Investigation of Emission Characteristics during Low Temperature Combustion using Multivariate Adaptive Regression Splines

Mario Velardi
West Virginia University

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation
https://researchrepository.wvu.edu/etd/480

This Dissertation is brought to you for free and open access by The Research Repository @ WVU. It has been accepted for inclusion in Graduate Theses, Dissertations, and Problem Reports by an authorized administrator of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.
Investigation of Emission Characteristics during Low Temperature Combustion using Multivariate Adaptive Regression Splines

by

Mario Velardi

Dissertation Thesis submitted to the College of Engineering and Mineral Resources at West Virginia University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

Mridul Gautam, Ph.D., Chair
Mario Perhinschi, Ph.D.
Andrew Nix, Ph.D.
Hailin Li, Ph.D.
Arvind Thiruvengadam, Ph.D.
Lorenzo Mariti, Ph.D.

Department of Mechanical and Aerospace Engineering

Morgantown, West Virginia 2014
ABSTRACT

Investigation of Emission Characteristics during Low Temperature Combustion using Multivariate Adaptive Regression Splines

Mario Velardi

Exhaust emissions from diesel engines operating in a low temperature combustion (LTC) regime are significantly affected by fuel composition and injection strategy. The starting point of this study is a collection of data correlating injection system parameters, and fuel characteristics, to response parameters such as engine-out emissions (oxides of nitrogen (NOx), total particulate matter (TPM), carbon monoxide (CO), hydrocarbons (HC)) and brake thermal efficiency (BTE).

The purpose of this work is to develop a statistical analysis tool to assist the emission analyst in modeling problems in which a response of interest is influenced by several variables and the objective is to optimize this response. The experimental data produced during LTC operation have been analyzed using an approach commonly known as Response Surface Methodology (RSM). Since the system under study may be responding to hidden inputs that are neither measured nor controlled, regression analysis must be performed via a flexible procedure. The methodology that will be used in this sense is called Multivariate Adaptive Regression Splines (MARS), which allows to approximate functions of many input variables given the value of the function at a collection of point in the input space.

Data was collected at West Virginia University's Engine and Emissions Research Laboratory for the project CRC AVFL-16. The test engine was a turbo-charged GM 1.9L operated in the LTC mode utilizing a split injection strategy. Main and pilot SOI timing and fuel split were varied per a 5 X 3 X 3 full factorial design. Advanced Vehicle Fuel Lubricants (AVFL) Committee of the Coordinating Research Council (CRC) defined a matrix of nine test Fuels for Advanced Combustion Engines (FACE) based on the variation of three properties: cetane number, aromatic content, and 90 percent distillation temperature. The experimental data was used has a platform for the code development, and for its validation.
Using multivariate data analysis is not only useful in visualizing correlations that otherwise would be hidden by the large amount of experimental data points, but it is also capable to predict the behavior of those points inside the domain where no data are available. As suggested by the name this is a regression methodology capable of adapting the shape of the regression splines to the data analyzed. Validation datasets which were independent of the calibration datasets were used to check the accuracy of the model predictions.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>I</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Literature Review</td>
<td>6</td>
</tr>
<tr>
<td>Low Temperature Combustion</td>
<td>7</td>
</tr>
<tr>
<td>Fuel Influence on Combustion</td>
<td>13</td>
</tr>
<tr>
<td>Emissions</td>
<td>17</td>
</tr>
<tr>
<td>Experimental Setup</td>
<td>20</td>
</tr>
<tr>
<td>Test Engine</td>
<td>21</td>
</tr>
<tr>
<td>Laboratory Instrumentation</td>
<td>21</td>
</tr>
<tr>
<td>Particulate Matter Sampling</td>
<td>28</td>
</tr>
<tr>
<td>Control of Engine Operating Parameters</td>
<td>28</td>
</tr>
<tr>
<td>Fuel Properties</td>
<td>29</td>
</tr>
<tr>
<td>Repeatability Study</td>
<td>31</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>33</td>
</tr>
<tr>
<td>Introduction</td>
<td>33</td>
</tr>
<tr>
<td>Two Simple Approaches to Prediction</td>
<td>33</td>
</tr>
<tr>
<td>Curse of Dimensionality</td>
<td>35</td>
</tr>
<tr>
<td>Tree Based Methods</td>
<td>35</td>
</tr>
<tr>
<td>MARS: Multivariate Adaptive Regression Splines</td>
<td>38</td>
</tr>
</tbody>
</table>
ANOVAs .................. 43
Case Study .................. 44

Preliminary Results 47
Experiment Reduction .................. 53

Results 58

\( NO_x \) .................. 60
Soot Concentration .................. 67
Carbon Oxide .................. 74
Hydrocarbon .................. 77
Verification Study .................. 82
Optimization .................. 86

Further studies based on MARS 92

Diesel Engine Modeling Development for ICCT Heavy-Duty Vehicle Simulation Tool 92
Analyze Dispersing Plume from Heavy Duty Diesel Trucks 95

Conclusions 101

APPENDIX A 104

APPENDIX B 110

NOx .................. 110
Soot Concentration .................. 112
CO .................. 114
HC .................. 118

APPENDIX C 120
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>NOx PM trade off curve [10]</td>
<td>7</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Emission characteristics at different injection timings [11]</td>
<td>8</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Performance and emissions comparison at different SOI [14]</td>
<td>10</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Injection rate and rate of heat release of conventional combustion vs. UNIBUS combustion [14]</td>
<td>12</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Operating region and IMEP for a compression ratio of 18 and intake charge temperature of 30 C [16]</td>
<td>13</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Schematic diagram of LTC combustion [17]</td>
<td>14</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Test engine for Cetane Number evaluation [20]</td>
<td>16</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Burner for distillation temperature evaluation [27]</td>
<td>18</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Heat Release Rate comparison between Diesel and Gasoline [32]</td>
<td>22</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Dependency of LTHR on Fuel Ignitability for Different Fueling Rates [34]</td>
<td>22</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Smoke production for different level of EGR [41]</td>
<td>24</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Schematic Overview of EERL Measurement System</td>
<td>26</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Test Engine GM Z19DTH in the EERL</td>
<td>28</td>
</tr>
<tr>
<td>Figure 14</td>
<td>CVS tunnel in the EERL</td>
<td>30</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Horiba Automotive Emission Analyzer System MEXA 7200D</td>
<td>31</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Example of NDIR Configuration</td>
<td>32</td>
</tr>
<tr>
<td>Figure 17</td>
<td>Example of FID Configuration</td>
<td>33</td>
</tr>
<tr>
<td>Figure 18</td>
<td>Horiba MEXA-720 NOx and principle of Zirconia Sensor</td>
<td>34</td>
</tr>
<tr>
<td>Figure 19</td>
<td>Experimental Setup for Nanoparticle Sampling</td>
<td>35</td>
</tr>
<tr>
<td>Figure 20</td>
<td>Air-Vac system</td>
<td>37</td>
</tr>
<tr>
<td>Figure 21</td>
<td>Schematic diagram of the EEPS spectrometer</td>
<td>38</td>
</tr>
<tr>
<td>Figure 22</td>
<td>Schematic diagram of the DMS classifier</td>
<td>38</td>
</tr>
<tr>
<td>Figure 23</td>
<td>Principle of Photoacoustic measurement</td>
<td>39</td>
</tr>
</tbody>
</table>
Figure 24  Medsker Electric Inc. alternate current dynamometer

Figure 25  FACE Diesel Fuels Design Matrix

Figure 26  Length of the subcube edge needed to capture a fraction r of the volume of the data, for different dimensions p.

Figure 27  Partition of a two dimensional space, and corresponding binary tree diagram

Figure 28  Basis functions

Figure 29  Schematic of the MARS forward model-building procedure

Figure 30  Block diagram of the MARS forward model-building procedure

Figure 31  Single variable plot and interaction surface

Figure 32  NOx emission from FACE 5

Figure 33  PM emission from FACE 5

Figure 34  CO emission from FACE 5

Figure 35  HC emission from FACE 5

Figure 36  BTE from FACE 5

Figure 37  NOx comparison between MARS results and experimental values

Figure 38  Comparison between experimental data and regression results

Figure 39  Dependence of NOx production from Cetane Number

Figure 40  Dependence of NOx production from Distillation Temperature

Figure 41  Dependence of NOx production from CN and Main SOI

Figure 42  Dependence of NOx production from T90 and Main SOI

Figure 43  Dependence of NOx production from CN, T90, and Main SOI

Figure 44  Dependence of NOx production from CN, Main SOI, and Pilot SOI

Figure 45  Comparison between experimental data and regression results

Figure 46  Dependence of soot production from Cetane number

Figure 47  Dependence of soot production from main SOI

Figure 48  Dependence of soot production from distillation temperature and main SOI

Figure 49  Dependence of soot production from Cetane number and aromatic content

Figure 50  Dependence of soot production from aromatic content and pilot SOI at low Cetane number
Figure 51  Dependence of soot production from aromatic content and pilot SOI at high Cetane number  
Figure 52  Comparison between experimental data and regression results  
Figure 53  Dependence of CO from Cetane number  
Figure 54  Dependence of CO from Cetane number and main SOI  
Figure 55  Dependence of CO from Cetane number, pilot SOI, and main SOI  
Figure 56  Comparison between experimental data and regression results  
Figure 57  Dependence of HC from Cetane number  
Figure 58  Dependence of HC from aromatic content  
Figure 59  Dependence of HC from main SOI  
Figure 60  Dependence of HC from main SOI, and cetane number  
Figure 61  Dependence of HC from main SOI, distillation temperature, and aromatic content  
Figure 62  Comparison between NOx experimental values and predicted results  
Figure 63  Comparison between experimental values and predicted results for CO, HC, PM, and BTE  
Figure 64  Distribution of the experimental points on a PM-NOx Pareto chart  
Figure 65  Distribution of the experimental points on a load-speed chart  
Figure 66  MY 2005 Mercedes engine  
Figure 67  Comparison between MARS results and ESC data  
Figure 68  Fuel consumption surface plot  
Figure 69  Lay out of the wind tunnel experimental setup  
Figure 70  Sampling pattern  
Figure 71  Residual sum of square for each sampling channel  
Figure 72  Comparison between MARS and experimental values  
Figure 73  Dependence of NOx from main SOI, and pilot SOI  
Figure 74  Dependence of NOx from main SOI, and fuel split  
Figure 75  Dependence of PM from cetane number, aromatic content, and main SOI  
Figure 76  Dependence of PM from cetane number, fuel split, and main SOI  
Figure 77  Dependence of CO from pilot SOI, and main SOI  
Figure 78  Dependence of CO from fuel split, and main SOI  
Figure 79  Dependence of CO from distillation temperature, pilot SOI, and main SOI
Figure 80  Dependence of CO from aromatic content, distillation temperature, and main SOI  
Figure 81  Dependence of HC from aromatic content, and main SOI  
Figure 82  Dependence of HC from cetane number, pilot SOI, and main SOI
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Hydrocarbon Terminology [24]</td>
<td>17</td>
</tr>
<tr>
<td>Table 2</td>
<td>Further ASTM Standard Test Methods used for FACE Diesel Fuels</td>
<td>20</td>
</tr>
<tr>
<td>Table 3</td>
<td>Test Engine Specifications</td>
<td>29</td>
</tr>
<tr>
<td>Table 4</td>
<td>Conditions and Parameters for Dilution with Air-Vac Model TD110H</td>
<td>35</td>
</tr>
<tr>
<td>Table 5</td>
<td>Residence time for dilution system components</td>
<td>36</td>
</tr>
<tr>
<td>Table 6</td>
<td>Analysis of FACE fuels characteristics</td>
<td>41</td>
</tr>
<tr>
<td>Table 7</td>
<td>Repeatability analysis</td>
<td>42</td>
</tr>
<tr>
<td>Table 8</td>
<td>ANOVA Decomposition</td>
<td>55</td>
</tr>
<tr>
<td>Table 9</td>
<td>Residual sum of squares for each characteristic predicted</td>
<td>66</td>
</tr>
<tr>
<td>Table 10</td>
<td>ANOVA Decomposition NOx</td>
<td>71</td>
</tr>
<tr>
<td>Table 11</td>
<td>ANOVA Decomposition PM</td>
<td>79</td>
</tr>
<tr>
<td>Table 12</td>
<td>ANOVA Decomposition CO</td>
<td>85</td>
</tr>
<tr>
<td>Table 13</td>
<td>ANOVA Decomposition HC</td>
<td>88</td>
</tr>
<tr>
<td>Table 14</td>
<td>Data points used for calibration</td>
<td>94</td>
</tr>
<tr>
<td>Table 15</td>
<td>Prediction model errors</td>
<td>96</td>
</tr>
<tr>
<td>Table 16</td>
<td>Set of parameters producing a global unconstrained optima in emission</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>characteristics</td>
<td></td>
</tr>
<tr>
<td>Table 17</td>
<td>Responses corresponding to the global unconstrained optima</td>
<td>100</td>
</tr>
<tr>
<td>Table 18</td>
<td>Set of parameters producing a global constrained optima in emission</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>characteristics</td>
<td></td>
</tr>
</tbody>
</table>
# Nomenclature

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ATDC</td>
<td>After top dead center</td>
</tr>
<tr>
<td>BMEP</td>
<td>Brake mean effective pressure</td>
</tr>
<tr>
<td>BTDC</td>
<td>Before top dead center</td>
</tr>
<tr>
<td>CA50</td>
<td>50 percent mass fraction burned</td>
</tr>
<tr>
<td>CFR</td>
<td>Code federal regulations</td>
</tr>
<tr>
<td>CI</td>
<td>Compression ignition</td>
</tr>
<tr>
<td>CLD</td>
<td>Chemiluminescence detector</td>
</tr>
<tr>
<td>EERL</td>
<td>Engine and emission research laboratory</td>
</tr>
<tr>
<td>EGR</td>
<td>Exhaust gas recirculation</td>
</tr>
<tr>
<td>FACE</td>
<td>Fuels for advanced combustion</td>
</tr>
<tr>
<td>DAQ</td>
<td>Data acquisition system</td>
</tr>
<tr>
<td>DMS</td>
<td>Differential Mobility Spectrometer</td>
</tr>
<tr>
<td>EEPS</td>
<td>Exhaust emissions particulate sizer</td>
</tr>
<tr>
<td>FID</td>
<td>Flame ionization detection</td>
</tr>
<tr>
<td>GC</td>
<td>Generalized Cross Validation</td>
</tr>
<tr>
<td>HC</td>
<td>Hydrocarbon</td>
</tr>
<tr>
<td>HFID</td>
<td>Heated flame ionization detector</td>
</tr>
<tr>
<td>HPN</td>
<td>Heptamethylnonane</td>
</tr>
<tr>
<td>HTHR</td>
<td>High temperature heat release</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Term</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>IC</td>
<td>Internal combustion</td>
</tr>
<tr>
<td>LOF</td>
<td>Lack of fitting</td>
</tr>
<tr>
<td>LTC</td>
<td>Low temperature combustion</td>
</tr>
<tr>
<td>LTHR</td>
<td>Low temperature heat release</td>
</tr>
<tr>
<td>MARS</td>
<td>Multivariate adaptive regression splines</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>NDIR</td>
<td>Non-dispersive infrared</td>
</tr>
<tr>
<td>NDIR</td>
<td>Non-dispersive infrared</td>
</tr>
<tr>
<td>NOx</td>
<td>Nitrogen oxide</td>
</tr>
<tr>
<td>OEM</td>
<td>Original equipment manufacturer</td>
</tr>
<tr>
<td>PID</td>
<td>Proportional-integral derivative</td>
</tr>
<tr>
<td>PM</td>
<td>Particulate matter</td>
</tr>
<tr>
<td>PRF</td>
<td>Primary reference fuels</td>
</tr>
<tr>
<td>PRR</td>
<td>Pressure rise rate</td>
</tr>
<tr>
<td>RSM</td>
<td>Response surface methodology</td>
</tr>
<tr>
<td>RSS</td>
<td>Residual sum of squares</td>
</tr>
<tr>
<td>SOI</td>
<td>Start of injection</td>
</tr>
<tr>
<td>SOI</td>
<td>Start of injection</td>
</tr>
<tr>
<td>SSV</td>
<td>Subsonic venturi</td>
</tr>
<tr>
<td>STD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>T90</td>
<td>Distillation temperature</td>
</tr>
<tr>
<td>TDC</td>
<td>Top dead center</td>
</tr>
<tr>
<td>UNIBUS</td>
<td>Uniform Bulky combustion System</td>
</tr>
<tr>
<td>VTG</td>
<td>Variable turbine geometry</td>
</tr>
</tbody>
</table>
Introduction

Diesel engines represent a convenient choice in terms of power density, reliability and durability compared to the gasoline counterpart. The combustion process in diesel engines on the other hand contributes to significantly high PM and NOx emissions, while HC and CO are relatively low. Diesel exhaust gases were identified in the past to be a serious threat to human health and the environment, and regarded as the main source of air pollution. On 17 October 2013, the World Health Organizations International Agency for Research on Cancer (IARC) announced that it has classified Outdoor Air Pollution as carcinogenic to humans [1]. After a review of available literature, IARCs expert group concluded that there is sufficient evidence that exposure to outdoor air pollution causes lung cancer [2]. They also noted a positive association with an increased risk of bladder cancer. Particulate matter, as a major component of outdoor air pollution, was evaluated separately and was also classified as carcinogenic to humans.

Due to tightening emissions regulations, both within the US and Europe, including concerns regarding greenhouse gases, next-generation combustion strategies for diesel engines have drawn increasing attention during recent years. One way to minimize NOx emissions is to limit the in-cylinder temperature during the combustion process by means of fuel injection strategies. However, fuel chemistry plays a significant role in the ignition delay; hence, influencing the overall combustion characteristics and the resulting emissions. Much of the existing technical literature has focused on the interaction of fuel properties and conventional diesel combustion, whereas studies of fuel property effects on advanced combustion are less prevalent.

The present study analyze the data collected during the project AVFL-16 [3]. These data refer to a light duty compression-ignition engine, operated at a fixed engine speed and load. Several split
injection strategy were investigated by varying the start of fuel injection and fuel injection quantity for two injection events. The fuel chemistry was taken into account by operating the engine with a matrix of nine different fuels. The matrix of nine fuels for advanced combustion engine research was defined by the Fuels for Advanced Combustion Engines (FACE) Working Group of the Advanced Vehicle, Fuel, and Lubricants Committee (AVFL) of the Coordinating Research Council (CRC) [4]. Development of the FACE diesel fuel matrix centered around three important fuel characteristics: auto-ignition quality, boiling range, and chemical composition. Respectively, these characteristics are represented by CN, T90, and AC.

The analysis of the data collected is particularly challenging due to the nature of the problem. Various engine operating parameters have conflicting effects on the NOx and PM emissions, for example retarding the injection timing to reduce NOx production will result in higher soot levels. The variations of NOx, and PM emissions introduced above involve tradeoffs that make achieving this goal especially difficult [5]. Considering also different fuel characteristics makes the phenomenon even more complex to analyze. Hence, only a systematic multivariate study can provide a clear understanding of the combustion characteristics of the engine. Such multivariate problems require a continuous quantitative mathematical model to link the factors to responses. In engine applications the relationship between the design or operating factors and the responses is highly nonlinear. The underlying mechanisms in most cases cannot be condensed to one equation that still reflects the physical laws and can be used as the objective or constraint function. Moreover, sometimes the true functional relationships or physical models are not even understood or available, and the governing equations of the processes are simply lacking.

A viable solution to approximate the underlying process is to create an empirical model. Such model does not contain any physical meaning; it is instead obtained through regression of the available data. This approach to the problem is commonly referred to as response surface methodology (RSM). This is a procedure in which the output variables are called responses and the independent input variables are called factors. The response function forms a surface or hyper-surface (if there are more than two factors) in the factor space. RSM refers to the analysis to be performed once that an emulator model is obtained by data fitting. The determination of an emulator model is not a trivial task in the case of multivariate, nonlinear problems.
Commonly used linear regression techniques fail to assess the actual value of the operating parameters in engine testing. This can be considered as one of the causes why the RSM approach is not a standard analysis procedure for emission testing. In the present study a code based on MARS algorithm, described by Friedman [6], has been developed in order to describe engine performance and emission characterization. The main advantage of MARS is that this is a nonparametric regression modeling procedure. Unlike better known linear regression techniques, MARS does not assume that coefficients are stable across the entire domain of each variable and instead uses splines to fit the response. By combining classical linear regression, and mathematical construction of splines, MARS is ideally suitable for problems with high input dimensions where the course of dimensionality would likely create problems for other techniques. This method has been successfully employed for various prediction and data mining applications both in recent years [7].

The use of multivariate data analysis is not just confined to visual correlations of hidden data, but also to predict behavior of data points within a domain. The MARS algorithm used for analyzing the data will be developed in this study, and the development of the algorithm will constitute a major portion and a significant outcome.

Objectives The focus of this work is to produce a statistical tool specifically designed to support the analysis of empirical data related to emissions and combustion. The main objective is to develop a regression tool capable to accommodating up to 6 engine and fuel parameters. In particular, the tool will address highly nonlinear high-dimensional problems and background noise, maintain a low computational cost and offer ease of interpretation. The analysis support code has to evaluate an equation for each emission or performance output that needs to be studied.

Once that the regression equations are proved to be reliable in describing the data, meaning that the fitting of the data is considered acceptable, these can be used to determine the importance of the governing factors. The evaluation of the importance of each factor over the response is both quantitative and graphical. The quantitative analysis is obtained by recasting the regression equation in a form that will allow to separate the contribution of each factor and their join interactions, the output of this analysis will be an index that will quantify the importance of the given independent variable (or group of variables) in relation to the given response. The data available on
LTC have been used has a platform to develop the analysis code described in this work. The main reasons that led to the choice of a set of data related to LTC, can be summarized in the following two points:

- A large set of data was available, in which multiple factors have been varied and different responses have been sampled. This offers the opportunity to challenge the regression code in a high dimensional domain, where the independent variables have conflicting effects on the responses. Moreover having to perform regression for several types of responses is useful in determining the flexibility of the modeling capability for different phenomenon.

- LTC represents a particularly challenging field of operation, characterized by high nonlinearity in the relationship between injection and fuel factors with emission responses. Considering that the purpose of the code described in this work will be to assess combustion and emission problems, proving its capability in analyzing LTC data will guarantee the applicability of the same method to conventional combustion problems.

The given data set has been used to perform an investigation on LTC and specifically the code described in this work determined regression equations correlating engine and fuel parameters to emission and performance values. In order to retain the investigation successful, and hence prove the capability of RSM has analysis methodology, the following specific objectives should be meet:

- To show the existing interaction between the fuel chemistry and the injection strategies, to further understand the phenomena that dominate Low Temperature Combustion.

- Represent the engine test data collected as surface plots that underline the correlations between the single variable and the output, and at the same time show the existing interactions between the different variables.

- As knowledge is gained about the response surface, major interest can be directed toward regions that appear to have greatest potential in terms of emission reduction. This will ultimately lead to the determination of the optimal operating conditions to minimize emissions while maintaining an acceptable thermal efficiency.
Literature Review

Introduction

Conventional diesel combustion is broadly used because of its high thermal efficiency, at the cost of high emissions in terms of PM and NOx. The significant NOx production is a direct consequence of the high temperature zones in the flame due to the necessity of spontaneous ignition which characterize diesel engines. Particulate matter formation instead is due to heterogeneous air fuel mixtures, which in turn leads to locally fuel rich zones. The majority of energy released during conventional combustion is at high temperatures, as a consequence of a diffusion flame [8]. Cooling the flame reduces the formation of NOx at the expenses of higher PM, HC, and CO production.

Studies have attempted to determine which is the minimum temperature required to achieve complete combustion [9]. Flyn et al. performed a study where the fuel specific NOx production was measured at different levels of exhaust gas recirculation (EGR). Their results show a decrease in NOx proportional to increased EGR rates, the limit is reached when combustion starts to deteriorate. The authors pointed out the minimum peak temperature for conventional diesel combustion is 2300K.

Approaches that allow engines to successfully operate at combustion temperatures lower than those encountered in conventional diesel are usually referred as LTC. Work on this family of strategies proved that the NOx-PM trade off curve can change dramatically at low temperature levels Figure 1 [10]. Altering the engine parameters that result in lower NOx and soot production will invariably result in lower combustion temperatures. The main idea behind LTC is to produce spontaneous ignition throughout the combustion chamber at low temperature, i.e. by means of very
lean homogeneous air-fuel mixtures. The low cylinder temperature is useful in reducing NOx and soot production, but, since it involves slow oxidation reactions that cause misfiring and quenching, will result in higher HC and CO emissions compared with conventional diesel combustion.

**Low Temperature Combustion**

In order to generate the conditions to perform LTC an increased premixing between air and fuel must be achieved, compared to conventional combustion. Higher levels of premixing can be obtained by increasing the injection delay, which can be controlled by increasing the amount of EGR. Longer ignition delay allows for more time available to fuel and air mixture to properly homogenize.

**Start of Injection Timing** One common method of promoting better mixing of the air and fuel charge is to advance the start of injection (SOI) timing. This will increase the homogeneity of the mixture allowing more time for air and fuel to mix. In conventional combustion NOx emissions increase with advanced SOI, but researchers have demonstrated that advanced SOI timing pared

![Figure 1: NOx PM trade off curve](image)

Figure 1: NOx PM trade off curve [10].
with increased injection pressure can reduce NOx emissions for SOI timings greater than 30° before top dead center (BTDC) [11]. This effect is likely a result of reduced diffusion flame zones, due to the absence of near stoichiometric air and fuel mixtures. Figure 2 shows the trend of NOx, CO, HC and fuel consumption with varying injection timing for different rail pressures. A decrease in NOx is obtained by anticipating the SOI timing at high pressure as a consequence of low temperature combustion. Marked increase in CO and HC is found under the same condition. The early fuel injection causes the spray interaction with the cylinder liners or piston walls; the result is wall quenching with resulting fuel problems and high concentrations of HC in the exhaust. Furthermore low temperature combustion is a major cause for high CO emissions.
**Fuel Split**  Modern compression ignition (CI) engines equipped with electronically controlled direct fuel injection system commonly use multiple pulse injections. The basic idea is to carry out a pilot injection of a small quantity of fuel before the main injection event. The primary advantage in using pilot injection is the ability to retard the main injection further than it would be possible with a single injection event. This is related to the reduction in ignition delay consequent to a higher temperature in the combustion chamber at the moment of the main injection. An important parameter in this sense is the time interval between the pilot and main injection event. When this time interval is too long the products of pilot injection mix sufficiently with the ambient gas in the cylinder, consequently no difference is produced with respect to the single injection case. On the other hand, when this interval is shorter a local high temperature zone is created near the fuel nozzle with a consequent shorter ignition delay [12]. It is still important to maintain a sufficient separation between pilot and main injection to ensure the main injection occurs after the injection delay of the pilot event. Under a NOx perspective the pilot injection can produce an effect similar to the one of an internal EGR, especially where the quantity of fuel injected during the pilot event is higher than 10% of the total fuel [13]. The burned gas produced from the combustion of the pilot injection will dilute the concentration of oxygen inside the combustion chamber for the main injection, acting basically like an EGR system.

Hasegawa and Yanagihara explored the effects of a multiple injection strategy on a concept engine named Uniform Bulky Combustion System (UNIBUS) [14]. During the development of the UNIBUS concept, a number of tests were performed to compare and quantify the effects of a double injection strategy with varying SOI timing versus a conventional diesel combustion strategy. For the double injection tests, the main SOI was fixed at 13° after top dead center (ATDC) while the pilot SOI was varied. The injection volume of fuel per cycle during the double injection strategy was held at 15 mm³/st for both injection events. The plot of brake mean effective pressure (BMEP) in Figure 3 shows that the UNIBUS strategy, as well as the majority of double injection strategy tests, are capable of achieving a BMEP close to that of conventional combustion with the same fuel quantity injected and while producing significantly lower NOX and smoke emissions. The main idea behind the UNIBUS system is to disperse fuel in small droplets that, upon evaporation, would form small fuel rich pockets uniformly distributed. An adequate spacing of these pockets will allow
Figure 3: Performance and emissions comparison at different SOI [14]
a first low temperature heat release (LTHR) stage in the compression stroke, during which the combustion of each pocket will not interact with the adjacent ones. Under these conditions the LTHR stage will be followed by a pause before the rest of the fuel is consumed. Further heating of the mixture, due to compression, will then produce a second more rapid high temperature heat release (HTHR) stage of combustion. The objective is to produce a premixed lean combustion during the HTHR that complete the oxidation of the remaining fuel hence producing little NOx and soot. An important feature to achieve the described combustion is early injection. It allows adequate time for LTHR to complete before the beginning of the second stage of combustion. It also ensures for maximum fuel penetration and dispersion as a consequence of the low density inside the cylinder at these early injection timings.

A way to promote the desired combustion process also at high engine loads, where early injection is not convenient, is using pilot injection. It was discovered that with the injection of about 30% of the fuel 30° BTDC, combined with a second injection around TDC, the combustion process remain premixed allowing for the tipical advantages of LTC without deficit in engine output [15]. The fuel injected in the pilot event undergoes LTHR and the main event will complete the combustion in the HTHR stage, as shown in Figure 4.

Figure 4 shows how the pilot injection allows postponing the main event, which in turn improves the premixing of air and fuel. Injection timings of the main fuel as late as 13° ATDC have been reported [14].

**Control of combustion** Early injection diesel LTC combustion has not been employed yet in production engines, despite of its low NOx and PM emissions, principally for the difficulty to control combustion under transient modes. Since the start of ignition is determined by in-cylinder conditions, there is no direct control on the combustion phasing. There are several indirect ways to influence the start of combustion, as excess air ratio, compress ratio, EGR rate and intake air charge temperature. However, those parameters are difficult to tightly control under transient modes.

**High load limit** High load conditions are critical for LTC combustion; this constrain is related to the premature start of heat release during compression stroke. A common approach to this problem is the so called dual mode operation, which consists in a switch between advanced combustion and
Figure 4: Injection Rate and Rate of Heat Release of Conventional Combustion vs. UNIBUS Combustion [14]

conventional combustion. It is still desirable to be able to perform a broad range of loads with the advanced combustion in order to avoid frequent transition between the two modes of operation. Several paths are viable to expand the range of operation of advanced combustion during high loads, for example applying high cooled EGR can postpone the ignition timing to crank angles near TDC as shown in Figure 5 [16].
Conceptual Combustion Model  Results obtained through optical engine testing are useful in visualizing of the combustion takes place during LTC [17]. The results showed in Figure 6 are obtained on a heavy-duty diesel engine operated with a single early fuel-injection and with 12.7% intake oxygen to simulate EGR. Panel 1 in Figure 6 shows how the low ambient temperature and density during the early-injection condition produces a long liquid jet penetration (blue) before a clear separation with the vapor phase becomes evident. The jet penetration during LTC can reach a length which is twice the one that occurs in conventional diesel engine, this may impinge in-cylinder surfaces and potentially degrade combustion efficiency and emissions. Panel 2 shows the vaporization due to the energy release as a consequence of the start of ignition, it proceeds up to panel 3. As the subsequent premixed combustion commences (panel 4), fluorescence appears in the laser sheet indicating that the mixture is almost stoichiometric. Panel 5 and 6 shows the areas where soot formation actually happens, i.e. in those areas where the fuel-air equivalence
1. Under the low-density conditions of early injection, liquid fuel (blue) penetrates much farther, even as it vaporizes (green).
2. Chemiluminescence from ignition reactions (green) envelopes the region containing liquid fuel (blue).
3. As the chemical energy of the fuel is released (green), the liquid fuel (blue) rapidly vaporizes.
4. OH (green) fills the jet cross-section shortly after ignition, indicating leaner mixtures than conventional diesel combustion.
5. Soot (red) is first detected near the head of the jet, in regions that are deficient in OH (green).
6. The strongest soot luminosity (red) is often observed in the roll-up vortices at the head of the jet, enveloped by OH (green).

Figure 6: Schematic diagram of LTC combustion [17]
ratio is greater than 2. The presence of soot in the head vortex of the jet indicates that the mixing is poorest in these regions. The long jet penetration contributes to a better mixing between fuel and air, which is the key feature leading to lower PM emissions compared to conventional combustion. The combustion is also more evenly distributed, leading to lower peak temperatures and consequently lower NOx emissions.

**Fuel Influence on Combustion**

Fuel properties are a key parameter in the combustion process, and consequently have a profound impact on the performance and emissions of engines employing LTC. Methods of control and necessary modification of engine hardware can be solely dependent on the properties of a selected fuel. This has resulted in a considerable amount of research to determine which fuels are best suited for advanced combustion [4]. Especially the capability of the fuel to mix with air and how easily the mixture will ignite affect the combustion. The ignitability is related to the chemical kinetic characteristics of the fuel when it mixes with air. CN is commonly used as a measure of fuel’s ignitability. The fuel volatility is the main index of how the fuel is able to vaporize in the air, combined with the temperature of the air flow interacting with the fuel. Higher levels of volatility guarantee a better mixing, and this parameter is described with distillation characteristics.

**Auto Ignition** In LTC engines the delay between injection of fuel into the air and ignition is much longer than that typically encountered in conventional diesel engines. A longer delay allows for higher rate of mixing and homogenizing, at the same time the temperature of the mixture reaches a point where it can auto-ignite making the process of conversion of the fuel air mixture into combustion product particularly rapid. The ignition delay is the main parameter of control for mixture homogeneity and combustion phasing, which has a strong impact on performance and emissions. The factors that mostly affect ignition delay are the temperature and pressure field, and the fuel characteristics. Figure 7 shows a comparison between diesel (CN=54) and gasoline (RON=95) engines in terms of heat release profiles [19]. It is interesting to notice how a LTHR stage is experienced by the diesel engine at 25°BTDC while gasoline does not. According to Shibata and Urushiahara [20] he LTHR stage can be inhibited, in some fuels, due to the following reasons:

1. Some fuels may contain only compounds that form stable radicals upon abstraction and do
not undergo LTHR.

2. Radicals that form fuel components, such as olefins and oxygenates, react with OH radicals that are important for the start of LTHR.

In most cases, LTHR can be attributed mainly to n-paraffins and to a lesser extent to branched paraffins. Aromatics have a LTHR inhibiting effect while compounds such as olefins and naphthenes both undergo LTHR and have a LTHR inhibiting effect [20]. Fuels that exhibit LTHR include diesel fuels with a cetane number above about 35-40. Fuels with a higher tendency to auto-ignite will generally have a more pronounced LTHR (Figure 8) [21].

While LTHR is not a requirement for successful premixed LTC operation, it can have a significant impact on the HTHR. Impacts include:

- a lower HTHR rates (see Figure 7). The HTHR rate can influence the combustion noise and the maximum load that can be achieved by the engine.
- a lower initial mixture temperature requirement to achieve optimal combustion phasing.
- a need for more advanced combustion phasing to achieve maximum power.
- increased tendency to misfire at retarded combustion phasing.

A fuel with a large LTHR tend to knock easily and is better suited for high speed and low torque operation. Fuels with a small LTHR are less likely to knock, hence they are better suited for low
speed and high torque operation [22].

**Measures of fuel Ignitability**  Cetane number is the most commonly used index to correlate well diesel-like fuels and ignition quality. Is widely proved that a decrease in Cetane number produce a delay in the ignition of the air-fuel mixture [23]. Other factors have strong influence on ignition delay and have to be considered when comparing fuel properties, for example [24] showed in their work that the relationship between different Cetane number diesel fuels and ignition delay was not constant at different temperatures and intake pressure. Other researchers [25] showed the ignition delay did not increase with low Cetane number when the engine is running at high loads without pilot injection.

**Fuel volatility**  Fuel volatility is another key parameter in LTC; it has to be high enough to guarantee fuel evaporation before impingement of the combustion chamber surfaces occurs. This explains the difference in behavior that occurs between gasoline and diesel when the start of injection is advanced. Emission strongly increase for early injection strategy that use diesel fuel because this kind of combustible is not enough volatile at low temperature. When the injection start earlier than 50° BTDC the temperature in the combustion chamber is not high enough to vaporize diesel fuels, this brings to impingement on the surfaces with consequent high values of emissions. On the
other hand gasoline fuels high volatility guarantee for vaporization also at the lower temperatures that characterize early injection [26].

**Emissions**

The main reasons that lead to the study of LTC engines is their capability to lead to significant reduction in thermal NO formation. While total NOx emissions from engines using LTC are lower than conventional diesel engines; it was reported that the fraction of $NO_2$ in NOx is consistently higher [27].

**Soot** Describing soot production in internal combustion engines can be extremely complex. Tao et al [28] were able to incorporate a soot generation model into KIVA-3V simulation code to model LTC engine operation with high EGR rates. The results suggested that the model is able to predict soot controlling mechanisms over a wide range of operating conditions including a EGR sweep from 0% to 68% combined with a split injection around TDC. As depicted from Figure 9; sufficiently high EGR rates will significantly lower the soot formation rate by mean of lower temperature combustion.

Increasing moderately EGR rates brings to increase in soot production especially in the final part of the crank motion. The increase is mainly due to lower soot oxidation, which is related to lower temperatures due to EGR flow. At EGR rates in excess of 65% the trend is reversed and a significant decrease in soot production is observed. This is probably due to lower rates of soot formation, as we can deduce observing that both the initial and final decrease has the same magnitude. Increase in the amount of EGR, combined with other measures, is used to lower emission, for example high injection pressure, high intake boost pressure, variable valve timing and variable compression ratio [29].

**Carbon Monoxide** LTC systems exhibit high carbon monoxide (CO) emissions, particularly at high dilution levels. CO is one of the intermediate species generated from the burning process of hydrocarbon in the fuel, when complete combustion is achieved this is oxidized to CO2. The efficiency of the oxidation depends on the local temperature and oxygen concentration [30].
Hydrocarbons  Like for CO the hydrocarbons (HC) produced by LTC are higher than those produced by conventional diesel operations [8]. The factors that cause this phenomenon can be listed as follows:

- The long spray penetration which characterize LTC operation bring fuel to accumulate on the combustion chamber surfaces.
- Fuel lean zones inside the combustion chamber are more common in LTC operations rather than in conventional diesel. Fuel in these areas is more prone to escape combustion leading to the production of unburned HC.
- Since the local peak temperatures are lower than conventional diesel, the fuel near the wall does not burn especially at low loads.
Experimental Setup

All measurements in this study were conducted at the Engine and Emission Research Laboratory (EERL) at West Virginia University for the project CRC AVFL-16 [31] [3]. The EERL test cell follows the recommendations outlined in the Code of Federal Regulations (CFR), Title 40, Part 1065 [32]. Figure 10 shows an overview of the EERL sampling capabilities for regulated and unregulated diesel exhaust emissions.

Figure 10: Schematic Overview of EERL Measurement System
The sampling system consists of a variable speed blower and a subsonic venture constant volume sampling system (CVS). Diluted gaseous sampling were analyzed using the Horiba® MEXA 7200D system. Primary data collected by the Horiba system comprised CO and CO\textsubscript{2} acquired by the non-dispersive infrared (NDIR) method, NO and NOx using the Chemi-luminescence (CLD) method, HC by using a heated flame ionization detector (HFID). Particulate matter characterization downstream the aftertreatment system was quantified both using the gravimetric method as outlined in Title 40 CFR Part 1065 [32], and by photo-acoustic method [33]. Diluted exhaust gas was sampled from the main dilution tunnel through a secondary dilution system and subsequently sampled onto a Pallflex 47 mm Teflon filter. Filters were weighed in an environmentally controlled clean room (Class 1000) using a Sartorius microbalance with an accuracy of 1 \( \mu g \).

**Test Engine**

The test engine used for this study was an in-line 4-cylinder common rail diesel engine model Z19DTH from General Motors, depicted in Figure 11. The intake air pressure was controlled via a variable turbine geometry (VTG) turbocharger. The engine was instrumented with thermocouples measuring engine lubricant, coolant, intake manifold and exhaust manifold temperatures. Inlet depression, intake manifold pressure and exhaust backpressure were measured as well by pressure transducers.

Since EGR rates that were adopted for the advanced combustion research on this engine were higher than the original equipment manufacturer (OEM) specifications, a larger EGR cooler was fitted to the engine in order to reduce inherent higher intake manifold temperatures. An overview of the test engine specifications are shown in Table 1.
Figure 11: Test Engine GM Z19DTH in the EERL

<table>
<thead>
<tr>
<th>Type</th>
<th>CDTi Diesel Engine</th>
<th>Bore</th>
<th>82mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer</td>
<td>General Motors</td>
<td>Stroke</td>
<td>90.4mm</td>
</tr>
<tr>
<td>Model</td>
<td>Z19DTH</td>
<td>Compression Ratio</td>
<td>17.5:1</td>
</tr>
<tr>
<td>Valve Configuration</td>
<td>4 Valves per Cylinder</td>
<td>Turbocharger</td>
<td>Garret VGT</td>
</tr>
<tr>
<td>Year</td>
<td>2005</td>
<td>Injection System</td>
<td>Common Rail</td>
</tr>
<tr>
<td>Configuration</td>
<td>In-Line 4 Cylinder</td>
<td>EGR</td>
<td>Cooled, External</td>
</tr>
<tr>
<td>Displacement</td>
<td>1.9L</td>
<td>Rated Power</td>
<td>110 kW at 4000 rpm</td>
</tr>
</tbody>
</table>

Table 1: Test Engine Specifications

Laboratory Instrumentation

This section describes the laboratory instrumentation in the EERL that was used for this study. The exhaust dilution system, gaseous emissions measurement instrumentation and particulate matter sampling system and technique, as well as in-cylinder pressure measurement and the engine and
dynamometer control system are presented.

**Constant Volume Sampling Dilution Tunnel** A total-exhaust dilution, CVS tunnel, designed to simulate the mixing of exhaust gas with ambient air conditions, maintains a nominally constant total molar flow rate of the diluted exhaust, as outlined in the Code of Federal Regulations (CFR), Title 40, Part 1065, Subpart 140 [32]. To accurately measure and actively control the flow rate maintaining proportional sampling of the exhaust constituents, a subsonic venturi (SSV) flow meter is used, see Figure 12. The SSV was calibrated for a Reynolds number at the throat greater than the maximum Reynolds number expected during testing and used only between the minimum and maximum calibrated flow rates.

![Figure 12: CVS tunnel in the EERL](image)

**Gaseous Emission Measurement** Diluted exhaust gas emissions extracted from the CVS tunnel were measured continuously using a HORIBA MEXA 7200D gaseous emission analyzer and included HC, CO as well as \( CO_2 \) and \( NO_x \) and \( O_2 \). The same emission characteristics were also sampled in the raw exhaust and intake manifold still using an HORIBA MEXA 7200D \( NO_x \) analyzer.
**Horiba Automotive Emission Analyzer System MEXA 7200D**

The Horiba automotive emission analyzer system MEXA 7200D is a modular components system consisting of a main control unit, an interface controller, an analyzer rack with up to five analyzer modules, a heated analyzer module for THC, THC/CH$_4$, NO$_x$ and NO/NO$_2$ analyzers, a power supply unit, as well as a solenoid supply unit, which routes to zero, span and samples gas to the analyzer modules, and a sample handling system dehumidifying the sample gas and directing it to the analyzers. A separate heated oven unit cart contains up to three heated analyzers (THC, NO$_x$ and CH$_4$) with heated lines, pumps and solenoid valves. The analyzer system is shown in Figure 13.

![Horiba Automotive Emission Analyzer System MEXA 7200D](image)

**Figure 13: Horiba Automotive Emission Analyzer System MEXA 7200D**

The analyzer modules measuring the concentration of CO and CO$_2$ (cold dry sample) use the non-dispersive infrared (NDIR) principle. Infrared energy at specific wavelengths is absorbed by a molecule consisting of different atoms and the degree of absorption is proportional to the concentration at constant pressure [34]. In a NDIR analyzer, an infrared beam is passed through a sample and a comparison cell by a light source. The comparison cell is filled with a gas non-absorbent to infrared radiation (such as nitrogen). Figure 14 shows an example of an NDIR
A sealed capacitor type detector, consisting of two cells separated by a movable membrane and filled with the gas to be measured, transforms the displacement of the membrane, which moves as a result of the differential pressure generated by the difference in quantity of radiation each gas received and therefore heated up and expanded, into an electrical output signal. To prevent interference with another gas component absorbing infrared radiation in the same wavelength range, an optical filter in front of the detector is used to eliminate the absorption area of the interfering component [35].

The concentration of total hydrocarbons is measured employing hydrogen ($H_2$) flame ionization detection (FID) (heated wet sample). Hydrocarbons introduced into a hydrogen flame generate ions, which are proportional to the number of carbon atoms in the sample. This measurement principle is sensitive to almost all hydrocarbon compounds [36]. Figure 15 shows a FID configuration.

The sample gas is mixed with $H_2$ and directed into the $H_2$ flame. Ions in the high-temperature area are generated according to the following the reaction:

$$\text{CH}^* + \text{O}^* \rightarrow \text{CHO}^* + e^- \quad (1)$$
where the stars (\(\ast\)) denote radicals. A DC voltage is applied on two electrodes in the ion collector which causes a migration of ions towards them and a current can be measured and amplified. Due to the proportionality of the current to the number of carbon atoms, this is a measure of the total hydrocarbons (THC), but no information of different hydrocarbon components can be obtained by this method [36]. The analyzer module measuring NO and \(NO_x\) concentration (heated dry atmospheric sample) uses a chemiluminescence detector (CLD). Sample gas containing NO is mixed with ozone gas (\(O_3\)) in a reactor to be oxidized into nitrogen dioxide (\(NO_2\)), whereas part of the \(NO_2\) is in excited state, releasing excited energy as light (radiation) when returning to the ground state, as shown in Equations 2 and 3:

\[
NO + O_3 \rightarrow NO_2^\ast + O_2
\]  
(2)

\[
NO_2^\ast \rightarrow NO_2 + hv
\]  
(3)

where the star (\(\ast\)) denotes the \(NO_2\) molecules in excited state. The light released is directly proportional to the NO molecule quantity before the reaction and therefore a measure of the NO concentration. This chemiluminescence signal is detected photo-electrically. Excited \(NO_2\) molecules can also return to ground state without radiation emission due to collision with other molecules (\(H_2O, CO_2, N_2\) or \(O_2\)). This interference can be reduced by reducing the pressure in the reaction chamber. Since there is also \(NO_2\) in the initial sample that does not have chemiluminescence, it has
to be converted to NO by means of a $NO_x$ converter. The measurement of the converted $NO_2$ and the aforementioned NO measurement can therefore be added up to yield the $NO_x$ concentration.

**Horiba MEXA-720 NO$_x$ Analyzer**  Two heated zirconia-ceramic ($ZrO_2$) sensors were installed in the intake manifold and exhaust pipe directly as parts of MEXA-720 NOx analyzer units (see Figure 16), which are capable of measuring $NO_x$ concentrations, air/fuel ratio, excess air ratio (lambda) and $O_2$ concentrations simultaneously. The measurement principle of a zirconia sensor is depicted in Figure 51. It is based on the oxygen conducting properties of zirconia [37]. Zirconia is used as an ion pump lowering the oxygen concentration from the sample gas to approximately 10ppm in the first internal cavity, where nitrogen dioxide is reduced to nitric oxide and oxygen. The pump current, depending on the amount of oxygen pumped, is measured and used to calculate the oxygen concentration in the sample gas. Further lowering the oxygen concentration to nearly zero ppm (1ppb) is achieved by an auxiliary oxygen pump (not showed in schema). Nitric oxide is further split into nitrogen and oxygen. A measure of the oxygen generated is the current created by the zirconia ion pump and can be used to calculate the NO concentration, which is an indicator of the $NO_x$ concentration in the exhaust stream, since $NO_2$ is reduced to NO in the first internal cavity [38]. Intake oxygen concentration was one of the primary controlled operating parameters in this study and raw exhaust $O_2$ concentration was used along with intake $O_2$ concentration to determine actual EGR fractions.
Particulate Matter Sampling

Nanoparticle number concentration and size distributions were determined using the Exhaust Emissions Particle Sizer (EEPS), a spectrometer from TSI Inc. (model 3090) as well as the Differential Mobility Spectrometer (DMS) from Cambustion (model DMS500). Continuous exhaust gas samples were extracted from the CVS tunnel (dilution ratio DR = 10) and routed through a double stage dilution system using ejector type dilutors, see Figure 17. The first stage was maintained at 140°C (DR = 6) in order to suppress condensation and particle nucleation phenomena, while the second stage utilized dilution air at ambient temperatures (25°C, DR = 11).

![Figure 17: Experimental Setup for Nanoparticle Sampling](image)

Control of Engine Operating Parameters

In order to have full control on the engine operating parameters, an open engine controller from Drivven Inc. was used. The controller is based on National Instruments hardware. Operating parameters such as main SOI, number and duration of fuel injection events, turbocharger boost (by controlling VGT vane position), EGR rate, rail pressure, throttle, and more, were accessed and controlled to obtain advanced combustion regimes.
Laboratory and Dynamometer Control  The GM Z19DTH engine was coupled to a Medsker Electric Inc. (MEI) alternate current (AC) dynamometer operated in speed mode (see Figure 18). Engine torque was controlled by means of a proportional-integral-derivative (PID) throttle controller integrated into the laboratory’s data acquisition system (DAQ). The EERLs DAQ system is an in-house solution using NI hardware and software developed by CAFEE with high-grade automation capabilities for engine testing as well as calibration and quality control. The DAQ system follows the recommendations outlined in 40 CFR, Part 1065 [32].

![Medsker Electric Inc. alternate current dynamometer](image)

Figure 18: Medsker Electric Inc. alternate current dynamometer

Fuel Properties

The Advanced Vehicles, Fuels, and Lubricants committee of the Coordinating Research Council specified and formulated a matrix of nine test fuels for advanced combustion engines (FACE) [4] based on the variation of three properties:

- Cetane number, a measure of ignition quality;
- Aromatic content, a measure of chemistry; and
- The 90 percent distillation temperature, a measure of volatility.

Table 2 displays the nine fuels used in this study.
These properties were obtained by fuel blending and determined to be of primary importance to the performance of advanced combustion engines. The target values built up a full factorial statistical design with a center run representing average marketplace values of the design variables. This design theoretically allows investigation of each combination of the fuel properties and the injection strategies. A graphical representation of the design matrix is shown in Figure 19, where target values (in blue) of the three factors at two levels build a design cube. The actual values of the formulated fuels are represented, as well (in red).

### Repeatability Study

As part of the AVFL-16 project [3] a study of the repeatability of test data generated from two split injection control strategies was performed with the ULSD fuel. The purpose of this effort
Figure 19: FACE Diesel Fuels Design Matrix: Target (Blue) vs. Actual (red)

was to develop a standard by which emissions and performance changes among the fuels could be attributed to fuel property differences and not to the variability associated with the equipment or control strategy.

The selected test was repeated three times in the morning and evening. This daily routine was repeated for three days providing data for 18 repeats of each test. From this data, the percent difference of the original test and average of the repeated tests was quantified. The standard deviation and coefficient of variation (COV) was also calculated for the repeat tests and is presented in the subsequent section. It should be noted that the original test was conducted four months prior to the 18 repeat tests and consisted of one test.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>NOx (g/kW-hr)</th>
<th>PM (mg/kW-hr)</th>
<th>CO (g/kW-hr)</th>
<th>HC (g/kW-hr)</th>
<th>BTE %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Test</td>
<td>0.352</td>
<td>104.5</td>
<td>14.5</td>
<td>2.74</td>
<td>29.7</td>
</tr>
<tr>
<td>Repeted Test</td>
<td>0.350</td>
<td>88.4</td>
<td>13.47</td>
<td>2.72</td>
<td>29.7</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.010</td>
<td>7.73</td>
<td>0.48</td>
<td>0.06</td>
<td>0.3</td>
</tr>
<tr>
<td>COV</td>
<td>2.81 %</td>
<td>8.75 %</td>
<td>3.58 %</td>
<td>2.35 %</td>
<td>0.99 %</td>
</tr>
<tr>
<td>Original vs Repeated</td>
<td>0.7 %</td>
<td>18.2 %</td>
<td>7.4 %</td>
<td>0.7 %</td>
<td>0.04 %</td>
</tr>
</tbody>
</table>

Table 3: Repetability analysis

Based on the results of this repeatability study, it was determined that any difference in HC emissions or NOX emissions greater than 3% could be deemed significant and thus attributed to differences in the fuels rather than variability associated with the data. While BTE should be greater than 1%. Similarly for CO emissions and soot emissions, a difference greater than 8% and 19% respectively, could also be deemed significant. These values were determined by selecting the maximum between COV and the percent difference between original and repeated study, and then rounded.
Regression Analysis

Introduction

When selecting the engine parameters for testing, it is not always known what impact each will have on performance and emissions, because the underlying function is implicit. Obtaining a regression model plays an important role in the data analysis, providing prediction and classification rules, and data analytic tools to understand the importance of different inputs. This may be a difficult task when the underlying function is non-smooth and highly non-linear. This is often the case when performing an experimental emission characterization in IC engines.

For each data point there is a set of variables that might be denoted as inputs (also called predictors or in general independent variables), which are measured or preset. These have some influence on one or more outputs (also called responses or dependent variables). The goal of regression analysis is to use the inputs to predict values of the outputs. The branch of statistics which study how to obtain prediction functions is also called machine learning or data mining.

This section gives a description of the theory behind the algorithm used in this work, based on [6] and [42]. In the statistical literature plenty of methods are described. In order to motivate why a certain approach have been preferred over others a brief description regarding the theory of machine learning is given in this chapter.

Two Simple Approaches to Prediction

Linear models and Least Square  Linear model has been a mainstay of statistics, and still remain the basic tool for many prediction models [43]. Given a vector of inputs $X$ the output $Y$ is
predicted using the following model:

\[ \hat{Y} = \hat{\beta}_0 + \sum_{j=1}^{p} X_m \hat{\beta}_j \]

(4)

The coefficients of equation 4 can be determined using several methods, but by far the most popular is the method of least squares. The idea is to pick the coefficients that minimize the residual sum of squares:

\[ RSS(\beta) = \sum_{i=1}^{n} (y_i - X_i^T \beta)^2 \]

(5)

Since this is a quadratic function the solution to the minimization problem always exist, even if might not be unique. The resulting equation will be characterized by the \( p \) parameters \( \hat{\beta} \). This method will produce smooth equations, but it does appear to rely on the assumption that the degree of the model is appropriate to fit the data. In the language of machine learning it is described as a low variance, and high bias method.

**Nearest-Neighbor Methods** The model function \( \hat{Y} \) is determined by the closest observation it he input space to \( x \). The k-nearest method is defined as follows:

\[ \hat{Y} = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i \]

(6)

Where \( N_k(x) \) is the set of \( k \) closest points to \( x_i \). The output is then the average of the \( k \) closest observations to \( x_i \) in the input space. The parameter to be selected when utilizing this method is the number of neighborhoods \( k \). It is clear that off the shelf least square methods cannot be used to determine \( k \), since we would always pick \( k = 1 \). The k-nearest-neighbor procedures do not appear to rely on any stringent assumption about the underlying data, and can adapt to any situation. The drawback of this approach is that every sub region depends strongly on a handful of input points and their positions, and is thus wiggly, unstable, high variance and low bias.

Each method has its own situation for which it works best. The linear models are suited for those scenarios where the data are affected by large variability; hence a general behavior wants to be detected at the cost of lost in local accuracy. Linear methods are often referred as model based, because the general model selection determines the fitting over the entire domain. The nearest neighbors methods are more appropriate when the information deriving from each training data
point is considered very relevant. Hence a locally accurate fitting is necessary, and that is why this family of mythology is often referred to as local. Large subsets of the most popular techniques are variants of these two simple procedures, and often are a combination of the two.

**Curse of Dimensionality**

Local methods like k-nearest-neighbor seem able to approximate the data optimally, as long as we are able to find a fairly large neighborhood of observations close to every x. It can be proved that with an infinite number of samples the nearest neighbor method will produce the best possible function in term of average squared error. In dealing with high dimensional problems this intuition breaks, as a consequence of the phenomenon commonly refereed to as course of dimensionality [39].

In a p-dimensional problem suppose we are interested in a fraction r of the observations. Assuming that the inputs are uniformly distributed in a p-dimensional unit hypercube, r will also be the fraction of volume captured by our neighborhood. The expected length of the edge we are sampling will be proportional to \( r^{1/p} \). So to cover just a small portion of the data in a high dimensional space we need to span on a very wide range of each input variable, as in Figure 20.

Another manifestation of the course is that the sampling density is proportional to \( N^{1/p} \), where \( N \) is the sample size. Thus in higher dimension the density of the data will be naturally lower. As the complexity of functions of many variables grow exponentially with the dimension, the number of responses necessary to determine those functions accurately grows exponentially too. We have seen that although local methods focus directly on estimating the function at a point, they face problems at high dimensions. It is also possible that local methods are inappropriate in low dimensions cases where a more efficient use of the data can be achieved by structured approaches.

**Tree Based Methods**

These, conceptually simple, family of methods partition the feature space into a set of rectangles, and then fit a model in each subspace. These methods use an approach which is intermediate between model based, and local methods. The simplest method belonging to this family is the recursive binary partitions method. The input space is first divided into two regions, and the
Figure 20: Length of the subcube edge needed to capture a fraction $r$ of the volume of the data, for different dimensions $p$. 
response is modeled by the mean $Y$ in each region. The variable and the split point are selected to achieve the best fit. Then one or both of these regions are split into two or more regions, and so on until some stopping rule is applied. A simplified bi-dimensional case is displayed in Figure (21). In the example the entire domain is divided into five regions, the diagram on the right side of Figure (21) shows the binary selection tree. The upper part of the binary tree is where the entire set is located. Observation satisfying at each junction are assigned to the left branch, and the sub division is arrested when the region satisfy a certain criteria. The partitioning algorithm needs to determine which splitting variable and splitting point to use, hence to determine the shape of the final tree. Starting from the entire data set a splitting variable and a splitting point will define the two regions:

$$R_1(j, s) = \{X|X_j \leq s\} \quad \text{and} \quad R_2(j, s) = \{X|X_j > s\}$$  (7)
Which are determined by solving:

$$\min_{j,s} \left[ \min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \quad (8)$$

Which consist in identifying two areas of the domain where the data are as less inhomogeneous as possible. For any choice of j and s, the inner minimization is solved by:

$$\hat{c}_1 = \text{ave}(y_i|x_i \in R_1(j,s)) \quad \text{and} \quad \hat{c}_2 = \text{ave}(y_i|x_i \in R_2(j,s)) \quad (9)$$

The tree size is determined by a tuning parameter, which represent the maximum number of branches.

**MARS: Multivariate Adaptive Regression Splines**

MARS is a nonparametric method for flexible regression modeling of multidimensional data, this means that the predictor does not take a predetermined form but is constructed according to information derived from the data. An expansion in product spline basis functions is carried over, but differently from other regression methodologies the number of basis functions as well as the parameters associated with each one are related to the data. Mars approximation takes the following form:

$$\hat{f}(x_1, x_2, ..., x_p) = \sum_{m=0}^{M} a_mB_m(x_1, x_2, ..., x_p) \quad (10)$$

Where $B_m(x_1, x_2, ..., x_p)$ are the basis functions. They take the form:

$$B_m(x_1, x_2, ..., x_p) = \prod_{k=1}^{K_m} b_{km}(x_{v(k,m)}|P_{km}) \quad (11)$$

This is the product of elementary functions $b_{km}()$, each one using just one input variable and characterized by a set of parameters $P_{km}$.

The basis functions in the MARS method have the following form:

$$b_{km}(x|s, t) = [s(x - t)]_+ \quad (12)$$

Where the subscript + indicates that only the positive part of the argument is taken, i.e.:

$$(x - t)_+ = \begin{cases} x - t, & \text{if } x > t \\ 0, & \text{if } x \leq t \end{cases} \quad \text{and} \quad (t - x)_+ = \begin{cases} t - x, & \text{if } x < t \\ 0, & \text{if } x \geq t \end{cases} \quad (13)$$
The sample case with $t = 0.5$ is shown in Figure 22. The goal of the algorithm is to produce a minimal set of basis functions $B_m$ for approximating each output function. This is accomplished through a two phase iterative approach. The first phase, also called forward part, generate a super set of basis functions. The data are then over-fitted and the number of basis functions is larger than optimal. The second phase, called backward procedure, selectively deletes basis functions with the goal of producing an acceptable fitting with the minimum number of basis.

**Forward Phase** A collection of basis functions is generated by assigning a reflected pair for each input $X_j$ with knots at each observed value $x_i^j$ of that input. This set of basis functions will be called $C$ and has the following form:

$$C = \{(X_j - t)_+, (t - X_j)_+\} \in \{x_{1j}, x_{2j}, \ldots, x_{N_j}\}, j = 1, 2, \ldots, p$$

Although each basis functions depends only on a single input it will represent a function over the entire domain $\mathbb{R}^p$. The model-building strategy adds one by one a function from the set $C$ and their products. Each term in equation 10 is a function in $C$, or a product of two or more functions.
contained in $C$. The main scope of the forward phase is to select which of those functions will be included in the model.

The selection is performed iteratively. Starting with a constant function $b_m$ and all the set $C$ as candidate as shown in Figure 23. The pair of basis function selected from the set $C$ will be inserted in the model set $M$. At each stage the candidate functions for $M$ are composed by all the products between the reflected functions in $C$ and the function already present in $M$. The term that will be added to $M$ will have the following form:

$$a_{M+1}b_l(X)(X_j - t)_+ + a_{M+2}b_l(X)(t - X_j)_+, \quad b_l \in M$$

(15)

The selection among all the possible outcomes is performed by minimization of the squared error. The coefficients that will multiply the new basis function are selected by least square method. Once the winning products are added to the model the process is continued until the model set $M$ reach the maximum number of terms. Figure 23 is a scheme of the forward model building procedure. On the left are the basis functions currently in the model $M$, at the first iteration this consists of only a constant. On the right are all the candidate basis functions belonging to $C$. These are reflected pairs of linear splines with knots at all observed values $x_{ij}$ of each predictor $X_j$. At each stage we consider all products of a candidate pair with a basis function already in the model. The logic for selecting which candidate function to use is described by the inner loop of the forward phase algorithm, Figure 24. Once a reflected pair is added to $M$ the algorithm must determine which of the basis functions already present in the model must be multiplied by the new term. The outer loop in Figure 24, determines which basis function $b_l(X)$ in equation 15 must be multiplied by the new term in the model. There is one restriction regarding the formation of the model set $M$; each input variable can appear at most once in a product. This is in order to avoid the formation of higher order powers of an input, which would be complex to govern in the forward model building phase. At the end of this process the result is a large model, which typically overfits the data, hence a deletion procedure is needed.

The MARS forward procedure approach can be seen as an evolution of the tree growing algorithm previously discussed. In the tree based methods at each iteration a node was split, a similar result is obtained in MARS multiplying a step function by a pair of reflected step functions. The
Figure 23: Schematic of the MARS forward model-building procedure.

The main improvement achieved by MARS with respect to tree based methods is the ability to capture additive effects. This is due to the fact that a knot location can be used as many times as necessary, while tree methods cannot split a node more than once. A key property of the functions shown in Figure 22 is their local behavior. They have a zero value over a large part of their range and nonzero only over a small part of the feature space. As a result the regression surface can be built up parsimoniously, because adding a basis function to improve the fitting in a certain area does not affect the entire domain. The use of higher order basis functions would produce a nonzero product everywhere, and would not work as well.
Figure 24: Block diagram of the MARS forward model-building procedure.
**Backward phase** The backward stepwise procedure is applied to the final set produced by the forward phase in order to selectively delete individual basis functions whose contribution to the fitting is judged (by a certain criterion) to be negligible.

The criterion used to judge the contribution of each basis function is the Generalized Cross Validation (GCV). This method is used instead of the more common Residual Sum of Squares (RSS) because otherwise the most complex model would always be considered the best. In any case the RSS is included in the GCV valuation, as follow:

\[
GCV = \frac{RSS}{f(N)}
\]

Where \(f(N)\) is a function function which value is proportional to the number of basis functions, and \(RSS\) is defined as follow:

\[
RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2
\]

In order to avoid an excessively noisy fit, i.e. roughness in the response function, an auxiliary term is added to the standard RSS. With this extra parameter the modified RSS function will look as follows:

\[
RSS^* = \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda J(f)
\]

Where \(J(f)\) is a function describing the roughness of the response function, and \(\lambda\) is a roughness penalty parameter which value can vary between 0 and 1. The roughness penalty parameter is left as a tuning degree of freedom to the user depending on the kind of fitting desired; in the following work it has been set to 0.5. At the end of the backward phase the model having the lowest GCV is selected as the final one.

**ANOVA**

The representation of the model given by equation 10 is useful in understanding the construction logic of the algorithm but does not provide very much insight into the nature of the approximation. The objective function can be rearranged into a form that reveals useful information about the predicted relationship between the response \(y\) and the covariates \(x\). This is done by collecting together all basis functions that pertain to the same predictor variable, so that the model can be
recast in the following form:

\[ \hat{f}(x_1, x_2, \ldots, x_p) = a_0 + \sum_{K_m=1} f_i(x_i) + \sum_{K_m=2} f_{ij}(x_i, x_j) + \]
\[ + \sum_{K_m=3} f_{ijk}(x_i, x_j, x_k) + \ldots \] (19)

In equation 19 the first sum includes all the basis functions that involve only a single variable. The second sum does the same thing for the two variables interaction, and so on until the maximum degree of interaction present in the model is represented. The ANOVA decomposition is then performed by grouping each of the particular variables that enter into the model. Depending on the degree of interaction present in the final model the output can be related to one or more variables. For example the first sum in equation 19 can be expressed as:

\[ f_i(x_i) = \sum_{K_m=1} a_m B_m(x_i) \] (20)

Equation 20 is a sum over a single variable basis functions involving only \(x_i\), similarly each bivariate function in the second sum of equation 19 can be expressed as:

\[ f_{ij}(x_i, x_j) = \sum_{K_m=2} a_m B_m(x_i, x_j) \] (21)

Equation 21 is a sum over all two variable basis functions involving the particular pair of variables \(x_i\) and \(x_j\). Adding this to the corresponding univariate functions provides the joint contribution of these variables to the model. Terms involving more variables can be collected together and represented similarly.

Interpretation of the MARS model is greatly facilitate through this ANOVA decomposition. This representation identifies the particular variables that enter into the model, whether they enter purely additively or are involved in interactions. The ANOVA tables presented in this work address the importance of the single variables, or of their interactions, composing the model by reporting the value of the coefficient \(a_m\) appearing in equations like 20 21.

**Case Study**

To determine that the code developed here works properly a case study has been performed. The purpose of this study is to test the ability of MARS algorithm, and in particular the specific code
here produced, to uncover interaction effects that are present in data. The test example is taken from [40]. They considered trying to model the function:

$$F(x) = 10 \cdot \sin(pi \cdot x_1 \cdot x_2) + 20(x_3 - 1/2)^2 + 10 \cdot x_4 + 5 \cdot x_5$$  \hspace{1cm} (22)

In the n=6 dimensional hypercube using N=200 points. The covariates were randomly generated from a uniform distribution and the responses were assigned using the given equation with the addition of a standard error. In this section we consider the same function but with an increase in dimensionality; n=10. Instead of one noise variable, there are now five such variables that are independent of f(x). The following table summarize the ANOVA decomposition obtained by modeling the given equation with MARS.

<table>
<thead>
<tr>
<th>Function</th>
<th>STD</th>
<th>GCV</th>
<th># of basis</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.761</td>
<td>62.542</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>5.512</td>
<td>153.263</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1.514</td>
<td>104.319</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2.739</td>
<td>33.806</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>1.415</td>
<td>3.164</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>5.155</td>
<td>44.224</td>
<td>5</td>
<td>1 2</td>
</tr>
</tbody>
</table>

Table 4: ANOVA Decomposition

The first column lists the function number. The second gives the value of standard deviation of the function. This gives one indication of its relative importance to the overall model and can be interpreted in a manner similar to the regression coefficient in a linear model. The third column gives another indication of the importance of the corresponding ANOVA function, by listing the GCV score for a model with the entire basis functions corresponding to that particular ANOVA function removed. This is used to judge if this function is making an important contribution to the model. The fourth columns gives the number of basis functions comprising the ANOVA function while the last column gives the particular predictor variables associated with the ANOVA function.
From Table 4 we see that the first five functions involve only one variable, and just one involves two variables. Judging from the second and third column of the ANOVA table all the indicated variables are important, and there is no indication in the model to the last five variables (pure noise). With respect to the originating function we see that MARS is able to uncover the presence of an interaction between the \( x_1 \) and \( x_2 \) variables. It also rank the variables in terms of importance, \( x_4 \) has twice a stronger effect than \( x_5 \). Figure 25 shows a graphical representation of the functions that have a stronger impact. The three additive contributions are plotted in the first three frames, while the joint contribution of \( x_1 \) and \( x_2 \) is plotted in the surface plot on the bottom. The graphical representation is a key feature of the MARS analysis because together with the ANOVA analysis allows to uncover the nature of the effect that each independent variable has on the analyzed output. For example we can determine that \( x_4 \) and \( x_5 \) have a linear effect while \( x_3 \) is parabolic.

![Figure 25: Single variable plot and interaction surface](image)

Comparing the results of the MARS fit to these data with the true underlying function, shows that the resulting model provides a fairly accurate and interpretable description. This is especially noteworthy given the high dimensionality (n=10) and the small sample size (N=100).
Preliminary Results

A preliminary version of the code has been developed using data that was generated on the FACE 5 fuel, and the experimental study was focused on injection strategy. The following plots (Figures 26, 27, 28, 29, 30) are produced to demonstrate the capability of this method to generate flexible functions that are capable to follow the trend of the experimental values even when the behavior is far from monotone. This result is a direct consequence of the nature of the functions used for approximation; unlike linear regression the objective function is composed of Heaviside functions rather than polynomial functions. A source of flexibility for this model is the possibility to perform tuning on the objective function. That is, a penalty parameter to avoid excessive roughness in the objective functions, and the maximum number of basis functions used for regression are set before performing the analysis.

Figure 26 shows the trend of NOx concentrations versus two injection characteristics at the time. The plots are organized in such a way that those in the same row are characterized by the same variables. The x and y axes represent two set of injection characteristics, with the third one used as a parameter to which are assigned three equally spaced values. Both the regression surfaces and the experimental points are displayed in order to visually evaluate the fitting capability of the MARS model. In all the panels composing Figure 26 the regression surfaces fit the experimental values except for the subplot (H), where the predicted value at 40 % fuel split and 40 Pilot SOI is not consistent with the experimental correspondent. The local lack of fitting (LOF) is a consequence of MARS being a compromise between a model based, and local methods. The model based algorithm are more oriented in finding a regression function which globally reproduce the trend of the data, consequently paying a price in terms of local fitting.
Surface plots reveal how retarding the main SOI is beneficial in terms of NOx reduction, due to the lower temperature reached in the combustion chamber. Varying the pilot injection we observe a less significant (compared to Main SOI) but more complex effect on NOx emissions. The key for analyzing this parameter is to look at the relative position of pilot and main injection. When the two are close the burning pilot fuel quantity will overlap the main injection event, leading to higher temperatures and consequently higher NOx emissions. A substantial separation between the pilot and main injection event may quench the pilot quantity ignition. Minami et al. [13] explained this phenomenon comparing the effect of pilot injection to an internal EGR, especially when the quantity of fuel injected during the pilot event is higher than 10% of the total fuel. The burned gas produced from the combustion of the pilot injection will dilute the concentration of oxygen inside the combustion chamber for the main injection, acting basically like an EGR system. The fuel split among pilot and main injection works as a scale effect for the considerations that have just been made, this can be easily observed by looking at the different planes in Figure 26.
Figure 27 shows the regression surface in regards to the PM emissions. The complex phenomena that bring to PM production and oxidation during combustion make the correlation with the injection strategy not as predictable as other emission characteristics, making the regression of this emission characteristic is the most challenging for the model.

Figure 27: PM emission from FACE 5 fuel as a function of pilot and main SOI on the first row, fuel split and main SOI on the second, fuel split and pilot SOI on the last row

The regression surfaces show the advantage in using pilot injection, which allows to retard the main injection further than it would be possible with a single injection event. This is related to the reduction in ignition delay consequent to a higher temperature in the combustion chamber at the moment of the main injection. An important parameter in this sense is the time interval between the pilot and main injection event. When this time interval is too long the products of pilot injection mix sufficiently with the ambient gas in the cylinder, consequently no difference is produced with respect to the single injection case. On the other hand, when this interval is shorter a local high temperature zone is created near the fuel nozzle with a consequent shorter ignition delay [13]. It is still important to maintain a sufficient separation between pilot and main injection to ensure the main injection occurs after the injection delay of the pilot event.
Looking at the plots in Figure 27, especially at those where pilot SOI is varied (first and last row), it can be noticed that the highest values of soot concentration are produced when the pilot injection is started at 30° BTDC. One way to interpret this result is by examining the mechanism of soot formation by fuel rich regions of burning diesel jets. Having the pilot injection close to the main injection event bring part of the fuel injected to be still unburned when the main injection starts. The fuel-rich pockets which do not have time to mix and burn prior to exhaust valve opening are among the main cause of soot generation. Another key parameter to consider, strongly related to injection timing, is the temperature evolution during combustion. While high temperatures at the end of the combustion enhance the oxidation of soot, having high temperatures at the time of injection reduce air entrainment and increase soot formation.

The importance of the pilot injection in PM production is further underlined by considering the fuel split; this is indeed a parameter that represents the intensity of the pilot injection. Looking at the second and third rows in Figure 27 we notice that larger amount of fuel delivered during the pilot injection brings to higher PM production. This can be related to an increase in temperature and decrease of oxygen concentration before the main injection event starts. These conditions usually lead to higher PM production. Generally an increase in temperature inside the combustion chamber leads to a reduction in lift-off length, which is well known to increase soot emissions [48].

Figures 29 and 28 show the behavior of HC and CO represented in the same manner as for the previous plots. These two characteristics show a similar trend with respect to injection timing and in both cases low temperature combustion represents a penalty. The low cylinder temperature is useful in reducing NOx and soot production, but, since it involves slow oxidation reactions that cause misfiring and quenching, will result in higher HC and CO emissions compared with conventional diesel combustion.
Figure 28: CO emission from FACE 5 fuel as a function of pilot and main SOI on the first row, fuel split and main SOI on the second, fuel split and pilot SOI on the last row.

Figure 29: HC emission from FACE 5 fuel as a function of pilot and main SOI on the first row, fuel split and main SOI on the second, fuel split and pilot SOI on the last row.
Figure 29 and 28 show the negative effect of advancing Pilot injection in terms of CO and HC production. The early fuel injection causes the spray interaction with the cylinder liners or piston walls; the result is wall quenching with resulting fuel problems and high concentrations of HC in the exhaust. Furthermore low temperature combustion is a major cause for high CO emissions.

The low temperature of the exhaust is also to be considered as an issue for the after-treatment of HC and CO. In general the factors that bring to higher production of CO and HC can be listed as follows [5]:

- The long spray penetrations which characterize LTC operation bring fuel to accumulate on the combustion chamber surfaces.
- Fuel lean zones inside the combustion chamber are more common in LTC operations rather than in conventional diesel. Fuel in these areas is more prone to escape combustion leading to the production of unburned HC.
- Since the local peak temperatures are lower than conventional diesel, the fuel near the wall does not burn.

The plots in Figure 30 show the trend of Brake Thermal Efficiency (BTE) with respect to the emission characteristics. The points analyzed in this study have a range of efficiency in the interval 29% – 31%, which makes the regression on this characteristic challenging to perform.

The pattern obtained using the regression surfaces regarding the BTE resemble the one obtained for NOx. This is physically meaningful considering that both NOx production and thermal efficiency increase with temperature, hence will reach higher value in correspondence of an injection strategy that guarantees high temperature.

When analyzing the best trade off in terms of emissions and performance it must be considered what is the threshold value to associate at each emission characteristics. Under this point of view the most critical characteristics among those that have been presented is the soot concentration. In this sense it is advantageous to inject a small quantity of fuel (30% of the total) at 40° BTDC as depicted from Figure 26. Advancing the start of the pilot injection event is beneficial in terms of PM but will produce an increase in HC and CO production for the reasons previously discussed.
Figure 30: BTE from FACE 5 fuel as a function of pilot and main SOI on the first row, fuel split and main SOI on the second, fuel split and pilot SOI on the last row.

Those characteristics still fall in the acceptable range. Regarding the start of the main injection event two choices are possible, both allowing for acceptable range of emission. It can be located at 4° BTDC or -2° BTDC, in the first case the engine shows better efficiency and lower PM emissions, while the latter case is more performing in terms of NOx reduction.

**Experiment Reduction**

In this section the potential use of MARS algorithm for optimizing the number of test necessary to characterize a given phenomenon is analyzed. The main idea is to use the regression capability of MARS within the classical Taguchi methodology.

**Taguchi** Standard Taguchi method applied to engine testing contemplates the selection of the levels of variation for the engine parameters. Three levels are the least number that can be used in order to study nonlinear responses of the parameter effects. Then a so called orthogonal matrix of tests is determined.
The orthogonal matrix testing makes use of an orthogonal array design to isolate the main effects of a single parameter on the response function. Orthogonality is intended in a combinatoric sense: for any pair of columns- in the array- all combinations of factors occur an equal number of times [50]. It is based on the assumption of no interactions between parameters, which in other terms implies that models used to fit the data collected do not include cross product terms between the parameters. However, the validity of this assumption needs to be proved, and this was done by comparing the model predictions with the measured points. The first step to estimate the factor effect is to calculate the overall means, which in the case of \( NO_x \) for a three level and three factors is:

\[
NO_{x_{\text{exp}}} = \frac{1}{9} \sum_{i=1}^{9} NO_{x_i}
\]  

(23)

The factor level response for each level is calculated using analysis of means (ANOM) as follows:

\[
NO_{x_{\text{MainSOI}}} = \frac{1}{3} \sum_{i=1}^{3} NO_{x_i}; NO_{x_{\text{PilotSOI}}} = \frac{1}{3} \sum_{i=4}^{6} NO_{x_i}; NO_{x_{\text{MainSOI}}} = \frac{1}{3} \sum_{i=1}^{3} NO_{x_i}
\]

(24)

Then the actual effect of the factor level is determined by its deviation from the overall mean, for instance:

\[
\Delta NO_{x_{\text{MainSOI}}} = NO_{x_{\text{MainSOI}}} - NO_{x_{\text{exp}}}
\]

(25)

The predictive model is derived for the data collected according to the orthogonal array design. This model consists of the simple sum of the individual factor effects and does not include cross-terms. The empirical model produced by this technique is compared to the results of a validation test. The verification test is performed to confirm the reliability of the prediction model. If the validation test confirms the model prediction then the set of engine tests selected is descriptive of the engine behavior.

**MARS** The approach suggested in this work makes use of the MARS algorithm to produce the predictive model. In this case the assumption of no interactions between the input parameters can be eliminated. The flexibility of the regression algorithm allows for use of scattered data for determining the prediction models. In selecting the test the entire range of variation of each parameter should be used, and like for the Taguchi method, at least three levels for each parameter...
should be considered. The approach suggested in this section is displayed by using an example based on the data presented in this work. The control parameters that we are considering are the three injection parameters (Main SOI, Pilot SOI, and Fuel Split). A subset of data points is considered, those are distributed according to a central composite design. This subset includes just half of the total data available, distributed according to the prescriptions described above. Figure 31 shows the distribution of the tests in a three dimensional space where each of the axis represent one of the input parameters, and the intensity of the output is described by the dimension and color intensity of the spheres.

Using this smaller data set an emulator model is generated using MARS to predict the $NO_x$ emissions. At this point the equation obtained is interrogated in all the points corresponding to the original dataset. The results are compared with the experimental values in the original complete
dataset. Figure 32 shows the comparison result.

![Figure 32: NOx comparison between MARS results and experimental values](image)

The difference between the emulator model and the actual values can be quantified by the residual sum of squares, which in this case is $R^2 = 0.93$. The same procedure is applied to soot concentration, CO, HC, and BTE. The results are summarized in table 5, while the rest of the describing plots are collected in Appendix. Table 5 also presents the results obtained using the classical Taguchi methodology.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>$R^2$ MARS</th>
<th>$R^2$ Taguchi</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOx</td>
<td>0.93</td>
<td>0.86</td>
</tr>
<tr>
<td>PM</td>
<td>0.69</td>
<td>0.29</td>
</tr>
<tr>
<td>CO</td>
<td>0.92</td>
<td>0.67</td>
</tr>
<tr>
<td>HC</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>BTE</td>
<td>0.73</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Table 5: Residual sum of squares for each characteristic predicted

The results in table 5 show a good prediction quality for NOx, CO, and HC. It is weaker for PM, and BTE which are also the characteristics with lower regression quality. This analysis proves
that with half of the tests a similar characterization of the phenomenon would have been possible.
Results

To achieve LTC the most important requirement is to generate homogeneous mixture formation prior to spontaneous ignition. This requires a longer ignition delay, which can be achieved or by proper selection of injection strategies or fuel selection strategy. The boiling ranges of the fuel and its volatility will strongly affect the outcome of a certain injection strategy, making the fuel characteristics a key parameter in the combustion process. Also the use of a split injection strategy is very important, because it will reduce the PRR and hence limit the soot production. The split injection strategy has to be developed accordingly to the CN category of the fuel under consideration.

The fuel injection parameters (namely main SOI timing and pilot SOI timing) comprising the split injection test matrices were then applied according to CN category. In the shown data, 50 percent mass fraction burned (CA50) was not held constant and varied among all fuels tested. A more advanced range of main SOI timing for the low CN fuels (30) (when compared to medium (40) and high (50) CN fuels) was established to limit misfire at retarded main SOI timing resulting from the longer ignition delay provided by these fuels. The shorter ignition delay exhibited by the medium and higher CN fuels required that the range of main SOI timing be retarded (from that of low CN fuels). Pilot SOI timings were also adjusted for each CN category to limit heat release before the main injection event; low CN fuels tolerated a more advanced range of pilot SOI timing, while the range of pilot SOI timing for medium and high CN fuels had to be retarded.

The approach used in this study to investigate these correlations is the one described in the Regression Analysis chapter. The performance parameters are then analyzed one at the time, in order to show how the variables of interests combine. The physical meaning of each performance
parameters can be very different from another, hence only a separate study can lead to an interpretability of the results. In a second stage a strategy to combine all the underlined outputs will be defined, in order to define an optimal set of variables for the multivariate optimization.

Independently from the parameters analyzed a preliminary manual screening of the training data is necessary in order to not spoil the model with unrealistic data. The experimental points where improper combustion was identified will not be used as training data. Only those tests where reliable and consistent data are produced will be fed in the model. Both the independent and dependent variables are adimensionalized based on the maximum and minimum value obtained in the entire range. The adimensionalization has to be carried out in order to allow model generation, and to simplify the optimization process.

The rest of this section will be organized as follows; one performance parameter at the time will be analyzed. For each characteristic a report on the quality of the regression will be provided, both in graphical and quantitative way, in order to evaluate how much we can relay on the regression results. Then an ANOVA analysis will be performed over the regressed equations, in order scale the input variables based on their effect on the specific characteristic under consideration, and to determine the key interactions between the given parameters. At this point a graphical evidence of those interactions will be produced; the response surface will be used to understand the correlation between the variables used to perform the study and the output under analysis.

After analyzing the single output one at the time, a global study will be performed. The objective is to perform a multivariate optimization study in order to determine the combination of engine and fuel parameters able to produce the best trade-off relatively to the parameters analyzed (namely NOx, PM, CO, HC, BTE). Before combining all the engine output analyzed it is necessary to determine which set of parameters should compose the computation domain. The input parameters are screened in order to make sure the combustion produced falls within the scope of this work. Hence all those combinations of fuel characteristics and injection timings that produce either misfire or a combustion clearly not definable as low temperature. To perform this kind of analysis a modification of the core of the MARS algorithm is needed. The regression in this case must be operated over categorical variables instead of numerical, to represent the nature of the problem.
where the combustion is either acceptable or not. The methodology leading to this study is reported in the optimization section.

\[ NO_x \]

Following the procedure described above the first characteristic to be analyzed is the NOx. Figure 33 compares the experimental points and the regression results obtained using the MARS algorithm. The plot wants to show the accuracy of the model to fit the training data. It does not lead to any physical interpretation, and the division in areas corresponding to different fuel characteristics is only produced to motivate the strong discontinuities in the data.

![Figure 33: Comparison between experimental data and regression results](image)

Once that it is established an acceptable level in the regression performance the equations obtained (in appendix A the explicit form is reported) are studied. The analysis method is based on ANOVA analysis, and the results are shown in table 6. The final equation is composed of 9 basis functions. These can be organized in the following table to underline the importance of the single variable on the final outcome.
The columns represent summary quantities for each one. The first column lists the function number. The second gives the standard deviation of the function. This gives one indication of its (relative) importance to the overall model and can be interpreted in a manner similar to a standardized regression coefficient in a linear model. The third column gives the associated value of GCV while the fourth column provides the number of basis functions comprising the ANOVA function. The last column gives the particular predictor variables associated with the ANOVA function (1: Main SOI, 2: Pilot SOI, 3: Fuel split, 4: Cetane number, 5: Aromatic content, 6: Distillation temperature).

The ANOVA table shows that the interaction effects among the independent variables have a stronger effect on the output than the same variables alone. These values are obtained analyzing the regression equation using the methodology described in the ANOVA section of the regression methods chapter. In particular entry 4, 5, and 8 indicate an important effect given by the joined contribution of fuel characteristics and engine parameters. The following plots give a graphical representation of the results presented in table 6.
Figure 34 represents the effect of Cetane number on NOx emissions, corresponding to entry 1 in Table 6. The plot includes three curves corresponding to three different values of Main SOI, this is done to account for the interaction effect between those variables depicted by the ANOVA analysis.

![Dependence of NOx production on CN number](image)

**Figure 34: Dependence of NOx production on Cetane Number**

From Figure 34 we see how there is a general increase in NOx production with CN. In order to understand the sharp change in slope of the regression plots in correspondence with high CN the characteristics of the regression algorithm should be considered. As a global method MARS tries to determine a general behavior of the objective function inside the domain. Where this global trend does not fit the data an extra basis function is added in order to reduce the lack of fit. In this case the extra basis function is activated at high CN to make up for the lack of fit that would be produced by a monotone function.

When analyzing the plots from figure 34 it must be underlined that dependently on the Cetane number a different injection strategy musty be selected, making the effect of these two characteristics possibly confounded. Low CN fuels require a more advanced range of main SOI timing, in order to limit the misfire resulting from a longer ignition delay. The shorter ignition delay exhibited by the medium and higher CN fuels required that the range of main SOI timing be retarded. Accordingly
to the main SOI also the pilot injection was adjusted, in order to avoid quenching phenomena between the two events. With respect to these considerations the results shown in Figure 34 gain even more importance. From what we have seen in Figure NOx, advancing injection does not have beneficial effect on NOx reduction. Since lower CN fuels show lower NOx production despite of a more advanced range of SOI timing, we can infer that low CN is beneficial for NOx production. This can be physically motivated assuming that, at equal conditions, higher ignition delays lead to lower combustion temperature, and consequently less NOx.

![Dependence of NOx production on Distillation Temperature](image)

**Figure 35: Dependence of NOx production on Distillation Temperature**

Figure 35 represents the NOx production as a function of distillation temperature. The plot shows clearly an increase in NOx production proportional to T90. A possible explication for this phenomenon is that higher distillation temperature involves a higher volatility of the fuels, which hence will lead to higher temperature during the combustion.

The third single variable with the strongest influence on NOx emission is the main SOI which has been broadly discussed in the chapter relative to the injection parameters.

In terms of interaction the ANOVA table suggests that the first two variables to variables to analyze are CN and main SOI. Figure 36 shows the surface plot relative to the interaction
between these two parameters in terms of NOx production. The regression surface obtained through regression is compared with the experimental data represented by the scatter points. The blue points are relative to the same set of parameters which characterize the plotted surface; the green points instead are relative to a different set of parameters. These extra points have been added in order to justify the trend of the regression surface.

The graphical representation of the interaction between CN and Main SOI shows that the second variable has a stronger impact on NOx emissions. Except for a peak corresponding to the most advanced Main SOI timing with the lowest CN fuel the trend of NOx emissions is rather monotone, showing a quadratic increase proportional to the distance from low CN and Main SOI. The surface in Figure 36 represents the regression equation obtained through MARS, where the variables that are not included in the plot assume a constant value. This value has been selected as the mean of the independent variable over its range of variation, and then these parameters are varied one at the time to check that they do not influence the morphology of the regression surface. In this case there is no relevant change in the shape of the surface but there is a change in the NOx scale. For this reason the surfaces corresponding to other set of parameters are not presented in this report, and

Figure 36: Dependence of NOx production on CN and Main SOI
the same presenting strategy is adopted for those circumstances were the parameterized variables do not influence the shape of the surface.

Figure 37: Dependence of NOx production on T90 and Main SOI

Figure 37 shows the contribution of distillation temperature and main SOI to NOx emissions. The blue scatter point represent the experimental data which share the same set of parameters used to produce the surface, green and red points are instead data points which characteristics are similar to the one in the equation but not equal. The surface obtained shows once again the dominant effect of Main SOI; the effect of distillation temperature is a light increase in NOx production.

The next two level interactions all refer to injection strategy parameters, which results are redundant with those shown in the preliminary results. The most important three level interactions according to the ANOVA table 6 is the one correlating the two fuel characteristics CN, and T90 with the main SOI timing. Figure 38 shows a three dimensional plot of the $NO_x$ emissions, where the three axis represent the independent variables, namely T90, Main SOI, and Cetane Number. The intensity of $NO_x$ production is represented by the color scale, and in order to make it visible inside the domain, it is sliced over five different planes judged to be representative of the overall behavior.
The results presented in Figure 38 show that among the variable presented Main SOI is the strongest, and can only slightly be tempered by fuel characteristics as Cetane Number and distillation temperature. This plot also shows how an increase in Cetane Number leads to higher $NO_x$ production. This statement is in contradiction with the conclusions drawn by recent studies [18]. The effect of CN on emissions has been misinterpreted as a consequence of the confounding effect of Main SOI. As we explained before the use of higher CN fuels imply the necessity of using more advanced injection timings. When the raw data are compared the effect of Main SOI can mask the other factors influence. The second strongest interaction depicted by the ANOVA table is between CN, main SOI and Pilot SOI. This three variable joint interaction is shown in Figure 39.
The response displayed in Figure 39 is the same of Figure 38 except for the pilot SOI instead of distillation temperature. The influence of pilot SOI on \( NO_x \) production is similar to the one of main SOI. The other interactions are either not very significant or already displayed in the section relative to injection characteristics, the plots are reported in appendix.

**Soot Concentration**

In this section the results relative to PM analysis will be displayed. Following the same structure of the previous section the first step is to analyze the quality of the regression. Figure 40 shows the experimental results compared with the emulator model. The first thing to notice is the high variability in the data, the analysis of the training data showed a standard deviation over 18%,
which clearly represent a challenge for the regression algorithm. The $R^2$ value obtained is 75% which is still in the range of acceptability.

![Figure 40: Comparison between experimental data and regression results](image)

Figure 40 shows how widely the soot emissions vary among the fuels tested, with FACE 8 and FACE 6 being the ones producing by far more soot. The emulator model is then analyzed using the ANOVA procedure; the results of such analysis are summarized in the following table 7.
The two single variables with the strongest impact on Soot emissions are Cetane number and main SOI as expect. Figure 41 and figure 42 show the contribution of those variables to the total PM emitted; in both cases three different trends have been produced.
The soot emissions increase with Cetane number, as depicted from figure 41. A noticeable
interaction appears at high Cetane number with the Main SOI. Figure 42 shows the dependence of soot production from main SOI, it shows that the lowest values of PM are achieved at the most advanced injection timing. The combination low Cetane number and advanced injection produce the minimal soot value. When analyzing join interaction between variables the one to show the strongest effect is between CN and T90. Several studies confirmed the observation that T90 plays a significant role in soot formation([44] [45] [46]); its increase leads to higher soot emissions. Figure43 shows the interaction between T90 and Main SOI.

![Figure 43: Dependence of soot production on distillation temperature and main SOI](image)

The second strongest interaction is between Cetane number and aromatic content. Figure 44 shows that the effect of CN is very marked, and it also strongly affects the way other parameters act on soot formation.
Figure 44: Dependence of soot production on Cetane number and aromatic content

The following two figures (45 46) show the behavior with respect to pilot SOI and aromatic content. A distinction has been made between results produced setting a high Cetane number (50) and low Cetane number (30). This is done because, as it is noticeable from figures 45 46, the interactions between those variables change significantly with different levels of CN. In both cases an advanced pilot injection is beneficial, while in general higher aromatic content lead to higher PM emissions. The influence of the aromatic content is stronger and less linear in the low CN number, and becomes more important for retarded pilot injections.
Figure 45: Dependence of soot production on aromatic content and pilot SOI at low Cetane number (CN=30)

Figure 46: Dependence of soot production on aromatic content and pilot SOI at high Cetane number (CN=50)
The figures shown in this section are the ones where more significant aspects of PM production are underlined, the rest of the plots are collected in the appendix.

**Carbon Oxide**

The quality of the fitting regarding the CO emissions is displayed in figure 47. The $R^2$ value obtained is of 71%, but as we can see from figure 47 the regression is optimal everywhere except for FACE 5 and FACE 6 fuels. Those fuels are characterized by a combination of high Cetane number and low aromatic content which generates a disagreement between the model results and the experimental data. Another source of uncertainty is non repeatability of the experimental data as depicted in section.

![Figure 47: Comparison between experimental data and regression results](image)

Table 8 summarize the results of ANOVA analysis regarding the CO emulator model.
<table>
<thead>
<tr>
<th>Function</th>
<th>STD</th>
<th>GCV</th>
<th># of basis</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.3</td>
<td>54.4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>5.8</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>2.6</td>
<td>2.6</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1.3</td>
<td>20.9</td>
<td>1</td>
<td>1 4</td>
</tr>
<tr>
<td>5</td>
<td>6.2</td>
<td>24.5</td>
<td>2</td>
<td>4 5</td>
</tr>
<tr>
<td>6</td>
<td>5.2</td>
<td>36.7</td>
<td>1</td>
<td>4 6</td>
</tr>
<tr>
<td>7</td>
<td>1.1</td>
<td>18.2</td>
<td>1</td>
<td>1 4 5</td>
</tr>
<tr>
<td>8</td>
<td>0.6</td>
<td>14.6</td>
<td>1</td>
<td>1 5 6</td>
</tr>
<tr>
<td>9</td>
<td>5.2</td>
<td>39.7</td>
<td>1</td>
<td>4 5 6</td>
</tr>
</tbody>
</table>

Table 8: ANOVA Decomposition CO

The fuel characteristic with the strongest impact on CO emissions appears to be CN. The following plot (figure 48) shows the response in terms of CO to different levels of CN. Lower CN lead to higher CO production, confirming the observations of other studies [3].

Figure 48: Dependence of CO from Cetane number
The other single variables to have an impact on CO emissions are all injection characteristics which have already been covered in the section. The two variable with the most relevant interaction are CN and main SOI, the surface plot is displayed in Figure 49.

Figure 49: Dependence of CO on Cetane number and main SOI

The same interaction including also pilot injection is visualized in figure 50. It is interesting to notice that pilot SOI has an impact which is stronger than main SOI.
Figure 50: Dependence of CO from Cetane number, pilot SOI, and main SOI

Hydrocarbon

This section describes the modeling of hydrocarbons. The regression quality for this variable is very accurate as depicted from figure 51, and also by a $R^2$ value of 90%.
Figure 51: Comparison between experimental data and regression results

Table 9: ANOVA Decomposition HC

<table>
<thead>
<tr>
<th>Function</th>
<th>STD</th>
<th>GCV</th>
<th># of basis</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.4</td>
<td>74.3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
<td>28.5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>5.6</td>
<td>12.6</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1.3</td>
<td>2.9</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>0.7</td>
<td>1.5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>4.2</td>
<td>36.7</td>
<td>1</td>
<td>1 4</td>
</tr>
<tr>
<td>7</td>
<td>1.1</td>
<td>18.2</td>
<td>1</td>
<td>2 4</td>
</tr>
<tr>
<td>8</td>
<td>5.6</td>
<td>14.6</td>
<td>2</td>
<td>2 3 4</td>
</tr>
<tr>
<td>9</td>
<td>4.5</td>
<td>39.7</td>
<td>1</td>
<td>1 4 5</td>
</tr>
</tbody>
</table>

The first characteristic listed in table 9 is the Cetane number, which is displayed in figure 52 for three different values. The plot shows a peak in HC at low CN numbers, which rapidly decreases.
The curves in figure 52 are representative of three values of main SOI, and it shows the impact of this factor on the response is nonlinear. According to the ANOVA analysis the second most influential parameter is the aromatic content. From figure 53 we notice the negative impact of high AC on HC. Considering that AC represent the specific amount of energy in the fuel and that HC are leftover of unburned fuel, this behavior can be considered to be as an incapacity of the combustion to utilize all the energy available in the fuel.
Figure 53: Dependence of HC on aromatic content

The deduction of a nonlinear effect of the main SOI is confirmed by plotting the HC response with respect to this factor (figure 54).

Figure 54: Dependence of HC on main SOI
It is now interesting to look at the join interaction between those two factors. Figure 55 shows how higher CN leads to lower emissions of HC and also to a less significant impact of the main SOI.

Figure 55: Dependence of HC on main SOI, and cetane number

From a three way prospective the most interesting results are obtained looking at the interaction between main SOI, distillation temperature and aromatic content as depicted in figure 56.
Verification Study

The model built to analyze the full interaction between fuel characteristics and injection strategy has been analyzed in comparison with the experimental data used to train the model itself. A rigorous validation of the model would require the comparison with experimental points not included in the training set. This kind of validation is aimed to judge the prediction capability of the model. Since the model is generated through regression of the data, it is necessary to determine if the behavior of the emission characteristics can be extrapolated or it is completely unrelated to the data surrounding it.
In this section the validation of the model is performed using an unconventional approach. Instead of building the model using less than all the data available, and then using the remaining for validation, here the model presented was trained using all the information available. A mockup model is then generated just for validation purposes excluding some of the available data, and in this section the prediction capability of this secondary model is investigated. The justification for this approach is related to the pursuit of the most refined model to describe the LTC process, while the exclusion of some of the data from the training process would have led to a lost in accuracy.

The mockup model was generated using all the data except for seven points that will be used for validation. The validation subset is determined by randomly selecting seven points belonging to the original dataset. The only constrain applied to the data selection is that each point should pertain to a different fuel in order to guarantee the generality of the validation, and to make sure that the regression is effective on each fuel. Table 10 lists the points selected for validation, and their respective factor levels, in the rest of this section each of this points will be referred to using the fuel name.

<table>
<thead>
<tr>
<th>CN</th>
<th>AC</th>
<th>T90</th>
<th>Main SOI</th>
<th>Pilot SOI</th>
<th>Fuel Split</th>
<th>Fuel name</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.93</td>
<td>26.1</td>
<td>269</td>
<td>8</td>
<td>50</td>
<td>40</td>
<td>FACE 1</td>
</tr>
<tr>
<td>32.02</td>
<td>50</td>
<td>270</td>
<td>6</td>
<td>50</td>
<td>40</td>
<td>FACE 3</td>
</tr>
<tr>
<td>28.44</td>
<td>40.7</td>
<td>337</td>
<td>4</td>
<td>45</td>
<td>35</td>
<td>FACE 4</td>
</tr>
<tr>
<td>54.2</td>
<td>22.2</td>
<td>279</td>
<td>-4</td>
<td>30</td>
<td>30</td>
<td>FACE 5</td>
</tr>
<tr>
<td>53.3</td>
<td>21.1</td>
<td>341</td>
<td>-4</td>
<td>30</td>
<td>40</td>
<td>FACE 6</td>
</tr>
<tr>
<td>50</td>
<td>43.5</td>
<td>342</td>
<td>0</td>
<td>30</td>
<td>30</td>
<td>FACE 8</td>
</tr>
<tr>
<td>44.95</td>
<td>37</td>
<td>321</td>
<td>2</td>
<td>235</td>
<td>40</td>
<td>ULSD</td>
</tr>
</tbody>
</table>

Table 10: Data points used for calibration

Figure 57 shows the comparison between the experimental points selected for the verification, and the values obtained using the mockup model. The response in figure 57 is NOx, and the plot clearly shows a good fitting between the predicted values and the experimental points.
Figure 57: Comparison between NOx experimental values and predicted results

Figure 58 displays the prediction accuracy for the other four responses.
Figure 58: Comparison between experimental values and predicted results for CO, HC, PM, and BTE

Table 11 summarize the prediction performance with reference to the results displayed above. For each characteristic the maximum difference between the predicted value and the experimental value is displayed. Max err. % is the ratio between the error and the range of variation of the response. \( R^2 \) is conventionally defined as the ratio between the residual sum of squares and the total sum of square. These two parameters are usually proportional one to the other. The last column collects the values determined during the repeatability study.

<table>
<thead>
<tr>
<th>Response</th>
<th>Max err</th>
<th>Max err. %</th>
<th>( R^2 )</th>
<th>Repeatability %</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOx (ppm)</td>
<td>4.8</td>
<td>7.6</td>
<td>0.91</td>
<td>3</td>
</tr>
<tr>
<td>PM (mg/m3)</td>
<td>23.3</td>
<td>20.6</td>
<td>0.81</td>
<td>19</td>
</tr>
<tr>
<td>BTE (%)</td>
<td>0.69</td>
<td>11.9</td>
<td>0.91</td>
<td>1</td>
</tr>
<tr>
<td>CO (ppm)</td>
<td>666.3</td>
<td>19.2</td>
<td>0.83</td>
<td>8</td>
</tr>
<tr>
<td>HC (ppm)</td>
<td>225.12</td>
<td>6.1</td>
<td>0.98</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 11: Prediction model errors
Comparing the results of this verification study with the threshold values defined during the repeatability study, we notice that there is a strong correlation between the quality of the model and the repeatability of the data. The repeatability indicates a variation in the output which is not controllable in the experimental apparatus used to collect the data. The regression model is generated starting from these same data; hence its prediction capability could never overcome the data quality.

**Optimization**

Once the emulator model is obtained by surface fit, it can be used to predict responses inside the factor space. It should be noted that the regression model should not be used for extrapolation outside the factor range. Most importantly, the model can be used to conduct optimization to search for the optima located on the response surface. In the optimization, the stationary point refers to the point of factor settings corresponding to zero partial derivatives of the response with respect to all the factors.

**Multi-Objective Optimization**  Engine optimization is a multi-objective problem (MOP), as several objective functions are of interest. Hence a multi-objective solution always represents a trade-off in the MOP. In this family of problem the notion of optimum is referred as the Pareto optimum [47].

A visual interpretation of the Pareto optima can be given by considering just two variables, so to be able to plot it on a plane. The data collected are plotted in Pareto chart (figure 59), the two axes represent the quantity to minimize, i.e. NOx and Soot concentration. This representation of the output is useful to have a rapid perception of the optima points, those will be the closest to the origin.
Figure 59: Distribution of the experimental points on a PM-NOx Pareto chart

**Weighted sum method**  A traditional method for multi-objective optimization is the weighted sum method, which seeks the Pareto optimal solution by combining several objective functions into one. We can formulate a general MOP in the following form:

\[
\begin{align*}
\min_Y (x, \beta) \\
g(x, \beta) < 0
\end{align*}
\]

(26)

Where \( Y = [y_1, ..., y_z]^T \) is an objective function vector, \( x \) is the factors vector, \( \beta \) is the regression
coefficients vector, and g is an inequality constraint vector. The weighted sum method consists in reducing the objective function vector to a scalar of the form:

$$\hat{Y} = \sum_{i=1}^{z} \frac{\lambda_i}{s_i f_i y_i}$$  \hspace{1cm} (27)

Where $s$, $f_i$, and $\lambda_i$ are the scale factors and the of the i-th objective, respectively. Typically, weights are chosen in such a way that their norm is one, and that none is negative.

The main drawback of this relatively simple method is that an even distribution of weights among objective functions does not always result in an even distribution of solutions on the Pareto front. This issue is particularly relevant in the problem studied in this work since the objective functions are expressed in different unit. In order to overcome this problem the weights have been chosen proportionally to the magnitude of the response for the given objective function, i.e.:

$$\lambda_i = \frac{\sum_{j=1}^{z} \max(y_j) - \max(y_i)}{\sum_{j=1}^{z} \max(y_j)}$$  \hspace{1cm} (28)

With this weight selection the two objective functions assume equal importance in $\hat{Y}$.

The parameter $s$ instead is chosen based on the results obtained from the compatibility study. A categorical function is associated with every set of independent variables in order to account for improper combustion. Its value is 1 in case of proper combustion and 0 when the combustion process is not achievable. By inserting this value at the denominator of the cumulative objective function it is guaranteed that the minimum will correspond to an allowable combustion process.

The weight factor $f_i$ is used to determine the relative importance of each emulator function in the global model. For each output the factor $f_i$ is determined by evaluating the density of each response above a certain threshold. For example the $f_i$ relative to the NOx response corresponds to the percentage of experimental data above the value of 50 ppm, which is determined to be the threshold for acceptable NOx emissions. This parameter is then used to express how critical is each response to the overall optimization, because those characteristics with a larger distribution of points above the allowable level will be considered more important.
Once that we defined a scalar function that accounts for both objective the optimum can be determined using standard multivariate analysis. The global optima is determined by the following factors:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cetane Number</td>
<td>30</td>
</tr>
<tr>
<td>Aromatic Content</td>
<td>26</td>
</tr>
<tr>
<td>Distillation temperature (°C)</td>
<td>276</td>
</tr>
<tr>
<td>Main SOI (°BTDC)</td>
<td>2</td>
</tr>
<tr>
<td>Pilot SOI (°BTDC)</td>
<td>45</td>
</tr>
<tr>
<td>Fuel Split</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 12: Set of parameters producing a global unconstrained optima in emission characteristics

The corresponding emission and performance responses are summarized in table 13. The % difference from the target represent the ratio between the result and the range of values that the given characteristic assume among the available data.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>MOP values</th>
<th>% difference from target</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOx (ppm)</td>
<td>30.5</td>
<td>6.1</td>
</tr>
<tr>
<td>Soot Concentration (mg/m3)</td>
<td>1.2</td>
<td>0.7</td>
</tr>
<tr>
<td>CO (ppm)</td>
<td>4000</td>
<td>0.7</td>
</tr>
<tr>
<td>HC (ppm)</td>
<td>2650</td>
<td>62</td>
</tr>
<tr>
<td>BTE %</td>
<td>29.1</td>
<td>63.8</td>
</tr>
</tbody>
</table>

Table 13: Responses corresponding to the global unconstrained optima

The values obtained represent the optima when no constrain is applied to the minimization method. This means that the characteristics which contribute the most to the global function have a stronger impact, namely NOx and PM, while the other are out of the range of acceptability. When describing low temperature combustion it is often accepted to produce high CO and HC, especially considering that those characteristics are easily retrofitted with a diesel oxidation catalyst (DOC). It is important to underline that the optimal point is refered to the load conditions reported in the
Constrained Optimization In the attempt to obtain an optima which will represent a better compromise among all the characteristics under examination a constrained optimization was performed. This means that the candidate set of factors for optimization must produce responses that lay within a range of acceptability. The threshold for the acceptability are reported in table 15. The set of factors corresponding to this optimization process is displayed in table 14. The CN in this case is much higher than the one obtained by unconstrained optimization, which according to the results displayed in this study leads to leaner HC and CO emissions. To compensate the increase in NOx and PM related to the selection of this fuel chemistry the Main SOI is retarded considerably.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Threshold</th>
<th>MOP values</th>
<th>% difference from target</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOx (ppm)</td>
<td>50</td>
<td>26.6</td>
<td>0.5</td>
</tr>
<tr>
<td>Soot Concentration (mg/m³)</td>
<td>10</td>
<td>4.1</td>
<td>3.5</td>
</tr>
<tr>
<td>CO (ppm)</td>
<td>3000</td>
<td>2410</td>
<td>48.6</td>
</tr>
<tr>
<td>HC (ppm)</td>
<td>1000</td>
<td>735</td>
<td>10.1</td>
</tr>
<tr>
<td>BTE %</td>
<td>29</td>
<td>29.5</td>
<td>67.2</td>
</tr>
</tbody>
</table>

Table 15: Responses corresponding to the global constrained optima

Table 15 shows the results of applying constrained optimization. Except for PM production all
the other characteristics lay closer to the target value. The soot formation is still largely below the threshold, making the result of constrained optimization a better candidate for this scope.
Further studies based on MARS

The code described in this work has been developed with the specific objective to help the emission analyst in studying a given problem in the field of engines and emissions. In this chapter two studies where MARS has been successfully applied are reported.

The first one consists in the development of a simulation tool for engine testing, aimed to uncover engine map behavior dependently on the operation regime. This scenario is not drastically different from the one addressed in this work, merely from a regression point of view. The output corresponding to a set of tests is analyzed and based on the outcome some deduction can be made.

The second study instead does not relate to engine lab, instead it is aimed to the description of plume emitted from heavy duty diesel truck. Data are collected in a wind tunnel specifically designed to conduct experiments on plume evolution in the atmosphere. Several points behind a mockup heavy-duty truck were sampled in a three dimensional domain. MARS is used to produce a map of particulate matter plume starting from experimental points. Once the discrete data set is converted into continuous functions a better visualization and analysis of the plume evolution is possible. The main focus of the analysis is on the influence that dilution ratio and cooling velocity have on the particle size and distribution.

Diesel Engine Modeling Development for ICCT Heavy-Duty Vehicle Simulation Tool

This study is aimed to the characterization of the fuel consumption in a MY 2005 Mercedes depending on operation parameters. Dynamometer testing was performed over four test cycles; namely
Federal Test Procedure (FTP) test cycle, European Stationary Cycle (ESC), and two space filling designs matrix. The space filling design matrixes were generated using respectively Latin Hypercube and Gaussian process. Figure 60 shows the tested points under the lug curve. Engine was instrumented for coolant temp, oil temp, oil pressure, EGR circuit temperature, coolant flow, in-cylinder pressure, turbo enthalpy drop. Figure 61 shows the laboratory setup.

Figure 60: Distribution of the experimental points on a load-speed chart

Figure 61: MY 2005 Mercedes engine
Using the data relative only to the FTP cycle an emulator equation was obtained using the MARS code. The equation was then interrogated over the points tested over all of the four test cycles. Figure 62 shows the comparison between the model results and the ESC data. The emulator equation fits 100% of the data, both those used to generate the model and the validation data.

![Comparison with ESC data](image)

**Figure 62: Comparison between MARS results and ESC data**

Figure 63 shows the behavior of fuel consumption with respect to engine load and speed. The blue scattered points represent the experimental data. The quality of the fitting suggests that MARS could be used to generate engine maps, by substituting lookup table with the emulator function.
The next step of this study will be to compare the MARS results obtained using stationary test points with transient results. The aim is to have a reverse engineering tool able to extrapolate the engine map used by the data. This phase of the project is still in progress.

**Analyze Dispersing Plume from Heavy Duty Diesel Trucks**

The objective of this work is to investigate evolution of particle number distributions in the tailpipe plume of a vehicle using experimental data collected at WVU’s wind tunnel (WT) facility coupled with the regression analysis method. Diesel PM remains in a state of continuous transformation (unstable) for some time after it is emitted into the atmosphere. The fate of these condensable organics/inorganics is significantly affected by the dilution and atmospheric aging of the exhaust
stream. A number of processes occur during atmospheric aging that can alter the size distribution of an aerosol, including homogeneous nucleation, binary homogeneous nucleation and coagulation.

**Experimental Setup** The WVU wind tunnel is a full-scale, open-circuit, suction tunnel. The open-circuit configuration is the key feature that guarantees fresh ambient background air flow over the vehicle and therefore, approximately constant dilution air composition and conditions for the entire test duration. It is capable to accommodate a full scale class 8 heavy-duty diesel truck, matching as closely as possible real world conditions. The test section is 16ft (5m) high, 16ft (5m) wide and 115ft (35m) long.

The instrumentation required to characterize the exhaust plume could not be placed within the flow field without strongly influencing it. However a single sampling probe could extract a localized fraction of the plume and redirect the sample to the instruments. In order to minimize the length of the transfer line connecting sampling probe and instrumentation and hence, the particulate matter losses, a novel solution has been adopted by implementing a cart carrying the entire suite of instruments and moving longitudinally within the ceiling of the tunnel.

In order to supply a continuous stream of exhaust for the exhaust plume interrogation within the tunnel, a vehicle is installed on a heavy-duty chassis-dynamometer, located outside the wind tunnel, and operated at a pre-defined vehicle speed. Inside the WVU wind tunnel a mock up cabin is used to generate the truck aerodynamics. Figure 64 shows the layout of the experimental setup.
Figure 64: Lay out of the wind tunnel experimental setup

The WVU wind tunnel is capable to capture the first seconds of plume formation and evolution, from a single vehicle with parameterized input, thus give the capability of discern the weight on real world emission of different aftertreatment technologies. To achieve a detailed plume investigation 130 sampling points divided into 10 planes were used to capture PM characteristics. Each plane presents the same sampling pattern, with different vertical dimension and position, to better focusing on the plume shape and dimension evolution. The pattern is composed of 2 nested hexagons, rotated by 90, and an additional center point (see Figure 65).
The experimental data were analyzed using a modified version of the MARS code presented in this work. The input variables in this case were the three Cartesian coordinates representing a point inside the wind tunnel. Several outputs were modeled, ranging from temperature to particle size distribution. A total of 48 different outputs were considered, 32 of them being the channels of the EEPS. Figure 66 shows the $R^2$ value of each variable.
Figure 66: Residual sum of square for each sampling channel

Figure 67 shows a comparison between two contour plots representing NOx values at 5 cm from the exhaust stack. The one on the left was obtained using MARS, the one on the right is obtained from the original data.

Figure 67: Comparison between MARS and experimental values
The goal of this analysis is to produce curves representing isovalues of nucleation PM inside this tunnel. The same typology of curves will be produced for those characteristics that are suspected to influence PM nucleation, such as temperature or turbulence intensity. In the opinion of the author this way to display the problem will help in determining the physical phenomena leading to nucleation.
Conclusions

The study presented in this work consisted in an analysis tool to support engine testing. Based on the study of regression modeling techniques presented in the regression analysis chapter, the MARS algorithm was reputed a viable basis for emission analysis. The features that led to the selection of MARS as the base algorithm can be listed as:

- **Accuracy:** When dealing with non-linearity in the training data, MARS is more accurate than polynomial regression. A comparison between these two methodologies is presented in the experiment reduction section.

- **Transparency:** The contribution of each input factor, and the interaction between them can be evaluated. The ANOVA tables are a synthetic way to display it.

- **Robustness:** The MARS model has been tested over different problem types and sampling sizes, proving a fair accuracy.

- **Efficiency:** The MARS algorithm does not require high computation effort to generate a model.

The main drawbacks of MARS compared to other regression algorithms are the conceptual complexity, and the constrain in the sample size. Compared to other regression strategies MARS require more time to be implemented, a version of the code produced is presented in Appendix C. The user of this analysis tool will not have to change the algorithm but just to recall the functions in Appendix C in a Matlab environment. The sample used to train the model must include at least 30 data points in order for MARS to operate correctly.

The analysis tool was developed and tested on data collected during the project AVFL-16 [3]. These experimental data describe the response of an engine to changes in the injection strategy and
fuel properties. To investigate the effect of fuel properties on advanced combustion a systematic multivariate analysis based on the MARS algorithm was carried on. The varying fuel characteristics were CN, aromatic content, and T90 coupled with a split injection control strategy on a GM Z19DTH light-duty compression-ignition engine.

A single engine operating condition consisting of a fixed engine speed of 2100 rpm and 3.5 bar BMEP was utilized. The split injection control strategy involved varying the start of the pilot injection, start of the main injection, and fuel split. A repeatability study was performed to develop a standard by which emissions and performance changes among the fuels could be attributed to fuel property differences and not to the variability associated with the equipment or control strategy. The analysis performed lead to the following conclusions regarding low temperature combustion:

- A quadratic correlation exists between NOx and main SOI. As expected, NOX decreased as the main SOI was retarded.
- Increasing the Cetane number per se leads to higher NOx emissions, despite the effect can be masked as a consequence of Main SOI.
- An increase in CN also leads to higher soot concentrations in the exhaust.
- Higher values of CN are beneficial in reducing CO and HC.
- Regarding CO emissions, CN has a stronger impact than injection strategies.
- Aromatic content and distillation temperature lead to less soot emissions. Their effect is less noticeable compared to CN and injection strategies.
- AC has a strong impact on hydrocarbons production.
- Efficiency is increased by advancing the injection timing, and higher CN.

Using the equation obtained through the regression analysis a simple optimization procedure was performed. Based on this study the fuel characteristics that best suit low temperature combustion are; very low CN, low aromatic content, and low distillation temperature. This combination of characteristics is such to inhibit the combustion to happen too fast, giving the mixture enough time to homogenize completely.
Regarding the methodology itself we can draw the following conclusions. Since the MARS model is not a physical model it strongly relies on the data used to educate it. The regression equations quality depends strongly on the accuracy of the data used to generate them. Overall, the technique applied to develop the illustrative conceptual model was useful in screening the data, in determining the optimal solution, and the minimum number of tests necessary to characterize the phenomenon. The use of MARS to generate regression equations can lead to much wider use of RSM in the field of engine testing, both for analysis purposes and procedure development. Below are some recommendations for additional testing and further analysis of the existing data.

- It would be beneficial to include engine modifications in a further study. Certain fuels lend themselves to different engine conditions. Engine conditions and hardware of special interest include intake temperature, intake pressure, intake oxygen, and compression ratio, EGR flow, and EGR temperature.

- Investigation of additional engine operating conditions would assist in better understanding the effects of fuel properties on engine emissions and performance while operating in advanced combustion regimes.
\[
NO_x = 0.27472 \cdot \text{heaviside}(0.9650 - 1.0 \cdot x) \cdot (x1 - 0.9650) - 0.7595 \cdot \text{heaviside}(x3 - 0.9333) \cdot (x3 - 0.9333) - 2.1436 \cdot \text{heaviside}(x1 - 0.9650) \cdot (x1 - 0.9650) + 2.5809 \cdot \text{heaviside}(0.42857 - 1.0 \cdot x) \cdot (x4 - 0.4285) - 2.4005 \cdot \text{heaviside}(0.85714 - 1.0 \cdot x4) \cdot (x4 - 0.8571) + 2.9457 \cdot \text{heaviside}(x4 - 0.4285) \cdot (x4 - 0.4285) + 0.78919 \cdot \text{heaviside}(x5 - 0.5) \cdot \text{heaviside}(0.85714 - 1.0 \cdot x4) \cdot (x4 - 0.8571) \cdot (x5 - 0.5) + 0.1814 \cdot \text{heaviside}(x5 - 0.25) \cdot \text{heaviside}(x6 - 0.5) \cdot (x5 - 0.25) \cdot (x6 - 0.5) - 11.9976 \cdot \text{heaviside}(x3 - 0.9333) \cdot \text{heaviside}(x4 - 0.4285) \cdot (x3 - 0.9333) \cdot (x4 - 0.4285) - 0.24074 \cdot \text{heaviside}(0.42857 - 1.0 \cdot x4) \cdot \text{heaviside}(0.98667 - 1.0 \cdot x3) \cdot (x3 - 0.98667) \cdot (x4 - 0.4285) - 0.1974 \cdot \text{heaviside}(1.0 - 1.0 \cdot x6) \cdot \text{heaviside}(0.85714 - 1.0 \cdot x4) \cdot (x4 - 0.8571) \cdot (x6 - 1.0) - 23.5635 \cdot \text{heaviside}(x4 - 0.8571) \cdot \text{heaviside}(0.057842 - 1.0 \cdot x1) \cdot (x1 - 0.0578) \cdot (x4 - 0.8571) - 1.0005 \cdot \text{heaviside}(0.75 - 1.0 \cdot x5) \cdot \text{heaviside}(x4 - 0.4285) \cdot (x4 - 0.4285) \cdot (x5 - 0.75) + 4.2008 \cdot \text{heaviside}(0.75 - 1.0 \cdot x5) \cdot \text{heaviside}(x4 - 0.8571) \cdot (x4 - 0.8571) \cdot (x5 - 0.75) - 8.0998 \cdot \text{heaviside}(x4 - 0.8571) \cdot \text{heaviside}(x5 - 0.75) \cdot (x4 - 0.85713) \cdot (x5 - 0.75) - 0.2378 \cdot \text{heaviside}(0.5 - 1.0 \cdot x5) \cdot \text{heaviside}(1.0 - 1.0 \cdot x6) \cdot \text{heaviside}(0.85714 - 1.0 \cdot x4) \cdot (x4 - 0.8571) \cdot (x5 - 0.5) \cdot (x6 - 1.0) - 2.548 \cdot \text{heaviside}(x2 - 0.8685) \cdot \text{heaviside}(x4 - 0.4285) \cdot \text{heaviside}(0.93333 - 1.0 \cdot x3) \cdot (x3 - 0.9333) \cdot (x2 - 0.8685) \cdot (x4 - 0.4285) - 3.8785 \cdot \text{heaviside}(x4 - 0.8571) \cdot \text{heaviside}(0.93333 - 1.0 \cdot x3) \cdot \text{heaviside}(x5 - 0.75) \cdot (x3 - 0.93332) \cdot (x4 - 0.8571) \cdot (x5 - 0.75) + 0.87591 \cdot \text{heaviside}(x5 - 0.5) \cdot \text{heaviside}(1.0 - 1.0 \cdot x6) \cdot \text{heaviside}(0.85714 - 1.0 \cdot x4) \cdot (x4 - 0.8571) \cdot (x5 - 0.5) \cdot (x6 - 1.0) - 13.0159 \cdot \text{heaviside}(x6 - 0.5) \cdot \text{heaviside}(0.25 - 1.0 \cdot x5) \cdot \text{heaviside}(x4 - 0.4285) \cdot (x4 - 0.4285) \cdot (x5 - 0.25) \cdot (x6 - 0.5) - 175.6666 \cdot \text{heaviside}(x1 - 0.96505) \cdot \text{heaviside}(x3 - 0.9333) \cdot \text{heaviside}(x4 - 0.4285) \cdot (x1 - 0.96505) \cdot (x3 - 0.9333) \cdot (x4 - 0.4285) - 154.7285 \cdot \text{heaviside}(0.5 - 1.0 \cdot x5) \cdot \text{heaviside}(x1 - 0.05784) \cdot \text{heaviside}(x4 - 0.8571) \cdot (x1 - 0.05784) \cdot (x4 - 0.8571) \cdot (x5 - 0.5) + 6.9722 \cdot \text{heaviside}(x6 - 0.5) \cdot \text{heaviside}(0.25 - 1.0 \cdot x5) \cdot \text{heaviside}(x4 -
0.4285 * heaviside(x2 - 0.0380) * (x4 - 0.42856) * (x2 - 0.03806) * (x5 - 0.25) * (x6 - 0.5) - 0.6137

(29)

\[
PM = 0.2306 * heaviside(0.96506 - 1.0 * x1) * (x1 - 0.9650) - 8.563 * heaviside(x1 - 0.9650) * 
(x1 - 0.9650) - 0.2494 * heaviside(0.28571 - 1.0 * x4) * (x4 - 0.2857) - 0.4434 * heaviside(0.25 - 1.0 * 
x5) * (x5 - 0.25) - 0.1739 * heaviside(x4 - 0.2857) * (x4 - 0.2857) - 0.1560 * heaviside(x5 - 0.25) * 
(x5 - 0.25) + 4.5442 * heaviside(x3 - 0.9333) * heaviside(x4 - 0.2857) * (x3 - 0.9333) * (x4 - 0.2857) - 
0.5418 * heaviside(0.77509 - 1.0 * x2) * heaviside(x1 - 0.6156) * (x2 - 0.7750) * (x1 - 0.6156) + 9.4347 * 
heaviside(x2 - 0.7750) * heaviside(x1 - 0.6156) * (x2 - 0.7750) * (x1 - 0.6156) + 6.2743 * heaviside(x5 - 
0.25) * heaviside(0.14286 - 1.0 * x4) * (x4 - 0.1428) * (x5 - 0.25) - 16.3151 * heaviside(x5 - 0.5) * 
heaviside(0.28571 - 1.0 * x4) * (x4 - 0.2857) * (x5 - 0.5) + 0.20789 * heaviside(x5 - 0.25) * heaviside(x4 - 
0.1428) * (x4 - 0.1428) * (x5 - 0.25) + 4.4402 * heaviside(0.25 - 1.0 * x5) * heaviside(0.038062 - 
1.0 * x2) * (x2 - 0.038) * (x5 - 0.25) + 2.3423 * heaviside(0.5 - 1.0 * x5) * heaviside(0.28571 - 1.0 * 
x4) * (x4 - 0.2857) * (x5 - 0.5) - 0.4680 * heaviside(0.5 - 1.0 * x6) * heaviside(0.28571 - 1.0 * x4) * 
(x4 - 0.2857) * (x6 - 0.5) + 66.319 * heaviside(0.25 - 1.0 * x5) * heaviside(0.57143 - 1.0 * x4) * 
heaviside(0.12 - 1.0 * x3) * (x3 - 0.1199) * (x4 - 0.5714) * (x5 - 0.25) - 22.6751 * heaviside(0.25 - 1.0 * 
x5) * heaviside(0.5 - 1.0 * x6) * heaviside(0.038062 - 1.0 * x2) * (x2 - 0.0380) * (x5 - 0.25) * (x6 - 0.5) - 
19.8129 * heaviside(0.75 - 1.0 * x5) * heaviside(0.42857 - 1.0 * x4) * heaviside(x1 - 0.6156) * (x4 - 
0.4285) * (x5 - 0.75) * (x1 - 0.6156) + 117.61 * heaviside(0.038062 - 1.0 * x2) * heaviside(x4 - 0.4285) * 
heaviside(x1 - 0.6156) * (x4 - 0.4285) * (x1 - 0.6156) * (x2 - 0.0380) - 197.8254 * heaviside(x5 - 
0.5) * heaviside(0.28571 - 1.0 * x4) * heaviside(0.61568 - 1.0 * x1) * (x1 - 0.6156) * (x4 - 0.2857) * 
(x5 - 0.5) + 605.14 * heaviside(0.42857 - 1.0 * x4) * heaviside(x5 - 0.75) * heaviside(x1 - 0.61567) * 
(x4 - 0.4285) * (x5 - 0.75) * (x1 - 0.6156) - 9.8699 * heaviside(x4 - 0.4285) * heaviside(x1 - 0.6156) * 
heaviside(x2 - 0.0380) * (x4 - 0.4285) * (x1 - 0.6156) * (x2 - 0.0380) - 2.7928 * heaviside(x6 - 0.5) * 
heaviside(x4 - 0.4285) * heaviside(x1 - 0.6156) * (x4 - 0.4285) * (x1 - 0.6156) * (x6 - 0.5) + 0.95225 * 
heaviside(x5 - 0.25) * heaviside(x6 - 0.5) * heaviside(x1 - 0.0578) * (x1 - 0.0578) * (x5 - 0.25) * 
(x6 - 0.5) + 8.3351 * heaviside(x5 - 0.5) * heaviside(x6 - 0.5) * heaviside(0.28571 - 1.0 * x4) * (x4 -
\[0.2857 \times (x_5 - 0.5) \times (x_6 - 0.5) + 1.5384 \times \text{heaviside}(0.25 - 1.0 \times x_5) \times \text{heaviside}(0.57143 - 1.0 \times x_4) \times \\
\text{heaviside}(x_3 - 0.1199) \times (x_3 - 0.1199) \times (x_4 - 0.5714) \times (x_5 - 0.25) - 29.0101 \times \text{heaviside}(0.72 - 1.0 \times x_3) \times \text{heaviside}(0.75 - 1.0 \times x_5) \times \text{heaviside}(0.42857 - 1.0 \times x_4) \times \text{heaviside}(x_1 - 0.61567) \times (x_4 - 0.4285) \times \\
(x_5 - 0.75) \times (x_3 - 0.7199) \times (x_1 - 0.6156) + 65.23 \times \text{heaviside}(0.75 - 1.0 \times x_5) \times \text{heaviside}(0.42857 - 1.0 \times x_4) \times \\
\text{heaviside}(x_3 - 0.7199) \times \text{heaviside}(x_1 - 0.6156) \times (x_4 - 0.42856) \times (x_5 - 0.75) \times (x_3 - 0.7199) \times (x_1 - 0.6156) + 188.45 \times \text{heaviside}(x_4 - 0.4285) \times \text{heaviside}(x_5 - 0.75) \times \text{heaviside}(x_1 - 0.6156) \times \text{heaviside}(x_2 - 0.0380) \times (x_4 - 0.4285) \times (x_5 - 0.75) \times (x_1 - 0.6156) \times (x_2 - 0.0380) + 0.31094 \times (30)
\]

\[CO = 0.5714 \times \text{heaviside}(0.75 - 1.0 \times x_5) \times (x_5 - 0.75) - 0.46230 \times \text{heaviside}(0.96506 - 1.0 \times x_1) \times \\
(x_1 - 0.9650) - 0.2902 \times \text{heaviside}(0.57143 - 1.0 \times x_4) \times (x_4 - 0.5714) - 4.4475 \times \text{heaviside}(x_1 - 0.9650) \times \\
(x_1 - 0.9650) - 0.3894 \times \text{heaviside}(x_5 - 0.75) \times (x_5 - 0.75) - 0.1014 \times \text{heaviside}(x_6 - 0.5) \times (x_6 - 0.5) - 0.0528 \times \text{heaviside}(0.96506 - 1.0 \times x_1) \times \text{heaviside}(x_3 - 0.0399) \times (x_1 - 0.9650) \times (x_3 - 0.0399) - \\
2.0822 \times x_1 \times \text{heaviside}(x_5 - 0.75) \times \text{heaviside}(x_1) \times (x_5 - 0.75) - 1.8265 \times \text{heaviside}(0.5 - 1.0 \times x_5) \times \text{heaviside}(x_4 - 0.5714) \times (x_4 - 0.5714) \times (x_6 - 0.5) + 0.78575 \times \text{heaviside}(x_6 - 0.5) \times \text{heaviside}(0.25 - 1.0 \times x_5) \times (x_5 - 0.25) \times (x_6 - 0.5) - 0.5787 \times \text{heaviside}(x_6 - 0.5) \times \text{heaviside}(0.5 - 1.0 \times x_5) \times (x_5 - 0.5) \times (x_6 - 0.5) + \\
2.0172 \times \text{heaviside}(x_6 - 0.5) \times \text{heaviside}(x_5 - 0.75) \times (x_5 - 0.75) \times (x_6 - 0.5) + 1.2819 \times \text{heaviside}(x_5 - 0.5) \times \text{heaviside}(x_4 - 0.5714) \times (x_4 - 0.5714) \times (x_5 - 0.5) - 0.9081 \times \text{heaviside}(x_6 - 0.5) \times \text{heaviside}(x_4 - 0.5714) \times (x_4 - 0.5714) \times (x_6 - 0.5) - 2.5193 \times \text{heaviside}(x_5 - 0.5) \times \text{heaviside}(x_6 - 0.5) \times (x_5 - 0.5) \times (x_6 - 0.5) - \\
1994.7080 \times \text{heaviside}(0.038062 - 1.0 \times x_2) \times \text{heaviside}(x_1 - 0.9650) \times (x_1 - 0.9650) \times \\
(x_2 - 0.0380) + 0.28194 \times \text{heaviside}(0.57143 - 1.0 \times x_4) \times \text{heaviside}(x_3 - 0.1199) \times (x_3 - 0.1199) \times (x_4 - 0.5714) + 1.6451 \times \text{heaviside}(0.04 - 1.0 \times x_3) \times \text{heaviside}(0.96506 - 1.0 \times x_1) \times (x_1 - 0.9650) \times (x_3 - 0.0399) + 0.5423 \times \text{heaviside}(0.85714 - 1.0 \times x_4) \times \text{heaviside}(0.96506 - 1.0 \times x_1) \times (x_1 - 0.9650) \times (x_4 - 0.8571) - 1.7001 \times \text{heaviside}(0.57143 - 1.0 \times x_4) \times \text{heaviside}(0.12 - 1.0 \times x_3) \times (x_3 - 0.1199) \times (x_4 - 0.5714) - 0.4711 \times \text{heaviside}(0.5 - 1.0 \times x_6) \times \text{heaviside}(0.64092 - 1.0 \times x_1) \times (x_1 - 0.6409) \times (x_6 - 0.5) + \\
5.5008 \times \text{heaviside}(0.75 - 1.0 \times x_5) \times \text{heaviside}(0.026667 - 1.0 \times x_3) \times (x_3 - 0.0266) \times (x_5 - 0.75) + 0.89697 \times (30)
\]
\[
\text{heaviside}(x_4 - 0.8571) \ast \text{heaviside}(0.96506 - 1.0 \ast x_1) \ast (x_1 - 0.9650) \ast (x_4 - 0.8571) - 2.25205 \ast \\
\text{heaviside}(0.75 - 1.0 \ast x_5) \ast \text{heaviside}(x_4 - 0.8571) \ast (x_4 - 0.8571) \ast (x_5 - 0.75) - 0.0680 \ast \text{heaviside}(0.75 - 1.0 \ast x_5) \ast \text{heaviside}(x_3 - 0.2066) \ast (x_3 - 0.2066) \ast (x_5 - 0.75) - 3.8196 \ast \text{heaviside}(0.85714 - 1.0 \ast x_4) \ast \\
\text{heaviside}(x_5 - 0.75) \ast (x_4 - 0.8571) \ast (x_5 - 0.75) - 3.8293 \ast \text{heaviside}(x_4 - 0.8571) \ast \text{heaviside}(x_5 - 0.75) \ast (x_4 - 0.8571) \ast (x_5 - 0.75) - 31.9873 \ast \text{heaviside}(x_1 - 0.9650) \ast \text{heaviside}(x_2 - 0.0380) \ast \\
(x_1 - 0.9650) \ast (x_2 - 0.0380) - 1.5163 \ast \text{heaviside}(0.61568 - 1.0 \ast x_1) \ast \text{heaviside}(0.75 - 1.0 \ast x_5) \ast \\
\text{heaviside}(0.85714 - 1.0 \ast x_4) \ast (x_4 - 0.85713) \ast (x_5 - 0.75) \ast (x_1 - 0.6156) - 54.16799 \ast \text{heaviside}(0.04 - 1.0 \ast x_3) \ast \text{heaviside}(x_4 - 0.8571) \ast \text{heaviside}(0.96506 - 1.0 \ast x_1) \ast (x_1 - 0.96505) \ast (x_4 - 0.8571) \ast (x_3 - 0.0399) + 10.595 \ast \text{heaviside}(0.5 - 1.0 \ast x_6) \ast \text{heaviside}(0.75 - 1.0 \ast x_5) \ast \text{heaviside}(x_4 - 0.85713) \ast \\
(x_4 - 0.8571) \ast (x_5 - 0.75) \ast (x_6 - 0.5) - 8.1382 \ast \text{heaviside}(0.5 - 1.0 \ast x_6) \ast \text{heaviside}(0.85714 - 1.0 \ast x_4) \ast \\
\text{heaviside}(x_5 - 0.75) \ast (x_4 - 0.85713) \ast (x_5 - 0.75) \ast (x_6 - 0.5) - 8.0255 \ast \text{heaviside}(0.5 - 1.0 \ast x_6) \ast \\
\text{heaviside}(x_4 - 0.85713) \ast \text{heaviside}(x_5 - 0.75) \ast (x_4 - 0.8571) \ast (x_5 - 0.75) \ast (x_6 - 0.5) - 0.1639 \ast \\
\text{heaviside}(0.85714 - 1.0 \ast x_4) \ast \text{heaviside}(0.96506 - 1.0 \ast x_1) \ast \text{heaviside}(x_2 - 0.1730) \ast (x_1 - 0.9650) \ast \\
(x_4 - 0.8571) \ast (x_2 - 0.1730) + 0.92503 \ast \text{heaviside}(x_4 - 0.8571) \ast \text{heaviside}(0.96506 - 1.0 \ast x_1) \ast \\
\text{heaviside}(x_3 - 0.0399) \ast (x_1 - 0.9650) \ast (x_4 - 0.8571) \ast (x_3 - 0.0399) - 11.2077 \ast \text{heaviside}(x_5 - 0.25) \ast \\
\text{heaviside}(0.5 - 1.0 \ast x_6) \ast \text{heaviside}(x_1 - 0.9650) \ast (x_1 - 0.9650) \ast (x_5 - 0.25) \ast (x_6 - 0.5) + 1.0864 \ast \\
\text{heaviside}(x_5 - 0.25) \ast \text{heaviside}(0.5 - 1.0 \ast x_6) \ast \text{heaviside}(0.96506 - 1.0 \ast x_1) \ast (x_1 - 0.9650) \ast (x_5 - 0.25) \ast \\
(x_6 - 0.5) + 1.3151 \ast \text{heaviside}(x_5 - 0.25) \ast \text{heaviside}(x_6 - 0.5) \ast \text{heaviside}(0.42857 - 1.0 \ast x_4) \ast (x_4 - 0.4285) \ast (x_5 - 0.25) \ast (x_6 - 0.5) - 5.9924 \ast \text{heaviside}(x_5 - 0.5) \ast \text{heaviside}(x_6 - 0.5) \ast \text{heaviside}(0.71429 - 1.0 \ast x_4) \ast (x_4 - 0.7142) \ast (x_5 - 0.5) \ast (x_6 - 0.5) + 3.735 \ast \text{heaviside}(x_5 - 0.25) \ast \text{heaviside}(x_6 - 0.5) \ast \\
\text{heaviside}(x_4 - 0.4285) \ast (x_4 - 0.4285) \ast (x_5 - 0.25) \ast (x_6 - 0.5) - 4.3952 \ast \text{heaviside}(x_6 - 0.5) \ast \\
\text{heaviside}(x_5 - 0.75) \ast \text{heaviside}(x_4 - 0.5714) \ast (x_5 - 0.75) \ast (x_4 - 0.5714) \ast (x_6 - 0.5) + 0.2122 \ast (31)
\]

\[
HC = 0.0693 \ast \text{heaviside}(0.8685 - 1.0 \ast x_2) \ast (x_2 - 0.8685) - 5.3652 \ast \text{heaviside}(0.13898 - 1.0 \ast x_1) \ast (x_1 - 0.1389) + 5.9501 \ast \text{heaviside}(0.57143 - 1.0 \ast x_4) \ast (x_4 - 0.5714) - 0.5384 \ast \text{heaviside}(0.25 - 1.0 \ast x_5) \ast (x_5 - 0.25) + 7.4159 \ast \text{heaviside}(0.75 - 1.0 \ast x_5) \ast (x_5 - 0.75) + 0.08863 \ast \text{heaviside}(0.5 - 1.0 \ast x_6) \ast (x_6 - 0.5) + 0.2122 \ast (31)
\]

106
$x6) \ast (x6 - 0.5) + 1.2671 \ast heaviside(x2 - 0.8685) \ast (x2 - 0.8685) - 1.0368 \ast heaviside(x4 - 0.8571) \ast (x4 - 0.8571) - 0.6098 \ast heaviside(x5 - 0.75) \ast (x5 - 0.75) + 5.6406 \ast heaviside(x4 - 0.5714) \ast (x4 - 0.5714) + 0.075051 \ast heaviside(x1 - 0.138) \ast (x1 - 0.1389) - 27.1994 \ast heaviside(0.5 - 1.0 \ast x5) \ast heaviside(x4 - 0.5714) \ast (x4 - 0.5714) + (x4 - 0.5714) - 0.0822 \ast heaviside(0.5 - 1.0 \ast x6) \ast heaviside(x1 - 0.1389) \ast (x1 - 0.1389) \ast (x6 - 0.5) - 0.5774 \ast heaviside(x5 - 0.25) \ast heaviside(0.13898 - 1.0 \ast x1) \ast (x1 - 0.13898) \ast (x5 - 0.25) - 0.3734 \ast heaviside(x5 - 0.25) \ast heaviside(0.85714 - 1.0 \ast x4) \ast (x4 - 0.8571) \ast (x5 - 0.5) + 24.629 \ast heaviside(x5 - 0.75) \ast heaviside(x4 - 0.5714) \ast (x4 - 0.5714) - 21.6816 \ast heaviside(x5 - 0.5) \ast heaviside(x4 - 0.5714) \ast (x4 - 0.5714) \ast (x5 - 0.5) \ast heaviside(x4 - 0.5714) \ast (x4 - 0.5714) \ast (x6 - 0.5) - 0.2914 \ast heaviside(x5 - 0.25) \ast heaviside(x1 - 0.1389) \ast (x1 - 0.1389) \ast (x5 - 0.25) + 1.13 \ast heaviside(0.75 - 1.0 \ast x5) \ast heaviside(x3 - 0.9333) \ast (x3 - 0.9333) \ast (x5 - 0.75) + 0.1953 \ast heaviside(0.8571 - 1.0 \ast x4) \ast heaviside(0.8685 - 1.0 \ast x2) \ast (x2 - 0.8685) \ast (x4 - 0.8571) - 1.2463 \ast heaviside(0.5 - 1.0 \ast x6) \ast heaviside(0.13898 - 1.0 \ast x1) \ast (x1 - 0.1389) \ast (x6 - 0.5) - 0.0869 \ast heaviside(0.75 - 1.0 \ast x5) \ast heaviside(0.9333 - 1.0 \ast x3) \ast (x3 - 0.9333) \ast (x5 - 0.75) - 24.7899 \ast heaviside(0.5 - 1.0 \ast x5) \ast heaviside(0.57143 - 1.0 \ast x4) \ast (x4 - 0.5714) \ast (x5 - 0.5) + 24.499 \ast heaviside(0.75 - 1.0 \ast x5) \ast heaviside(0.85714 - 1.0 \ast x4) \ast (x4 - 0.8571) \ast (x5 - 0.75) + 5.8 \ast heaviside(0.13898 - 1.0 \ast x1) \ast heaviside(x4 - 0.4285) \ast (x4 - 0.4285) \ast (x1 - 0.1389) + 17.358 \ast heaviside(0.75 - 1.0 \ast x5) \ast heaviside(x4 - 0.8571) \ast (x4 - 0.8571) \ast (x5 - 0.75) + 112.32 \ast heaviside(0.85714 - 1.0 \ast x4) \ast heaviside(0.86851 - 1.0 \ast x2) \ast heaviside(0.057842 - 1.0 \ast x1) \ast (x1 - 0.0578) \ast (x2 - 0.8685) \ast (x4 - 0.8571) - 0.9164 \ast heaviside(x5 - 0.5) \ast heaviside(x6 - 0.5) \ast heaviside(x4 - 0.5714) \ast (x4 - 0.5714) \ast (x5 - 0.5) \ast (x6 - 0.5) - 0.2957 \ast heaviside(x5 - 0.25) \ast heaviside(x6 - 0.5) \ast heaviside(x1 - 0.1389) \ast (x1 - 0.1389) \ast (x5 - 0.25) \ast (x6 - 0.5) + 2.3854 \ast heaviside(x5 - 0.25) \ast heaviside(0.5 - 1.0 \ast x6) \ast heaviside(0.1389 - 1.0 \ast x1) \ast (x1 - 0.1389) \ast (x5 - 0.25) \ast (x6 - 0.5) + 0.3531 \ast heaviside(x5 - 0.25) \ast heaviside(0.8571 - 1.0 \ast x4) \ast heaviside(0.933 - 1.0 \ast x3) \ast (x3 - 0.933) \ast (x4 - 0.8571) \ast (x5 - 0.25) - 1.3942 \ast heaviside(x5 - 0.5) \ast heaviside(0.5714 - 1.0 \ast x4) \ast heaviside(0.72 - 1.0 \ast x3) \ast (x3 - 0.7199) \ast (x4 - 0.5714) \ast (x5 - 0.5) + 0.8340 \ast heaviside(x5 - 0.25) \ast heaviside(0.2857 - 1.0 \ast x4) \ast heaviside(x1 - 0.1389) \ast (x4 - 0.2857) \ast (x1 - 0.1389) \ast (x5 - 0.25) + 0.1671 \ast heaviside(x5 - 0.25) \ast heaviside(0.5 - 1.0 \ast x6) \ast heaviside(x1 - 0.1389) \ast (x1 - 0.1389) \ast (x5 - 0.25) \ast (x6 - 0.5) - 0.2448 \ast heaviside(x1 - 0.0578) \ast heaviside(0.8571 - 1.0 \ast x4) \ast heaviside(0.8685 - 1.0 \ast x2) \ast (x1 - 0.0578) \ast (x2 - 0.8685) \ast (x4 - 0.8571) + 1440.3 \ast heaviside(0.5 -
\[ BTE = 0.20364 \cdot \text{heaviside}(x1 - 0.0578) \cdot (x1 - 0.0578) + 6.282 \cdot \text{heaviside}(0.0578 - 1.0 \cdot x1) \cdot (x1 - 0.0578) + 0.0708 \cdot \text{heaviside}(0.7750 - 1.0 \cdot x2) \cdot (x2 - 0.7750) + 1.0573 \cdot \text{heaviside}(0.4285 - 1.0 \cdot x4) \cdot (x4 - 0.4285) - 0.5192 \cdot \text{heaviside}(0.75 - 1.0 \cdot x5) \cdot (x5 - 0.75) - 0.1017 \cdot \text{heaviside}(0.5 - 1.0 \cdot x6) \cdot (x6 - 0.5) - 1.309 \cdot \text{heaviside}(x2 - 0.7750) \cdot (x2 - 0.7750) - 0.3092 \cdot \text{heaviside}(x6 - 0.5) \cdot (x6 - 0.5) - 0.4131 \cdot \text{heaviside}(x6 - 0.5) \cdot \text{heaviside}(0.8369 - 1.0 \cdot x1) \cdot (x1 - 0.8369) \cdot (x6 - 0.5) - 0.8066 \cdot \text{heaviside}(x6 - 0.5) \cdot \text{heaviside}(0.5 - 1.0 \cdot x5) \cdot (x5 - 0.5) \cdot (x6 - 0.5) - 0.4730 \cdot \text{heaviside}(x5 - 0.25) \cdot \text{heaviside}(x4 - 0.4285) \cdot (x4 - 0.4285) \cdot (x5 - 0.25) + 0.8492 \cdot \text{heaviside}(x3 - 0.1199) \cdot \text{heaviside}(x4 - 0.4285) \cdot (x3 - 0.1199) \cdot (x4 - 0.4285) + 0.57063 \cdot \text{heaviside}(0.7750 - 1.0 \cdot x2) \cdot \text{heaviside}(0.4285 - 1.0 \cdot x4) \cdot (x4 - 0.4285) \cdot (x2 - 0.7750) - 0.5143 \cdot \text{heaviside}(0.25 - 1.0 \cdot x5) \cdot \text{heaviside}(0.7750 - 1.0 \cdot x2) \cdot (x2 - 0.7750) \cdot (x5 - 0.25) - 5.3919 \cdot \text{heaviside}(x4 - 0.4285) \cdot \text{heaviside}(0.12 - 1.0 \cdot x3) \cdot (x3 - 0.1199) \cdot (x4 - 0.4285) + 4.8437 \cdot \text{heaviside}(0.25 - 1.0 \cdot x5) \cdot \text{heaviside}(x4 - 0.4285) \cdot (x4 - 0.4285) \cdot (x5 - 0.25) - 5.4912 \cdot \text{heaviside}(0.4285 - 1.0 \cdot x4) \cdot \text{heaviside}(x2 - 0.7750) \cdot (x4 - 0.4285) \cdot (x2 - 0.7750) - 16.1963 \cdot \text{heaviside}(0.4285 - 1.0 \cdot x4) \cdot \text{heaviside}(x5 - 0.75) \cdot (x4 - 0.4285) \cdot (x5 - 0.75) + 11.712 \cdot \text{heaviside}(x6 - 0.5) \cdot \text{heaviside}(0.5 - 1.0 \cdot x5) \cdot \text{heaviside}(x4 - 0.5714) \cdot (x4 - 0.5714) \cdot (x5 - 0.5) \cdot (x6 - 0.5) + 0.3896 \cdot (33) \]
APPENDIX B

NOx

Figure 68: Dependence of NOx from main SOI, and pilot SOI
Figure 69: Dependence of NOx from main SOI, and fuel split
Soot Concentration

Figure 70: Dependence of PM from cetane number, aromatic content, and main SOI
Figure 71: Dependence of PM from cetane number, fuel split, and main SOI
Figure 72: Dependence of CO from pilot SOI, and main SOI
Figure 73: Dependence of CO from fuel split, and main SOI
Figure 74: Dependence of CO from distillation temperature, pilot SOI, and main SOI
Figure 75: Dependence of CO from aromatic content, distillation temperature, and main SOI
Figure 76: Dependence of HC from aromatic content, and main SOI
Figure 77: Dependence of HC from cetane number, pilot SOI, and main SOI
APPENDIX C

1 clc
2 close all
3 clear all
4
5
6 load FuelComp1.csv %Studio Caratteristiche combustibile e parametri di ignizione
7 x1=FuelComp1(:,1); %Cetane Number
8 x2=FuelComp1(:,2); %Aromatic Content
9 x3=FuelComp1(:,3); %T90
10 x4=FuelComp1(:,4); %Main SOI
11 x5=FuelComp1(:,5); %Pilot SOI
12 x6=FuelComp1(:,6); %Fuel Split
13 y1=FuelComp1(:,7); %NOx
14 y2=FuelComp1(:,8); %Soot Concentration
15 y3=FuelComp1(:,9); %BTE (%)
16 y4=FuelComp1(:,10); %CO (ppm)
17 y5=FuelComp1(:,11); %HC (ppm)
18
19 %Store the uncoded variable
20 Xp = [x1 x2 x3 x4 x5 x6];
21 Yp=y5; %!!!!!!!!!!!!!!!!!!!!!!
%Coded Variables

coded = struct('x1l', {min(x1)}, 'x1h', {max(x1)}, 'x2l', {min(x2)}, 'x2h', {max(x2)}, 'x3l', {min(x3)}, 'x3h', {max(x3)}, 'x4l', {min(x4)}, 'x4h', {max(x4)}, 'x5l', {min(x5)}, 'x5h', {max(x5)}, 'x6l', {min(x6)}, 'x6h', {max(x6)}, 'y1l', {min(y1)}, 'y1h', {max(y1)}, 'y2l', {min(y2)}, 'y2h', {max(y2)}, 'y3l', {min(y3)}, 'y3h', {max(y3)}, 'y4l', {min(y4)}, 'y4h', {max(y4)});

x1 = (x1 - coded.x1l) ./ (coded.x1h - coded.x1l);
x2 = (x2 - coded.x2l) ./ (coded.x2h - coded.x2l);
x3 = (x3 - coded.x3l) ./ (coded.x3h - coded.x3l);
x4 = (x4 - coded.x4l) ./ (coded.x4h - coded.x4l);
x5 = (x5 - coded.x5l) ./ (coded.x5h - coded.x5l);
x6 = (x6 - coded.x6l) ./ (coded.x6h - coded.x6l);

%Coded Output

y1 = (y1 - coded.y1l) ./ (coded.y1h - coded.y1l);
y2 = (y2 - coded.y2l) ./ (coded.y2h - coded.y2l);
y3 = (y3 - coded.y3l) ./ (coded.y3h - coded.y3l);
y4 = (y4 - coded.y4l) ./ (coded.y4h - coded.y4l);
y5 = (y5 - coded.y5l) ./ (coded.y5h - coded.y5l);

X = [x1 x2 x3 x4 x5 x6];
Y = y5;

%Use MARS model
%Set the parameters
params = params(56, [], [], 0, [], 4);

%Build the model
model = build(X, Y, params)

% Evaluate the equations
% X=allunga(X);

yp=PROVAequationBis(X, model);

% UNCODE THE VARIABLES
X(:,1)=X(:,1).*((coded.x1h-coded.x1l)+coded.x1l);
X(:,2)=X(:,2).*((coded.x2h-coded.x2l)+coded.x2l);
X(:,3)=X(:,3).*((coded.x3h-coded.x3l)+coded.x3l);
X(:,4)=X(:,4).*((coded.x4h-coded.x4l)+coded.x4l);
X(:,5)=X(:,5).*((coded.x5h-coded.x5l)+coded.x5l);
X(:,6)=X(:,6).*((coded.x6h-coded.x6l)+coded.x6l);

yp(:)=yp(:).*((coded.y5h-coded.y5l)+coded.y5l); %THE UNCODING PARAMETERS HAVE TO CORRESPOND TO THE VARIABLE OF INTEREST

%Plot pair-wise contribution of variables
% Graphics(X,yp,Xp,Yp)

save('FuelComp1HC')
anova(model,X,Yp)
% Check
plot(Yp, 'DisplayName', 'Yp', 'YDataSource', 'Yp'); figure(gcf)
hold on
plot(yp, 'r', 'DisplayName', 'yp', 'YDataSource', 'yp'); figure(gcf)

function yp=PROVAequationBis(X, model)

% Create the equation using the coefficients and the knot determined by
% the
% MARS algorithm
%
yp=zeros(length(X),length(model.coefs));
yp(:,1)=model.coefs(1);
i=1;
papa=[0;model.parents(1)];

for i=1:length(model.coefs)-1 % I coefficienti sono uno in piu
    rispetto alle BF
    %Scrie la ultima BF, che non e' mai imparentata
    x = X(:,model.knotdims{i})(length(model.knotdims{i})));
    BF(:,i)= max(0, (x - model.knotsites{i})(length(model.knotsites{i}))
                     )*model.knotdirs{i}(length(model.knotdirs{i}));

% Studia le parentele

fifi(i)=length(model.knotsites{i});
\[ pip_i = \text{length}(\text{model.knotsites}_i)'; \]

\[ \text{if } \text{fifi}_i > 1 \]

\[ \text{if } \text{model.parents}_i > 0 \]
\[
\begin{align*}
  x &= X(:, \text{model.knotdims}_i)(\text{length}(\text{model.knotdims}_i)) ; \\
  \text{BF}(:, i) &= \text{BF}(:, \text{papa}(i+1)). \times \text{max}(0, (x - \text{model.knotsites}_i(\text{fifi}_i))) \times \text{model.knotdirs}_i(\text{fifi}_i) ; \\
\end{align*}
\]

%Se due BF sono imparentatati il BF parente si prende tale e quale senza cambiare neanche la direzione

\[ \text{pipi}(i) = \text{pipi}(i) - \text{fifi}(	ext{papa}(i+1)) ; \]

\[ \%\text{papa}(i+1) = 0 ; \]

\[ \text{end} \]

\[ \text{if } \text{pipi}(i) > 1 \]

\[ \text{if } \text{papa}(	ext{model.parents}(i)+1) == 0 \]
\[
\begin{align*}
  \%\text{Studia il caso in cui non c'è parentela ma ci sono più variabili} \\
  \text{for } j = 1 : \text{pipi}(i) - 1 ; \%\text{Ci possono essere più variabili dentro la BF}(i) \\
\end{align*}
\]

\[ x_2 = X(:, \text{model.knotdims}_i(j)) ; \]
\[
\begin{align*}
  \text{BF}(:, i) &= \text{BF}(:, i). \times \text{max}(0, (x_2 - \text{model.knotsites}_i(j))) \times \text{model.knotdirs}_i(j) ; \\
  \%\text{BF}(:, i) &= \text{max}(0, (x - \text{model.knotsites}_i(\text{fifi}_i))) \times \text{model.knotdirs}_i(\text{fifi}_i) \times \text{max}(0, (x_2 - \text{model.knotsites}_i(1))) \times \text{model.knotdirs}_i(1)) ; \\
\end{align*}
\]
yp(:,i+1) = yp(:,i) + model.coefs(i+1).*BF(:,i);
end

yp = yp(:,length(model.knotdims)+1);

return

function trainParams = params(maxFuncs, c, cubic, cubicFastLevel, ... selfInteractions, maxInteractions, threshold, prune, useMinSpan, ... useEndSpan, maxFinalFuncs)

if (nargin < 1) || (isempty(maxFuncs))
    trainParams.maxFuncs = 21;
else
    trainParams.maxFuncs = maxFuncs;
end

if (nargin < 2) || (isempty(c))
    trainParams.c = 3;
else
    trainParams.c = c;
end

if (nargin < 3) || (isempty(cubic))
trainParams.cubic = true;
else
    trainParams.cubic = cubic;
end

if (nargin < 4) || (isempty(cubicFastLevel))
    trainParams.cubicFastLevel = 2;
else
    trainParams.cubicFastLevel = cubicFastLevel;
end

if (nargin < 5) || (isempty(selfInteractions))
    trainParams.selfInteractions = 1;
else
    trainParams.selfInteractions = selfInteractions;
end

if (trainParams.cubic) && (trainParams.selfInteractions > 1)
    trainParams.selfInteractions = 1;
end

if (nargin < 6) || (isempty(maxInteractions))
    trainParams.maxInteractions = 1; % applicable maximum is d *
    trainParams.selfInteractions
else
    trainParams.maxInteractions = maxInteractions;
end

if (nargin < 7) || (isempty(threshold))
    trainParams.threshold = 1e-4;
else

trainParams.threshold = threshold;

end

if (nargin < 8) || (isempty(prune))
    trainParams.prune = true;
else
    trainParams.prune = prune;
end

if (nargin < 9) || (isempty(useMinSpan))
    trainParams.useMinSpan = -1; % default = -1 = automatic
else
    if useMinSpan == 0
        trainParams.useMinSpan = 1; % 1 and 0 is the same here (no endspan)
    else
        trainParams.useMinSpan = useMinSpan;
    end
end

if (nargin < 10) || (isempty(useEndSpan))
    trainParams.useEndSpan = -1; % default = -1 = automatic
else
    if useEndSpan == 0
        trainParams.useEndSpan = 1; % 1 and 0 is the same here (no endspan)
    else
        trainParams.useEndSpan = useEndSpan;
    end
end
if (nargin < 11) || (isempty(maxFinalFuncs))
    trainParams.maxFinalFuncs = Inf;
else
    trainParams.maxFinalFuncs = maxFinalFuncs;
end

return

function [model, time] = build(Xtr, Ytr, trainParams, weights)

if trainParams.maxInteractions >= 2
    trainParams_actual_c = trainParams.c;
else
    trainParams_actual_c = 2*trainParams.c/3; % penalty coefficient for additive modelling
end

if trainParams.useMinSpan == 0
    trainParams.useMinSpan = 1; % 1 and 0 is the same here (no endspan)
end

if trainParams.useEndSpan == 0
    trainParams.useEndSpan = 1; % 1 and 0 is the same here (no endspan)
end

if (nargin < 4)
    weights = [];
else
wd = diag(weights);
if nargin < 5
    modelOld = [];
end

fprintf('Building model...
');
ws = warning('off');
tic;

maxIters = floor(trainParams.maxFuncs / 2); % because basis functions are added two at a time
YtrMean = mean(Ytr);
YtrSS = sum((Ytr - YtrMean).^2);
minX = min(Xtr);
maxX = max(Xtr);

if trainParams.useEndSpan < 0
    endSpan = getEndSpan(d);
else
    endSpan = trainParams.useEndSpan;
end

if isempty(modelOld)
    X = ones(n,1);
    err = 1; % errore normalizzato al livello costante
    model.coefs = YtrMean;
    model.knotdims = {};
else
    % code for adding basis functions
end
model.knotsites = {};  
model.knotdirs = {};  
model.parents = [];  
model.trainParams = [];  
model.MSE = Inf;  
model.GCV = Inf;  

else  
    model = modelOld; % Usa modelOld come modello iniziale  
end  

if endSpan*2 >= n  
    if isempty(modelOld)  
        model.MSE = YtrSS / n;  
        model.GCV = gcv(model, model.MSE, n, trainParams.actual_c);  
        if trainParams.cubic  
            model.t1 = [];  
            model.t2 = [];  
        end  
    end  
else  

    % FORWARD PHASE  

    if isempty(modelOld) % no forward phase when modelOld is used  
        fprintf('Forward phase .');  

        % crea una lista ordinata di elementi  
        [ordinatiXtr ordinatiXtrInd] = sort(Xtr);  
        if trainParams.useEndSpan ~= 1
% elimina i dati iniziali e finali
ordinatiXtr = ordinatiXtr(endSpan:end-(endSpan-1),:);
ordinatiXtrInd = ordinatiXtrInd(endSpan:end-(endSpan-1),:);
end

if trainParams.cubic
tmp_t1 = [];
tmp_t2 = [];
end

basisFunctionList = []; % basis functions candidate
BasisAggiunte = 0; % basis functions aggiunte all ultima iterazione

% the main loop of the forward phase
for depth = 1 : maxIters
    basisFunctionList = createList(basisFunctionList , Xtr ,
        ordinatiXtr , ordinatiXtrInd , ...
        n, d, model, BasisAggiunte ,
        trainParams, endSpan);
end

% ferma la forward phase se non ci sono candidati
if isempty(basisFunctionList)
    if trainParams.cubic
        t1 = tmp_t1;
t2 = tmp_t2;
    end
    break;
end

tmpErr = inf(1,size(basisFunctionList ,2));
tmpCoefs = \inf(length(model.coefs) + 2, size(basisFunctionList, 2));

Xtmp = zeros(n, size(X, 2) + 2);
if ~trainParams.cubic
    Xtmp(:,1:end−2) = X;
end

% prova tutte le accoppiate di basis functions
for i = 1 : size(basisFunctionList, 2)
    if trainParams.cubic
        [t1 t2 dif] = KnotsEstremi(model, basisFunctionList
          {1, i}, basisFunctionList{2, i}, ...
          d, minX, maxX, tmp_t1, tmp_t2);
        Xtmp(:,1:end−2) = X;
        \% update basis functions with the updated side
        knots
        for j = 1 : length(model.knotdims)
            if dif(j)
                Xtmp(:,j+1) = createbasisfunction(Xtr, Xtmp
                  , model.knotdims{j}, model.knotsites{j}, ...
                  model.knotdirs{j}, model.
                  parents(j), minX, maxX, t1
                  (j,:), t2(j,:));
            end
        end
        \% New basis function
        dirs = basisFunctionList{3, i};
        Xtmp(:,end−1) = createbasisfunction(Xtr, Xtmp,
          basisFunctionList{1, i}, basisFunctionList{2, i},
          ...}
352 dirs, basisFunctionList{4,i}, minX, maxX, t1(end,:), t2(end,:));
353 if isnan(Xtmp(1,end−1)), Xtmp(:,end−1) = []; end

% Reflected partner
354 dirs(end) = −dirs(end);
355 Xtmp(:,end) = createbasisfunction(Xtr, Xtmp,
356     basisFunctionList{1,i}, basisFunctionList{2,i},
357     ..
358 dirs, basisFunctionList{4,i}, minX, maxX, t1(end,:), t2(end,:));
359 if isnan(Xtmp(1,end)), Xtmp(:,end) = []; end

else
% New basis function
360 dirs = basisFunctionList{3,i};
361 Xtmp(:,end−1) = createbasisfunction(Xtr, Xtmp,
362     basisFunctionList{1,i}, ..
363     basisFunctionList{2,i}, dirs,
364     basisFunctionList{4,i}, minX, maxX);
365 if isnan(Xtmp(1,end−1)), Xtmp(:,end−1) = []; end

% Reflected partner
366 dirs(end) = −dirs(end);
367 Xtmp(:,end) = createbasisfunction(Xtr, Xtmp,
368     basisFunctionList{1,i}, ..
369     basisFunctionList{2,i}, dirs,
370     basisFunctionList{4,i}, minX, maxX)
371);
372 if isnan(Xtmp(1,end)), Xtmp(:,end) = []; end

end
if size(Xtmp, 2) == size(X, 2) + 2, create a pair of basis functions

\[
[\text{coeffs } \text{tmpErr}(i)] = \text{lreg}(Xtmp, Ytr, \text{weights}, \text{wd});
\]

\[
\text{tmpErr}(i) = \text{tmpErr}(i)/\text{YtrSS};
\]

\[
\text{tmpCoefs}(:,i) = \text{coeffs};
\]

elseif size(Xtmp, 2) == size(X, 2) + 1, create a single basis function

\[
[\text{coeffs } \text{tmpErr}(i)] = \text{lreg}(Xtmp, Ytr, \text{weights}, \text{wd});
\]

\[
\text{tmpErr}(i) = \text{tmpErr}(i)/\text{YtrSS};
\]

\[
\text{tmpCoefs}(:,i) = [\text{coeffs}; \text{NaN}];
\]

\[
\text{Xtmp} = [\text{Xtmp} \text{zeros}(n,1)];
\]

else, no basis function created (size(Xtmp, 2) == size(X, 2))

\[
\text{tmpErr}(i) = \text{Inf};
\]

\[
\text{Xtmp} = [\text{Xtmp} \text{zeros}(n,2)];
\]

end

eend

[newErr, ind] = \text{min}(\text{tmpErr}); \% analizza il contributo della base aggiunta

% stop the forward phase if no significant contribution

if (\text{isnan}(\text{newErr})) || (\text{err(end)} - \text{newErr} < \text{trainParams.threshold})

if trainParams.cubic

\[
\text{t1} = \text{tmp}_{-}\text{t1};
\]

\[
\text{t2} = \text{tmp}_{-}\text{t2};
\]

end

break;

eend
% Update the model with new basis function

if trainParams.cubic

    [t1 t2 dif] = KnotsEstremi(model, basisFunctionList{1, ind}, basisFunctionList{2,ind}, ...
        d, minX, maxX, tmp_t1, tmp_t2);

    d, minX, maxX, tmp_t1, tmp_t2);

% update basis functions with the updated side knots

    for j = 1 : length(model.knotdims)
        if dif(j)
            X(:,j+1) = createbasisfunction(Xtr, X, model.
                knotdims{j}, model.knotsites{j}, ...
                model.knottedirs{j}, model.parents(j),
                minX, maxX, t1(j,:), t2(j,:));

        end
    end

% Add the new basis function

dirs = basisFunctionList{3, ind};

    Xn = createbasisfunction(Xtr, X, basisFunctionList{1, ind}, basisFunctionList{2,ind}, ...
        dirs, basisFunctionList{4,ind}, minX, maxX, t1(end ;), t2(end ;));

    if isnan(Xn(1)), Xn = []; end

% aggiunge una coppia di basis functions

dirs(end) = -dirs(end);

    Xn2 = createbasisfunction(Xtr, X, basisFunctionList{1, ind}, basisFunctionList{2,ind}, ...
        dirs, basisFunctionList{4,ind}, minX, maxX, t1( end ;), t2( end ;));

    if isnan(Xn2(1)), Xn2 = []; end

    X = [X Xn Xn2];
if ~isempty(Xn) && ~isempty(Xn2) % aggiunge una sola basis function
    t1(end+1,:) = t1(end,:); 
    t2(end+1,:) = t2(end,:); 
end 
else 
    dirs = basisFunctionList{3, ind}; 
    % Add the new basis function 
    Xn = createbasisfunction(Xtr, X, basisFunctionList{1, ind}, 
    basisFunctionList{2,ind}, dirs, basisFunctionList 
    {4,ind}, minX, maxX); 
    if isnan(Xn(1)), Xn = []; end 
    % Add the reflected partner 
    dirs(end) = -dirs(end); 
    Xn2 = createbasisfunction(Xtr, X, basisFunctionList{1, ind}, 
    basisFunctionList{2,ind}, dirs, basisFunctionList 
    {4,ind}, minX, maxX); 
    if isnan(Xn2(1)), Xn2 = []; end 
    X = [X Xn Xn2]; 
end 

model.coefs = tmpCoefs(:,ind); 

% add the basis functions to the model 
BasisAggiunte = 0; 
dirs = basisFunctionList{3, ind}; 
if ~isempty(Xn)
    model.knotdims{end+1,1} = basisFunctionList{1, ind};
model.knotsites{end+1,1} = basisFunctionList{2, ind};
model.knotdirs{end+1,1} = dirs;
model.parents{end+1,1} = basisFunctionList{4, ind};
BasisAggiunte = BasisAggiunte + 1;

else
    model.coefs{end} = [];
end

if ~isempty(Xn2)
    dirs{end} = −dirs{end};
    model.knotdims{end+1,1} = basisFunctionList{1, ind};
    model.knotsites{end+1,1} = basisFunctionList{2, ind};
    model.knotdirs{end+1,1} = dirs;
    model.parents{end+1,1} = basisFunctionList{4, ind};
    BasisAggiunte = BasisAggiunte + 1;
else
    model.coefs{end} = [];
end

err{end+1} = newErr;
if (newErr < trainParams.threshold) || ...
    (length(model.coefs) + 2 > n)
    break;
end

if trainParams.cubic
    tmp_t1 = t1;
    tmp_t2 = t2;
end

basisFunctionList(:, ind) = [];
end % end of the main loop

if verbose, fprintf('
'); end
end % end of isempty(modelOld)

% BACKWARD PHASE

if trainParams.prune

fprintf('Backward phase .');

if ~isempty(modelOld) % Se non c'è un modello gia scremato rianalizza tutte le basi

if (doCubicFastLevel == -1) || (doCubicFastLevel >= 2)

X = ones(n, length(model.knotdims)+1);
for i = 1 : length(model.knotdims)

X(:,i+1) = createbasisfunction(Xtr, X, model.

    knotdims{i}, model.knotsites{i}, ...

    model.knotdirs{i}, model.parents(i),

    minX, maxX);

end

[model.coefs model.MSE] = lreg(X, Ytr, weights, wd);
model.MSE = model.MSE / n;

end % end of the main loop
model.GCV = gcv(model, model.MSE, n,
    trainParams_actual_c);

else
    % create all the basis functions (cubic) from scratch
    t1 = model.t1;
    t2 = model.t2;
    X = ones(n,length(model.knotdims)+1);
    for i = 1 : length(model.knotdims)
        X(:,i+1) = createbasisfunction(Xtr, X, model.
            knotdims{i}, model.knotsites{i}, ...
            model.knotdirs{i}, model.parents(i),
            minX, maxX, t1(i,:), t2(i,:));
    end
end

end

models = {model};
mses = model.MSE;
gcvs = model.GCV;

% the main loop of the backward phase
for j = 1 : length(model.knotdims)
    tmpErr = inf(1, length(model.knotdims));
    tmpCoefs = inf(length(model.coefs)-1, length(model.knotdims ));
    % cancella le basis functions una alla volta
    for k = 1 : length(model.knotdims)
        Xtmp = X;
Xtmp(:,k+1) = [];

if trainParams.cubic

    % create temporary t1, t2, and model with a deleted basis function
    tmp_t1 = t1;
    ttmp_t1(k,:) = [];
    tmp_t2 = t2;
    ttmp_t2(k,:) = [];
    tmp_model.knotdims = model.knotdims;
    tmp_model.knotdims(k) = [];
    tmp_model.knotsites = model.knotsites;
    tmp_model.knotsites(k) = [];
    tmp_model.knotdirs = model.knotdirs;
    tmp_model.knotdirs(k) = [];
    tmp_model.parents = model.parents;
    tmp_model.parents(k) = [];
    tmp_model.parents = updateParents(tmp_model.parents, k);
    [tmp_t1 tmp_t2 dif] = KnotsEstremi(tmp_model, [], [], d, minX, maxX, tmp_t1, tmp_t2);

    % update basis functions with the updated side knots
    for i = 1 : length(tmp_model.knotdims)
        if dif(i)
            Xtmp(:,i+1) = createbasisfunction(Xtr, Xtmp
                                                , tmp_model.knotdims{i}, tmp_model.
                                                knotsites{i}, ...
                                                tmp_model.knotdirs{i},
                                                tmp_model.parents(i), minX
                                                , maxX, tmp_t1(i,:)),
        end
    end

end
\[
\text{tmp}_2(i,:) = \text{coeffs} \text{tmpErr}(k) = \text{lreg}(\text{Xtmp}, \text{Ytr}, \text{weights}, \text{wd});
\]

end

end

end

[\text{coeffs tmpErr}(k)] = \text{lreg}(\text{Xtmp}, \text{Ytr}, \text{weights}, \text{wd});

\text{tmpCoefs}(:,k) = \text{coeffs};

end

[\text{dummy, ind}] = \text{min}(\text{tmpErr}); \text{find out the best modification}

\text{X}(:,\text{ind+1}) = [];\text{model.coefs} = \text{tmpCoefs}(:,\text{ind});

\text{model.knotdims(\text{ind})} = [];\text{model.knotsites(\text{ind})} = [];\text{model.knotdirs(\text{ind})} = [];\text{model.parents(\text{ind})} = [];

\text{model.parents} = \text{updateParents}(\text{model.parents},\text{ind});

if \text{trainParams.cubic}

\text{t1(\text{ind,:})} = [];\text{t2(\text{ind,:})} = [];

[\text{t1 t2 dif}] = \text{KnotsEstremi}(\text{model}, [], [], \text{d}, \text{minX}, \text{maxX}, \text{t1(i,:)}, \text{t2(i,:)});

% update basis functions with the updated side knots

for \text{i} = 1 : \text{length}(
\text{model.knotdims})

if \text{dif}(\text{i})

\text{X(\text{:},i+1)} = \text{createbasisfunction}(\text{Xtr}, \text{X}, \text{model}.\text{knotdims}\{\text{i}\}, \text{model.knotsites}\{\text{i}\}, \ldots
\text{model.knotdirs}\{\text{i}\}, \text{model.parents}\{\text{i}\}, \text{minX}, \text{maxX}, \text{t1(i,:)}, \text{t2(i,:)})

end

end
end
end
model.t1 = t1;
model.t2 = t2;
end
models{end+1} = model;
mses(end+1) = tmpErr(ind) / n;
gcvs(end+1) = gcv(model, mses(end), n, trainParams_actual_c);

% end of the main loop

if trainParams.maxFinalFuncs >= length(models{1}.coefs)
    [g, ind] = min(gcvs);
elseif trainParams.maxFinalFuncs > 1
    [g, ind] = min(gcvs(end-trainParams.maxFinalFuncs+1:end));
    ind = ind + length(gcvs) - trainParams.maxFinalFuncs;
else
    g = gcvs(end);
    ind = length(gcvs);
end
model = models{ind};

if doCubicFastLevel >= 2
    % turn the cubic modelling on
    trainParams.cubic = true;
    [t1 t2] = KnotsEstremi(model, [], [], d, minX, maxX, [], []);
    % update all the basis functions
X = ones(n, length(model.coefs));

for i = 1 : length(model.knotdims)
    X(:, i+1) = createbasisfunction(Xtr, X, model.knotdims{i}, model.knotsites{i}, ...
    model.knotdirs{i}, model.parents(i), minX, maxX, t1(i,:), t2(i,:));
end

model.t1 = t1;
model.t2 = t2;

[model.coefs model.MSE] = lreg(X, Ytr, weights, wd);
model.MSE = model.MSE / n;
model.GCV = gcv(model, model.MSE, n, trainParams.actual.c);
else
    model.MSE = mses(ind);
    model.GCV = g;
end

if verbose, fprintf('
'); end

end % end of "trainParams.prune"

end % end of "if endSpan*2 >= n"

model.trainParams = trainParams;
model.minX = minX;
model.maxX = maxX;
model.endSpan = endSpan;

time = toc;
if verbose
fprintf('Number of basis functions in the final model: %d\n',
    length(model.coefs));
fprintf('Total effective number of parameters: %0.1f\n', ...
    length(model.coefs) + model.trainParams.c * length(model.
    knotdims) / 2);
maxDeg = 0;
if length(model.knotdims) > 0
    for i = 1 : length(model.knotdims)
        if length(model.knotdims{i}) > maxDeg
            maxDeg = length(model.knotdims{i});
        end
    end
end

warning(ws);
return

% % % % % % % % % % % % % % % % Auxiliary functions

function g = gcv(model, MSE, n, c)
% Calculates GCV

enp = length(model.coefs) + c * length(model.knotdims) / 2; % parametri
    nel modello
if enp >= n
    g = Inf;
else
    p = 1 - enp / n;
end
\[ g = \text{MSE} / (p \times p); \]

end

return

function parents = updateParents(parents, deletedInd)
% Updates direct parent indexes after deletion of a basis function.
parents(parents == deletedInd) = 0;
tmp = parents > deletedInd;
parents(tmp) = parents(tmp) - 1;
return

function basisFunctionList = createList(basisFunctionList_old, Xtr, ... ordinatiXtr, ordinatiXtrInd, n, d, model, BasisAggiunte, trainParams, endSpan)
% Takes the old list of basis functions and adds new ones according to the
% current model. If the old list is empty, adds only linear basis
% functions. If it is non-empty, adds only basis functions with
% interactions which result from the last BasisAggiunte basis functions

% Create linear basis functions

if (isempty(basisFunctionList_old)) && (BasisAggiunte == 0)
basisFunctionList = cell(3,0);
counter = 1;
if trainParams.useMinSpan ~= 1
    % get the list of knot sites allowed due to minSpan
\text{massimi} = \text{mod}(1:n, \text{getMinSpan(d, n, trainParams.useMinSpan)}) == 0;

\text{end}

\text{for} \ di = 1 : d \ %for each dimension
\text{if} \ \text{trainParams.useMinSpan} \neq 1
\ \text{massimiPos} = \text{ordinatiXtr(:,di)};
\text{else}
\ \text{massimiPos} = \text{unique(ordinatiXtr(:,di))};
\text{end}
\text{lastknot} = \text{Inf};
\text{%aggiunge le basis functions alla lista}
\text{for} \ i = 1 : \text{size(massimiPos,1)}
\text{if} \ (\text{trainParams.useMinSpan} == 1) \ || \ ...
\ (((\text{massimi}(i + \text{endSpan}-1)) \&\& (\text{lastknot} \neq \text{massimiPos}(i))
\text{)}
\text{lastknot} = \text{massimiPos}(i);
\text{basisFunctionList\{1, counter\}} = \text{di};
\text{basisFunctionList\{2, counter\}} = \text{lastknot};
\text{basisFunctionList\{3, counter\}} = 1;
\text{basisFunctionList\{4, counter\}} = 0;
\text{counter} = \text{counter} + 1;
\text{end}
\text{end}
\text{return}
\text{end}

\text{if} \ (\text{trainParams.maxInteractions} < 2) \ || \ (\text{BasisAggiunte} < 1)
\text{basisFunctionList} = \text{basisFunctionList\_old};
\text{return}
% Create basis functions with interactions
basisFunctionList = basisFunctionList_old;
counter = size(basisFunctionList_old,2) + 1;
start = length(model.knotdims)–(BasisAggiunte – 1);

% loop through all the basis functions already in the model
for j = start : length(model.knotdims)
    if length(model.knotdims{~}) < trainParams.maxInteractions
        alloweddims = 1 : d;
        if trainParams.selfInteractions <= 1
            % will not consider the already used dimensions
            alloweddims = setdiff(alloweddims, model.knotdims{~});
        else
            for i = 1 : d
                if length(find(model.knotdims{~} == i)) >= trainParams.
                    selfInteractions
                    alloweddims = setdiff(alloweddims, i);
                end
            end
        end
    end
end
if isempty(alloweddims)
    continue
end

if trainParams.useMinSpan ~= 1

nonzero = listNonZero(Xtr, model.knotdims[j], model.knotsites[j], model.knotdirs[j]);

minSpan = getMinSpan(d, length(find(nonzero)), trainParams.useMinSpan);

if ~isfinite(minSpan)
    continue
end

massimi = mod(1:n, minSpan) == 0;

for di = alloweddims %for each dimension
    lastknot = Inf;

    ind = ordinatiXtrInd(nonzero(ordinatiXtrInd(:, di)), di);
    ind = ind(massimi(ind))';

    for i = ind
        if lastknot ~= Xtr(i, di)
            lastknot = Xtr(i, di);
            basisFunctionList{1, counter} = [model.knotdims{j} di];
            basisFunctionList{2, counter} = [model.knotsites{j} lastknot];
            basisFunctionList{3, counter} = [model.knotdirs{j} 1];
            counter = counter + 1;
        end
    end
end

else

end

147
nonzero = listNonZero(Xtr, model.knotdims{j}, model.knotsites{j}, model.knotdirs{j});

for di = alloweddims

massimiPos = unique(ordinatiXtr(nonzero(ordinatiXtrInd(:,di)),di));

for i = 1:size(massimiPos,1)

basisFunctionList{1, counter} = [model.knotdims{j} di];

basisFunctionList{2, counter} = [model.knotsites{j} massimiPos(i)];

basisFunctionList{3, counter} = [model.knotdirs{j} 1];

basisFunctionList{4, counter} = j;

counter = counter + 1;

end

end

end

end

end

end

return

function nonzero = listNonZero(Xtr, knotdims, knotsites, knotdirs)

% Lists nonzero (according to the parent basis function) sites where
% knots % may be placed.

nonzero = true(size(Xtr,1),1);

for j = 1:size(Xtr,1)
for i = 1 : length(knotdims)
    z = Xtr(j,knotdims(i)) - knotsites(i);
    if ((z >= 0) && (knotdirs(i) < 0) || ...
        ((z <= 0) && (knotdirs(i) > 0))
        nonzero(j) = false;
        break;
    end
end
end
return

function s = getEndSpan(d)
    % Calculation of endSpan so that potential knot sites that are too close to
    % the ends of data intervals are not considered.
    % s = floor(3 - log2(0.05/d));
    s = floor(7.32193 + log(d) / 0.69315); % precomputed version
    if s < 1, s = 1; end
    return

function s = getMinSpan(d, nz, param)
    %Determina la posizione per i nuovi knot
    if nz == 0
        s = Inf;
    else
        if param < 0 % automatic
            % s = floor(-log2(-log(1-0.05)/(d*nz)) / 2.5);
            s = floor((2.9702 + log(d*nz)) / 1.7329);
        else
            s = param;
        end
if s > nz
    s = Inf;
end

if s < 1, s = 1; end

return

function [coefs err] = lreg(x, y, w, wd)
% Linear regression (unweighted and weighted)
if isempty(wd)
    coefs = (x’ * x) \ (x’ * y);
    err = sum((y-x*coefs).^2);
else
    x_wd = x’ * wd;
    coefs = (x_wd * x) \ (x_wd * y);
    err = sum((y-x*coefs).^2.*w);
end
return

function grafica(X, yp, Xp, Yp)
% Grafica

i = 1:11:length(yp);
y = yp(i);
x1 = X(1:121:length(yp),1);
x2 = X(1:11:121, 2);
YYY = reshape(y, 11, 11);

figure(1)
surf(x1, x2, YYY)
hold on

j = 1:3:length(Yp); % Takes only the 30% fuel split points
Xs = Xp(j, :);
Ys = Yp(j);

scatter3(Xs(:, 1), Xs(:, 2), Ys, 'filled')
title('HC (ppm) for 30% fuel split');
xlabel('Main SOI (BTDC)')
ylabel('Pilot SOI (BTDC)')
zlabel('HC (ppm)')

% 35% fuel split
clear y x1 x2 YYY Xs Ys i j

i = 6:11:length(yp);
y = yp(i);
x1 = X(6:121:length(yp), 1);
x2 = X(6:11:121, 2);
YYY = reshape(y, 11, 11);
figure(2)
surf(x1, x2, YYY)
hold on
j = 2:3:length(Yp); % Takes only the 35% fuel split points
Xs = Xp(j,:);
Ys = Yp(j);

scatter3(Xs(:,1),Xs(:,2),Ys,'filled')
title('HC (ppm) for 35% fuel split');
xlabel('Main SOI (BTDC)');
ylabel('Pilot SOI (BTDC)');
zlabel('HC (ppm)');

% 40% fuel split

clear y x1 x2 YYY Xs Ys i j

i = 11:11:length(yp);
y = yp(i);
x1 = X(11:121:length(yp),1);
x2 = X(11:11:121,2);
YYY = reshape(y,11,11);

figure(3)
surf(x1,x2,YYY)
hold on

j = 3:3:length(Yp); % Takes only the 40% fuel split points
Xs = Xp(j,:);
Ys = Yp(j);

scatter3(Xs(:,1),Xs(:,2),Ys,'filled')
title('HC (ppm) for 40% fuel split');
xlabel( 'Main SOI ( BTDC)' )
ylabel( 'Pilot SOI ( BTDC)' )
zlabel( 'HC (ppm)' )

%%% MAIN SOI & FUEL SPLIT

clear y x1 x2 YYY Xs Ys i j
clear i k j l d A

%generate the index for selecting the appropriate points
i =12:121:length(yp);
   i=i ';
k=0:10;
k=k ';
k=repmat(k , length(i) ,1) ;
l=1:length(k) ;
l=l ';
j(1 ,1)=i(floor((1-1)/11+1))+k(1); 

%select the points
y=yp(j) ;
x1=X(1:121:length(yp) ,1) ;
x3=X(1:11 ,3) ;
YYY=reshape(y,11,11) ;

figure(4)
surfc(x1 ,x3 ,YYY)
hold on

clear j

b=1:9:length(Yp) ; %Takes only the lower Pilot injection points
b=b ‘;

v=0:2;

v=v ‘;

v=repmat(v,length(b),1);

m=1:length(v);

m = m ‘;

j(m)=b(floor((m−1)/3+1))+v(m);

Xs=Xp(j,:);

Ys=Yp(j);

scatter3(Xs(:,1),Xs(:,3),Ys,'filled')

title('HC (ppm) for 40 BTDC pilot injection');

xlabel('Main SOI (BTDC)')

ylabel('Fuel Split (%)')

zlabel('HC (ppm)')

% % % % % 45 deg BTDC Pilot Injection

clear y x1 x2 YYY Xs Ys i j

clear i k j l d A

generate the index for selecting the appropriate points

i=56:121:length(yp);

i=i ‘;

k=0:10;

k=k ‘;

k=repmat(k,length(i),1);

l=1:length(k);

l=l ‘;

j(1,1)=i(floor((1−1)/11+1))+k(1);
%select the points

y=yp(j);
x1=X(56:121:length(yp),1);
x3=X(1:11,3);
YYY=reshape(y,11,11);

figure(5)
surf(x1,x3,YYY)
hold on
clear j

b=4:9:length(Yp); %Takes only the medium Pilot injection points
b=b';
v=0:2;
v=v';
v=repmat(v,length(b),1);
m=1:length(v);
m=m';
j(m)=b(floor((m-1)/3)+1)+v(m);

Xs=Xp(j,:);
Ys=Yp(j);

scatter3(Xs(:,1),Xs(:,3),Ys,'filled')
title('HC (ppm) for 45 BTDC pilot injection');
xlabel('Main SOI (BTDC)')
ylabel('Fuel Split (%)')
997 zlabel('HC (ppm)')

998

999 \%50\,deg\ BTDC\ Pilot\ Injection

1000 clear y x1 x2 YYY Xs Ys i j

1001 clear i k j l d A

1002 \%generate\ the\ index\ for\ selecting\ the\ appropriate\ points

1003 i=111:121:length(yp);

1004 i=i';

1005 k=0:10;

1006 k=k';

1007 k=repmat(k,length(i),1);

1008 l=1:length(k);

1009 l=l';

1010 j(l,1)=i(floor((l-1)/11+1))+k(l);

1011 %select\ the\ points

1012 y=yp(j);

1013 x1=X(111:121:length(yp),1);

1014 x3=X(1:11,3);

1015 YYY=reshape(y,11,11);

1016 %figure(6)

1017 surf(x1,x3,YYY)

1018 hold on

1019 clear j

1020 %Takes\ only\ the\ medium\ Pilot\ injection\ points

1021 b=7:9:length(Yp);

1022 b=b';

1023 v=0:2;
v=v';

v=repmat(v,length(b),1);

m=1:length(v);
m=m';

j(m)=b(floor((m-1)/3+1))+v(m);

Xs=Xp(j,:);
Ys=Yp(j);

scatter3(Xs(:,1),Xs(:,3),Ys,'filled')
title('HC (ppm) for 50 BTDC pilot injection');
xlabel('Main SOI (BTDC)');
ylabel('Fuel Split (%)');
zlabel('HC (ppm)');

%%%%% 2D Parametric plot

clear y x1 x2 YYY Xs Ys i j
clear i k j l d A

figure(7)
i=11:121:length(yp);
i=i';
y1=yp(i);
x=X(i,1);
i2=56:121:length(yp);
i2=i2';
y2=yp(i2);
i3=121:121:length(yp);
i3=i3';
y3 = yp(i3);
plot(x, y1, '-g+');
title('Experimental vs Fitted data points');
xlabel('Main SOI');
ylabel('HC (ppm)');
hold on
plot(x, y2, '-ro');
hold on
plot(x, y3, '-ms');
hold on
legend('40 pilot SOI and 40% fuel split', '45 pilot SOI and 40% fuel split', '50 pilot SOI and 40% fuel split')
hold on
j = 3:9:length(Yp);
Xs = Xp(j, 1);
Ys = Yp(j);
scatter(Xs, Ys, 100, 'g+');
hold on
j = 6:9:length(Yp);
Xs = Xp(j, 1);
Ys = Yp(j);
scatter(Xs, Ys, 100, 'ro');
hold on
j = 9:9:length(Yp);
Xs = Xp(j, 1);
Ys = Yp(j);
scatter(Xs, Ys, 100, 'ms');
References


160

[23] P. Risberg, *Describing the Auto-Ignition Quality of Fuels in HCCI Engines*, Doctoral thesis, School of Industrial Engineering and Management, Royal Institute of Technology, S-100 44 Stockholm


[38] Horiba *Non-Sampling Type NOx-A/F Analyzer MEXA-720NOx, User Manual* 2003


[57] J. Friedman *Fitting functions to noisy data in high dimensions* Computer Science and Statistics. 1988


