analysis of *TRACS.Pav_Texture* (SMTD, mm) for the heavy trucks App 2.

Figure 229 shows variation of fuel consumption due to changes of *TRACS.Pav_Texture* (SMTD). For this variable, apart from the ANN1, which, again, shows to be prone to overfit the single predictors, LR, RF and ANN seem to agree on an increment of fuel consumption due to higher values of SMTD. Although the SVR disagree on this, showing a reduction of fuel consumption overall, it is possible to see that after approximately 1.0 mm of SMTD also this model follows a trend similar to that identified by LR, RF and ANN.

Figure 230 shows variation of fuel consumption due to changes of *Altitude.AVG*. The ANN1 keeps to overfit the data, providing poorly interpretable results. The ANN seems to overfit the data around 90 m of *Altitude.AVG*, however, it usually tends to follow the same trend of the LR. For the SVR and RF models, the second seem to provide an approximation of the first, with an almost constant trend identified. Due to the different responses given by the different models, it is fair to say

that inclusion of *Altitude.AVG* in the developed models probably does not help in improving their accuracy.

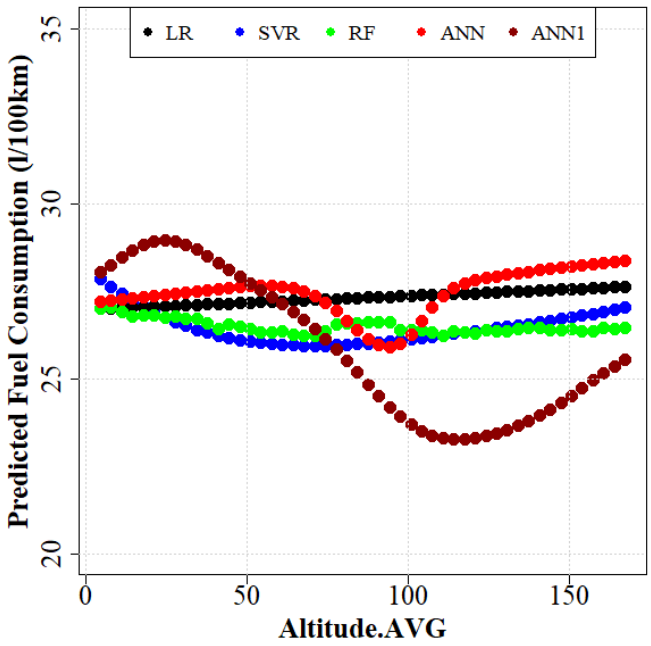


Figure 230. Parametric analysis of *Altitude.AVG* (mamls) for the heavy trucks App 2.

Figure 231 shows variation of fuel consumption for changes of *Atm_Pressure_hPa*. All model show similar trends and identify a reduction of fuel consumption due to higher *Atm_Pressure_hPa*. Again, that might be due to better efficiency of the engine due to higher air density or higher temperature.

Figure 232 shows variation of fuel consumption due to changes of *Temperature_deg_C*. The figure shows that all models agree on the identified trends, with a detected reduction of fuel consumption due to higher air temperature. This can be as high as 3 l/100 km over a range from 0 ℃ to 20 ℃.

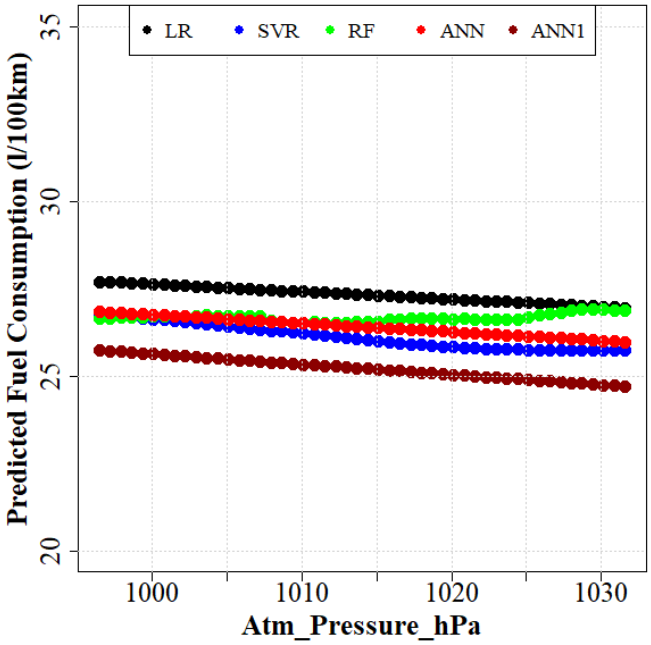


Figure 231. Parametric analysis of *Atm_Pressure_hPa* (hPa) for the heavy trucks App 2.

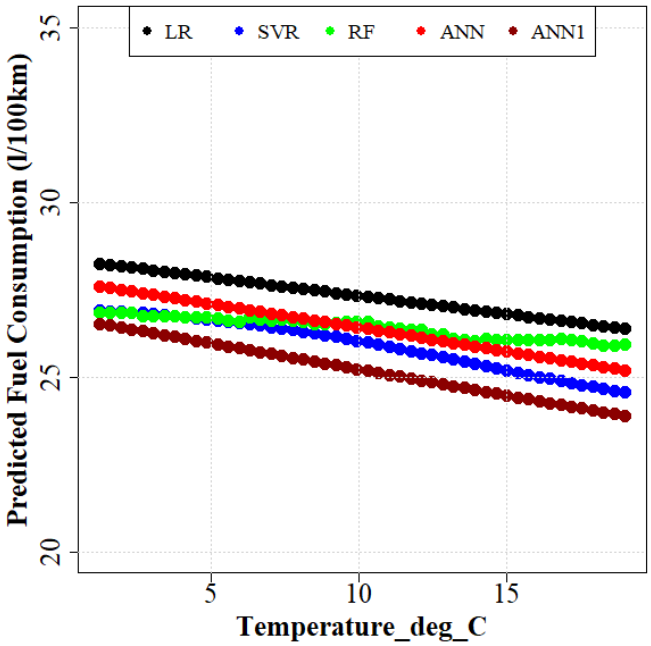


Figure 232. Parametric analysis of *Temperature_deg_C* (°C) for the heavy trucks App 2.

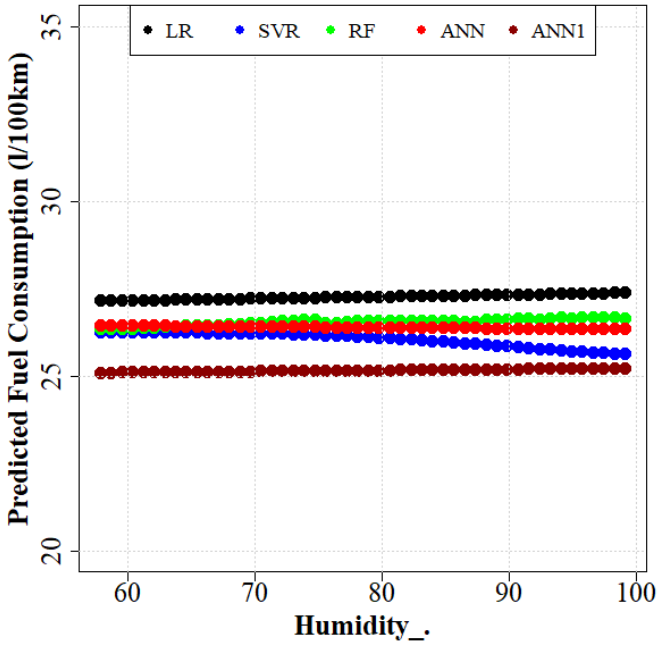


Figure 233. Parametric analysis of *Humidity_*. (%) for the heavy trucks App 2.

Figure 233 shows variation of fuel consumption due to changes of *Humidity_*. The figure shows that, while the SVR shows a reduction of fuel consumption for higher humidity and the ANN no apparent impact, LR, RF and ANN1 seem to detect a slight increment with similar trends. However, this impact would be very small, lower than the average error that the developed models have.

Figure 234 shows variation of fuel consumption due to changes of *WindSpeedDir_kn*. The figure shows that all models developed for the considered type of truck detect an increment of fuel consumption due to higher values of this variable, which is physically reasonable. From the figure, this can be estimated at approximately 1 - 2 l/100 km between a range of ± 100 kn.

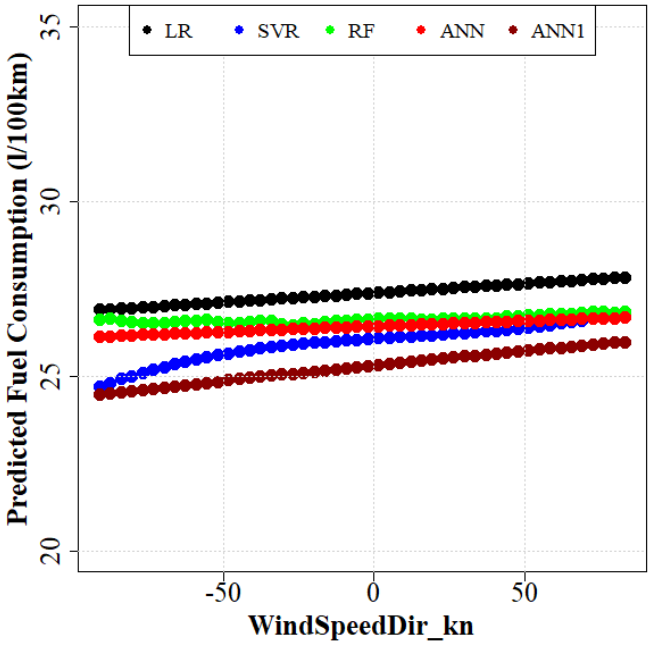


Figure 234. Parametric analysis of *WindSpeedDir_kn* (kn) for the heavy trucks App 2.

Table 67 and 68 report impacts on fuel consumption calculated between the 25^{th} - 75^{th} and 5^{th} - 95^{th} percentiles of the considered variables, respectively. This provides a numerical estimation of how each of the developed models estimates the impact of each of the considered variables on fuel consumption, it helps in the interpretation of the parametric analysis, and provides with values to compare with results from the studies conducted in the past, draw conclusions and make recommendations. The tables (Table 67 and 68) report results of the LR, SVR, RF and ANN only. This is because, in the parametric analysis, it was seen that the ANN1 although more accurate than the ANN at predicting the overall fuel consumption of the considered fleet of trucks, it showed to be prone to overfitting most of the single variables considered. Therefore, although less accurate, when it comes to interpretability of the results, there is more confidence in using a smaller neural network (i.e. ANN) than the larger (i.e. ANN1) and, for this reason, results of the

parametric analysis conducted on the ANN1 are finally not reported in the tables above.

Table 67 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 25th and 75th percentiles of each of the considered variable, for the heavy trucks App 2.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	3.64	13.33	3.51	12.87	4.58	16.79	3.11	11.38
Geom.Gradient	12.06	44.17	18.02	65.99	15.10	55.29	21.96	80.43
Torque.AVG	7.73	28.31	2.06	7.55	2.83	10.35	5.68	20.79
Revs.AVG	0.95	3.47	0.85	3.11	0.64	2.35	1.48	5.43
Speed.AVG	-0.62	-2.26	0.22	0.79	0.30	1.11	-0.67	-2.45
Acceleration	-1.74	-6.37	-7.38	-27.03	-1.48	-5.41	-2.71	-9.93
Traveled.Time	-0.49	-1.81	-0.89	-3.26	0.17	0.62	-1.33	-4.86
Geom.Abs_Radius	-0.82	-2.99	0.21	0.77	-0.33	-1.22	-2.13	-7.80
Geom.Crossfall	-0.29	-1.07	-3.63	-13.29	0.25	0.90	-2.53	-9.28
TRACS.LPV03m	-0.03	-0.10	0.87	3.18	1.00	3.68	1.36	4.97
TRACS.LPV10m	0.02	0.07	-0.05	-0.19	0.09	0.32	0.15	0.54
TRACS.LPV30m	0.27	0.98	-1.67	-6.12	-0.44	-1.59	0.55	2.03
TRACS.Pav_ Texture	0.82	3.01	-1.35	-4.96	0.12	0.44	-1.01	-3.69
Altitude.AVG	0.32	1.17	0.17	0.61	-0.01	-0.05	0.32	1.18
Atm_Pressure_hPa	-0.38	-1.41	-0.68	-2.48	-0.10	-0.37	-0.44	-1.62
Temperature_deg_C	-0.95	-3.47	-1.31	-4.82	-0.58	-2.12	-1.23	-4.49
Humidity_.	0.12	0.43	-0.34	-1.25	0.15	0.54	-0.05	-0.20
WindSpeedDir_kn	0.48	1.75	0.68	2.49	0.10	0.35	0.28	1.01

As mentioned, two calculations were made. One for the $25^{th} – 75^{th}$ percentiles (see Table 67) and another for the $5^{th} – 95^{th}$ percentiles (see Table 68). This is because although there might be more confidence in the estimates made within the first range, the latter might be of better interest for the considered study, as this might include the

extreme conditions of the road surface, and their impact on the fuel consumption of the considered fleet of trucks.

Table 68 - Summary of the impacts (Δ) on fuel consumption (l/100 km), calculated between the 5th and 95th percentiles of each of the considered variable, for the heavy trucks App 2.

Variables	Δ LR	Δ% LR	Δ SVR	Δ% SVR	Δ RF	Δ% RF	Δ ANN	Δ% ANN
Gross.Weight_kg	6.55	24.00	6.57	24.07	6.69	24.51	5.63	20.60
Geom.Gradient	21.71	79.51	31.09	113.9	28.71	105.1	34.33	125.7
Torque.AVG	13.92	50.96	2.39	8.76	4.27	15.64	10.12	37.06
Revs.AVG	1.71	6.24	1.56	5.71	2.09	7.64	2.68	9.81
Speed.AVG	-1.11	-4.06	-0.23	-0.84	0.38	1.40	-1.22	-4.46
Acceleration	-3.13	-11.46	-12.85	-47.06	-1.92	-7.04	-5.43	-19.87
Traveled.Time	-0.89	-3.25	-1.41	-5.15	0.55	2.02	-0.70	-2.56
Geom.Abs_Radius	-1.47	-5.37	0.85	3.12	-0.52	-1.92	-1.56	-5.71
Geom.Crossfall	-0.53	-1.93	-5.17	-18.94	0.21	0.78	-4.22	-15.47
TRACS.LPV03m	-0.05	-0.17	1.56	5.72	1.31	4.80	2.27	8.33
TRACS.LPV10m	0.03	0.13	-0.09	-0.33	-0.16	-0.57	1.22	4.48
TRACS.LPV30m	0.48	1.76	-2.89	-10.58	-0.58	-2.14	-0.06	-0.23
TRACS.Pav_ Texture	1.48	5.42	-2.53	-9.27	0.08	0.29	0.55	2.00
Altitude.AVG	0.57	2.10	-0.76	-2.79	-0.66	-2.43	1.05	3.86
Atm_Pressure_hPa	-0.69	-2.53	-0.98	-3.61	0.24	0.88	-0.79	-2.91
Temperature_deg_C	-1.71	-6.25	-2.17	-7.94	-0.89	-3.28	-2.21	-8.08
Humidity_.	0.21	0.78	-0.56	-2.05	0.34	1.26	-0.10	-0.36
WindSpeedDir_kn	0.86	3.15	1.83	6.69	0.20	0.75	0.50	1.82