



FINAL REPORT

AVFL-23: DATA MINING OF FACE DIESEL ENGINE FUELS

R. Gieleciak, C. Fairbridge
NATURAL RESOURCES CANADA, CanmetENERGY in DEVON

Work performed for:
COORDINATING RESEARCH COUNCIL

JULY 2016

NATURAL RESOURCES CANADA
DIVISION REPORT CDEV-2016-0038-RE

DISCLAIMER

This report and its contents, the project in respect of which it is submitted, and the conclusions and recommendations arising from it do not necessarily reflect the views of the Government of Canada, its officers, employees, or agents.

EXECUTIVE SUMMARY

This report summarizes the results of applying advanced statistical techniques to identify trends and correlations of the engine performance and emissions of the Fuels for Advanced Combustion Engines (FACE) diesel fuels with fuel properties and operating conditions (CRC Project AVFL-23: Data Mining of FACE Diesel Fuels). In 2009 the Coordinating Research Council (CRC) FACE Working Group designed the FACE research matrix of nine diesel fuels to enable the investigation of the influence of cetane number, aromatic content, and 90% distillation temperature (T90) on advanced combustion strategies. These parameters: ignition quality, chemistry, and volatility, were carefully selected in order to represent key diesel fuel characteristics.

In a previous CRC project (FACE-1), the physical and chemical properties of the FACE diesel fuels were extensively characterized. In addition to standard ASTM tests, advanced analytical methods, such as comprehensive two-dimensional gas chromatography (GC×GC) and nuclear magnetic resonance (NMR) spectroscopy, were employed to better elucidate the chemistry of these fuels and thus enable engine and combustion researchers to determine the effects of FACE diesel fuel hydrocarbon composition on combustion and emissions characteristics of advanced combustion engines.

Since then, the FACE diesel fuels have been used in many engine experiments, including a homogeneous charge compression ignition (HCCI) single-cylinder engine (SCE) and a high-speed direct-injection (HSDI) engine at Oak Ridge National Laboratory; a DI diesel SCE and Cooperative Fuel Research (CFR) engine at National Research Council Canada (NRCC); a DI diesel SCE at Navistar; a light-duty compression-ignition engine at West Virginia University (CRC AVFL-16 project), and a heavy-duty optical engine for fuel-effects studies at Sandia National Laboratories.

In the present study, to further correlate the enormous quantities of fuel property and engine data, ‘smart’ chemometrics strategies were developed and applied. Principal component analysis (PCA), which was used to explore and visualize the information included in the chemical and engine database space,

highlighted the similarities and differences in fuel behaviours in the various engines. The partial least squares (PLS) regression technique was also used to create robust and statistically-reliable models correlating engine input parameters and fuel characteristics with engine output parameters (i.e., CO, HC, NO_x, and soot emissions; indicated specific fuel consumption (ISFC); and indicated thermal efficiency (ITE)). The investigation of the regression coefficients (b), variance in projection (VIP) or selectivity ratio (SR) calculated by the PLS procedure provided information about the signs and influences of engine control input variables on the engine-related output parameters.

The results demonstrate that engine-out emissions and efficiency profiles depend strongly on engine operating parameters and to some extent on the chemical and physical properties of the diesel fuels. A summary of the PLS models containing the 3 fuel design + engine control parameters (model 2a) is tabulated below. The filled circles (•) represent highly influential parameters and the open circles (◦) represent moderately influential parameters. Plus (+) symbols indicate parameters are directly correlated while negative (-) symbols indicate inverse correlations. Of the 3 FACE Diesel Fuel design parameters, CN appears to have the most influence on emissions and engine performance. More complex models containing 40 fuel parameters plus the engine control parameters (designated as models 3a and 3b) indicate that the number of secondary carbons in alkyl chains (determined by NMR) and aromatic carbons attached to alkyl groups are influential parameters on emissions.

HCCI-SCE 0.5L Engine (ORNL):

Parameter	HC	CO	NOx	ISFC	ITE
CN	● (-)	● (-)	● (+)	● (+)	● (-)
T90	○ (+)	(+)	(-)	● (+)	● (-)
Aromatics	○ (-)	(-)	(-)	(+)	(-)
fuel rate	(+)	(+)	● (+)	(+)	(-)
airflow	○ (+)	(+)	(-)	● (+)	(-)
lambda	● (+)	● (+)	● (-)	(+)	(-)
intake T	● (+)	● (-)	(+)	● (+)	● (-)
intake P	● (-)	○ (-)	○ (+)	(-)	(+)

LTC MCE 1.9L Engine (HECC-ORNL):

Parameter	HC	CO	NOx	PM	ITE
CN	● (-)	● (-)	● (+)	● (+)	(-)
T90	(-)	(-)	(+)	(+)	(+)
Aromatics	(-)	(-)	(+)	(+)	(-)
aicDynoTorque	● (-)	● (-)	○ (+)	(+)	(-)
aicEGRMassGM	○ (-)	(+)	● (-)	● (+)	● (+)
aicINTO2GM	(-)	(-)	(+)	(+)	(+)
MainSOI(ATDC)	● (+)	● (+)	● (-)	● (+)	● (-)

CDC-SCE 0.5L Engine (CDC-ORNL):

Parameter	HC	CO	NOx	Soot	ISFC	ITE
CN	● (-)	● (-)	(+)	(+)	● (-)	● (+)
T90	(+)	(+)	(-)	(-)	(-)	(+)
Aromatics	(+)	○ (+)	(+)	(+)	(+)	(-)
Throttle (mA)	● (-)	(+)	● (-)	● (-)	● (-)	● (+)
TCHAR (°C)	○ (+)	● (+)	(+)	(+)	○ (+)	● (-)
Airflow	(+)	● (-)	○ (+)	(+)	○ (-)	○ (+)
Fuelflow	● (-)	(+)	● (+)	● (+)	● (+)	● (-)
AFR Flow	● (-)	○ (+)	● (-)	● (+)	● (+)	● (-)
Vol.Effic.	(+)	● (-)	○ (+)	(+)	○ (-)	○ (+)

LTC Split Injection Engine (WVU-AVFL-16):

Parameter	HC	CO	NOx	Soot	BTE
CN	● (-)	● (-)	● (+)	● (+)	● (+)
T90	(+)	(-)	(+)	(+)	(-)
Aromatics	(-)	(-)	(+)	(+)	(-)
Main SOI (BTDC)	● (-)	● (-)	● (+)	● (-)	● (+)
PilotSOI (BTDC)	● (+)	● (+)	● (-)	● (-)	● (-)
Fuel Split(%Pilot)	(+)	(-)	(+)	(+)	(-)

Sandia Heavy Duty Optical Engine:

Parameter	HC	CO	NOx	FSN	ETAF	ETAC
CN	● (-)	● (-)	(-)	● (+)	(-)	● (+)
T90	(-)	(+)	(+)	(-)	(+)	(-)
Aromatics	○ (-)	(-)	(+)	○ (+)	(+)	○ (+)
inj rail P	(-)	(-)	(+)	● (-)	● (+)	(+)
O2 mole fraction	(-)	● (-)	● (+)	○ (-)	(-)	○ (+)

● - highly influential parameter

○ - moderately influential parameter

(+) - directly correlated

(-) - inversely correlated

CONTENTS

DISCLAIMER	I
EXECUTIVE SUMMARY	II
ABBREVIATIONS	XVII
1.0 INTRODUCTION	1
2.0 NUMERICAL ANALYSIS	2
2.1 PRINCIPAL COMPONENT ANALYSIS	2
2.2 MULTILINEAR REGRESSION.....	3
2.3 PARTIAL LEAST-SQUARES REGRESSION	3
2.4 MODEL VALIDATION	4
2.5 IMPORTANCE OF VARIABLES IN REGRESSION EQUATIONS.....	5
3.0 CHEMICAL COMPOSITION AND PHYSICAL PROPERTIES OF FACE DIESEL FUELS – CURRENT STATUS	5
4.0 EXPLORATION OF CHEMICAL/PHYSICAL SPACE OF FACE DIESEL FUELS.....	15
5.0 FINDING RELATIONSHIPS BETWEEN PHYSICAL PROPERTIES AND COMPOSITIONAL DATA	29
6.0 ENGINE DATASETS	35
6.1 MODELING STRATEGIES	36
7.0 ENGINE DATASETS – ANALYSIS	40
7.1 HCCI-SCE 0.5L (ORNL).....	40
7.1.1 DATA DESCRIPTION	40
7.1.2 EXPLORATORY ANALYSIS	40
7.1.3 MODELING APPROACH.....	47
7.1.4 HYDROCARBON EMISSIONS	48
7.1.5 CARBON MONOXIDE EMISSIONS	50
7.1.6 NITROGEN OXIDES EMISSIONS	52
7.1.7 SOOT EMISSIONS	54
7.1.8 INDICATED SPECIFIC FUEL CONSUMPTION (ISFC).....	54

7.1.9	THERMAL EFFICIENCY	55
7.1.10	SUMMARY OF PLS MODEL 2A RESULTS	56
7.2	LTC – MCE 1.9L (HECC-ORNL)	57
7.2.1	DATA DESCRIPTION	57
7.2.2	EXPLORATORY ANALYSIS	57
7.2.3	MODELING APPROACH.....	62
7.2.4	HYDROCARBON EMISSIONS	64
7.2.5	CARBON MONOXIDE EMISSIONS	67
7.2.6	NITROGEN OXIDES EMISSIONS	70
7.2.7	SOOT EMISSIONS	73
7.2.8	INDICATED SPECIFIC FUEL CONSUMPTION.....	75
7.2.9	THERMAL EFFICIENCY	75
7.2.10	SUMMARY OF PLS MODEL 2a RESULTS.....	78
7.3	CDC – SCE 0.5L (CDC-ORNL)	79
7.3.1	DATA DESCRIPTION	79
7.3.2	EXPLORATORY ANALYSIS	80
7.3.3	MODELING APPROACH.....	83
7.3.4	HYDROCARBON EMISSIONS	84
7.3.5	CARBON MONOXIDE EMISSIONS	86
7.3.6	NITROGEN OXIDES EMISSIONS	88
7.3.7	SOOT EMISSIONS	89
7.3.8	INDICATED SPECIFIC FUEL CONSUMPTION.....	91
7.3.9	THERMAL EFFICIENCY	92
7.3.10	SUMMARY OF PLS MODEL 2a RESULTS.....	93
7.4	LTC – SPLIT INJECTION (LTC-AVFL-16).....	94
7.4.1	DATA DESCRIPTION	94
7.4.2	EXPLORATORY ANALYSIS	95
7.4.3	MODELING APPROACH.....	97
7.4.4	HYDROCARBON EMISSIONS	98
7.4.5	CARBON MONOXIDE EMISSIONS	100
7.4.6	NITROGEN OXIDES EMISSIONS	102

7.4.7	SOOT EMISSIONS	104
7.4.8	INDICATED SPECIFIC FUEL CONSUMPTION.....	106
7.4.9	THERMAL EFFICIENCY	106
7.4.10	SUMMARY OF PLS MODEL 2a RESULTS.....	108
7.5	SANDIA HEAVY-DUTY OPTICAL ENGINE	110
7.5.1	EXPLORATORY ANALYSIS	111
7.5.2	MODELING APPROACH.....	115
7.5.3	INDICATED SPECIFIC HYDROCARBON EMISSIONS (ISHC).....	116
7.5.4	INDICATED SPECIFIC CARBON MONOXIDE EMISSIONS (ISCO).....	118
7.5.5	INDICATED SPECIFIC NITROGEN OXIDES EMISSIONS (ISNOX).....	120
7.5.6	SMOKE EMISSIONS	122
7.5.7	INDICATED FUEL-CONVERSION EFFICIENCY (ETAF)....	123
7.5.8	SUMMARY OF PLS MODEL 2a RESULTS.....	124
8.0	SUMMARY AND CONCLUSIONS	126
9.0	RECOMMENDATIONS	130
10.0	ACKNOWLEDGEMENTS.....	130
11.0	REFERENCES	131
	APPENDIX A: PCA INTRODUCTION AND EXAMPLE	134
	APPENDIX B: GC×GC ANALYSIS OF FACE DIESEL FUELS (TAKEN FROM REF 25).....	140
	APPENDIX C: PCAVISTOOL TUTORIAL	145
	C1. INTRODUCTION.....	147
	C2. INSTALLATION.....	147
	C3. WORKING WITH PCAVISTOOL	148
	C3.1. PCA MODELS CONTAINER.....	148
	C3.2. LOADING PCA MODELS	149
	C3.3. PLOTTING OPTIONS	150
	C3.4. FIGURE ACTIONS	154

C3.5. LIST OF VARIABLES	158
C3.6. SAVING IMAGES	161
C4. SOFTWARE SUPPORT	162
APPENDIX D: BILOTS	163
APPENDIX E: MATRIX CORRELATION (RV-COEFFICIENT)	166
APPENDIX F: PREPROCESSING METHODS	169
APPENDIX G: BOX-PLOTS	173
APPENDIX H: PLS MODELS DETAILS	176
APPENDIX I: THE LIST OF INPUT/OUTPUT PARAMATERES COLLECTED DURING SANDIA HEAVY-DUTY OPTICAL ENGINE EXPERIMENTS	195

TABLES

Table 1 – Thermophysical properties of FACE diesel fuels: thermal diffusivity (D) and temperature coefficient of refractive index ($-dn/dT$)	15
Table 2 – List of 28 compositional, chemical, and physical datasets available for the FACE diesel fuels	17
Table 3 – PCA projection results for selected compositional, chemical, and physical datasets available for FACE diesel fuels	19
Table 4 – Correlation results for selected compositional, chemical, and physical datasets available for the FACE diesel fuels	31
Table 5 – Engine specifications and operating conditions (based on the table presented in (35))	35

FIGURES

Figure 1 – FACE cube with GC×GC data plotted on the edges (Fuel FD9 not shown here but is presented in Appendix B). Hydrocarbon types are colour-coded. The size of the bubble (area) represents the concentration of a particular compound. On each GC×GC bubble
--

plot the x-axis corresponds to the boiling point scale (i.e. the boiling point increases from left to right). For higher resolution chromatograms please see figures in Appendix B.....	7
Figure 2 – Comparison of saturate and aromatic contents reported by different chromatographic methods for the FACE diesel fuels. The two numbers (xxx/yyy) located above each bar set correspond to calculated average and median concentrations, respectively, for particular hydrocarbon groups	9
Figure 3 – Comparison of paraffinic, cycloparaffinic, mono, and diaromatic contents for the FACE diesel fuels. The two numbers (xxx/yyy) located above each bar set correspond to calculated average and median concentrations, respectively, for particular hydrocarbon groups	10
Figure 4 – Comparison of iso- and normal paraffin contents reported by different chromatographic methods for the FACE diesel fuels. The two numbers (xxx/yyy) located above each bar set correspond to calculated average and median concentrations, respectively, for particular hydrocarbon groups	11
Figure 5 – Molecular representation (aka super-molecule) of carbon types quantified using NMR data	12
Figure 6 – FACE fuels according to ‘super-molecule’ representation	13
Figure 7 – Comparison of NMR results reported by the concentration (mol%) of carbon types (CTi) for FACE diesel fuels. The CTi labels correspond to the labels in Figure 5.....	13
Figure 8 – NIR spectra (3400-10800 cm ⁻¹) of FD1-FD4 (black, low CN fuels) and FD5-FD9 (red, high CN fuels). The inset presents expanded NIR region with depicted C-H stretching bands; sym (symmetric), asym (asymmetric), arom (aromatic).	14
Figure 9 – Matrix of FACE diesel fuels. The blue cube shows the target property values. The red cube presents the measured property values.	16

Figure 10 – Hydrocarbons by FIA plotted in three-dimensional space described by aromatics, saturates, and olefins content	18
Figure 11 – FACE-cube-like PCA representation of the major datasets discussed in this section, where a) represents the FACE cube from experimental design, b) its real-world version, and 8 cubes below show 3D scores plots made by plotting each fuel according to the values of its first three PCs calculated for eight selected datasets.	27
Figure 12 – Correlation map of RV-coefficients calculated for all 23 datasets presented in Table 2 (dataset 1-23). Datasets 24 and 25 are the original FACE diesel fuels target property data and experimentally measured property values (see Figure 9 for reference), respectively. In this plot, bubble size and colour indicate the strength of the correlation between respective pairs of datasets.....	28
Figure 13 – The correlation matrix between all physical properties measured for FACE diesel fuels (a). Selected significant correlations ($p < 0.05$) (b).	29
Figure 14 – Examples of strong positive (a) and negative (b) correlations between physical properties	30
Figure 15 – External validation of selected regression models presented in Table 4 on two datasets of fuels (33, 34) for selected physical properties: Net heat of combustion (a), API gravity (b), bromine number (c), cloud point (d), kinetic viscosity (e), cetane number by engine method (f), flash point (g), and HFRR lubricity (h).....	34
Figure 16 – Set of six modeling strategies used for evaluation of engine control parameters and fuel properties effects on engine emission and efficiency. Where m - total number of experimental points collected for a particular engine, n - the number of engine control parameters, k - total number of variables including cross terms for all variables, Y - dependent variable (engine emission or efficiency).	36

Figure 17 – The list and pictorial description of the 40 selected fuel variables used in modeling approach	38
Figure 18 – PLS modeling results – an example	39
Figure 19 – Scheme presenting list of control input (I*) and output (O*) engine parameters for HCCI engine dataset	40
Figure 20 – Scatter representation of cross-correlation between control input engine parameters where (I1) fuel rate (g/min), (I2) airflow (g/min), (I3) lambda (corrected), (I4) actual intake temp (°C), (I5) intake pressure (psi). Histograms for each variable are located on the diagonal. The inset in the top-right corner presents the relation between lambda and actual intake temperature. Colour of the points in the inset represents the respective FACE fuel.	41
Figure 21 – Scatter representation of cross-correlation between output engine parameters obtained for engine run with fuel rate ~ 6.2 g/min. The meaning of each output (O*) is presented in Figure 19. The inset located in the top-right corner magnifies the existing relations between combustion phasing for O8, the 50% burn point (MFB50 aka CA50) and NOx emissions (O4).	42
Figure 22 – PCA scores (a) and loadings (b) plot (PC1-PC2) obtained after running PCA for HCCI ORNL engine input (control) parameters. On the scores plot, bubble colour and size are associated with the respective fuels (i.e., smallest dark blue dots – FD1, largest brown bubbles – FD9).....	43
Figure 23 – PCA scores plots (a) and loadings plots (b) (PC1-PC2) obtained after running PCA for the HCCI ORNL engine output parameters for all events. On the scores plots, bubble colour and size are associated with the respective fuels (i.e., smallest dark blue dots – FD1, largest brown bubbles – FD9).....	44
Figure 24 – PCA scores plots for engine output parameters measured for fuel flow rates (a) ~6.3 g/min, (b) ~8.4 g/min, (c) ~10.5 g/min, (d)	

~12.5 g/min. The objects were assigned to three clusters (represented by color coded circles).	45
Figure 25 – Scheme illustrating influence of engine output (ISFC) and chemistry of the fuels (T30) on observed engine object splitting in PCA scores plot for engine output parameters measured for fuel rate ~12.5 g/min	46
Figure 26 – Scatter representation of cross-correlation between output engine parameters (O^*) and engine control variables (I^*) obtained for engine run with a fuel flow rate ~ 6.3 g/min. The insets on the right side present two examples of the bivariate relationships coloured by fuel number.	47
Figure 27 – Scheme presenting list of control input (I^*) and output (O^*) engine parameters for LTC- MCE 1.9L (as provided by ORNL researchers and placed in FACE diesel engine database)	57
Figure 28 – Scatter representation of cross-correlation between control input engine parameters, where meanings of the labels (I^*) are given in Figure 27. Histograms for each variable are located on the diagonal. The inset in the top-right corner magnifies the existing relations between <i>aic EGR Mass GM</i> parameter and start of injection (SOI) timing.	58
Figure 29 – Scatter representation of cross-correlation existing among output engine parameters obtained for HECC engine, where numbers indicate consecutive output parameters (O^* - see Figure 27). The inset located in the top-right corner magnifies the existing relations between ignition delay and NO _x emissions.	58
Figure 30 – PCA scores plots (a) and loadings plot (b) (PC1-PC2) obtained after running PCA for HECC ORNL engine input (control) parameters. On the scores plot, bubble colour and size are associated with respective fuels (i.e., smallest dark blue dots – FD1, largest brown bubbles - FD9).	59

- Figure 31 – PCA score (a) and loading (b) plots (PC1-PC2) obtained after running PCA for ORNL HCCI engine output parameters for all events. On the scores plot, bubble colour and size are associated with respective fuels (i.e., smallest dark blue dots – FD1, largest brown bubbles – FD9).60
- Figure 32 – Projection of select engine output parameters into objects located in the scores plots where scores plot objects are coloured by values of (a) ignition delay, (b) CO emissions, (c) CA50, (d) particulate matter, (e) brake thermal efficiency, and (f) NOx emissions61
- Figure 33 – After removing events located in groups 1 and 2 (two groups on the top – behind the colour bar) and calculating two PCA models for both groups 4 and 3, the resulting scores plots show a characteristic distributions of events. Points on the plot are coloured by fuel type (see respective colour bar).62
- Figure 34 – Scatter representation of cross-correlations between output engine parameters (O^*) and engine control variables (I^*) obtained for HECC engine (for meaning of O^* and I^* see Figure 27). I4 is a SOI timing control variable and plots similar to these presented may be found in the original paper (5). Note the strong correlation between O11 (BTE) vs I1(aicDynoTorque).63
- Figure 35 – Scheme presenting list of control input (I^*) and output (O^*) engine parameters for CDC- SCE 0.5L (as provided by ORNL researchers and located in FACE diesel engine database).....79
- Figure 36 – Scatter representation of cross-correlation between control input engine parameters. For meaning of labels (I^*) on both x and y axes please refer to the Figure 35. Histograms for each variable are located on the diagonal. The inset in the top-right corner magnifies the existing relations between air flow and fuel flow. ..80
- Figure 37 – The scatter representation of cross-correlation existing among selected output engine parameters obtained for CDC engine, where

- numbers indicate consecutive output parameters (O^* - see Figure 35). The inset located in the top-right corner magnifies the existing relations between SOI_c -SOC (aka ignition delay) and IMEP.81
- Figure 38 – The PCA scores (a) and loadings (b) plot (PC1-PC2) obtained after running PCA for HECC ORNL engine input (control) parameters. On scores plot bubbles colour and size are associated with respective fuels (i.e., smallest dark blue dots – FD1, largest brown bubbles – FD9).82
- Figure 39 – The PCA score (a) and loading (b) plots (PC1-PC2) obtained after running PCA for CDC ORNL engine output parameters for all events. On scores plot bubble colour and size are associated with respective fuels (i.e., smallest dark blue dots – FD1, largest brown bubbles – FD9).82
- Figure 40 – Scatter representation of cross-correlation between output engine parameters obtained for CDC engine run where I^* are control variables and O^* indicate output engine parameters.83
- Figure 41 – Scheme presenting list of control input and output engine parameters for light-duty compression-ignition engine run in split injection control strategy.....95
- Figure 42 – (a) Scatter representation of cross-correlation between control input engine parameters. For meaning of labels (I^*) on both x and y axes please refer to the Figure 41. Histograms for each variable are located on the diagonal. (b) The design cube for split injection strategy, where points indicate specific conditions used for each engine run with a particular fuel (categorized as low-, medium-, and high-CN fuels).....95
- Figure 43 – Scatter representation of cross-correlation existing among output engine parameters obtained for CDC engine, where numbers indicate consecutive output parameters (O^* - see Figure 41). Histograms for each variable are located on the diagonal. The

inset in the top-right corner magnifies the high correlation between hydrocarbon and carbon monoxide emissions.....	96
Figure 44 – PCA scores plot (PC1-PC2-PC3) obtained after running PCA for LTC-AVFL engine output parameters for all events (a), and events characterized by the same control input parameter values (b). On the scores plots, colour and marker type are associated with respective fuel set (low-, medium-, or high-CN fuel).....	97
Figure 45 – Scatter representation of cross-correlation between output engine parameters obtained for LTC-AVFL engine runs, where I* are control variables and O* indicate output engine parameters (see Figure 41).....	98
Figure 46 – The Sandia engine dataset cleaning process results in classification of all 238 parameters into four groups of variables. Examples of parameters classified in each set of variables are presented in colour-coded boxes.....	111
Figure 47 – Sandia experimental design. Six measurement points for each of five fuels resulted in 30 test conditions.	111
Figure 48 – The PCA scores plots (PC1-PC2-PC3) obtained after running PCA for Sandia engine input (control) parameters. The points on the dataset are coloured according to intake oxygen mole fraction (X_{O_2}) (a) and fuel type (b). The PCA scores plots (PC1-PC2) (c) present two main groups of points: low (LP) and high (HP) injection pressure. The PCA scores plots (PC1-PC3) (d) groups engine inputs based on fuel type and oxygen dilution.	112
Figure 49 – PCA scores (a) and loadings plots (b) (PC3-PC4) of the Sandia engine input dataset, where pLPfuel=low-pressure fuel pressure [psig], TuAir/N2=air/nitrogen temperature upstream of orifice [°C], fueldens=fuel density [kg/m ³], pdCO2=CO2 pressure downstream of orifice [bar], injlag=injector lag [s], DOIaCA=actual duration of injection [CAD], fuelLHV=lower heating value of fuel [MJ/kg].....	113

- Figure 50 – Bar plots of parameters having the greatest influence on the observed grouping of fuels in the PCA scores plot (PC3-PC4)...114
- Figure 51 – PCA scores plots (PC1-PC2-PC3) obtained after running PCA for Sandia engine output parameters, coloured according to intake oxygen mole fraction (X_{O_2}) (a) and PCA scores plots (PC1-PC2), coloured according to fuel type (b). HP and LP denote high and low injector pressures, respectively.115
- Figure 52 – Scatter representation of cross-correlation between output and experimentally designed engine input parameters obtained for Sandia engine runs. On the right is an example of the relationship between charge-dilution and indicated specific NO_x emissions for engine events run with high injection-pressure.115
- Figure 53 – PCA projection of input engine data (IN) and output engine data (OUT) into plane spanned by first two principal components. Engine events operating on low CN (CN<40) are plotted as blue points, and engine events operating on high CN (CN>40) are plotted as red points.127

ABBREVIATIONS

Aopt	the optimal number of PLS components
Aro	aromatic content
AVFL	Advanced Vehicle, Fuel, and Lubricants
ATDC	after top-dead-center
<i>b</i>	PLS/MLR regression coefficient
B(I)SFC	break (indicated) specific fuel consumption
B(I)TE	break (indicated) thermal efficiency
BTDC	before top-dead-center
CAD	crank angle degree
CDC	conventional diesel combustion
CFA	2007 #2 ULSD emissions certification fuel, batch A
CN	cetane number
(IS)CO	indicated specific carbon monoxide emissions
CFR	cooperative fuel research
CR	compression ratio
CRC	Coordinating Research Council (www.crcao.com)
CV	cross-validation
DI	direct injection
DM	data mining
FACE	fuels for advanced combustion engines
FID	flame ionization detection
FIMS	field ionization mass spectrometry
GC×GC	comprehensive two-dimensional gas chromatography
(IS)FC	Indicated specific fuel consumption
(IS)HC	indicated specific hydrocarbon emissions
HCCI	homogenous charge compression ignition
HECC	high-efficiency clean combustion
HSDI	high-speed direct-injection
IR	infrared spectroscopy
IMEP	indicated mean effective pressure
LLNL	Lawrence Livermore National Laboratory

LOO-CV	leave-one-out cross-validation
LTC	low temperature combustion
LTHR	low temperature heat release
MCE	multi-cylinder engine
MFBXX	crank angle degree at which XX% of mass of fuels has been burned
MLR	multiple linear (or multilinear) regression
NIR	near-infrared spectroscopy
(IS)NO _x	indicated specific nitrogen oxides emissions
NMR	nuclear magnetic resonance spectroscopy
NRC	National Research Council of Canada
ORNL	Oak Ridge National Laboratory
PCA	principal component analysis
PCCI	premixed charge compression ignition
PLS	partial least-squares regression
PM	particulate matter
PNNL	Pacific Northwest National Laboratory
R	correlation coefficient
RMS	root mean square error
RMSECV	root mean square error of cross-validation
RMSEP	root mean square error of prediction
RV	multivariate generalization of squared correlation coefficient
SCE	single-cylinder engine
SOI	start of injection
SOC	Start of combustion
SR	selectivity ratio
TOF-MS	time-of-flight mass spectrometry
TXX	distillation characteristics. Temperature at which XX% of fuel is distilled
ULSD	ultra-low sulphur diesel fuel
UV-VIS	ultraviolet-visible spectroscopy
VIP	variable importance in projection

1.0 INTRODUCTION

In 2009 the Coordinating Research Council (CRC) Fuels for Advanced Combustion Engines (FACE) working group designed the FACE research matrix of nine diesel fuels to enable the investigation of the influence of cetane number (CN), aromatic content, and 90% distillation temperature (T90) on advanced combustion strategies (1, 2). These parameters were carefully selected to represent key diesel fuel characteristics: ignition quality, chemistry, and volatility, respectively.

In a previous CRC project (FACE-1) (1) the physical and chemical properties of the FACE diesel fuels were extensively characterized. Advanced analytical methods, such as comprehensive two-dimensional gas chromatography (GC×GC) and nuclear magnetic resonance (NMR) spectroscopy, were employed to enable engine and combustion researchers to determine the effects of diesel fuel hydrocarbon composition on combustion and emissions characteristics of advanced combustion engines.

Since then, the FACE diesel fuels have been used in many engine experiments, including a heavy-duty optical engine for fuel-effects studies at Sandia National Laboratories (3); a homogeneous charge compression ignition (HCCI) single-cylinder engine (SCE) (4), a high-speed direct-injection (HSDI) engine (5) and a conventional diesel compression ignition engine at Oak Ridge National Laboratory (6); a direct injection (DI) diesel SCE and Cooperative Fuel Research (CFR) engine at National Research Council Canada (NRCC) (7, 8); a DI diesel SCE at Navistar (9); and a light-duty compression-ignition engine at West Virginia University (10).

The amount of data generated over the years for the FACE diesel fuels by both physical/compositional analysis and engine-based research is enormous. In the current study we collected much of the available and disseminated datasets and information in order to develop correlations for those data using tools developed in the data mining (DM) field.

DM is commonly defined as a process of finding and extracting relationships or patterns from dataset repositories. Such patterns and models derived from high-dimensional data summarize low-level information and convert it to high-level, potentially useful knowledge. DM has been growing in popularity and has been extensively applied in many fields of science, including agriculture, meteorology, industry, and engineering (11, 12).

The current investigation had two main goals. The first goal was to perform ‘data mining’ and statistical analysis of the massive amount of both standard and advanced analytical data existing for the nine FACE diesel fuels. The second goal was to identify relationships that may exist between engine output properties (i.e., NO_x, CO, HC, soot, fuel consumption, and thermal efficiency) and engine control parameters alone or combined with the physical and chemical properties of the investigated FACE fuels.

Principal component analysis (PCA), multilinear regression (MLR), and partial least-squares (PLS) regression enhanced and extended the interpretation of the data. The results from these numerical analyses confirmed the existence of meaningful relationships between fuel composition and engine output properties.

2.0 NUMERICAL ANALYSIS

The statistical calculations were performed using CanmetENERGY in-house programs developed using MATLAB (13). All MATLAB functions and m-scripts are available from the authors upon request.

2.1 PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a statistical method to extract the systematic variance from a large data set into a few coherent parameters called principal components (PCs) (14, 15). The data are treated as a matrix (**X**), composed of m objects (samples, molecules, or engine events) and n variables (fuel descriptors, engine control or output parameters). PCA decomposes this matrix into k score vectors and k loading vectors, **t** ($m \times k$) and **p** ($k \times n$), respectively, plus the unexplained variance, the residuals, **E** ($m \times n$). Normally only a few PCs are scientifically explainable, the rest usually describe noise or other analytical variations. The equation for the PCA model is:

$$\mathbf{X} = \mathbf{t} \times \mathbf{p} + \mathbf{E} \quad \text{Eq 1}$$

A practical example of the application of PCA with an explanation of score and loading plots, is presented in Appendix A.

2.2 MULTILINEAR REGRESSION

Multilinear regression (MLR) is a common and well-known method used in statistical studies. Equations linking the explanatory variables to the experimental response property are developed with the following format (Eq 2):

$$Y = b_0 + b_1x_1 + \dots + b_nx_n \quad \text{Eq 2}$$

where Y is the modeled property (dependent variable), x_n are the independent variables, and n is the total number of independent variables. The intercept (b_0) and the regression coefficients (b_i) are determined using the least-squares method.

2.3 PARTIAL LEAST-SQUARES REGRESSION

Partial least-squares (PLS) regression is a well-recognized statistical method widely used to solve multivariate calibration problems in analytical chemistry, econometrics, bioinformatics, and engineering. PLS is a generalization of MLR and can handle data having correlated, noisy, or numerous variables that make up a matrix X (16). In the work presented in this report, X consists of either physicochemical properties, engine control parameters, or both fuel chemistry and engine parameters. Y contains one of the engine output parameters (i.e., CO/HC/NO_x/soot emission, ISFC, etc.) that has been transformed.

As the PLS algorithm is well established, and is one of the most commonly used in multivariate calibration problems, only the method will be described here in detail. For further reading refer to references 17 and 18.

PLS gives a compact solution that is statistically more robust than that provided by MLR. The linear PLS model finds “new variables” (PLS factors), which are linear combinations of the original variables. In contrast to PCA, PLS factors are created in such a way as to maximize the covariance between variables and the correlated property (dependent variable).

To avoid overfitting, a strict test for the significance of each consecutive PLS component is necessary; the PLS procedure is stopped when the components are non-significant. Cross-validation is a practical and reliable method for testing this significance (19). Application of PLS thus allows the construction of larger regression model equations while eliminating most variables to avoid overfitting. PLS is typically used in combination with cross-validation to

obtain the optimum number of components. This ensures that the PLS models are selected based on their ability to predict the data rather than to fit the data.

2.4 MODEL VALIDATION

The calibration and predictive capability of a PLS (or MLR) model must be tested through model validation. To achieve validation, the available dependent and independent data were divided into multiple training and validation sets using the Monte-Carlo simulation approach (20). Then PLS models were established on the training dataset and internal validation (based on cross-validation) was performed, together with external validation (based on the test set) to test the predictive capability of the resulting PLS models.

The internal predictive ability of a model was evaluated by “leave-one-out” cross-validation (LOO-CV) (q^2_{model}) on the training set, which was calculated using Eq 3, as well as in terms of RMSECV (root mean square error of cross-validation), (Eq 4):

$$q^2_{\text{model}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{Eq 3}$$

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad \text{Eq 4}$$

where y_i , \hat{y}_i , and \bar{y} are the experimental, LOO-CV predicted, and averaged property values of the objects for the training set, respectively; n is the number of objects in the model set used for generating the PLS (or MLR) model.

The predictive capability of a model on the external prediction set can be expressed by the q^2_{test} using Eq 5; it can also be evaluated by means of RMSEP (root mean square error of prediction), see Eq 6:

$$q^2_{\text{test}} = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad \text{Eq 5}$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m}} \quad \text{Eq 6}$$

where y_i , \hat{y}_i , and \bar{y}_i are the experimental, predicted, and averaged property values of the objects for the test set, respectively; m is the number of objects in the test set – not involved in the model construction.

The resulting PLS models were tested for their statistical validity and robustness by applying a Y-scrambling technique (21). About 1000 random permutations of the response variables were computed and used to generate PLS pseudomodels. In this procedure, complete scrambling of the elements of Y while keeping the corresponding data in X fixed destroys any relationships between the X- and Y-variables. Therefore, Y-scrambling yields PLS models that reflect the absence of a real association between X and Y data, thus producing insignificant models. The resulting PLS models should be characterized by poor statistical parameters such as the q^2 model, RMSECV, RMSEP etc. Y-scrambling was performed to check the possibility of random correlation and robustness of the final PLS models. The statistical characteristics of Y-randomization routines are provided for each engine response model generated in this report.

2.5 IMPORTANCE OF VARIABLES IN REGRESSION EQUATIONS

In order to estimate the importance of variables included in the regression model (especially PLS model) we used three parameters that are calculated during each run of the PLS algorithm: regression coefficients (b), variable importance in projection (VIP) (22) and target projection (TP) with selectivity ratio (SR) (23, 24). These parameters are known to be useful tools for interpretation of PLS models and may allow for better quantification of the contribution of variables to the regression model.

3.0 CHEMICAL COMPOSITION AND PHYSICAL PROPERTIES OF FACE DIESEL FUELS – CURRENT STATUS

In previous studies both standard ASTM tests and advanced analytical techniques were employed for in-depth characterization of the physical and chemical properties of the FACE diesel fuels (1, 2). The ASTM standard tests included measurements of physical properties (e.g., specific gravity, viscosity, cetane number, etc.), distillation profiles, aromatics by supercritical fluid chromatography, elemental analysis, and mass spectrometry, among others. Advanced methods consisted of gas chromatography-field ionization mass spectrometry (GC-FIMS), comprehensive two-dimensional gas chromatography (GC×GC) using both mass spectrometer

(MS) and flame ionization (FID) detectors, and nuclear magnetic resonance (NMR) spectroscopy.

In addition to the aforementioned analyses, researchers from CanmetENERGY repeated the GC×GC analyses using two optimized column combination configurations, and applied enhanced visualization tools to present chromatographic data. As a result, both ‘normal’ (a non-polar column followed by a polar column - nGC×GC) and ‘reverse’ (a polar column followed by non-polar column – rGC×GC) GC×GC experiments were performed to characterize the hydrocarbon composition of the FACE diesel fuels. The experimental results were presented in a CanmetENERGY report (25) which is available for download from:

<http://www.crcao.org/Publications/advancedVehiclesFuelsLubricants/FACE/index.html>).

In contrast to other advanced analytical methods, such as GC-FIMS, the revised GC×GC analysis revealed more detailed chemical structure which can be visually presented and could assist in more comprehensive correlation analyses in the future.

Figure 1 shows GCxGC results from the most recent report in the form of “bubble” plots overlaid on the FACE design cube. In the plots, each hydrocarbon class is represented by a different color and the size of each bubble represents the mass concentration of all of the compounds of a given hydrocarbon class. This drawing is for reference only and illustrates the compositional complexity of the fuels described here by only three descriptors: aromatics, T90, and CN (for higher-resolution GC×GC plots see Appendix B). Even this simplified cartoon-like images enable a quick comparison summary of differences in fuel composition:

- Alkylbenzenes for six fuels (FD1-FD5 and FD7) have similar carbon number distributions (see green bubbles in the plots)
- FD7 and FD8 are the only fuels that have a significant concentration of diaromatics, especially biphenyls (see big cyan bubbles). The biphenyls appear to be from the same source, but the concentration in FD8 is higher.
- To increase the CN in FD5, FD7, and FD8 the same n-paraffinic stream (notice large red bubbles) was apparently blended into the fuels.
- Fuels FD3, FD6, and FD9 (see Appendix B) seem to have the same hydrocarbon matrix fingerprint. The difference between these samples is limited to a varied

content of light alkylbenzenes (green bubbles) and heavy saturates (blue and yellow bubbles close to the right-bottom corner).

- To attain higher T90 temperatures, the same stream (i.e. distillation cut) of saturates was apparently added to fuels FD2, FD4, FD6, FD8 and FD9 (same pattern of blue and yellow bubbles close to the right-bottom corner).

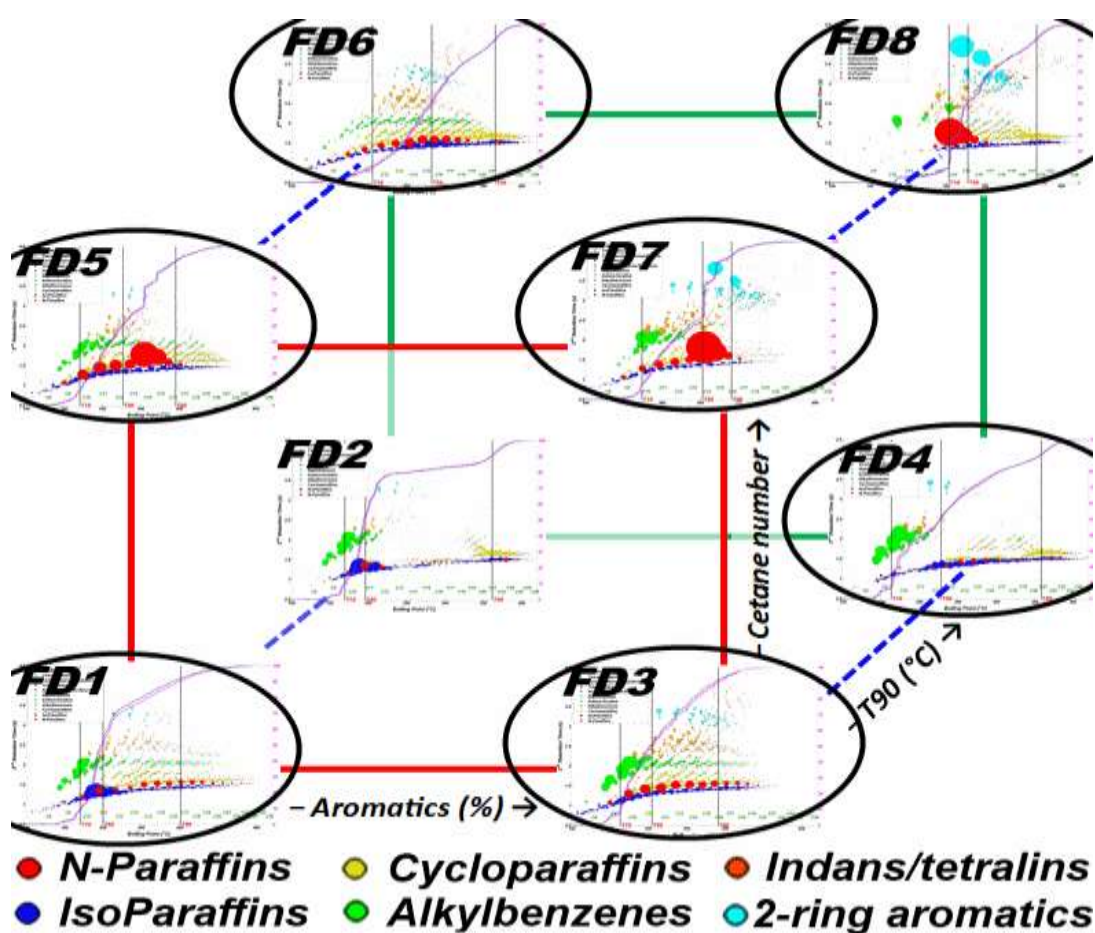


Figure 1 – FACE cube with GCxGC data plotted on the edges (Fuel FD9 not shown here but is presented in Appendix B). Hydrocarbon types are colour-coded. The size of the bubble (area) represents the concentration of a particular compound. On each GCxGC bubble plot the x-axis corresponds to the boiling point scale (i.e. the boiling point increases from left to right). For higher resolution chromatograms please see figures in Appendix B.

Those and other conclusions can be quickly drawn from the GC×GC chromatograms depicted in Figure 1. The following bar plots (Figures 2 – 4) were taken from the CanmetENERGY report (21) in order to summarize all the compositional analyses carried out on the FACE diesel fuels and that were used in the statistical analysis conducted in the current study. For example, Figure 2 presents a comparison of the saturate and aromatic contents reported by different chromatographic methods. In Figure 3 a more detailed breakdown of the hydrocarbon type composition is shown. Figure 4 presents a breakdown between the n-paraffins and isoparaffins. Note that for many cases, the GC-FIMS and GC-MS techniques report slightly different values than other techniques. The two numbers (xxx/yyy) located above each bar set corresponds to the calculated average and median concentrations, respectively for specific hydrocarbon groups of a particular fuel.

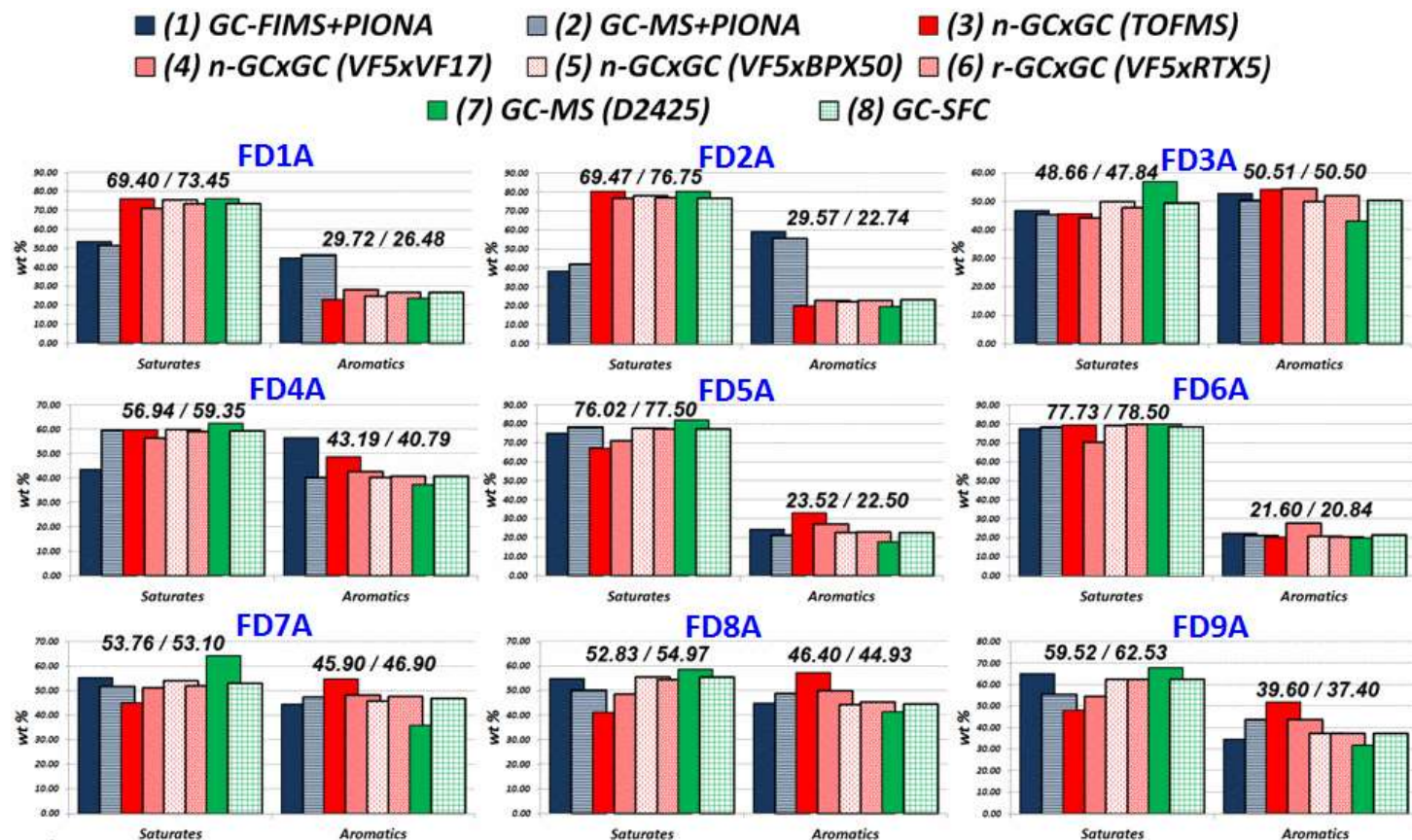


Figure 2 – Comparison of saturate and aromatic contents reported by different chromatographic methods for the FACE diesel fuels. The two numbers (xxx/yyy) located above each bar set correspond to calculated average and median concentrations, respectively, for particular hydrocarbon groups

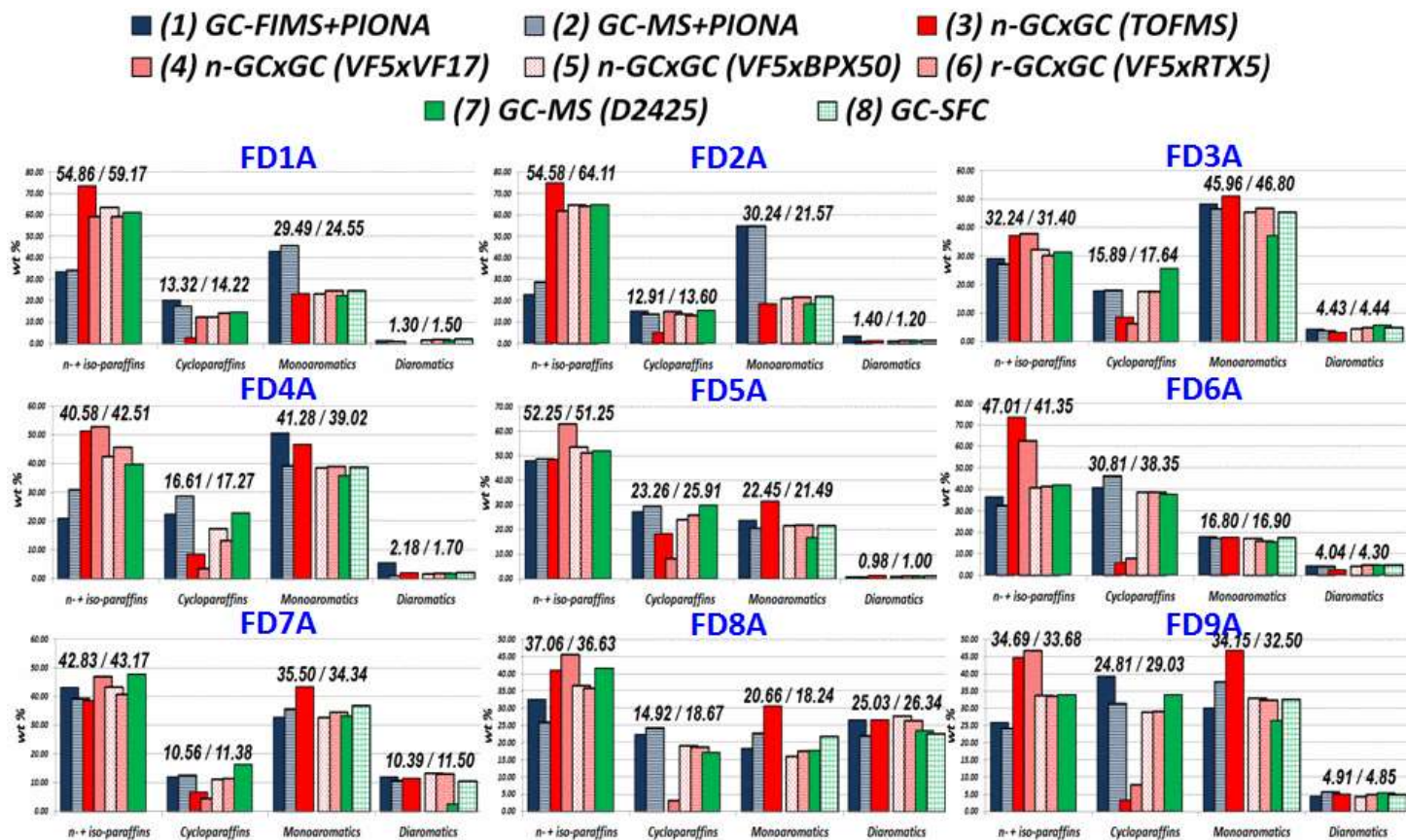


Figure 3 – Comparison of paraffinic, cycloparaffinic, mono, and diaromatic contents for the FACE diesel fuels. The two numbers (xxx/yyy) located above each bar set correspond to calculated average and median concentrations, respectively, for particular hydrocarbon groups

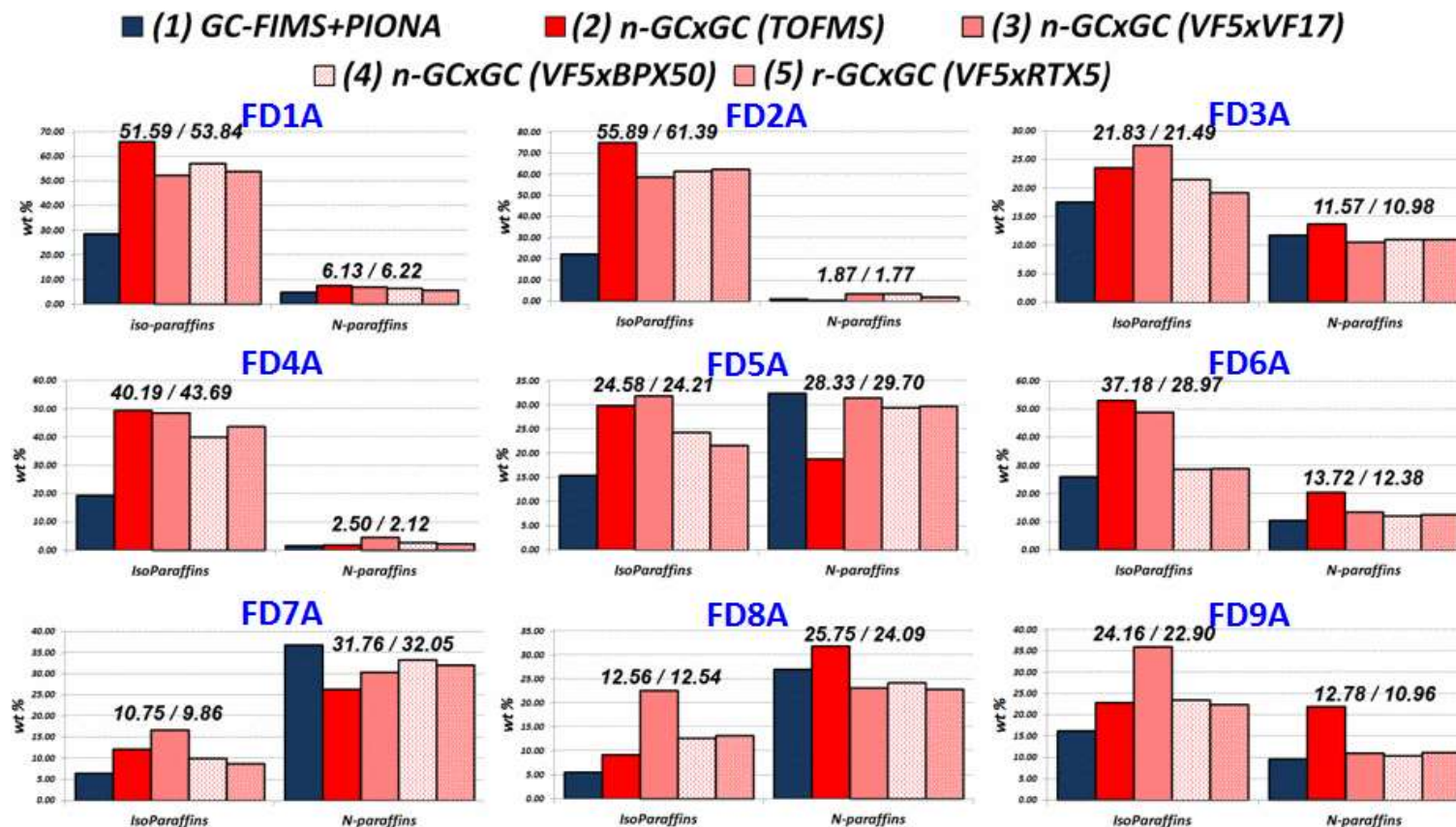


Figure 4 – Comparison of iso- and normal paraffin contents reported by different chromatographic methods for the FACE diesel fuels. The two numbers (xxx/yyy) located above each bar set correspond to calculated average and median concentrations, respectively, for particular hydrocarbon groups

In contrast to the chromatographic analytical results presented earlier, NMR provides fuel composition by quantifying specific carbon structural groups instead of giving information about the entire molecule. NMR compositional data may correlate better with engine emission characteristics such as soot formation since combustion depends upon the breaking and forming of carbon bonds. The ~90 distinct carbon substructures found in NMR data were narrowed to 10 carbon types (CTs) (26). In the current study, a graphical molecular structure was created to represent the resulting CT fragments as shown in Figure 5. The resulting ‘super-molecule’ was constructed to include all ten carbon types in one chemical structure. Each atom of the molecule is colour-coded and its size is related to the molar concentration (mol%) of its respective CT found by NMR analysis. NMR results for the FACE diesel fuels are presented in Figure 6 as ‘super-molecules,’ which enables fuel comparison in a simple pictorial manner. For example, FD1 has no CT9 or CT10 carbon atoms, very few CT3, CT5, and CT6 atoms, and significant amounts of CT1, CT2, and CT4 atoms. The NMR molecular representation presented here are original and have not been shown earlier in a report or other communications. Figure 7 shows the NMR CT results in a more traditional fashion.

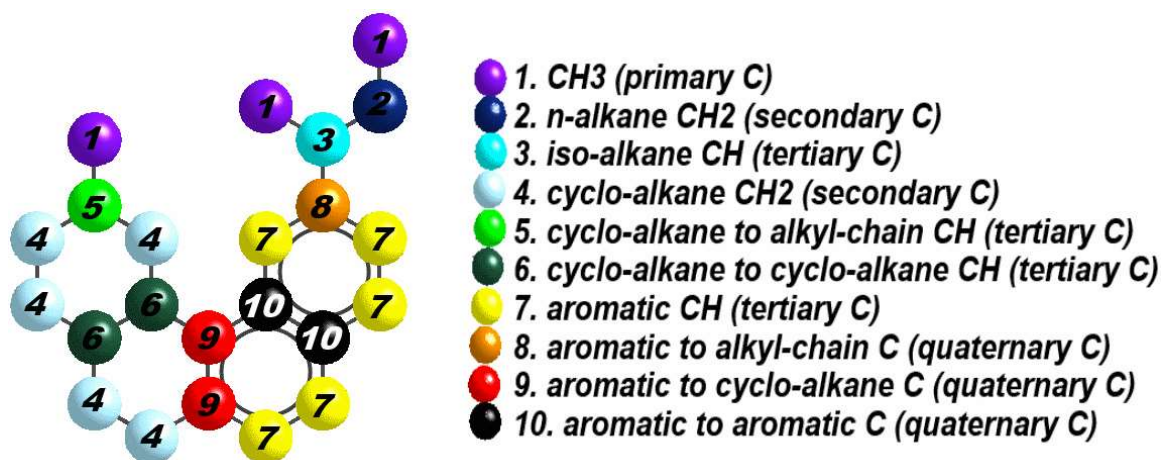


Figure 5 – Molecular representation (aka super-molecule) of carbon types quantified using NMR data

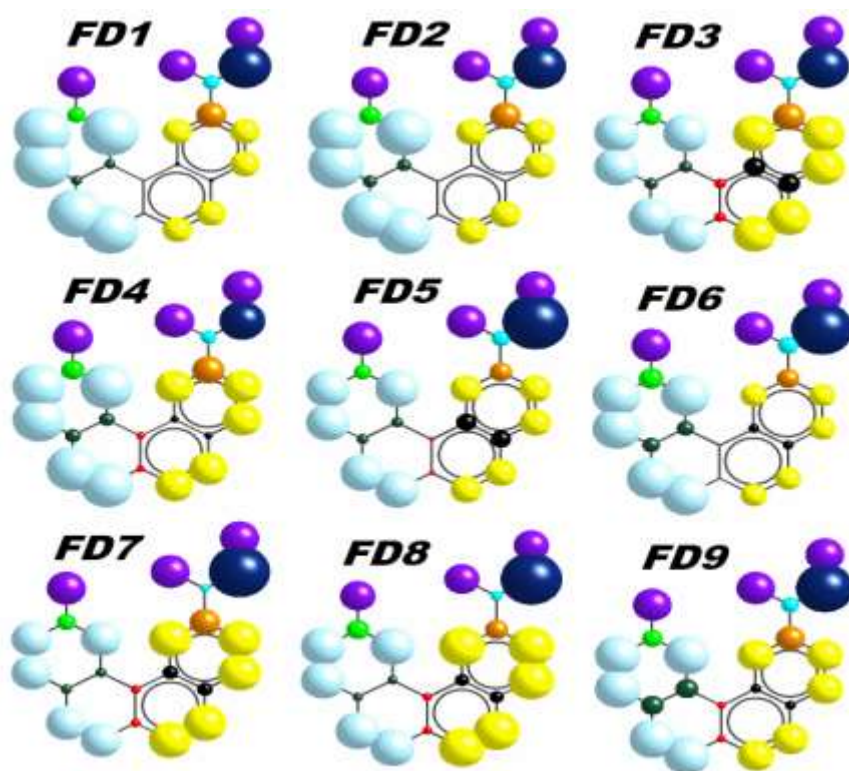


Figure 6 – FACE fuels according to ‘super-molecule’ representation

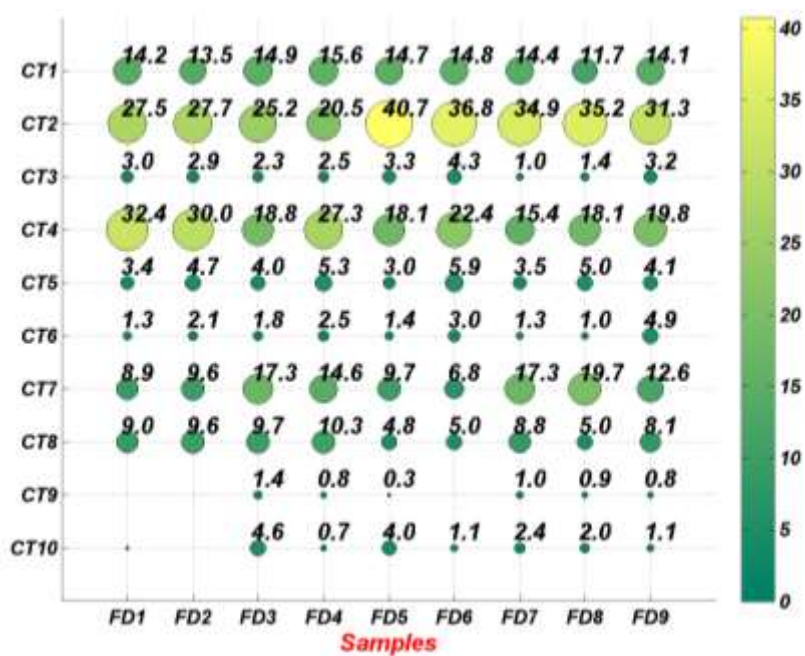


Figure 7 – Comparison of NMR results reported by the concentration (mol%) of carbon types (CTi) for FACE diesel fuels. The CTi labels correspond to the labels in Figure 5

In addition to gas chromatographic and NMR compositional characteristics presented in many reports and publications, CanmetENERGY has previously performed a suite of spectroscopic experiments (27) on FACE diesel fuels including UV-VIS, Raman, NIR, and mid-IR. Both the report and spectroscopic datasets are available from the authors upon request.

Raman, IR, and UV spectra display a number of well-defined bands associated with structural fragments of the molecules in the FACE diesel fuel blends. Several features, including the location of the absorption edge (UV), are variable among different fuels; information about various aliphatic and aromatic contents was extracted by numerical analyses. Figure 8 presents an example of NIR spectra of the FACE diesel fuels. The presented spectral region (4,000 to 11,000 cm^{-1}) covers the absorption bands corresponding to overtones and combinations of vibrations of C-H bonds. The inset in Figure 8 displays the expanded NIR spectra of the 5300–6100 cm^{-1} region. Bands in this region are mainly due to overtones of the C-H stretching and provide estimates of the relative concentrations of the diverse CH_n ($n=1, 2$, and 3) molecular fragments in the molecules. Notice that the intensity of the C-H overtone bands shows significant variations among the different samples and separates the FACE diesel fuels into two groups characterized by low (i.e. FD1-FD4) and high (i.e. FD5-FD9) CN fuels, respectively.

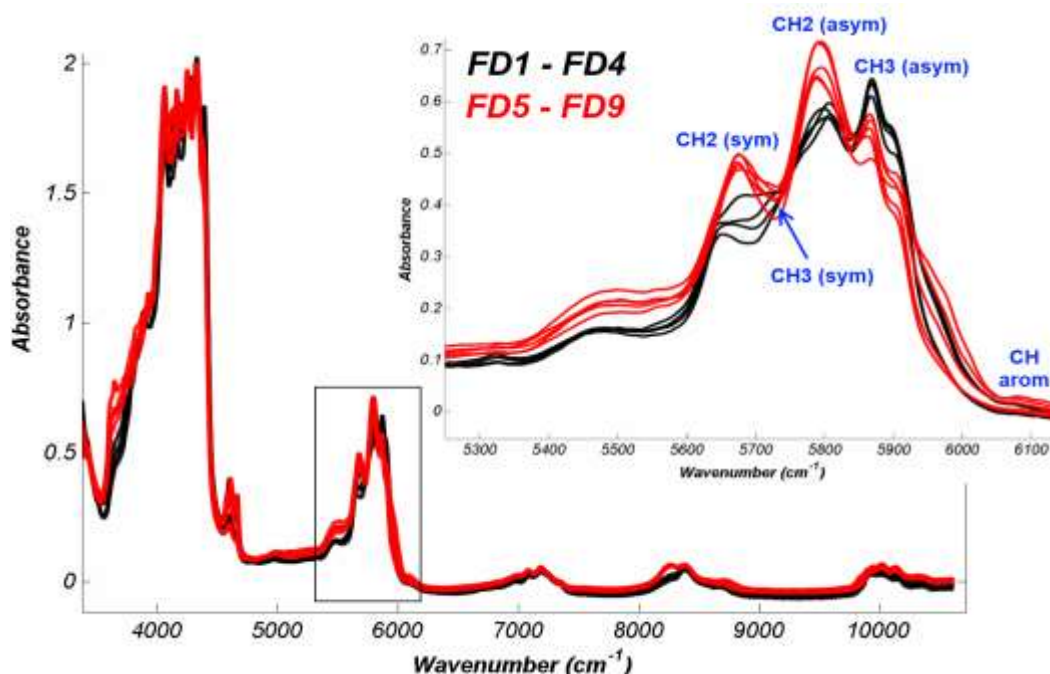


Figure 8 – NIR spectra (3400-10800 cm^{-1}) of FD1-FD4 (black, low CN fuels) and FD5-FD9 (red, high CN fuels). The inset presents expanded NIR region with depicted C-H stretching bands; sym (symmetric), asym (asymmetric), arom (aromatic).

Recently, thermophysical measurements were utilized by the National Research Council of Canada (NRC) to investigate the nine FACE diesel fuels. Specifically, thermal lens and optical interferometer techniques were used to determine the thermal diffusivity (D) and temperature coefficient of the refractive index ($-dn/dT$), respectively (see Table 1). Details of the experimental procedures are given in references 28 and 29. Thermal diffusivity quantifies the rate at which a temperature disturbance at one point in a body propagates to another point. The information about a material's ability to store and transfer heat is utilized, for example, in heat transfer models. Previously it was shown that $-dn/dT$ decreases as molecular volume or molecular weight increases (30). Such simple relations were found only for pure hydrocarbons and were not deeply investigated for organic compounds containing heteroatoms. The theory of the relationships between the thermophysical properties and chemical compositions of FACE fuels will be developed in future work.

Table 1 – Thermophysical properties of FACE diesel fuels: thermal diffusivity (D) and temperature coefficient of refractive index ($-dn/dT$)

FACE Fuel	D (10^{-3} cm ² /s)	$-dn/dT$ (10^{-4} K ⁻¹)
FD1	0.85±0.02	4.38±0.05
FD2	0.82±0.02	4.34±0.05
FD3	0.98±0.02	4.46±0.02
FD4	0.90±0.03	4.31±0.02
FD5	0.87±0.02	4.31±0.03
FD6	0.84±0.04	4.06±0.02
FD7	0.92±0.02	4.38±0.02
FD8	0.93±0.02	4.13±0.02
FD9	0.98±0.03	4.22±0.02

In the following section (section 4.0) the chemical space of the FACE diesel fuels as represented by both advanced and standard ASTM experimental datasets is explored.

4.0 EXPLORATION OF CHEMICAL/PHYSICAL SPACE OF FACE DIESEL FUELS

In the literature the FACE diesel fuels are often represented by the FACE fuel design matrix cube (similar to the blue cube in Figure 9) built in a three-dimensional coordinate system based on three fuel properties: CN, aromatic content, and T90 distillation. In general, the

measured fuel properties (red polygon in Figure 9) were within reasonable proximity of the target values; however, for some fuels there were noticeable shifts between measured and target properties (see fuel No. 7).

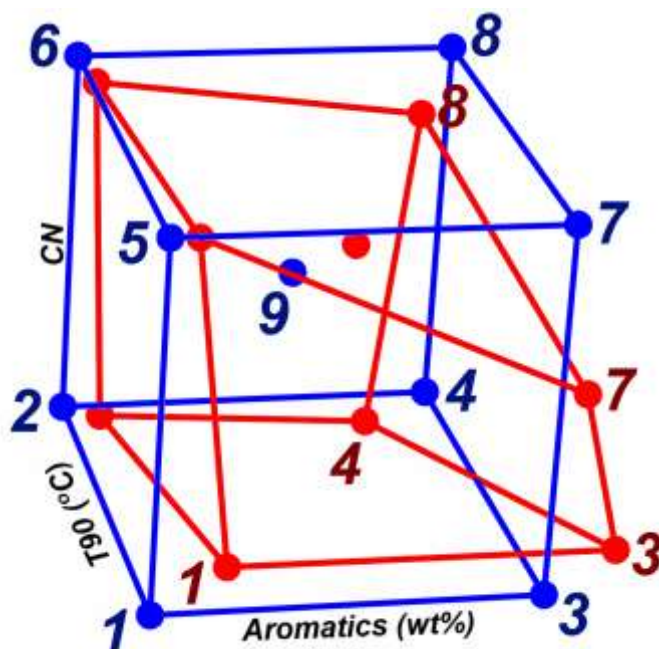


Figure 9 – Matrix of FACE diesel fuels. The blue cube shows the target property values. The red cube presents the measured property values.

For the current data mining study, all the analytical measurements for the FACE diesel fuels, both from standard ASTM analyses and advanced characterizations, were collected from the sources mentioned in section 2.0 and are located in the EXCEL file titled 'FACEfuels-chemistry.xlsx' attached to this report. The list of all fuel physical/chemical properties and HC compositional data information including size (i.e. the number of measured variables) and a short description is presented in Table 2.

Table 2 – List of 28 compositional, chemical, and physical datasets available for the FACE diesel fuels

No	Dataset tag	Dataset size ^{a)}	Description
1	PhysChem	10	Physical/chemical properties such as specific gravity, kinematic viscosity, cetane number/index, flash point, etc.
2	FIA	3	Hydrocarbons by ASTM D1319 (Fluorescent indicator adsorption)
3	SFC	3	Aromatics by ASTM D5186 (Supercritical fluid chromatography)
4	MS	16	Hydrocarbons by ASTM D2425 (Mass spectrometry)
5	Elemental	4	Carbon, hydrogen, nitrogen, and sulphur
6	Distill86	13	Distillation by ASTM D86 and measured by two laboratories
7	SimDis	101	Simulated distillation by ASTM D2887
8	HNMRa	6	¹ HNMR results provided by CanmetENERGY
9	HNMRb	23	¹ HNMR results provided by PNNL
10	CNMRa	15	¹³ CNMR results provided by CanmetENERGY
11	CNMRb	20	¹³ CNMR results provided by PNNL
12	NMRdetailed	40	Table provided by CanmetENERGY
13	NMRct	10	Carbon types by NMR
14	GCFIMSmain_hc	31	Hydrocarbons by GC-FIMS + PIONA
15	GCFIMSmain_cno	19	Carbon number distribution by GC-FIMS + PIONA
16	GCFIMSdetailed	558	GC-FIMS by carbon number and hydrocarbon type + PIONA
17	GCMS	33	Hydrocarbons by GCMS (saturates by ASTM D2786 / aromatics by ASTM D3239) + PIONA)
18	nGCxGCmain	7	Hydrocarbon types by GCxGC-FID with normal column configuration (non-polar × polar)
19	nGCxGCdetailed	134	Hydrocarbon distribution by carbon atom (normal column configuration)
20	rGCxGCmain	10	Hydrocarbon types by GCxGC-FID with reversed column configuration (polar × non-polar)
21	rGCxGCdetailed	112	Hydrocarbon distribution by carbon atom (reversed column configuration)
22	AllMed	18	The median value of concentration for hydrocarbon class calculated on all GCxGC, GC-FIMS, GC-MS, SFC and FIA results.
23	AllAvg	18	Average value of concentration for hydrocarbon class calculated on all GCxGC, GC-FIMS, GC-MS, SFC, and FIA results.
24	Thermo	2	Thermophysical properties (D) and dn/dT
25	nearIR	7779	NIR spectra (raw data)
26	midIR	4149	Mid-IR spectra (raw data)
27	raman	3609	Raman spectra (raw data)
28	uvvis	3101	UV-VIS spectra (raw data)

^{a)} Number of variables/parameters

Datasets with three or fewer variables can be plotted in the Cartesian system. For example, the number of variables measured by fluorescent indicator adsorption (FIA) is 3 (aromatics, olefins, and saturates) so the FIA can be depicted in three-dimensional space (see Figure 10). Based on the relative location of the objects in the plot, the FACE fuels can be described with respect to their similarities in the FIA space. For example, FD2 and FD6 are very close to each other, which means that they have a very similar content of HC types characterized by FIA. The point for the FD4 fuel is far away from the other points, which indicates that it is very different compositionally from the other fuels.

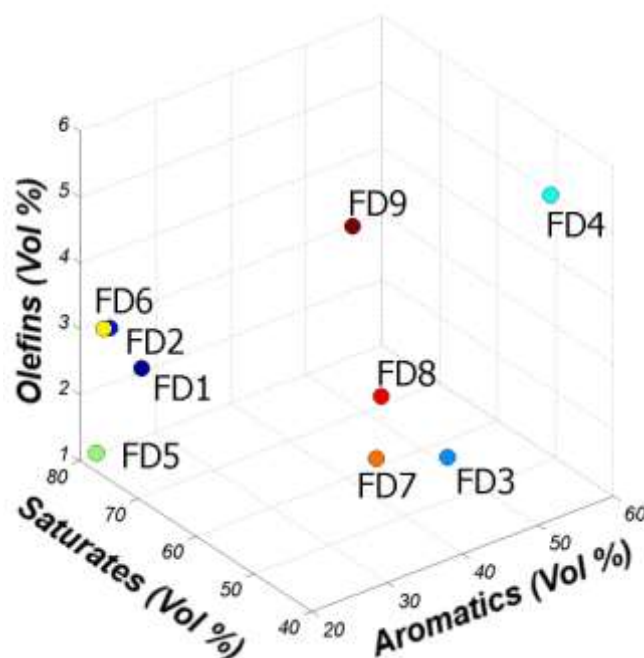


Figure 10 – Hydrocarbons by FIA plotted in three-dimensional space described by aromatics, saturates, and olefins content

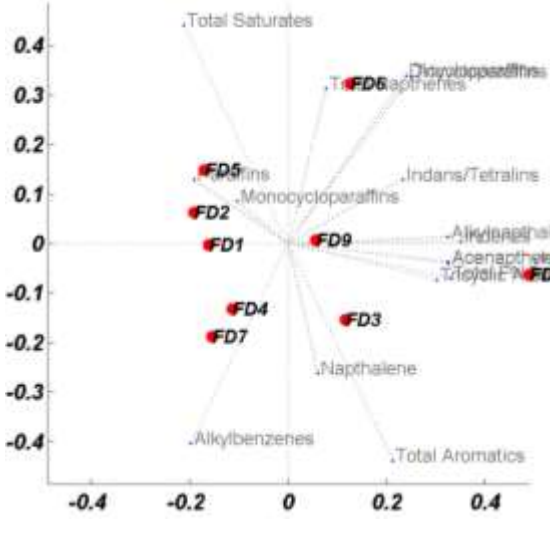
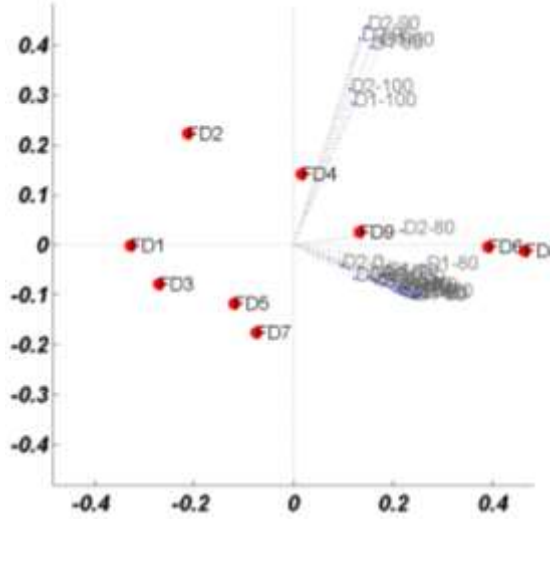
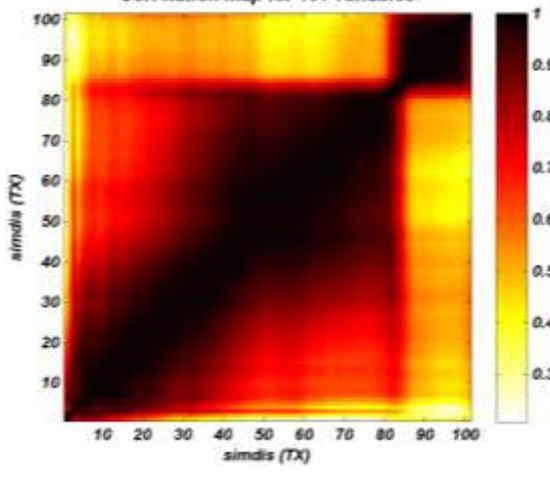
However, the number of variables in most of the datasets is larger than three, and so we are unable to depict them in the same way as for the FIA set. For example, the FACE diesel fuels are described by 10 physical properties, so objects would need to be placed in ten-dimensional space, for which graphical presentation is impossible. For such multidimensional cases, the experimental results were analyzed by one of the most popular methods of exploratory analysis: principal component analysis (PCA). Exploration of chemical/physical space using PCA projection is of great help in data analysis, providing information that would be too difficult to extract from the original data matrix.

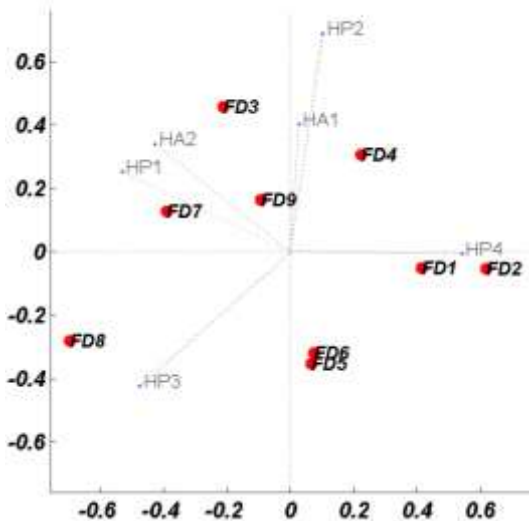
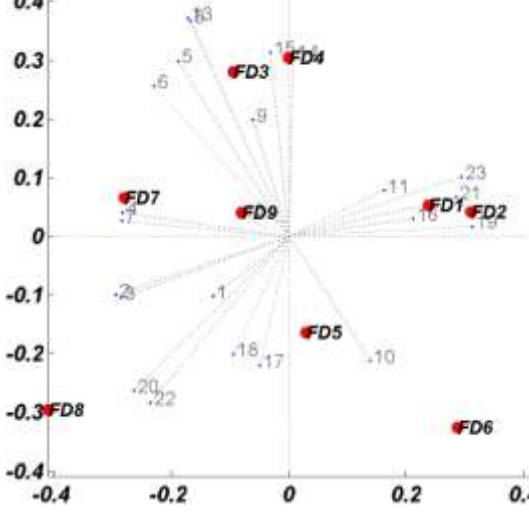
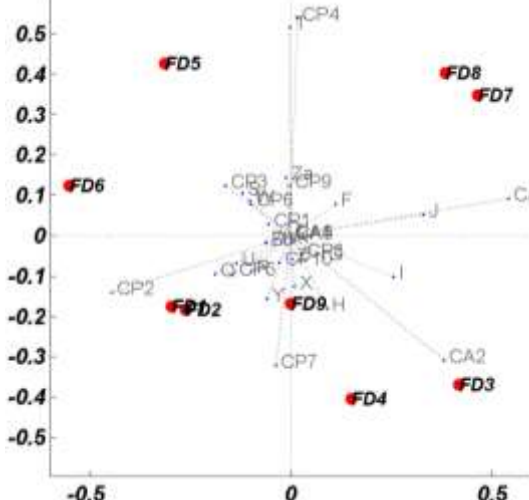
PCA models were generated in MATLAB and saved in structures that can be represented graphically and interpreted using PCAVisTool software written in-house (see Appendix C for more details). All the plots presented in this section (see Table 3) were created and evaluated using this software. Here we report only concluding figures and remarks based on PCA for FACE diesel fuels for a particular dataset.

Once scores (linking to the samples) and loadings (linking to the variables) have been calculated by the PCA, there are several ways of visually representing the results, including bar charts with consecutive score plotted against sample number, 2-D and 3-D scores plots, 2-D and 3-D loading plots, and biplots. In this section, we decided to use 2-D biplots (see Appendix D for more details) since they can convey PCA results in a simple and comprehensive way.

Table 3 – PCA projection results for selected compositional, chemical, and physical datasets available for FACE diesel fuels

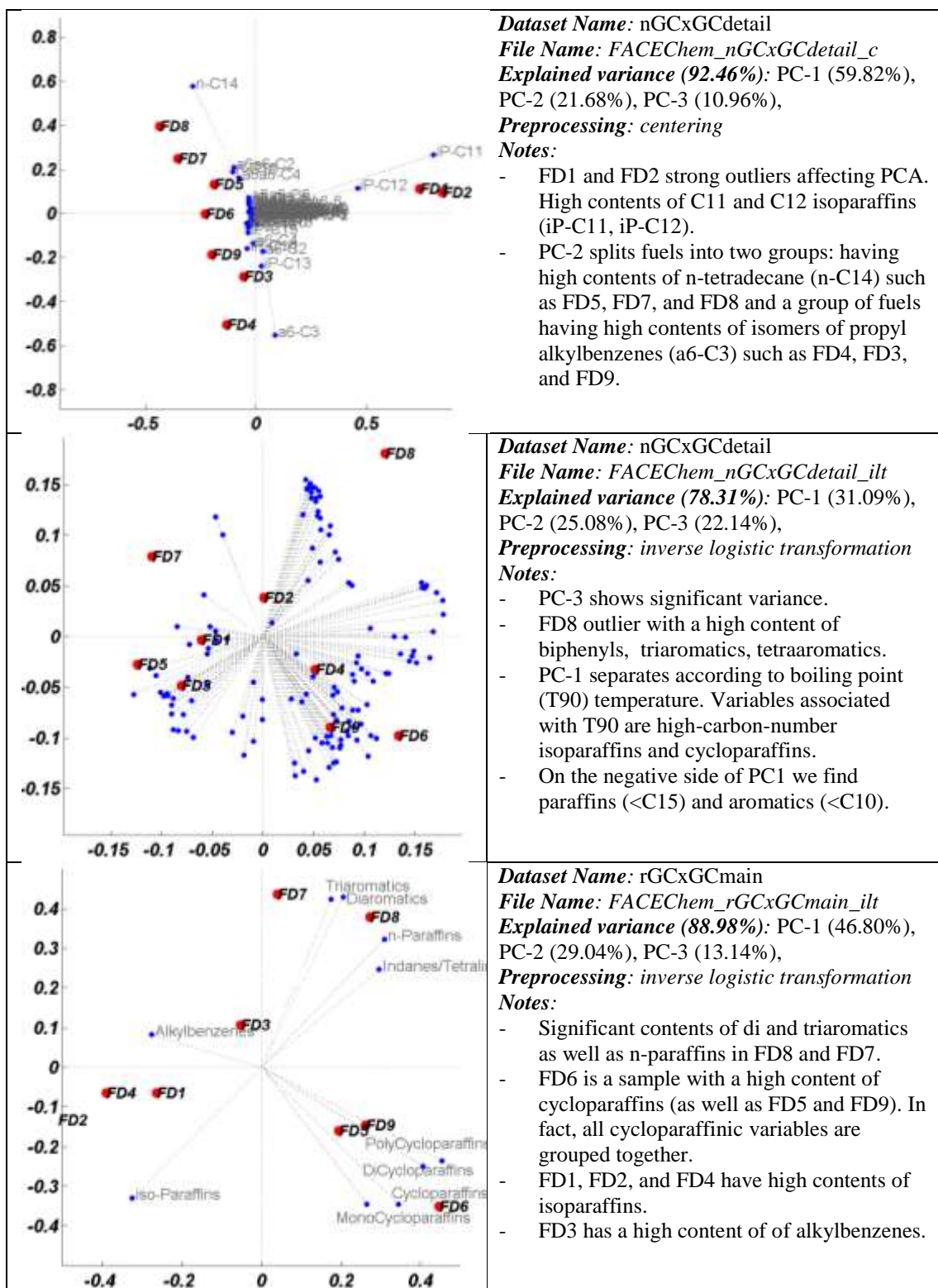
<i>Graphic representation of the dataset</i>	<i>Description</i>
	<p>Dataset Name ^{a)}: PhysChem File Name ^{b)}: FACEChem_prop Explained variance (89.80%) ^{c)}: PC-1 (52.9%), PC-2 (25.57%), PC-3 (11.33%), Preprocessing ^{d)}: autoscaling Notes:</p> <ul style="list-style-type: none"> - Fuels FD9 and FD7 are very similar. - Clear separation of the low (FD1-FD4) and high (FD5-FD9) cetane number fuels. - FD3 has outstanding Bromine Number much greater than other fuels. - Fuels FD6 and FD8 have high viscosity and flashpoint values. - Derived CN and CN by Engine methods are highly correlated (as expected).

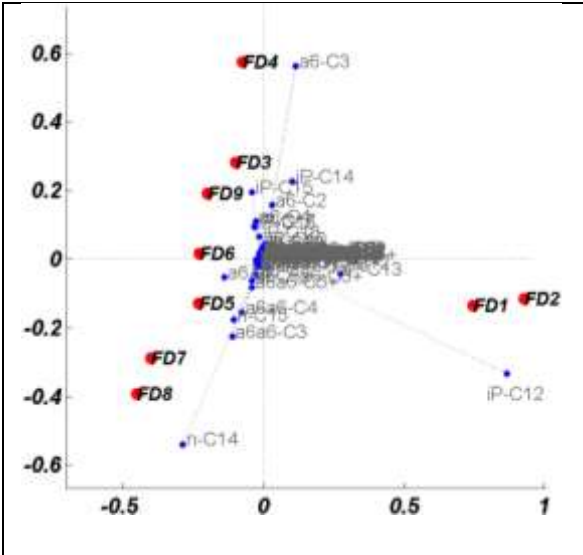
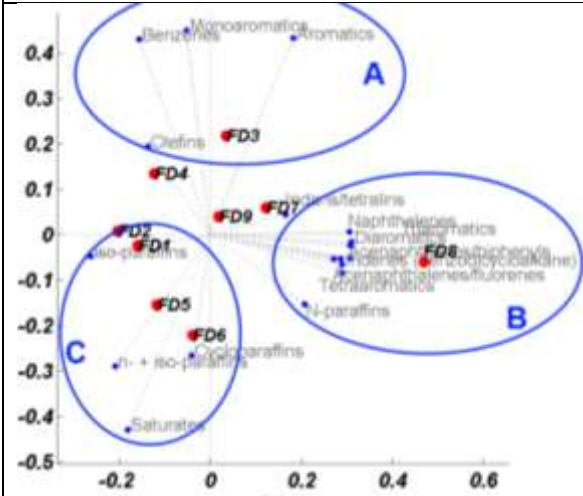
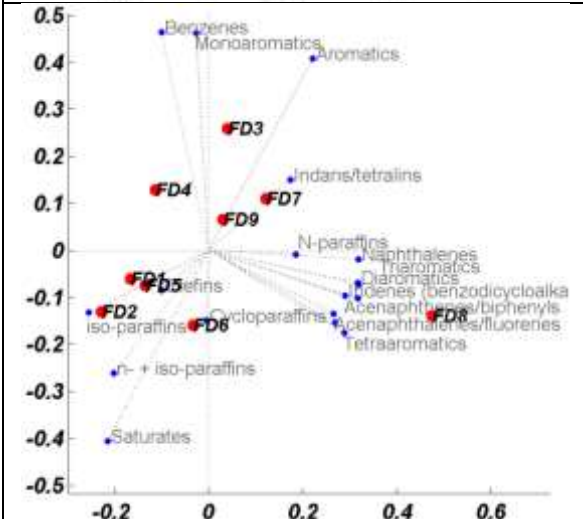
	<p>Dataset Name: MS</p> <p>File Name: FACEChem_ms_ilt.pcmat</p> <p>Explained variance (87.98%): PC-1 (45.15%), PC-2 (23.74%), PC-3 (19.10%),</p> <p>Preprocessing: inverse logistic transformation</p> <p>Notes:</p> <ul style="list-style-type: none"> - MS dataset split on low and high aromatic fuels on PC-2. - FD8 and FD6 fuels have a high di- and tricyclopentadiene contents. - Fuel FD8 is a clear outlier due to high content of tricyclic aromatics, acenaphthenes, acenaphthylenes, and total PNAs - 1st and 2nd quadrants are occupied by saturates, 3rd and 4th by aromatics - Highly correlated variables: tricyclic aromatics, acenaphthenes, acenaphthylenes
	<p>Dataset Name: Distill86</p> <p>File Name: FACEChem_distill86_c</p> <p>Explained variance (95.92%): PC-1 (74.77%), PC-2 (14.35%), PC-3 (6.80%),</p> <p>Preprocessing: centering</p> <p>Notes:</p> <ul style="list-style-type: none"> - First two PCs separate samples on low (FD1/FD3/FD5/FD7) and high (FD2/FD4/FD6/FD8) T90 fuels. - Notice two groups of variables; one group comprising of T90 and T100 (aka. D-90, D-100) (top right corner) and the other group (bottom right corner) comprising of the rest of the distillation temperatures. - There is a high correlation between close points on distillation curves (for example T60 vs. T70).
<p>Correlation map for 101 variables</p> 	<p>Dataset Name: SimDis</p> <p>File Name: FACEChem_simdis_c</p> <p>Explained variance (97.50%): PC-1 (77.89%), PC-2 (11.92%), PC-3 (7.69%),</p> <p>Preprocessing: centering</p> <p>Notes:</p> <ul style="list-style-type: none"> - The PCA results are very similar to those presented in the Distill86 dataset. - Figure (on the left) presents the correlation coefficients map calculated between all SimDis variables (in this case TXs). Using such a map we can investigate the correlation between variables. As can be seen, there are two groups of highly correlated variables (two distinct dark shapes on the map).

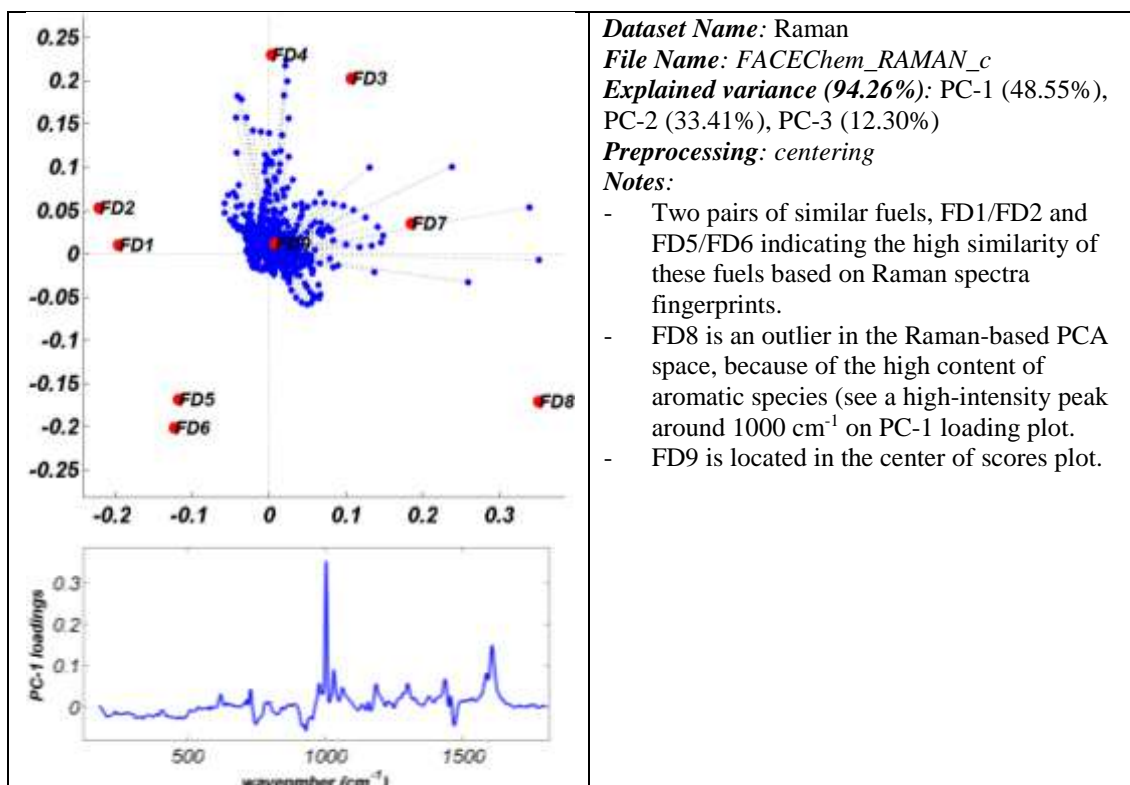
	<p>Dataset Name: HNMRA File Name: FACEChem_HNMR1_ilt Explained variance (94.94%): PC-1 (54.03%), PC-2 (27.14%), PC-3 (13.78%), Preprocessing: inverse logistic transformation Notes:</p> <ul style="list-style-type: none"> - The first PC is strongly influenced by HP4 HNMR atom type (a paraffinic CH₃). - Fuels FD1 and FD2 have a high molar percentage of paraffinic CH₃ protons (branched isoparaffins or cycloparaffins). - Fuel FD8 has a high concentration of paraffinic CH₂ fragments (long paraffinic chain) (HP3).
	<p>Dataset Name: HNMRA File Name: FACEChem_HNMR2_ilt Explained variance (83.35%): PC-1 (35.95%), PC-2 (28.94%), PC-3 (17.47%), Preprocessing: inverse logistic transformation Notes:</p> <ul style="list-style-type: none"> - In comparison to the previous CanmetENERGY HNMR dataset, PNNL HNMR dataset is described by a larger number of atom types (23 vs. 6). - The PC-1 splits FACE HNMR dataset on low and high paraffinic samples. - FD1 is similar to FD2 and both have a high content of CH₃ groups of paraffins (n- and iso-) (var 23). - FD8 seems to be an outlier (high content of 'Di-ring aromatics' (variables 2 and 3) and 'CH₂ groups' (variables 20 and 22). - FD3 and FD4 are fuels with highly 'substituted mono-ring aromatics' (variables 5, 6, 8, 13-15). - FD9 is closest to the center.
	<p>Dataset Name: CNMRall File Name: FACEChem_CNMRall_ilt Explained variance (94.94%): PC-1 (49.18%), PC-2 (38.48%), PC-3 (7.27%), Preprocessing: inverse logistic transformation Notes:</p> <ul style="list-style-type: none"> - Two CNMR datasets (PNNL and CanmetENERGY) were joined and analyzed using PCA. - The PC-1 separates fuels on low and high aromatic samples. The PC-2 separates fuels on low and high-CN fuels. - FD1 and FD2 are very similar fuels. - High correlation between respective CNMR atom types reported by both laboratories. For instance 'Aromatic C-H' (CA3) and 'Substituted carbon in alkenes (RHC=CR2)

	and para CH in toluene' (J) and $R=0.9945$. This observation is expected as the paired NMR atom types usually share a similar NMR spectra region.
	<p>Dataset Name: NMRct File Name: FACEChem_NMRct_ilt Explained variance (79.50%): PC-1 (36.87%), PC-2 (25.25%), PC-3 (15.43%), Preprocessing: inverse logistic transformation Notes:</p> <ul style="list-style-type: none"> - FD1 and FD2 are very similar and both have high contents of 'cyclo-alkane CH2' carbon type (CT4). However, based on other analytical techniques the concentration of cycloparafins is low for these two fuels. - FD9 is closest to the center which is expected position for this averaged fuel. - PC-2 seems to split FACE diesel fuels according to CN value.
	<p>Dataset Name: GCFIMSmains_hc File Name: FACEChem_GCFIMShc_ilt Explained variance (80.09%): PC-1 (40.27%), PC-2 (24.99%), PC-3 (14.84%), Preprocessing: inverse logistic transformation Notes:</p> <ul style="list-style-type: none"> - FD8 seems to be an outlier in this space with a high content of benzodicycloalkanes (or biphenyls); FD7 also trends in this direction but with a lower content of biphenyls. - PC-1 more or less divides low/high CN fuels. - FD3 and FD7 are closest to the center.
	<p>Dataset Name: GCFIMSmains_CN File Name: FACEChem_GCFIMScn_ilt Explained variance (87.23%): PC-1 (49.83%), PC-2 (24.36%), PC-3 (13.05%), Preprocessing: inverse logistic transformation Notes:</p> <ul style="list-style-type: none"> - FD8 seems to be an outlier in carbon number space due to a high content of C14, C18-C20 compounds. - FD2 has high contents of C8 and C9 compounds.

	<p>Dataset Name: GCFIMSDetailed File Name: FACEChem_GCFIMSDet_ilt Explained variance (70.20%): PC-1 (25.23%), PC-2 (24.09%), PC-3 (20.88%), Preprocessing: inverse logistic transformation Notes:</p> <ul style="list-style-type: none"> - FD8 seems to be an outlier - high contents of aromatics, benzodicycloalkanes, biphenyls (C14:2:C20).
	<p>Dataset Name: GCMS File Name: FACEChem_GCMS_ilt Explained variance (88.71%): PC-1 (50.31%), PC-2 (30.78%), PC-3 (7.62%), Preprocessing: inverse logistic transformation Notes:</p> <ul style="list-style-type: none"> - FD8 seems to be an outlier - high contents of polyaromatics, benzodicycloalkanes, and biphenyls. - Variable space split into three regions: saturates, monoaromatics, and polyaromatics.
	<p>Dataset Name: nGCxGCmain File Name: FACEChem_nGCxGCmain_ilt Explained variance (86.79%): PC-1 (50.55%), PC-2 (20.73%), PC-3 (15.51%), Preprocessing: inverse logistic transformation Notes:</p> <ul style="list-style-type: none"> - FD1 and FD2 contain significant amount of isoparaffins. - FD6 has a significant content of cycloparaffins. - FD8 has high contents of di and triaromatics.



	<p>Dataset Name: rGCxGCdetail File Name: FACEChem_nGCxGCdetail_c Explained variance (89.85%): PC-1 (58.16%), PC-2 (21.85%), PC-3 (9.84%), Preprocessing: centering Notes:</p> <ul style="list-style-type: none"> - Percent variance explained by first three PC components are similar to that obtained for PCA on the nGCxGCdetail dataset. Also the same conclusions. - Very similar FD1 and FD2: these two fuels are strong outliers affecting PCA results. FD1 and FD2 have a high content of isoparafrins having 12 carbon atoms (iP-C12 ; dodecane isomers). - Unlike normal GCxGC, reverse phase can better separate homologues of isoparafrinic compounds.
	<p>Dataset Name: AllGCavg File Name: FACEChem_AllGCavg_ilt Explained variance (89.04%): PC-1 (52.11%), PC-2 (23.01%), PC-3 (13.91%), Preprocessing: inverse logistic transformation Notes:</p> <ul style="list-style-type: none"> - Three groups of variables (A, B, C). - FD8 fuel (outlier in the AllGCavg space) because of the high concentration of compounds located in the group B (i.e. Diaromatics, naphthalenes, acenaphthenes/biphenyls etc.).
	<p>Dataset Name: AllGCmed File Name: FACEChem_AllGCmed_ilt Explained variance (89.04%): PC-1 (52.11%), PC-2 (23.01%), PC-3 (13.91%) Preprocessing: inverse logistic transformation Notes:</p> <ul style="list-style-type: none"> - Same results as for AllGCavg dataset.



^{a)} Dataset tag was taken from Table 2; ^{b)} *.pcmat File Name used for PCAVisTool; ^{c)} summed explained variance for the first three PCs; ^{d)} centering, autoscaling, or inverse logistic transformation (see Appendix F for details).

In order to visualize and evaluate the hydrocarbon composition (as determined by different analytical methods) and physical properties of the FACE diesel fuels, PCA analysis was performed on the respective datasets (results presented in Table 3). In most cases the first three components carried more than 90% of the total variance, hence chemical and physical space used in modeling can be summarized as 3D space spanned by the first three PCs. Figure 11 shows selected examples of 3D scores plots presented in a similar way to the original FACE design cube space where fuels located in the cube corners are interconnected. Location and distribution of FACE diesel fuels in the PCA space depends on the specific dataset analyzed. According to this plot we can see a different degree of similarity between the PCA space cubes calculated for a particular dataset. For example, nGC×GCmain (set 18) and AllMed (set 22) are like the original FACE cube with FD9 located in the center, which means that these two datasets may be strongly correlated with the FACE cube and with each other.

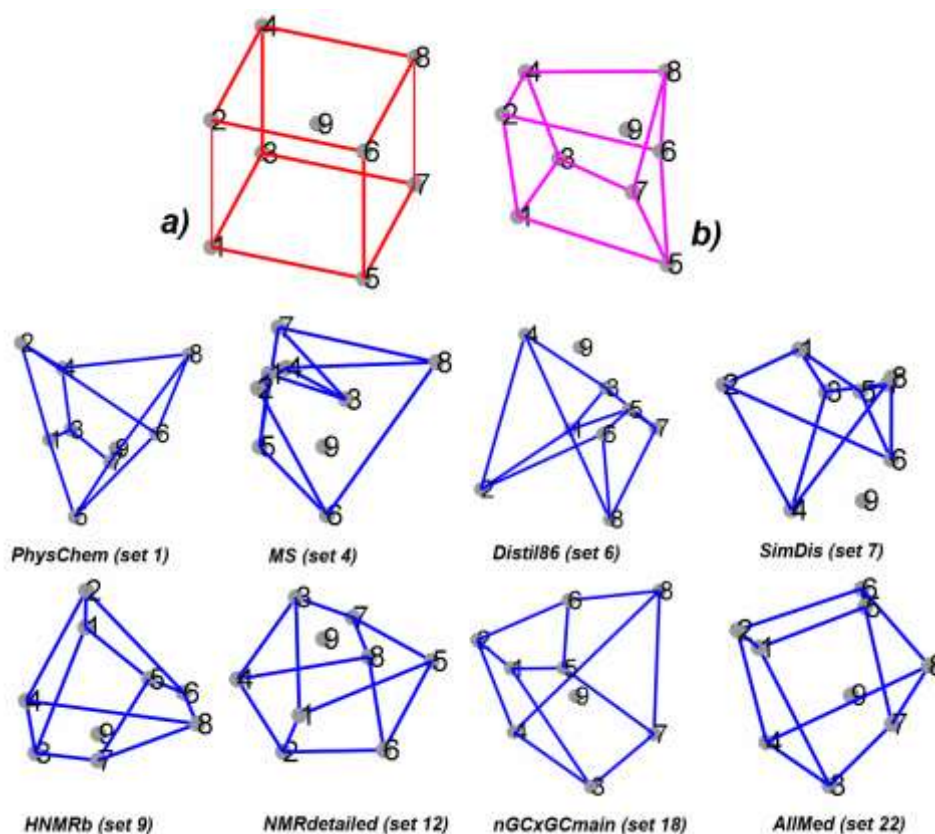


Figure 11 – FACE-cube-like PCA representation of the major datasets discussed in this section, where a) represents the FACE cube from experimental design, b) its real-world version, and 8 cubes below show 3D scores plots made by plotting each fuel according to the values of its first three PCs calculated for eight selected datasets.

The simplest and most convenient way of determining whether two shapes (data blocks or matrices) are similar is by estimating how well they are correlated. Matrix correlation tests can be used for this purpose (31, 32). Most such tests provide a correlation coefficient parameter, which can be interpreted in the same way as Pearson's correlation coefficient. In this report RV-correlation coefficient (defined in Appendix E) is used to report correlations among different FACE compositional, chemical, and physical datasets. The RV-coefficient has values between zero and one [0,1], where a value of 1 is assigned in the case of a perfect correlation. Figure 12 reports on the RV-coefficient values calculated between respective datasets including the original FACE design cube. The calculations of RV-coefficients were done on at least the three first PCs obtained after PCA. In that figure datasets no. 1-23 are reported in Table 2, datasets 24, 25 are the original FACE diesel fuels target property data and experimental measured property values, respectively (see Figure 9 for reference). The dark red bubbles on the RV coefficient map

indicate the high correlation between respective datasets. For example, dataset 25 is correlated with a few other datasets: 1-PhysChem (0.72), all NMR datasets (0.68-0.77), 18-nGC×GCmain (0.71), 22-allMed (0.75), and obviously 24-FACE fuels target properties. The high RV-correlation coefficient between the PhysChem dataset and compositional data indicates which type of analytical data is strongly correlated with physical properties and may be useful for modeling these properties. Among others simdis (dataset #7), GCFIMSmain_hc (dataset #14), nGC×GCmain (dataset #18), and AllAvg (dataset #23) seem to be the most correlated with physical properties measured on the FACE fuel set. The significance of the RV-coefficient can be assessed empirically using Monte Carlo permutations (see Appendix E).

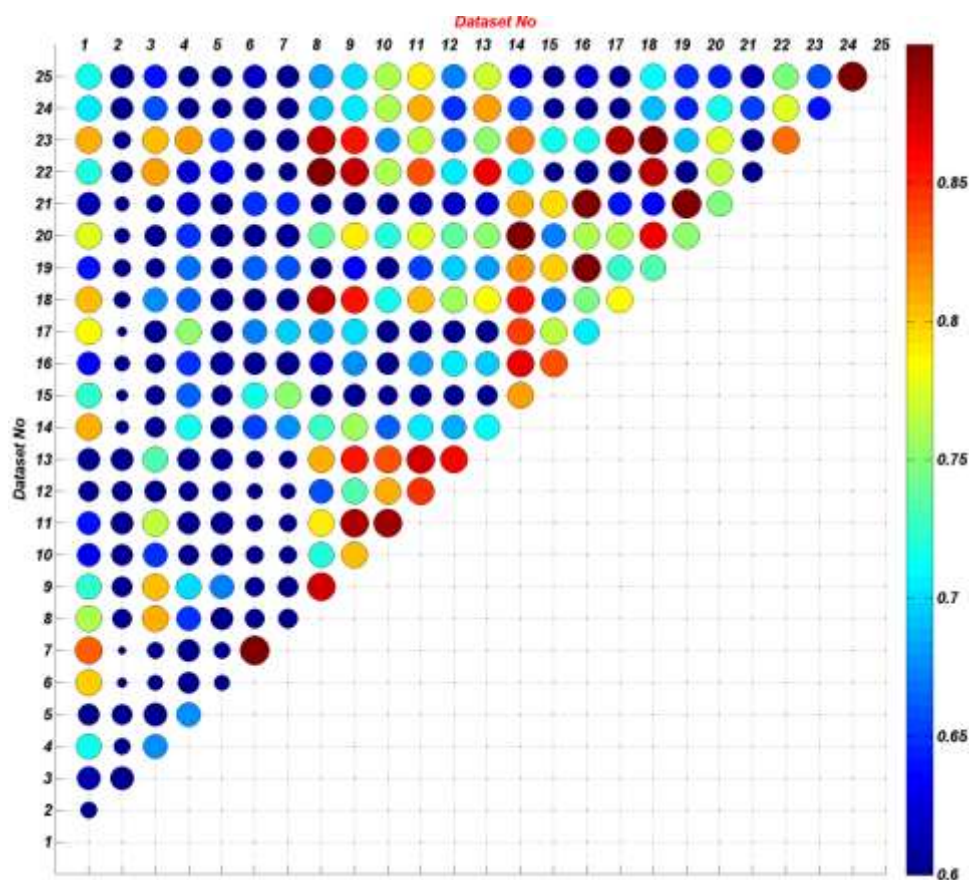


Figure 12 – Correlation map of RV-coefficients calculated for all 23 datasets presented in Table 2 (dataset 1-23). Datasets 24 and 25 are the original FACE diesel fuels target property data and experimentally measured property values (see Figure 9 for reference), respectively. In this plot, bubble size and colour indicate the strength of the correlation between respective pairs of datasets.

5.0 FINDING RELATIONSHIPS BETWEEN PHYSICAL PROPERTIES AND COMPOSITIONAL DATA

In this section trends between compositional information and physicochemical properties measured for the FACE diesel fuels are investigated.

A correlation coefficient (R) measures the strength and direction of a linear relationship between two variables. If the correlation coefficient is lower than zero the trend between variables is inverse. A correlation matrix is a square table showing the correlations between all pairs of variables in the dataset. A correlation matrix is always symmetrical. Figure 13 shows the lower part of the correlation matrix calculated for the physical parameters measured for FACE diesel fuels.

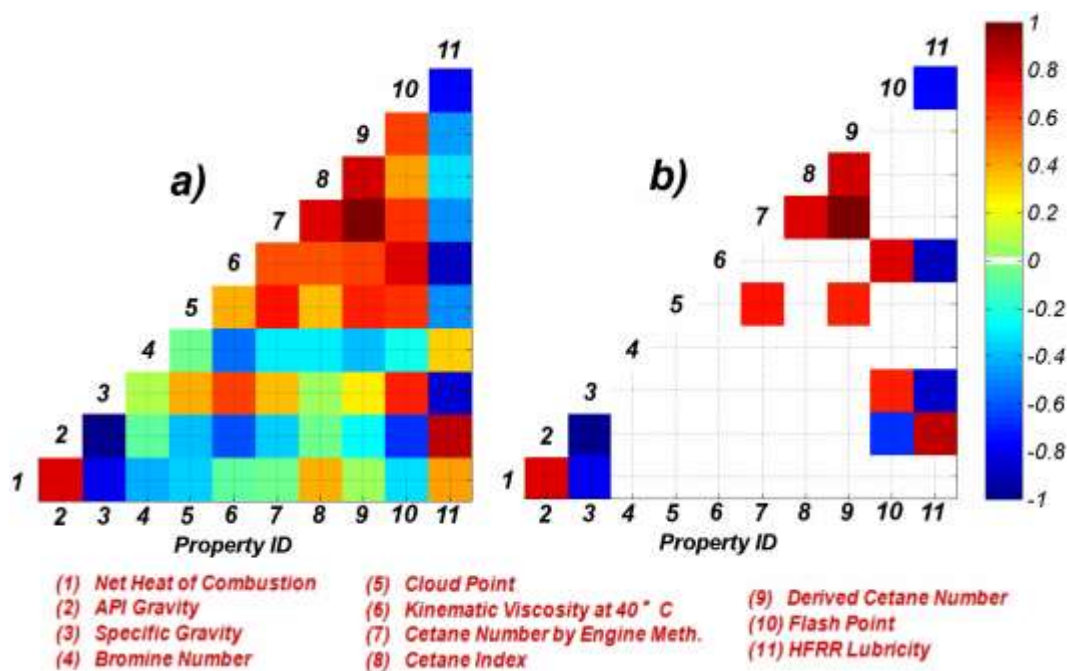


Figure 13 – The correlation matrix between all physical properties measured for FACE diesel fuels (a). Selected significant correlations ($p < 0.05$) (b).

In Figure 13 the dark red boxes indicate strong positive correlations between two physical properties. Examples include: API gravity/lubricity (2 & 11); CN/DCN (7 & 9), viscosity/flash point (6 & 10). The dark blue boxes show strong inverse correlations between two physical properties. Examples: specific gravity/lubricity (3 & 11), specific gravity/API gravity (3 & 2), viscosity/lubricity (6 & 11), net heat of combustion/specific gravity (1 & 3). Figure 14 presents examples of positive and negative correlations between selected fuel physical properties.

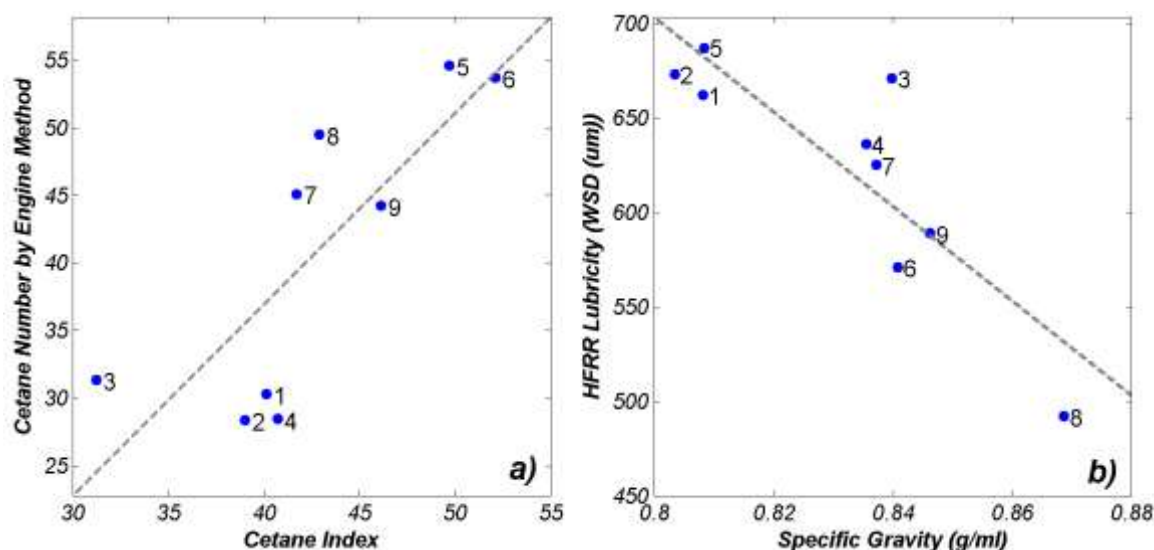
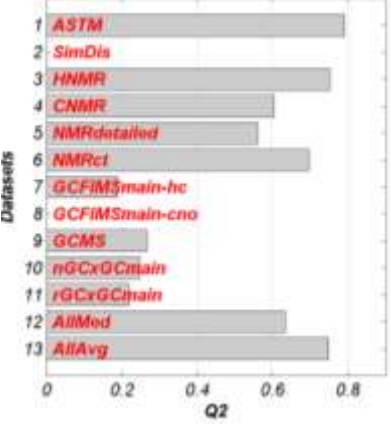
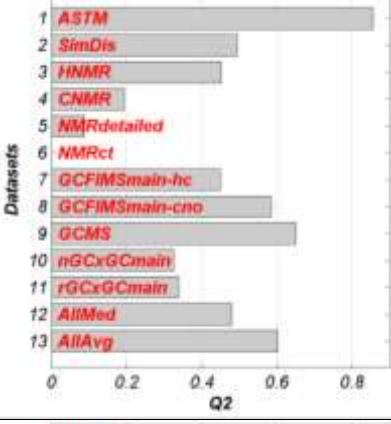
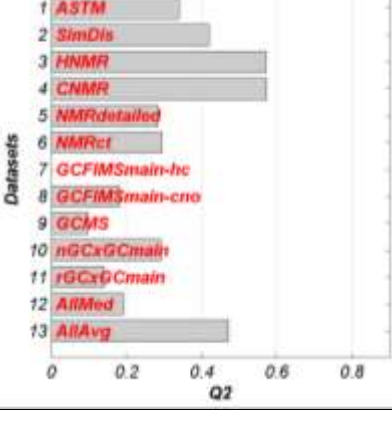


Figure 14 – Examples of strong positive (a) and negative (b) correlations between physical properties

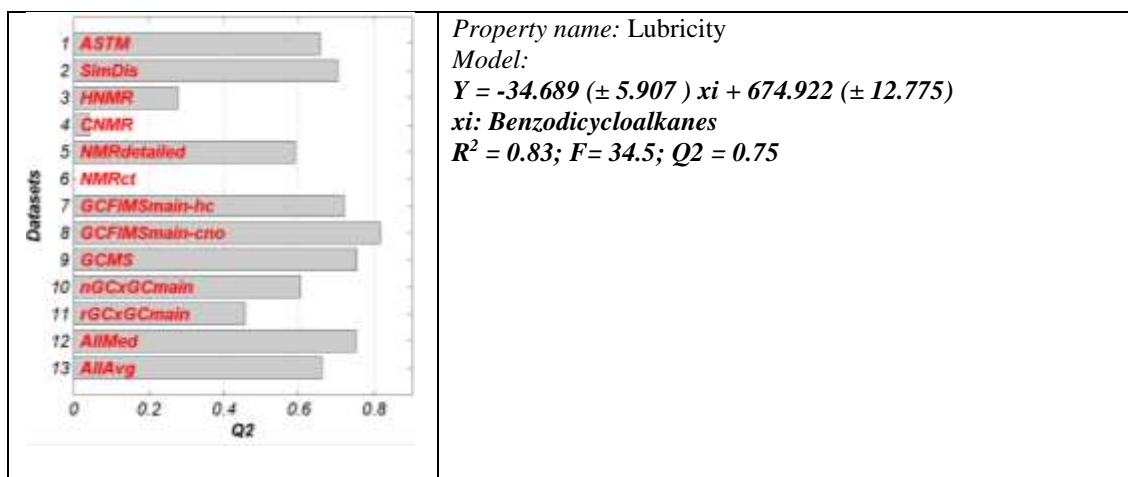
The next objective was to identify possible existing relationships between FACE fuel properties and compositional information provided by selected datasets. The results are presented in Table 4, where all the details including the bar plot (built on model's performance for 13 datasets) and the 'best' regression model information are provided. We used leave-one-out validation to validate and test the robustness of the models. This technique involves using a single observation from the original sample as the validation data, and the remaining observations as the training data (see section 2.4). Quality of the models is estimated based on the *q2model* (aka Q2) parameter, where higher value of that parameter indicates on better quality models.

Note: The reader should approach the calculated correlations carefully and should not generalize to other sets beyond this particular set of the FACE diesel fuels, as they have been computed for a low number of objects.

Table 4 – Correlation results for selected compositional, chemical, and physical datasets available for the FACE diesel fuels

<i>The simple regression results</i>	<i>Description</i>
	<p>Property name ^{a)}: Net Heat Of combustion</p> <p>Model:</p> $Y = 0.033 (\pm 0.005) xi + 43.610 (\pm 0.149)$ <p><i>xi</i>: Total Aromatics by MS D2425</p> <p>$R^2 = 0.88$; $F=50.00$; $Q2 = 0.78$</p>
	<p>Property name: API gravity/Specific gravity</p> <p>Model:</p> $Y = -6.864 (\pm 0.8678) xi + 632.042 (\pm 75.019)$ <p><i>xi</i>: Carbon</p> <p>$R^2 = 0.90$; $F=62.6$; $Q2 = 0.86$</p>
	<p>Property name: Bromine number</p> <p>Model:</p> $Y = 0.029 (\pm 0.007) xi + 0.123 (\pm 0.034)$ <p><i>xi</i>: Tertiary carbon in naphthalene units and methyl substituted aromatics</p> <p>$R^2 = 0.69$; $F= 15.8$; $Q2 = 0.57$</p>

<p>Horizontal bar chart showing Q2 values for 13 datasets for Cloud point property. The y-axis lists datasets 1-13, and the x-axis shows Q2 from 0 to 0.8. Datasets 1-12 are red, and dataset 13 is grey.</p>	<p>Property name: Cloud point Model: $Y = 22.213(\pm 4.608) xi - 51.507(\pm 4.437)$ <i>xi: n-paraffins-C16</i> $R^2 = 0.77$; $F = 23.2$; $Q2 = 0.65$</p>
<p>Horizontal bar chart showing Q2 values for 13 datasets for Cetane number by engine method property. The y-axis lists datasets 1-13, and the x-axis shows Q2 from 0 to 0.8. Datasets 1-12 are red, and dataset 13 is grey.</p>	<p>Property name: Cetane number by engine method Model: $Y = -0.743 (\pm 0.083) xi + 67.018(\pm 3.14)$ <i>xi: Monoaromatics (GCFIMSmmain-hc)</i> $R^2 = 0.92$; $F = 80.2$; $Q2 = 0.87$</p> <p>$Y = 0.995 (\pm 0.118) xi - 7.229 (\pm 5.806)$ <i>xi: paraffinic CH2 (HNMR)</i> $R^2 = 0.91$; $F = 70.8$; $Q2 = 0.86$</p>
<p>Horizontal bar chart showing Q2 values for 13 datasets for Flash point property. The y-axis lists datasets 1-13, and the x-axis shows Q2 from 0 to 0.8. Datasets 1-12 are red, and dataset 13 is grey.</p>	<p>Property name: Flash point Model: $Y = 2.758(\pm 0.534) xi + 138.619(\pm 2.152)$ <i>xi: dicycloparaffins by MS D2425</i> $R^2 = 0.79$; $F = 26.6$; $Q2 = 0.66$</p>
<p>Horizontal bar chart showing Q2 values for 13 datasets for kinematic viscosity property. The y-axis lists datasets 1-13, and the x-axis shows Q2 from 0 to 0.8. Datasets 1-12 are red, and dataset 13 is grey.</p>	<p>Property name: kinematic viscosity Model: $Y = 0.021 (\pm 0.002) xi - 2.2611 (\pm 0.447)$ <i>xi: T24 Simdis</i> $R^2 = 0.93$; $F = 94.2$; $Q2 = 0.89$</p>



Regression models (or trends) were further validated on external datasets extracted from two other CRC reports (33, 34). Models used for prediction of the selected properties of the fuels are taken from Table 4 and the results of the predictions for 16 fuels are shown in Figure 15. Based on the presented results we can conclude:

- Properties such as net heat of combustion, kinematic viscosity, and density are relatively easy to model based on provided chemical information encoded in the set of nine FACE fuels.
- CN by engine method predictions based on $^1\text{HNMR}$ data are useful, especially for ‘regular’ fuels such as ULSDs, SO1, SO2, OS1, FD4b, and FD7b. It is interesting that only one HNMR-based fuel descriptor produces such good results.
- Based on simdis data and after implementing the PLS algorithm we were able to predict kinematic viscosity of unique fuels such as GTL and RDs.
- Properties such as cloud point or bromine number are difficult to model. These and similar quality models need to be re-evaluated on larger populations of fuel samples.

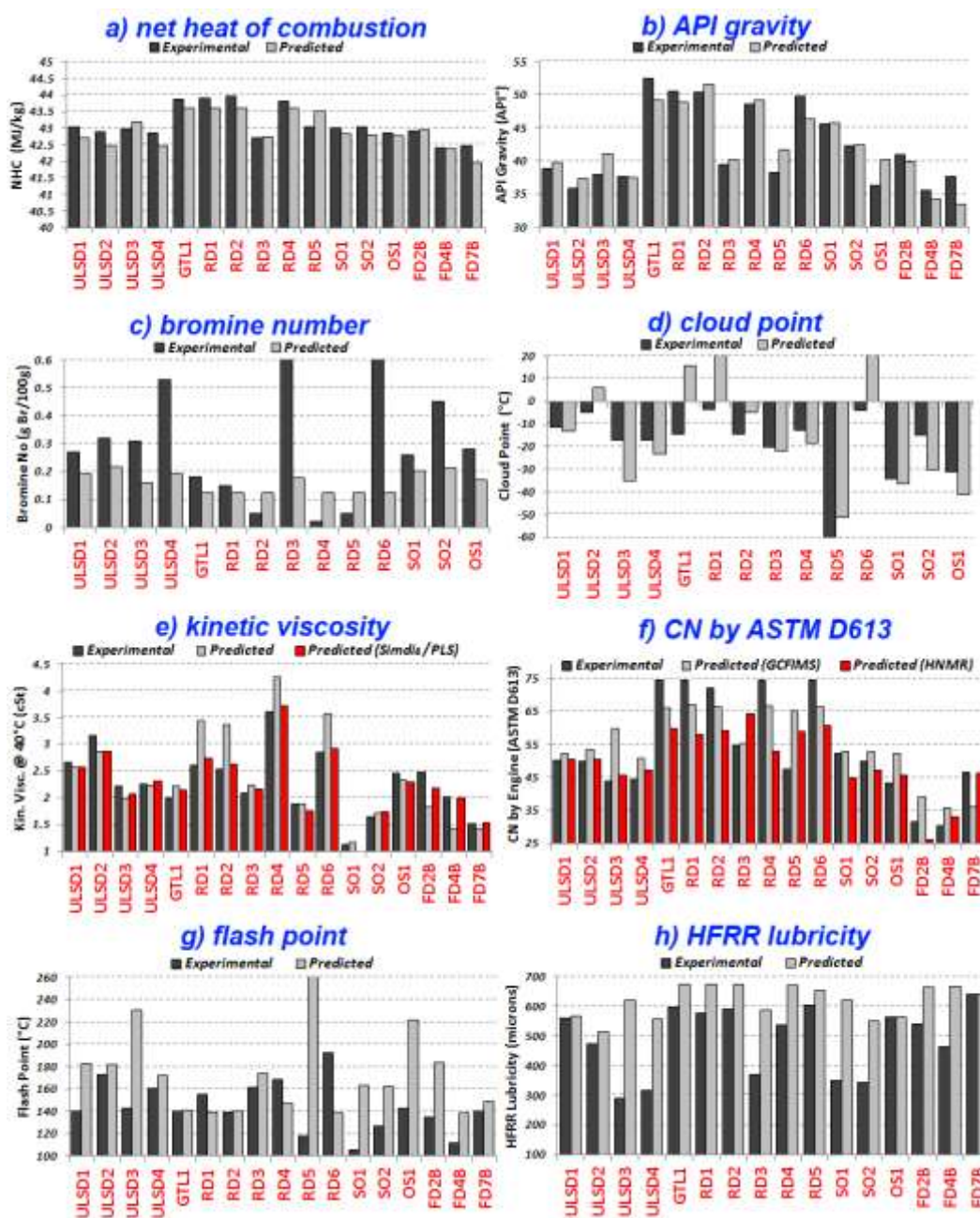


Figure 15 – External validation of selected regression models presented in Table 4 on two datasets of fuels (33, 34) for selected physical properties: Net heat of combustion (a), API gravity (b), bromine number (c), cloud point (d), kinetic viscosity (e), cetane number by engine method (f), flash point (g), and HFRR lubricity (h).

6.0 ENGINE DATASETS

To date, eight research engine experimental studies have been conducted with the FACE diesel fuels by research and academic institutions in the US and Canada (3-10, 35). In those studies, various types of engines and operating conditions were investigated, including conventional diesel combustion (CDC), homogenous charge compression ignition (HCCI), and low-temperature combustion (LTC). Engine specifications and operating conditions for the aforementioned studies are listed in Table 5 together with references to the original papers where the studies are reported.

Table 5 – Engine specifications and operating conditions (based on the table presented in (35))

<i>Engine tag</i>	<i>Case</i>	<i>Cylinder displacement (l)</i>	<i>Compr. Ratio</i>	<i>Fueling</i>	<i>Nozzle</i>	<i>Speed (RPM)</i>	<i>MEP (bar)</i>	<i>EGR (%)</i>	<i>Intake O₂ (%)</i>	<i>Models in this work</i>	<i>Ref.</i>
<i>HCCI_ORNL</i>	<i>HCCI (SCE 0.5L)</i>	0.517	10.5	Atomizer	N/A	1800	2~6	0	N/A	YES	(4)
<i>HECC_ORNL</i>	<i>LTC (MCE 1.9L)</i>	0.477	17.5	CRDI	7-H 148°	1500	2.6	41	N/A	YES	(5)
<i>CDC_ORNL</i>	<i>CDC (SCE 0.5L)</i>	0.517	10.5	Unit injector	N/A	1800	2~6	0	N/A	YES	(6)
<i>LTC_SplitInj (AVFL16)</i>	<i>LTC (MCE 1.9L)</i>	0.477	17.5	CRDI	N/A	2100	3.5	42- 49	16	YES	(10)
<i>Sandia Optical</i>	<i>LTC</i>	1.72	12.3	CRDI	2-H 140°	1500	variable	variable	variable	YES	(3)
<i>LTC_SingleInj (AVFL16)</i>	<i>LTC (MCE 1.9L)</i>	0.477	17.5	CRDI	N/A	2100	3.5		variable	NO	(10)
<i>HCCI_NRC</i>	<i>HCCI (SCE CFR)</i>	0.612	9~15	Air Assist	N/A	900	4	60	N/A	NO	(7)
<i>LTC_NRC</i>	<i>LTC (SCE 2.44L)</i>	2.44	16.25	CRDI	N/A	900	3.5	60	N/A	NO	(8)
<i>LTC_Navistar</i>	<i>LTC (SCE 0.744L)</i>	0.744	16.2	CRDI	6-H 156°	1671	5.5~1 4.6	N/A	13~17	NO	(9)

In the current study, not all the engine studies listed in Table 5 were analyzed. We were limited to the studies where a complete engine dataset is available. In a few cases (indicated by “NO” in the table) we had no access to the original data (only to selected values provided by Kim *et al.* (35). We restricted the study to engine data generated by Oak Ridge National Laboratory (ORNL), West Virginia University (CRC AVFL-16) and Sandia National Laboratories. All available engine data were correctly formatted and are ready for further statistical studies.

Not all the engine studies presented in Table 5 evaluated the complete set of the 9 FACE diesel fuels. For instance, the Sandia study examined only four of the nine FACE fuels and a

certification ULSD#2 (aka CFA fuel). In the LTC engine (West Virginia University – AVFL-16) case, FD2 fuel was replaced by a certification ULSD#2.

6.1 MODELING STRATEGIES

In this study the engine emissions and efficiency properties are modeled using several strategies presented in Figure 16; here “strategy” means the way in which an input data matrix was constructed.

- Strategy - 1a: only selected engine control parameters,
- Strategy - 1b: Strategy 1a + cross terms for all variables,
- Strategy - 2a: Strategy 1a + FACE fuel CN/T90/aromatic content,
- Strategy - 2b: Strategy 2a + cross terms for all variables,
- Strategy - 3a: Strategy 1a + 40 FACE fuel descriptors (see Figure 17),
- Strategy - 3b: Strategy 3a + cross terms for all variables,

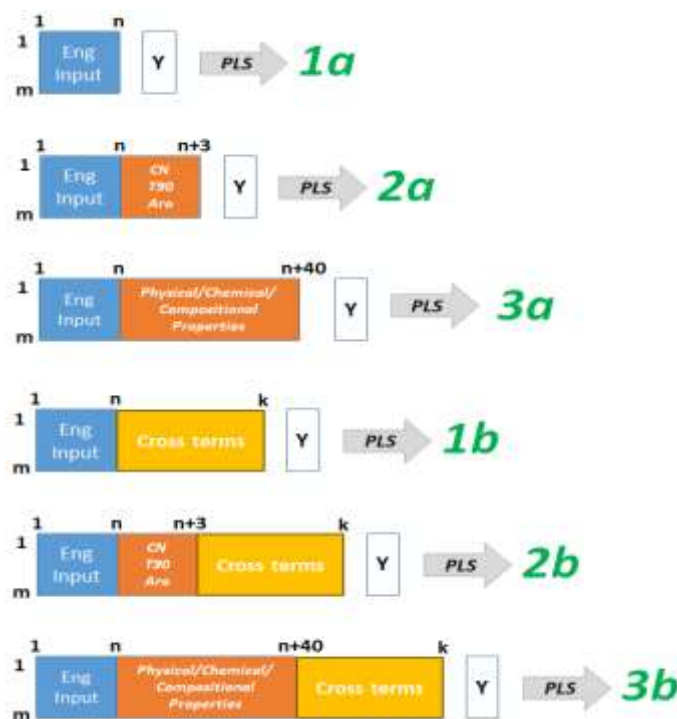


Figure 16 – Set of six modeling strategies used for evaluation of engine control parameters and fuel properties effects on engine emission and efficiency. Where m - total number of experimental points collected for a particular engine, n - the number of engine control parameters, k - total number of variables including cross terms for all variables, Y - dependent variable (engine emission or efficiency).

For each strategy, the relationship between input data and the dependent variable (here engine emissions or engine efficiency) was calculated using the PLS modeling (see section 2.0 for more details). We used PLS because it is less restrictive than MLR regarding assumptions such as a normal distribution of variables and no collinearity between variables. In this sense, PLS is an ideal method for analysis of data with many, noisy, and collinear variables in the dataset.

Strategies 2a and 2b include only three fuel descriptors (CN, T90, and aromatic content) as these were the target fuel properties selected in the FACE diesel fuel matrix design. Moreover, in previous studies that evaluated the influence of fuel chemistry on engine behaviour, these three fuel parameters were identified as the most important.

Strategies 3a and 3b include the set of 40 fuel descriptors (depicted in Figure 17), including CN, T90, and aromatic content. This extended list of fuel descriptors should provide deeper insight in fuel chemistry of analyzed fuels and should help identify which other properties (aside from CN/T90/Aro) may be important in modeling engine output parameters.

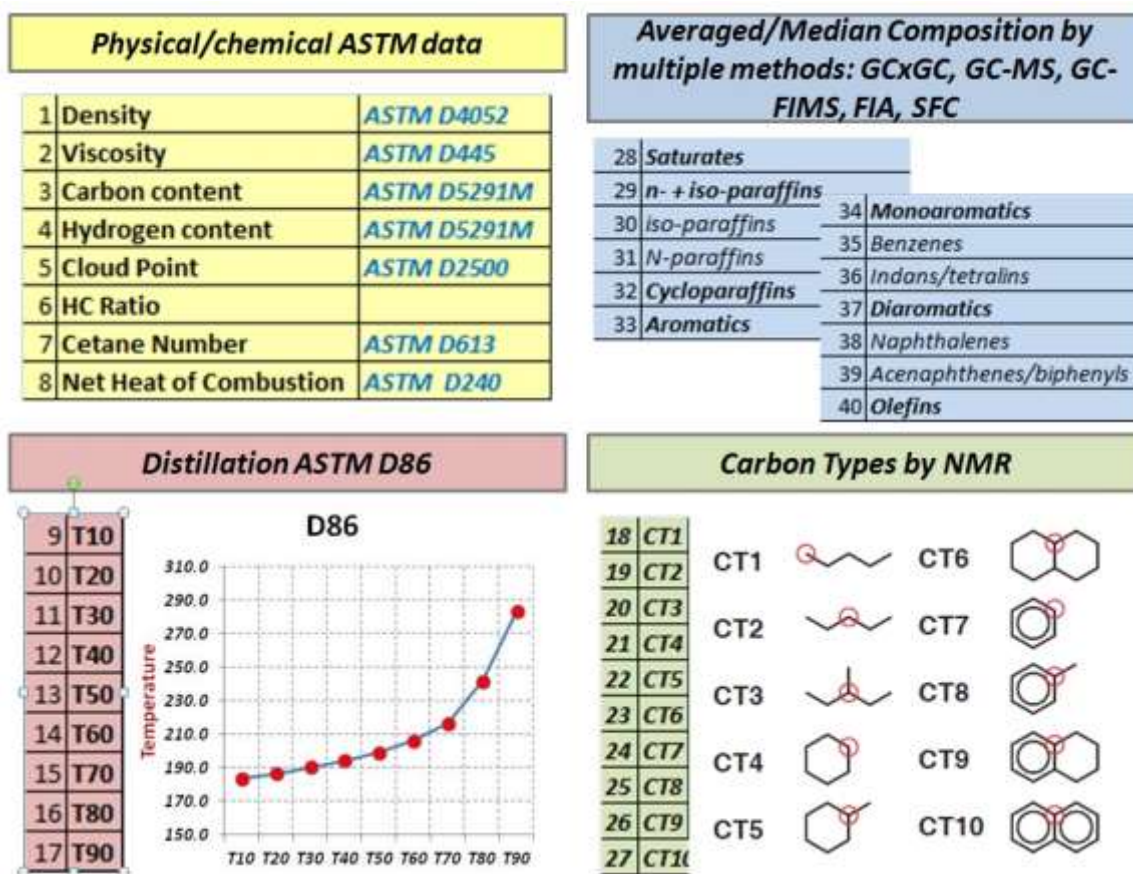


Figure 17 – The list and pictorial description of the 40 selected fuel variables used in modeling approach

For each strategy and engine the input data matrices were auto-scaled according to the scheme presented in Appendix F.

The following ‘engine output’ profiles were modeled for engine datasets available for this study: engine-out emissions (NO_x/HC/CO emission and soot/particulate matter) and combustion/fuel conversion efficiencies. The emission profiles were subjected to logarithmic transformation before analysis (see Appendix F).

All the modeling results are presented in report tables similar to that in Figure 18. Regression equations for models built around strategies 1a and 2a consisting of a low number of parameters are presented in Appendix H. The models developed for other strategies will not be presented in this report, because of their complexity.

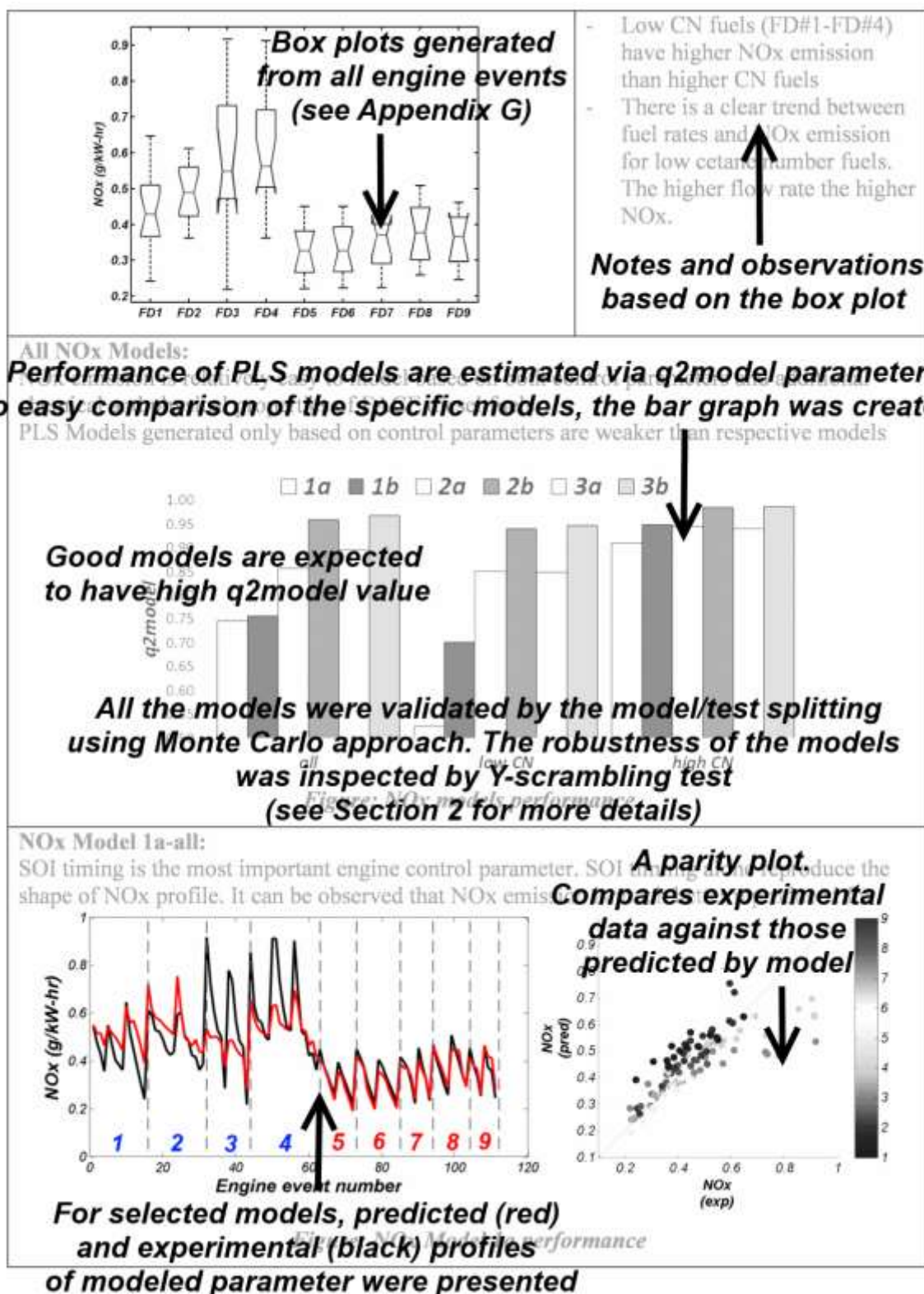


Figure 18 – PLS modeling results – an example

7.0 ENGINE DATASETS – ANALYSIS

7.1 HCCI–SCE 0.5L (ORNL)

The nine FACE diesel fuels were evaluated at ORNL in a premixed HCCI engine under varying conditions of fuel rate, air/fuel ratio, and intake temperature (4). Authors of that work found that engine performance expressed by parameters such as NO_x, CO, HC, ISFC, ITE, MB50, etc. was dependent on fuel CN and engine control variables.

7.1.1 DATA DESCRIPTION

The scheme in Figure 19 depicts the list of input control and output parameters from the HCCI - SCE 0.5L study (as provided by ORNL researchers and placed in our FACE diesel engine database). For the nine fuels, a total of 567 experimental data points (aka engine events) were taken. Four fueling rates were run for each fuel.

No	Tag	In/Out parameter	unit
v1		Test Description1	
v2		Test Description2	
v3		Test Description3	
v4	I1	fuel rate	g/min
v5	I2	airflow	g/min
v6	I3	lambda	corrected
v7	I4	int temp actual	C
v8	I5	intake press	psi gage
v9	I6	intake press	kPa

No	Tag	In/Out parameter	unit
v10	O1	exh temp	C
v11	O2	HC	ppm
v12	O3	CO	ppm
v13	O4	NO _x	ppm
v14	O5	IMEP	bar
v15	O6	COV IMEP	%
v16	O7	MFB10	deg ATDC
v17	O8	MFB50	deg ATDC
v18	O9	MFB90	deg ATDC
v19	O10	CA10-90	deg CA
v20	O11	dP/dCA max	bar/deg
v21	O12	LTHR	%
v22	O13	ISFC	g/kwh
v23	O14	ITE	fraction

Figure 19 – Scheme presenting list of control input (I*) and output (O*) engine parameters for HCCI engine dataset

7.1.2 EXPLORATORY ANALYSIS

In the first step of analysis of engine data the correlations among input engine control and engine output parameters were performed and presented in scatter plot format. The results are shown in Figure 20 and Figure 21, respectively. In the case of a strong one-to-one relationship, a plot of the two variables should yield a more-or-less straight line.

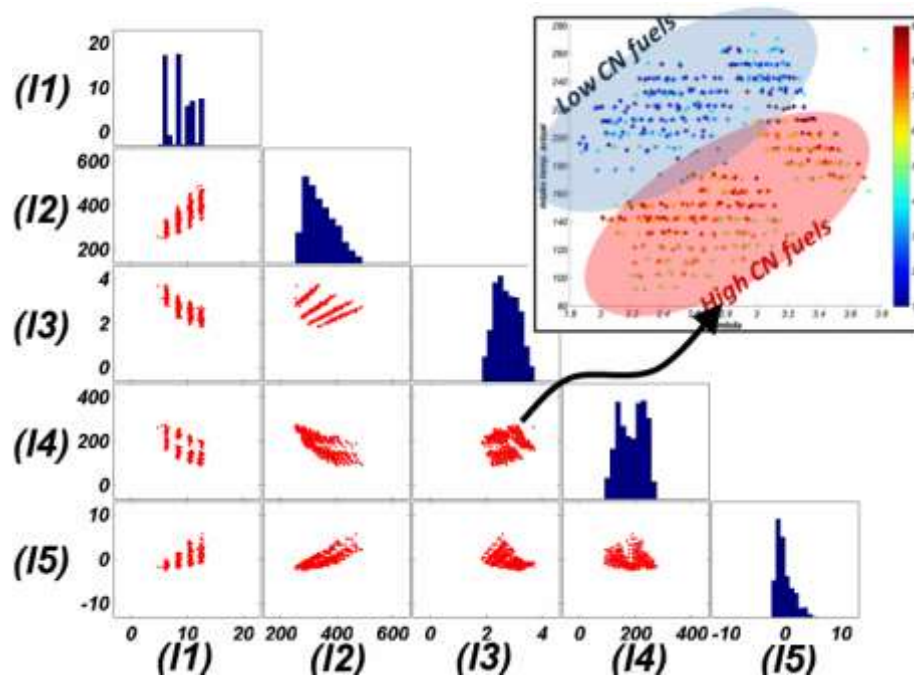


Figure 20 – Scatter representation of cross-correlation between control input engine parameters where (I1) fuel rate (g/min), (I2) airflow (g/min), (I3) lambda (corrected), (I4) actual intake temp (°C), (I5) intake pressure (psi). Histograms for each variable are located on the diagonal. The inset in the top-right corner presents the relation between lambda and actual intake temperature. Colour of the points in the inset represents the respective FACE fuel.

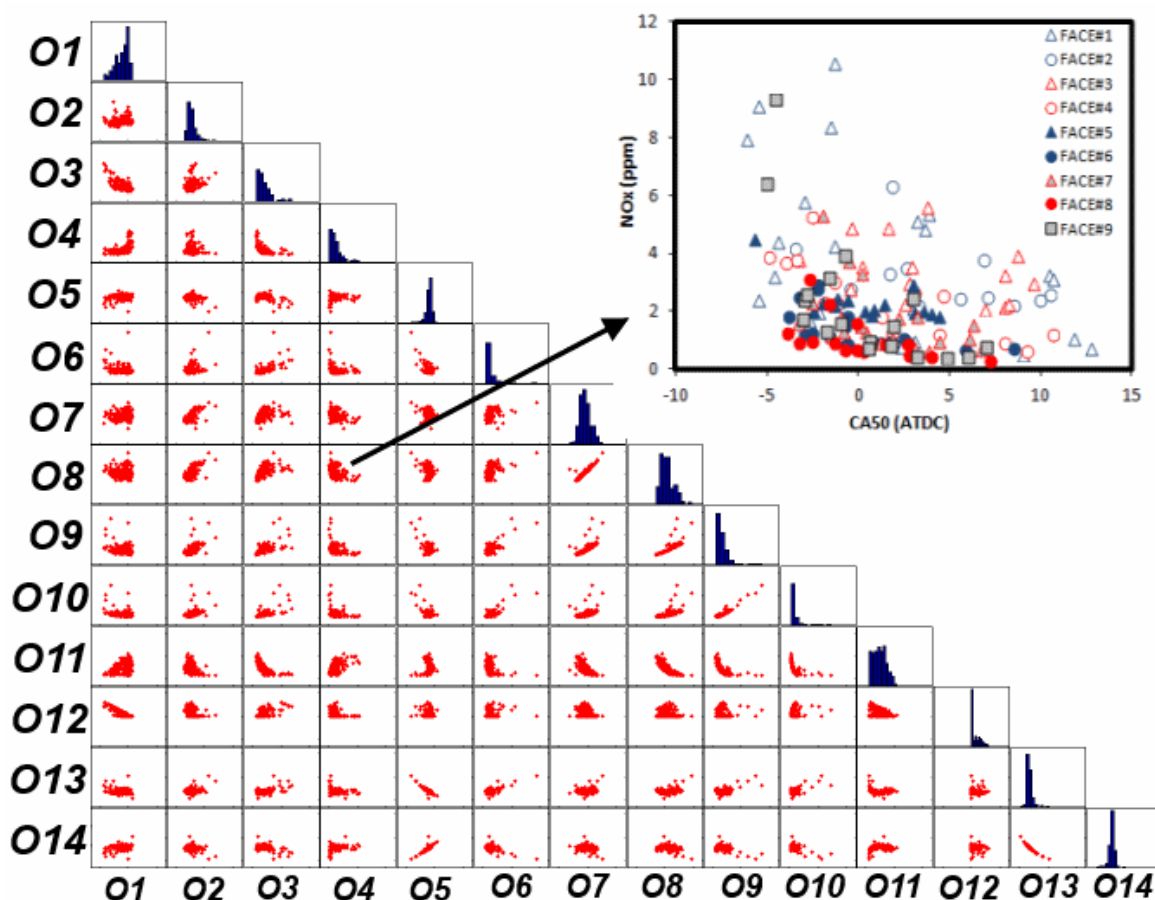


Figure 21 – Scatter representation of cross-correlation between output engine parameters obtained for engine run with fuel rate ~ 6.2 g/min. The meaning of each output (O*) is presented in Figure 19. The inset located in the top-right corner magnifies the existing relations between combustion phasing for O8, the 50% burn point (MFB50 aka CA50) and NOx emissions (O4).

PCA was performed on both input and output data provided to CanmetENERGY. The scores plot for input data (Figure 22) shows clear separation of engine control data points into two groups: low and high CN. This means that the selected engine control parameters were strongly influenced by fuel chemistry, in this case CN. In addition to a clear separation based on the fuel cetane, the scores plot reveals four groups of points aligned in parallel to each other. Separation of points in the PCA space corresponds to the fact that the engine was run with four different fueling rates.

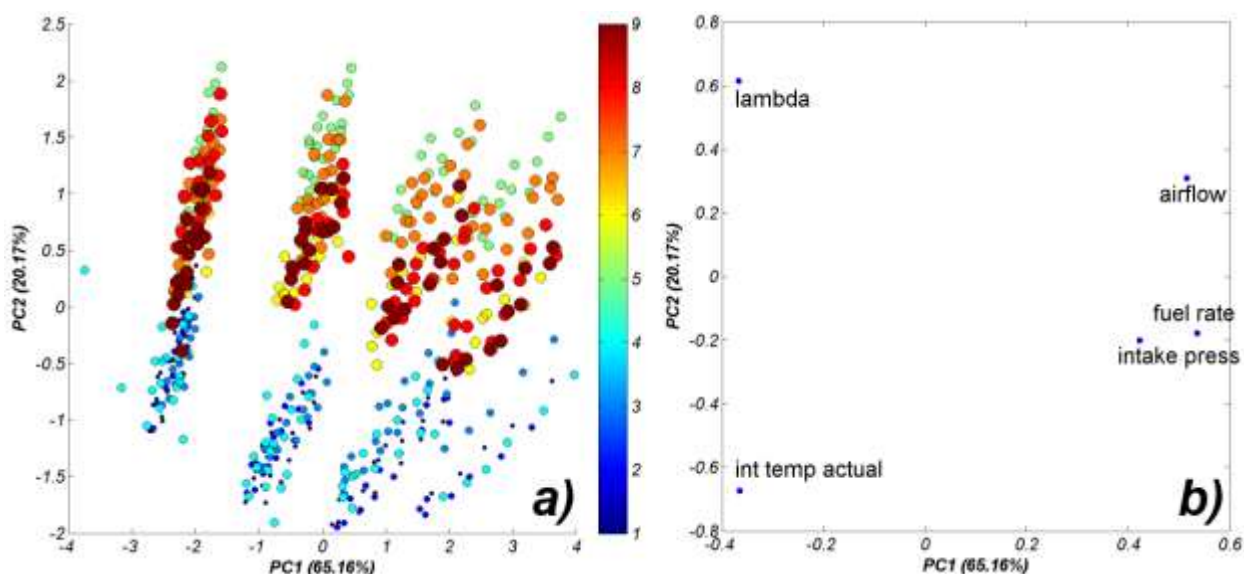


Figure 22 – PCA scores (a) and loadings (b) plot (PC1-PC2) obtained after running PCA for HCCI ORNL engine input (control) parameters. On the scores plot, bubble colour and size are associated with the respective fuels (i.e., smallest dark blue dots – FD1, largest brown bubbles – FD9).

Figure 23 presents results of the PCA calculated on all engine output parameters obtained for all 567 engine scenarios in the form of a scores plot (Figure 23a) and a loadings plot (Figure 23b). In the scores plot, bubbles are differentiated by colour and size, both of which are associated with the fuel number – shown in the colour bar (i.e., smallest dark blue dots – FD1, largest brown bubbles – FD9). On the scores plot we can see clustering of points into three groups. The first group (blue bubbles) consists of events obtained for the low-CN fuels (FD1-FD4). The second group consists of events run on high-CN fuels (FD5-FD9). Within the latter group we can observe further splitting of objects into two groups. One of them contains events run on three FACE fuels (FD5, FD7, and FD9) and the other consists of fuels FD6 (yellow) and FD8 (red). It is worth mentioning that the groupings do not appear to be entirely exclusive, but overlap to some extent. The loadings plot (Figure 23b) explains the observed separation of objects seen in the scores plot. Based on the distribution of variables in the loadings plot we observe that events derived on low-CN fuels produce more NO_x and are characterized by higher thermal efficiency (furthest to the right in the plot). High-CN fuels produce more CO and HC emissions. Indicated specific fuel consumption (ISFC) and low temperature heat release (LTHR)

are the highest for FD6 and FD8 (yellow and red bubbles in the plot). The loadings plot provides information about correlations among engine output parameters:

- ITE and ISFC are inversely correlated: these two parameters are on opposite sides of the loadings plot.
- Exh. Temp, NO_x, IMEP parameters are correlated to a certain degree.
- HC and CO are inversely correlated with exh. temperature and NO_x emissions.
- All the MFBs are correlated.

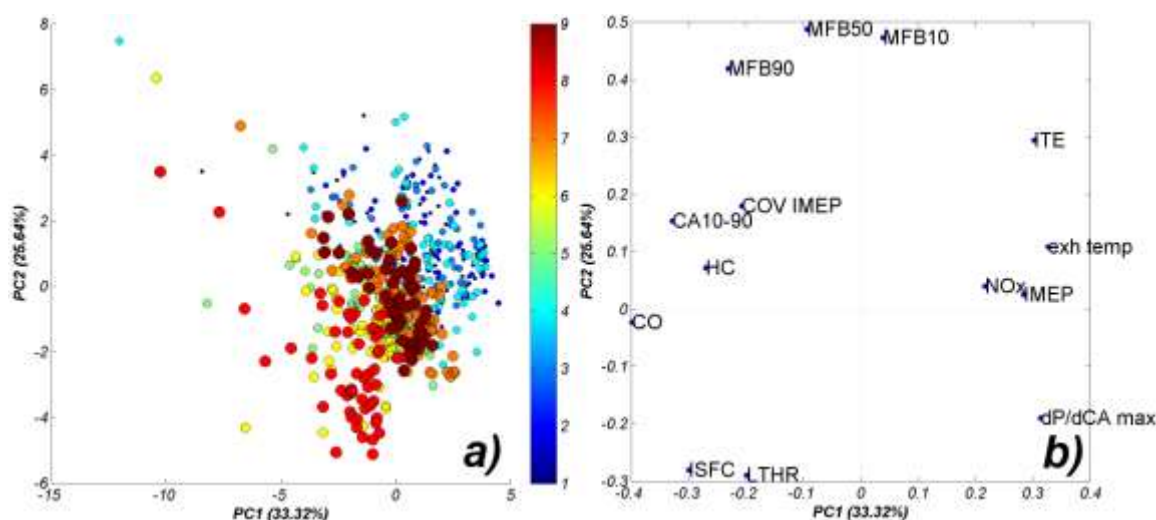


Figure 23 – PCA scores plots (a) and loadings plots (b) (PC1-PC2) obtained after running PCA for the HCCI ORNL engine output parameters for all events. On the scores plots, bubble colour and size are associated with the respective fuels (i.e., smallest dark blue dots – FD1, largest brown bubbles – FD9).

Figure 24 presents the PCA scores plots calculated for engine output parameters run at specified fuel rates for all fuels. The scores plots for the analysis indicates that the objects are not arranged in a random way. By moving from left to right along the x-axis (especially for higher fuel rates, i.e. Figure 24b - d), a simple pattern can be observed whereby engine events run on low-CN FACE fuels are mainly located on the left side of the scores plot. In this case we can observe a tendency for increased separation of objects with increasing fuel rate. For example, for low fuel rate (Figure 24a) the separation of objects for low- and high-CN fuels is not as clear as seen for high fuel flow rate in Figure 24d. Moreover, in a group of events derived from higher-CN fuels we observe further splitting between the groups of medium-CN fuels (FD7, FD9, and FD5) and high-CN fuels (FD6 and FD8). Interestingly, for the lowest fuel rate (~6.3 g/min)

events run on fuel FD5 are located within group of FD6 and FD8 fuels. This is not the case in higher fuel rates experiments, where FD5 objects are always grouped together with FD7 and FD9 objects.

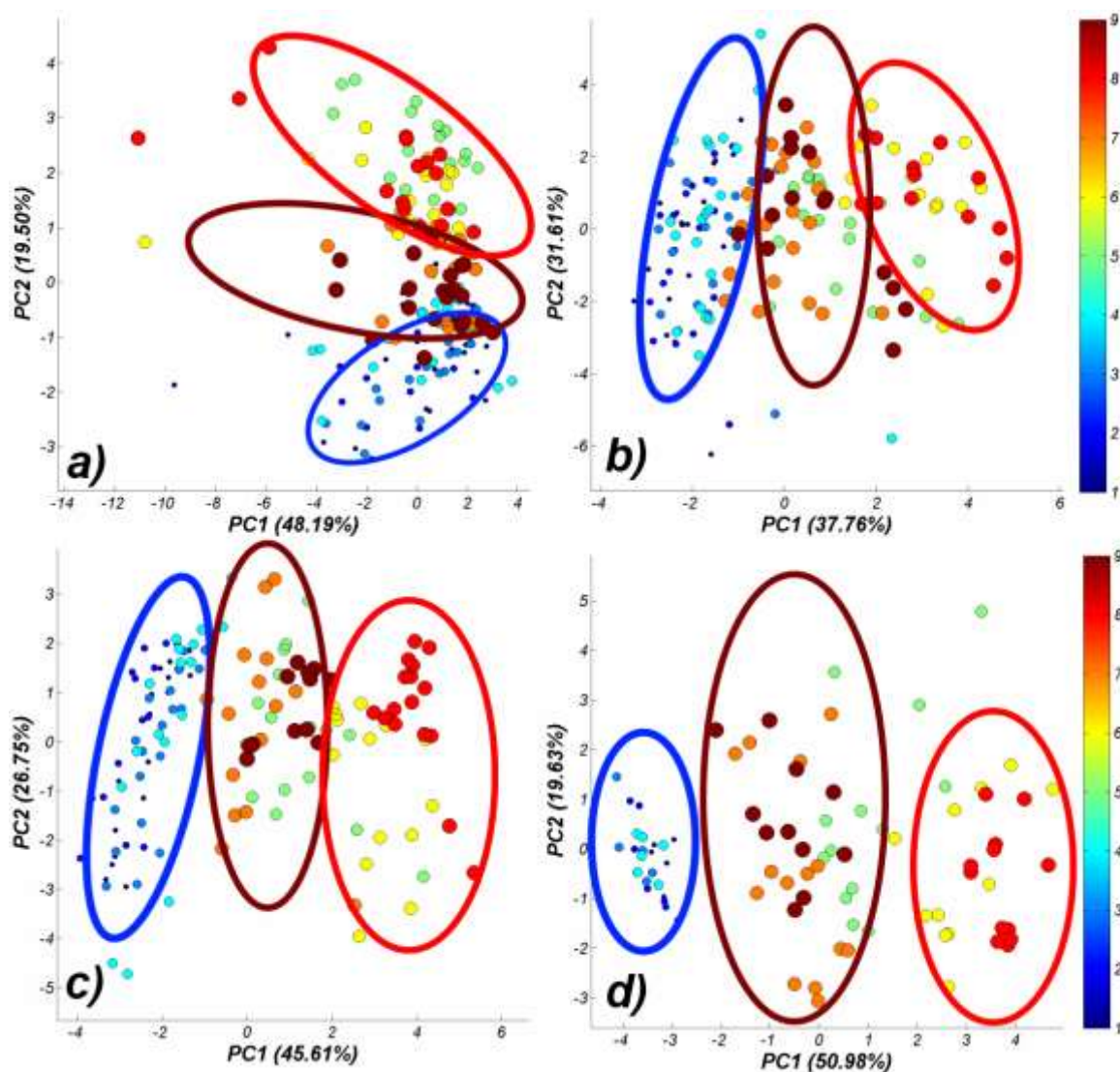


Figure 24 – PCA scores plots for engine output parameters measured for fuel flow rates (a) ~6.3 g/min, (b) ~8.4 g/min, (c) ~10.5 g/min, (d) ~12.5 g/min. The objects were assigned to three clusters (represented by color coded circles).

We noted previously that during experimental design the control space was divided into low ($CN < 40$) and high ($CN > 40$) cetane fuels (see Figure 22). Similar separation is observed in the PCA space based on engine output parameters. However, here events derived from two high-CN fuels i.e. FD6 and FD8 (see yellow and red bubbles in Figure 23a and Figure 24b – d) show

an additional degree of separation. The obvious conclusion can be summarized as: the distribution of points generated by this ORNL SCE 0.5L engine operated in HCCI mode depends not only on the control variables but also on the fuel properties. As fuel flow rate changes, FD6 and FD8 remain overlapped, suggesting that differences in outputs from these two fuels, as a function of flow rate should come from chemistry differences only. This is where there appears to be only major differences in aromatic carbon types.

Next, we investigated which physical and chemical properties (out of 40 selected in section 6.0) may have affected the behaviour of events in the PCA space. We found that only a few fuel descriptors follow the trends observed in the PCA scores plots; specifically, viscosity and the T10-T30 values differentiate the FD6/FD8 fuel pair from the other fuels. CN, CT2, CT8, and *n*-paraffinic content affect the separation based on low- and high-CN events. Figure 25 presents a simple depiction of the aforementioned conclusions.

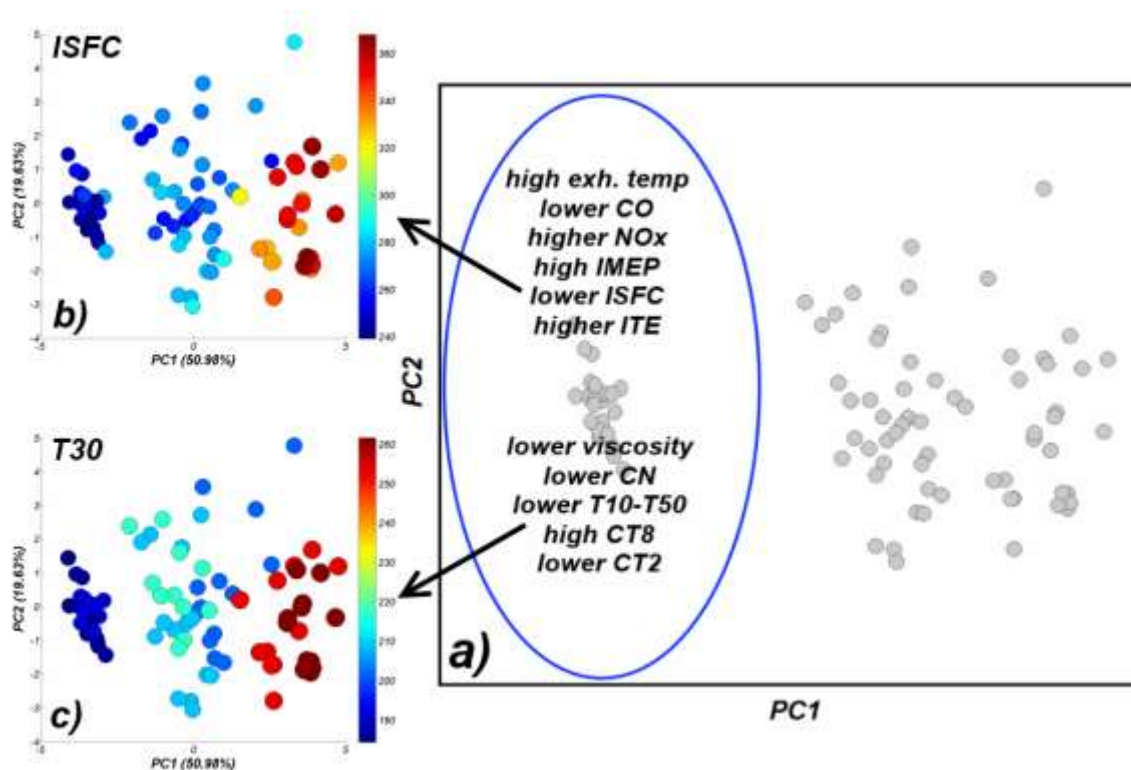


Figure 25 – Scheme illustrating influence of engine output (ISFC) and chemistry of the fuels (T30) on observed engine object splitting in PCA scores plot for engine output parameters measured for fuel rate ~12.5 g/min

7.1.3 MODELING APPROACH

Simple one-to-one correlations between engine control parameters and selected engine output parameters are shown in Figure 26.

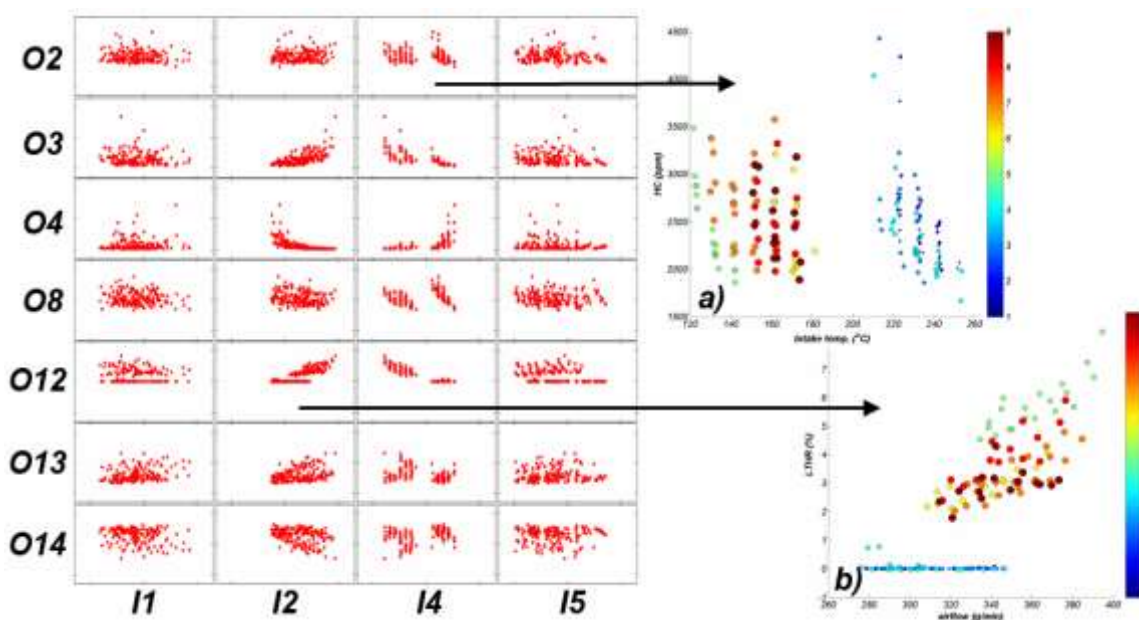


Figure 26 – Scatter representation of cross-correlation between output engine parameters (O*) and engine control variables (I*) obtained for engine run with a fuel flow rate ~ 6.3 g/min. The insets on the right side present two examples of the bivariate relationships coloured by fuel number.

In this section the results of modeling of emission and efficiency parameters will be presented. The engine was modeled using strategies described in section 6:

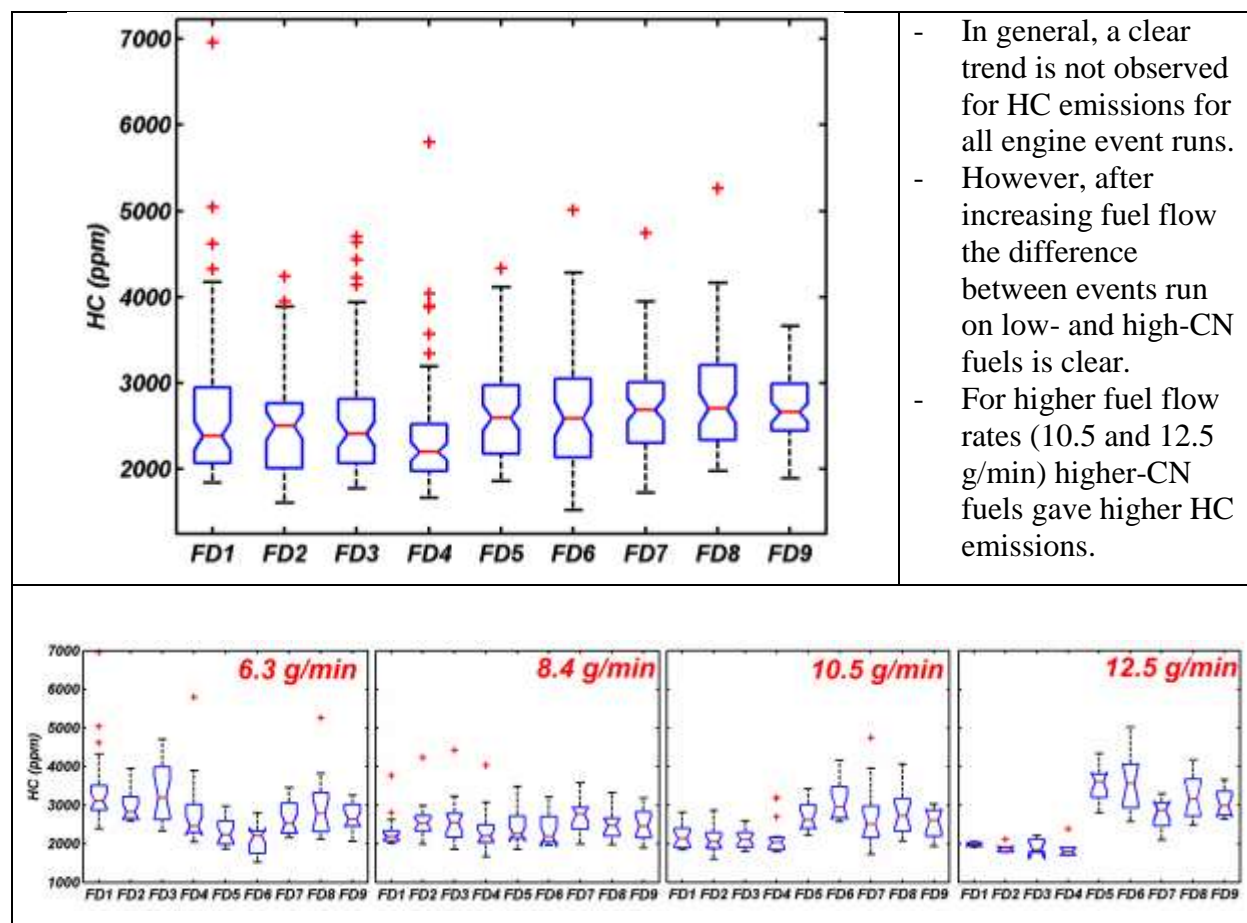
- (strategy: 1a): fuel rate (I1), air flow (I2), lambda (I3), intake temperature (I4), intake pressure (I5),
- (strategy: 1b): strategy 1a + cross terms for all variables,
- (strategy: 2a): strategy 1a + FACE fuel CN/T90/aromatic content,
- (strategy: 2b): strategy 2a + cross terms for all variables,
- (strategy: 3a): strategy 1a + 40 FACE fuel descriptors (see section 6),
- (strategy: 3b): strategy 3a + cross terms for all variables.

PLS models were built for: 1) all events; 2) events run with selected fuel rates: ~6.3 g/min, ~8.4 g/min, ~10.5 g/min, ~12.5 g/min; and finally 3) events performed only on high-CN

fuels (optionally) and/or low-CN fuels (optionally). The last two sets of models for high/low CN were used only in NO_x and ISFC modeling.

For each modeled property at least 30 PLS models were calculated and summary of the results is presented in following sections (7.1.4 – 7.1.8).

7.1.4 HYDROCARBON EMISSIONS



- In general, a clear trend is not observed for HC emissions for all engine event runs.
- However, after increasing fuel flow the difference between events run on low- and high-CN fuels is clear.
- For higher fuel flow rates (10.5 and 12.5 g/min) higher-CN fuels gave higher HC emissions.

All HC models:

Hydrocarbons are difficult to predict using any proposed modeling strategy. In general, PLS models built on combined sets of engine control parameters and fuel descriptors (models 2a, 2b, 3a, and 3b) are stronger than those based only on engine control parameters (1a models). Good PLS models are obtained for engine events run with the highest fuel flow rate (see 12.5 g/min bar plot), where additional improvement may be achieved after adding information about fuel properties/composition.

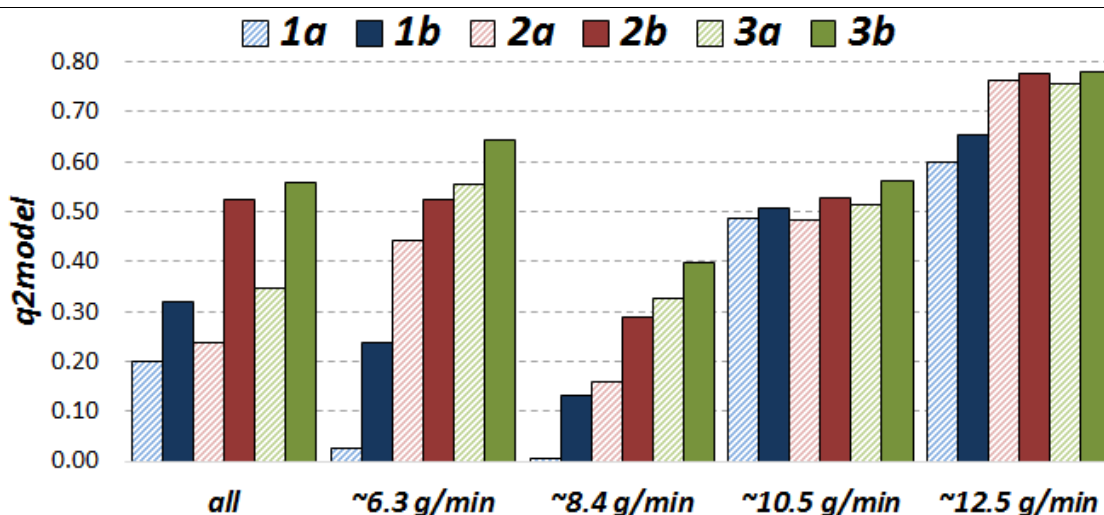


Figure: Figure: HC models performance

HC model 1a-all:

Bad capability to model hydrocarbon emission. Does not improve after combining fuel descriptors in HC modeling.

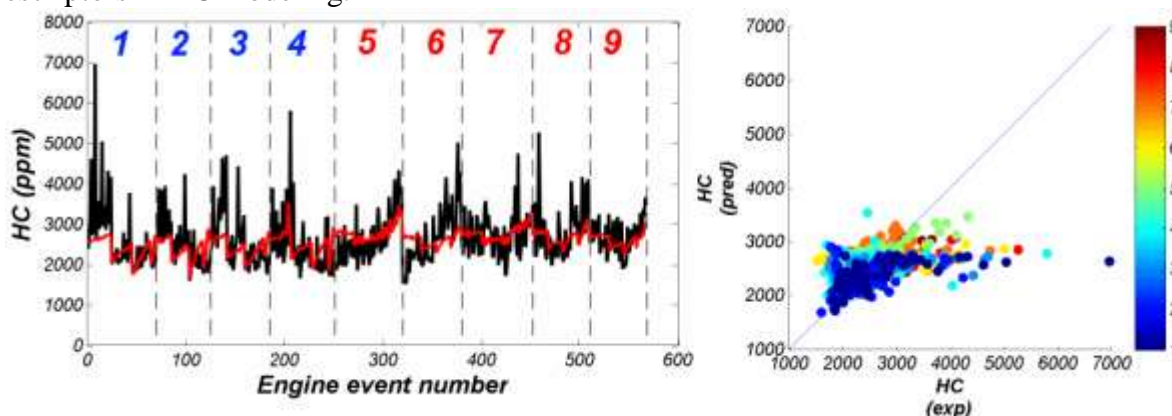
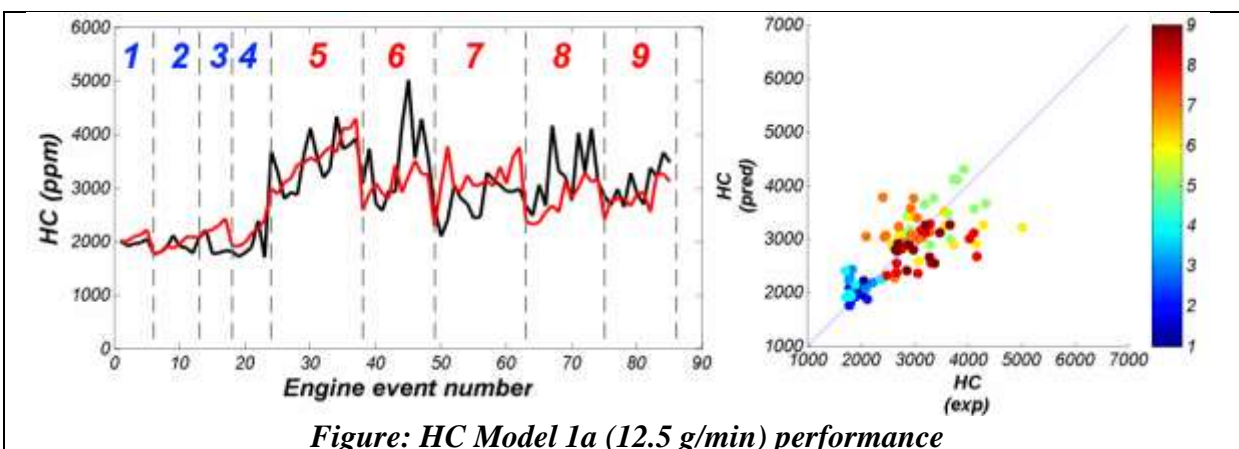


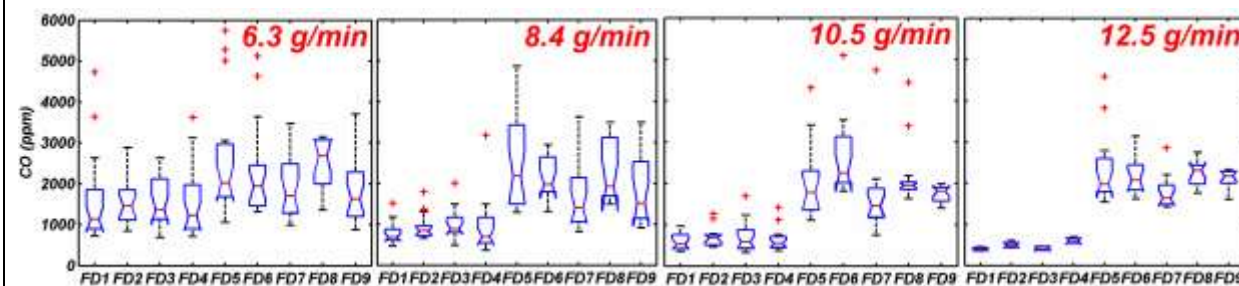
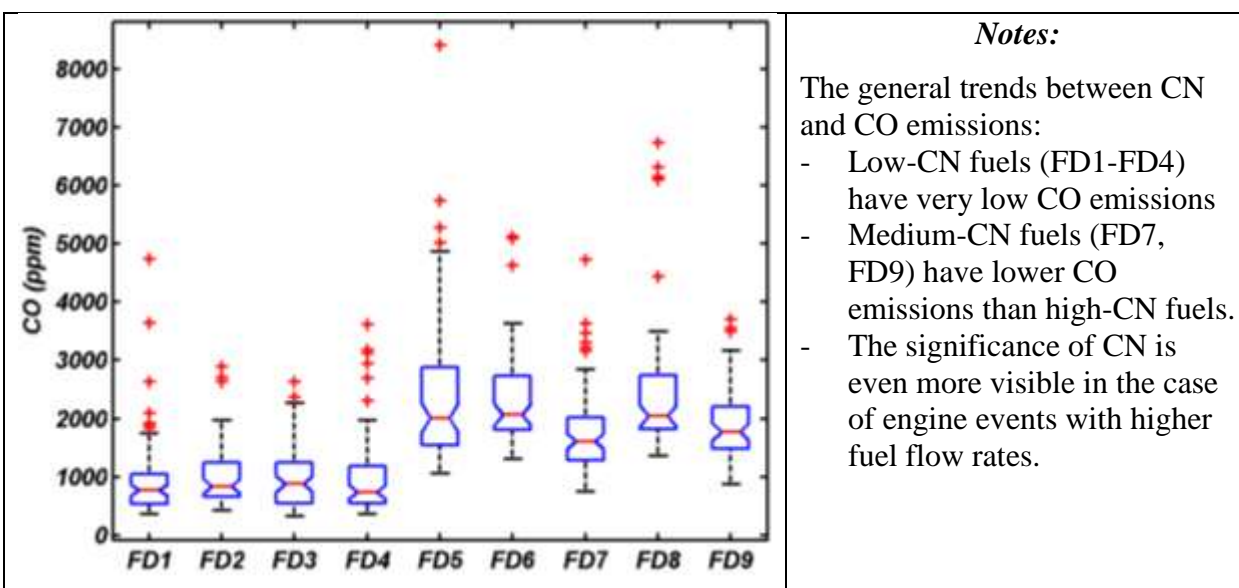
Figure: HC Model 1a performance

HC models for fuel rate 12.5 g/min:

The predicted profile for HC emissions calculated for events run with the highest fuel flow rate follows the experimental trend relatively well (see red vs. black curves in figure below). Important control parameters: Intake temperature (-) and intake pressure (+). HC models improve after including CN/T90/Aro in the modeling. CN (+) is the most important fuel descriptor in this new improved model (2a).



7.1.5 CARBON MONOXIDE EMISSIONS



All CO models:

Models built for all engine events has good CO predictability, which further improve after adding fuel descriptors.

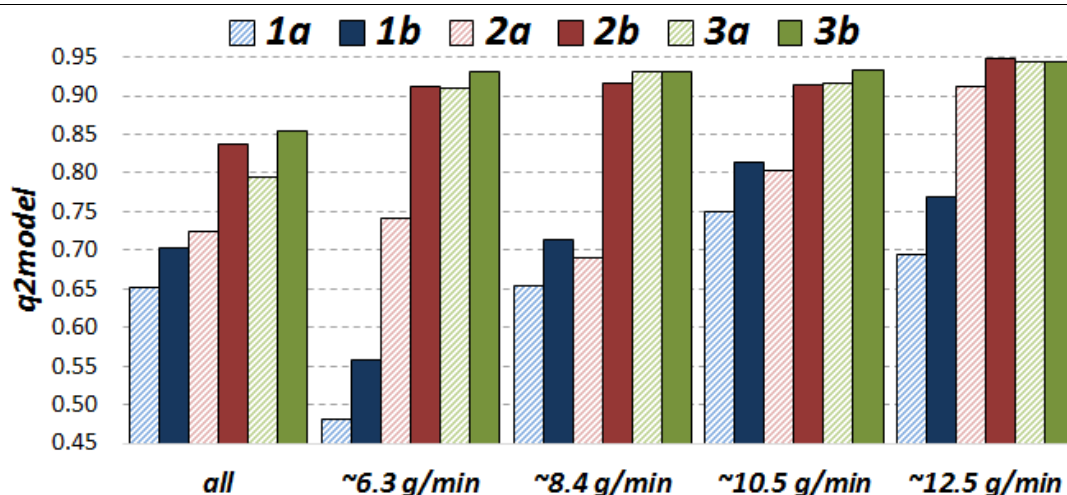


Figure: CO models performance

CO models 1a-all and 2a-all:

Notice that the model 1a-all predicted profile for CO emission (red line) follows experimental CO values (black line) better for low-CN fuels. Note the poor CO emission prediction for FD6 and FD8 fuels.

Important control parameters: Intake temperature (-), intake pressure (-) and lambda (+)

CO model improves after including fuel descriptor parameters (model 2a-all), especially CN/T90/Aromatics, with CN being the most important.

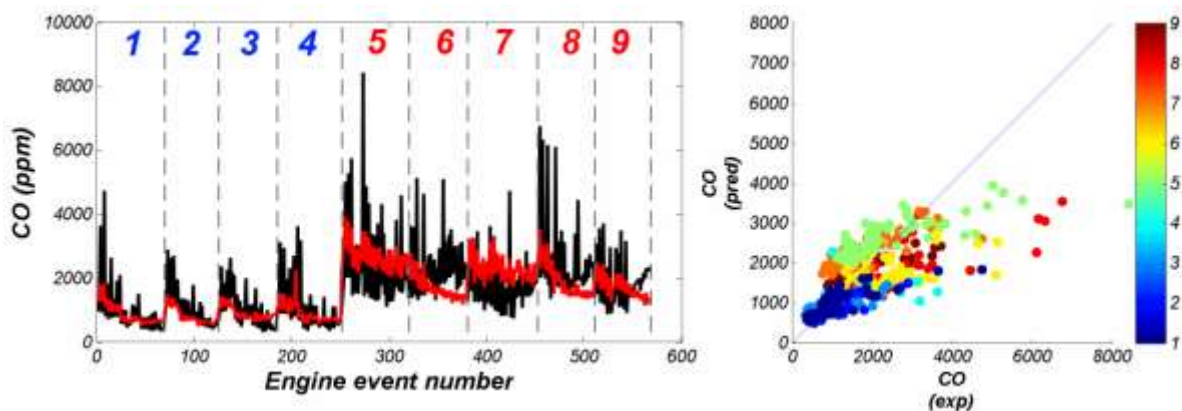
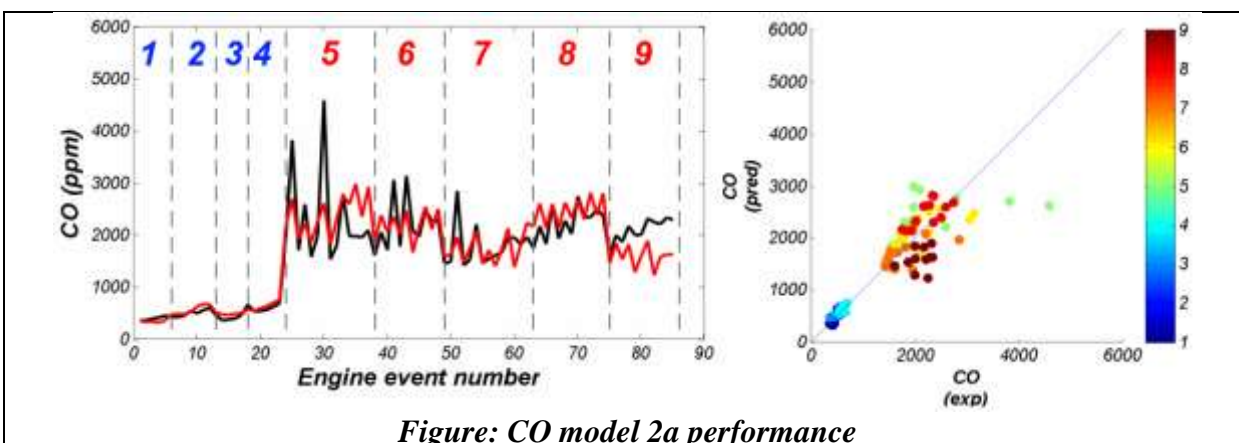


Figure: CO Model 1a performance

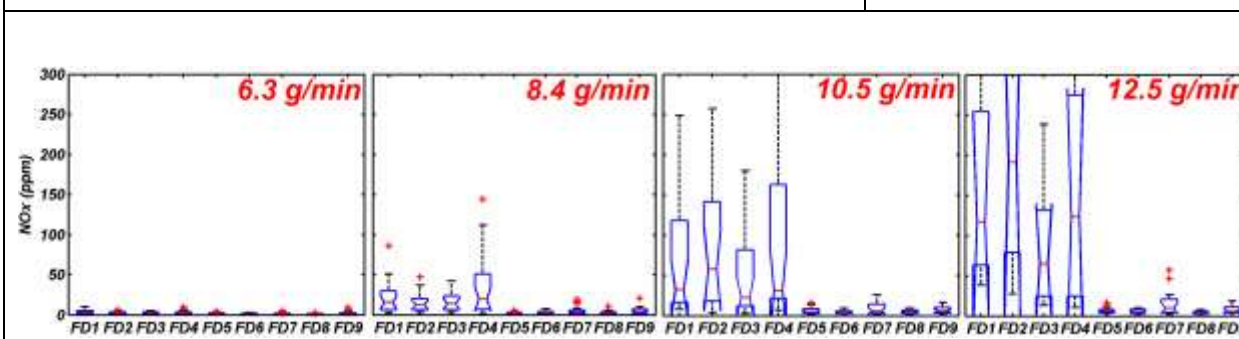
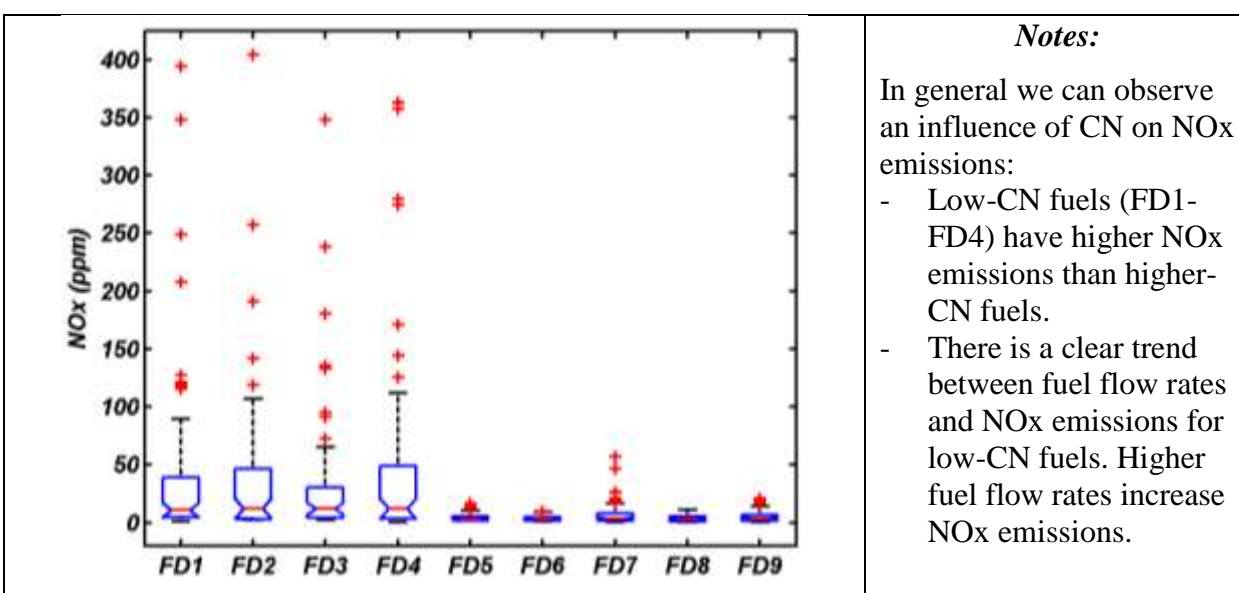
CO Model 2a for fuel rate 12.5 g/min

Important control parameters: Intake temperature (-), intake pressure (-)

Important fuel parameters: cetane number (+), T90 (+)



7.1.6 NITROGEN OXIDES EMISSIONS



All NOx models:

Good PLS models for all events.

The best NOx models were calculated using combined data (control parameters + fuel descriptors, i.e., Models 2a and 3a). Models for low-CN fuels were better than for high-CN fuels.

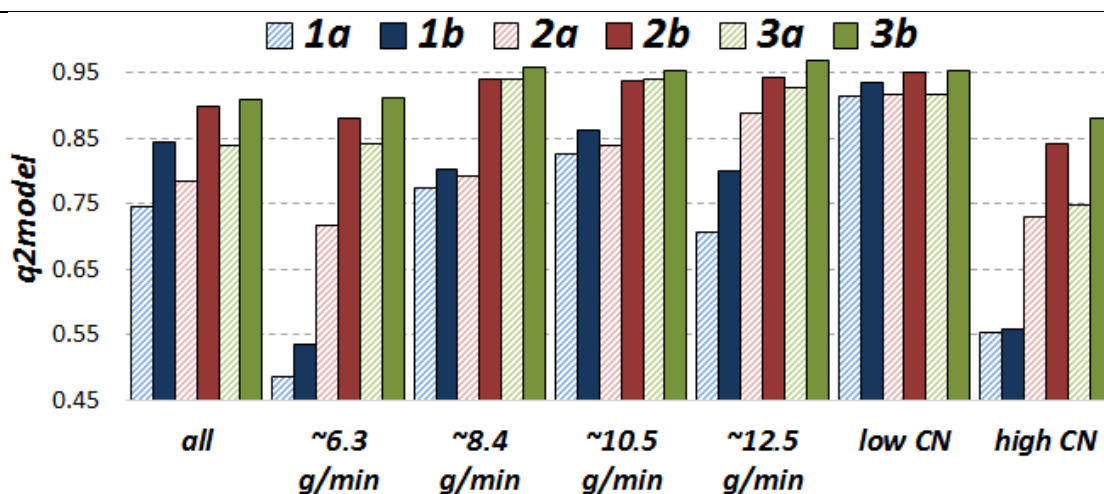


Figure: NOx models performance

NOx model 1a-all:

Observed inability of the NOx model to properly predict high NOx emissions for low-CN fuels.

Important control parameters: lambda (-), fuel rate (+), air flow (-).

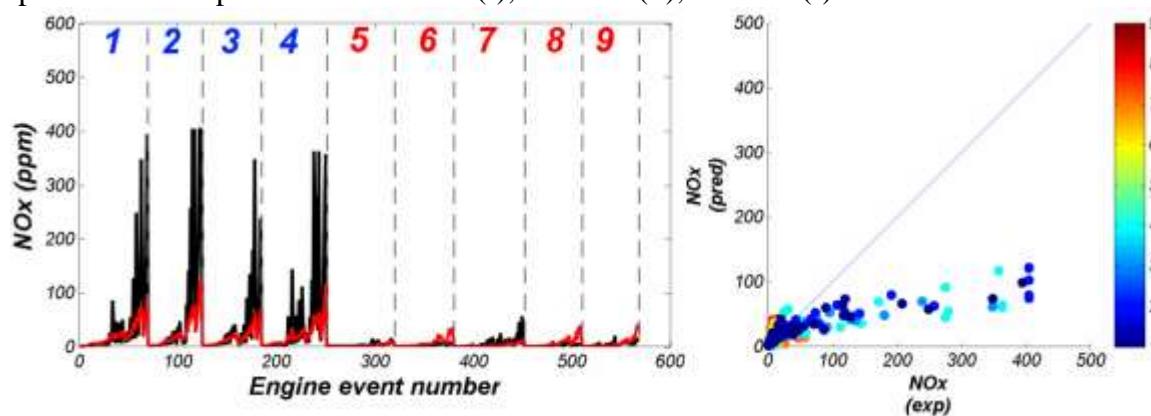


Figure: NOx Model 1a performance

NOx models for low-CN fuels

Good models based only on engine control parameters.

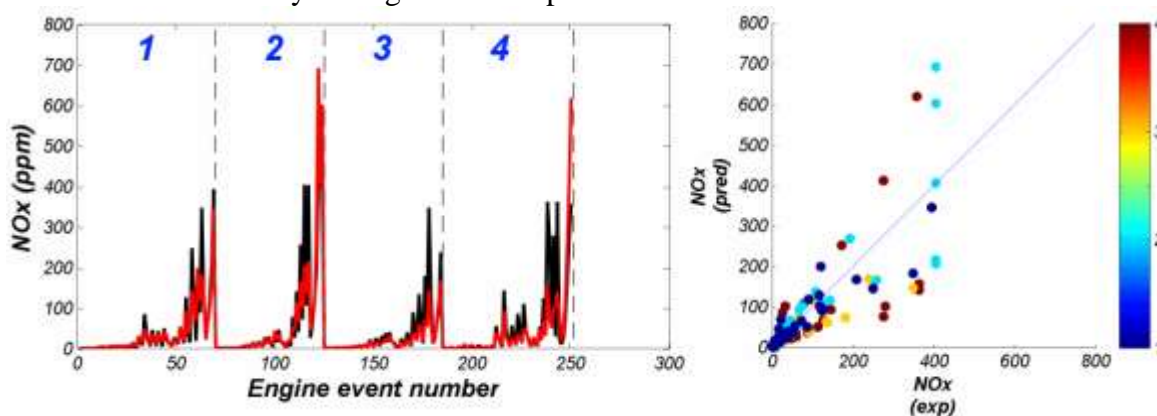
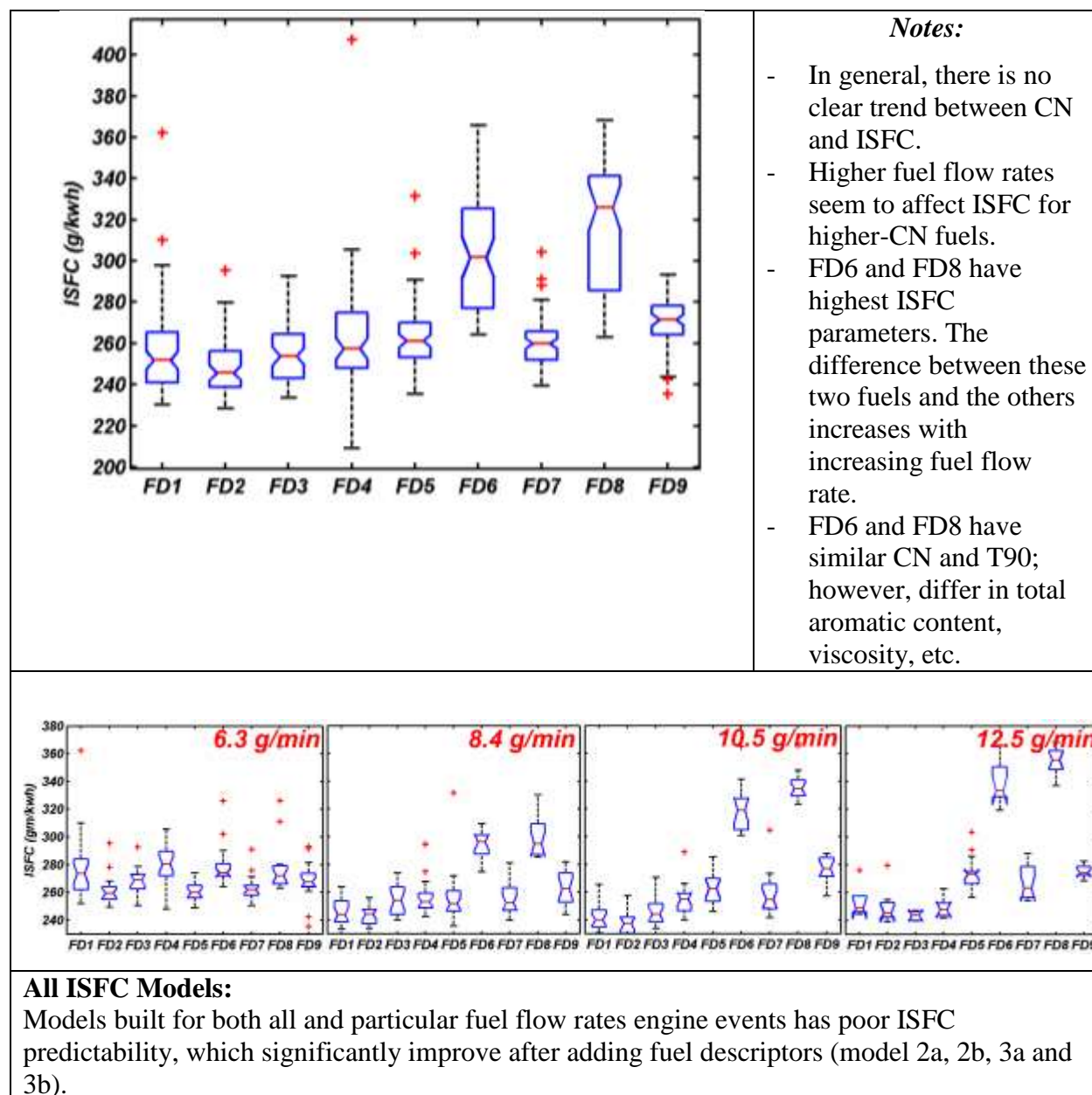


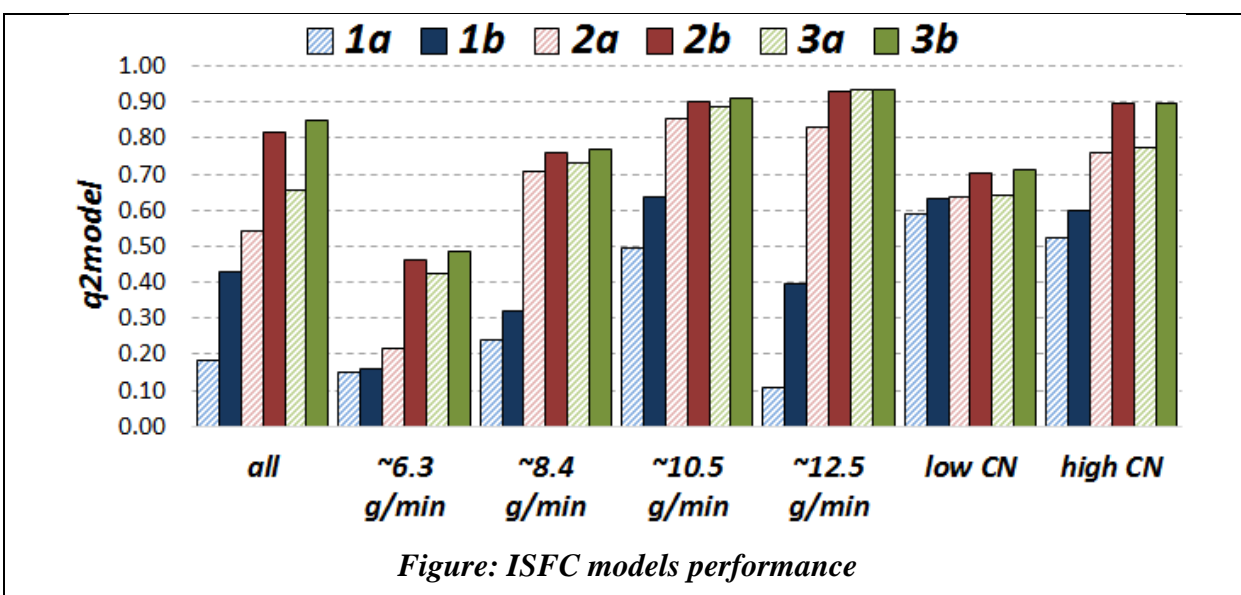
Figure: NOx Model 1a performance

7.1.7 SOOT EMISSIONS

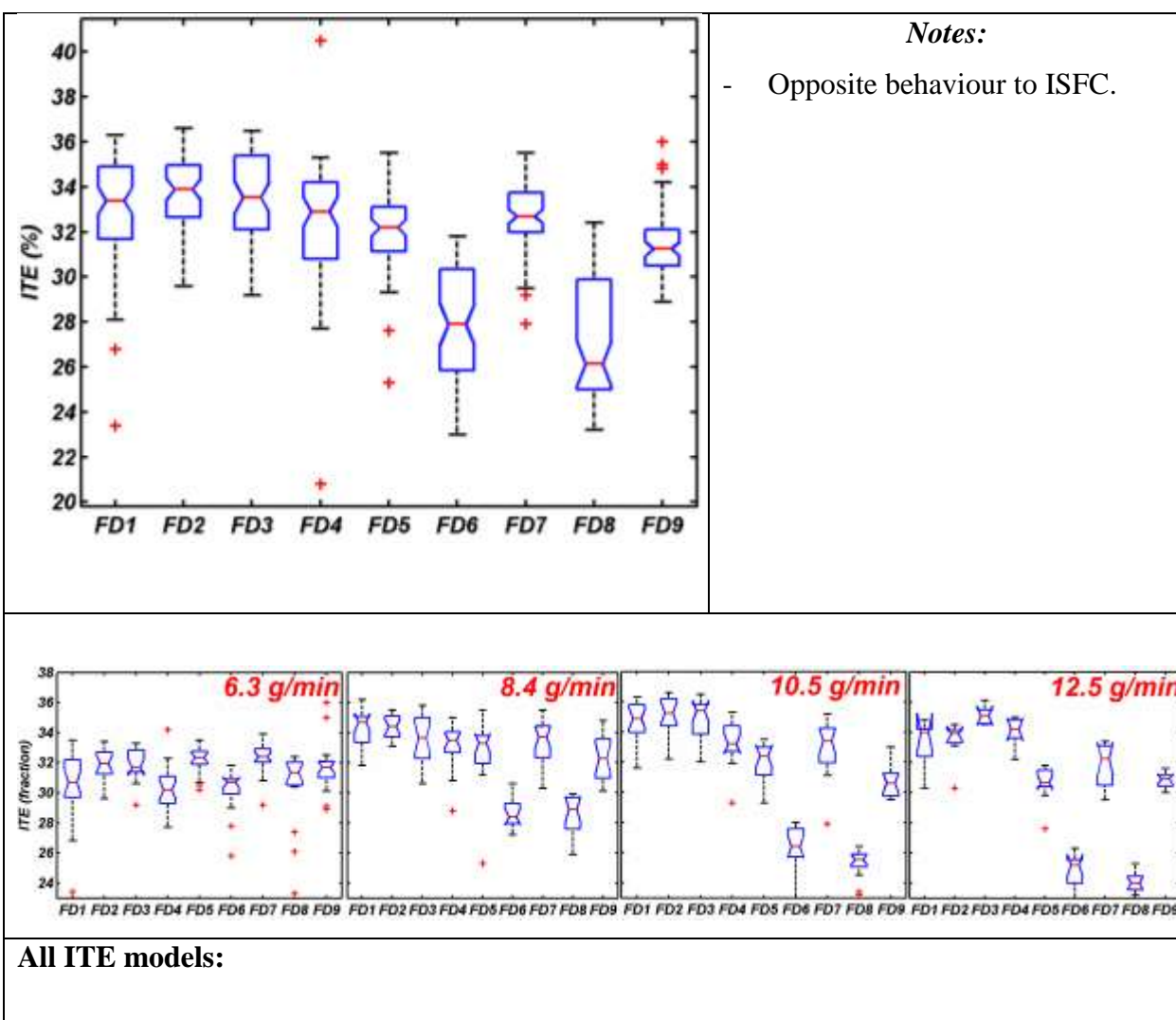
Neither PM nor soot emission were measured in the HCCI study performed by ORNL.

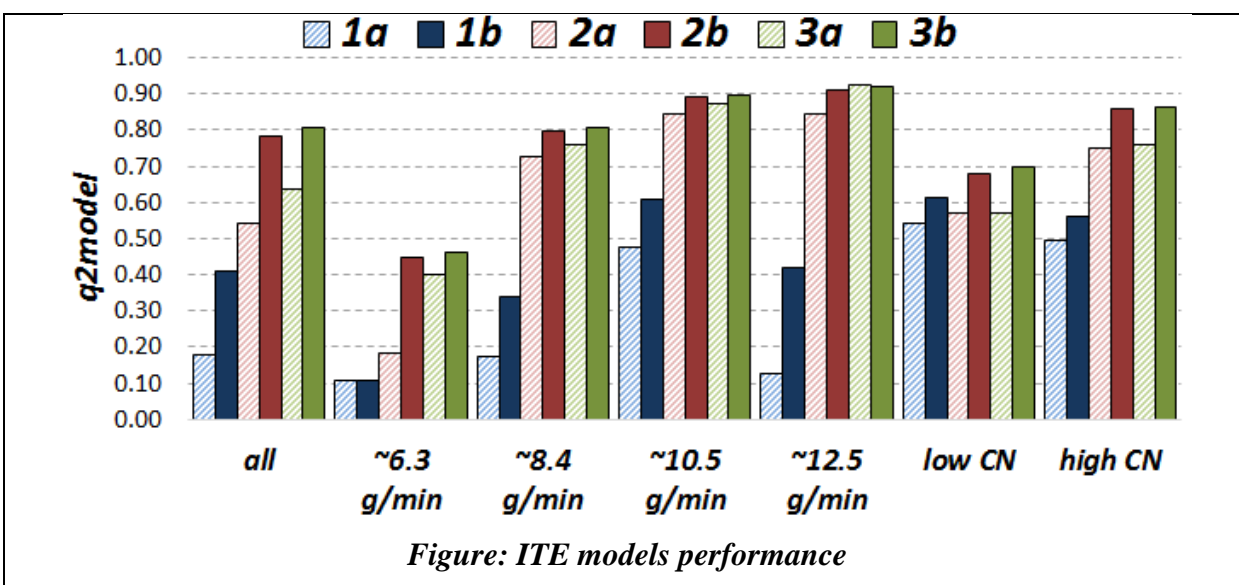
7.1.8 INDICATED SPECIFIC FUEL CONSUMPTION (ISFC)





7.1.9 THERMAL EFFICIENCY





7.1.10 SUMMARY OF PLS MODEL 2A RESULTS

A summary of the PLS models containing the 3 fuel design + engine control parameters (model 2a) for this engine is tabulated below. The filled circles (•) represent highly influential parameters and the open circles (◦) represent moderately influential parameters. Plus (+) symbols indicate parameters are directly correlated while negative (-) symbols indicate inverse correlations.

HCCI-SCE 0.5L Engine (ORNL):

Parameter	HC		CO		NOx		ISFC		ITE	
CN	•	(-)	•	(-)	•	(+)	•	(+)	•	(-)
T90	◦	(+)		(+)		(-)	•	(+)	•	(-)
Aromatics	◦	(-)		(-)		(-)		(+)		(-)
fuel rate		(+)		(+)	•	(+)		(+)		(-)
airflow	◦	(+)		(+)		(-)	•	(+)		(-)
lambda	•	(+)	•	(+)	•	(-)		(+)		(-)
intake T	•	(+)	•	(-)		(+)	•	(+)	•	(-)
intake P	•	(-)	◦	(-)	◦	(+)		(-)		(+)

• - highly influential parameter
(+) - directly correlated

◦ - moderately influential parameter
(-) - inversely correlated

7.2 LTC – MCE 1.9L (HECC-ORNL)

An experimental study was performed at Oak Ridge National Laboratory using a 1.9-L GM DI diesel engine in high-efficiency clean combustion (HECC) mode (5). The nine FACE diesel fuels were evaluated in order to understand fuel property effects on low-temperature combustion (LTC) in a light-duty diesel engine.

7.2.1 DATA DESCRIPTION

The scheme in Figure 27 presents the list of input and output control parameters from the LTC (HECC) experiment. For the nine fuels, a total of 111 experimental data points (aka. engine events) were taken.

No	Tag	In/Out parameter	unit	No	Tag	In/Out parameter	unit
v1		aicDynoSpeed	RPM	v22	O9	THC Emissions	g/kW-hr
v2	I1	aicDynoTorque	ft-lb	v23		NOx Emissions	g/kg-fuel
v3	I2	aic_EGR_Mass_GM	%	v24		CO Emissions	g/kg-fuel
v4	I3	aic_INT_O2_GM	%	v25		THC Emissions	g/kg-fuel
v5	I4	Main SOI	ATDC	v26	O10	Fuel	g/kW-hr
v6		Exhaust Rate	g/s	v27		Torque	N-m
v7	O1	Exhaust Rate	ACFM	v28	O11	BTE	%
v8	O2	Exhaust Rate	SCFM	v29		SOI_1	deg
v9		NOx Emissions	mg/s	v30	O12	CA05_1	deg
v10		CO Emissions	mg/s	v31	O13	CA10_1	deg
v11		THC Emissions	mg/s	v32	O14	CA50_1	deg
v12	O3	NMHC Emissions	mg/s	v33	O15	CA90_1 2	deg
v13	O4	CO2 Emissions	mg/s	v34	O16	IgnitionDelay_1	deg
v14	O5	O2 Emissions	mg/s	v35	O17	Combustion Noise	dB
v15		Fuel	g/s	v36	O18	EXH_PRT1_T	C
v16		Raw Smoke	FSN	v37	O19	TUR_IN_T	C
v17		Raw Smoke	mg/m3	v38		CA05-90	deg
v18	O6	PM (from Smoke Meter)	g/kW-hr	v39	O20	CA05-90_2	deg
v19		PM (from Smoke Meter)	mg/s	v40	O21	CA10-90	deg
v20	O7	NOx Emissions	g/kW-hr	v41	O22	CA50-90	deg
v21	O8	CO Emissions	g/kW-hr				

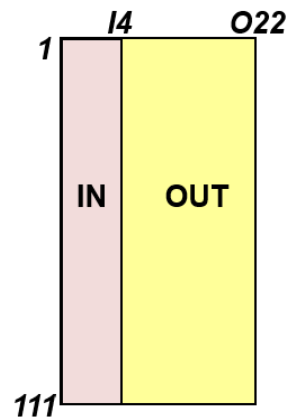


Figure 27 – Scheme presenting list of control input (I*) and output (O*) engine parameters for LTC- MCE 1.9L (as provided by ORNL researchers and placed in FACE diesel engine database)

7.2.2 EXPLORATORY ANALYSIS

In the first step of the analysis of the engine data, correlations of input engine control and engine output parameters were performed and presented in scatter plot form. The results are shown in Figure 28 and Figure 29. In the case of a strong one-to-one relationship, a plot of the two variables should yield a more-or-less straight line.

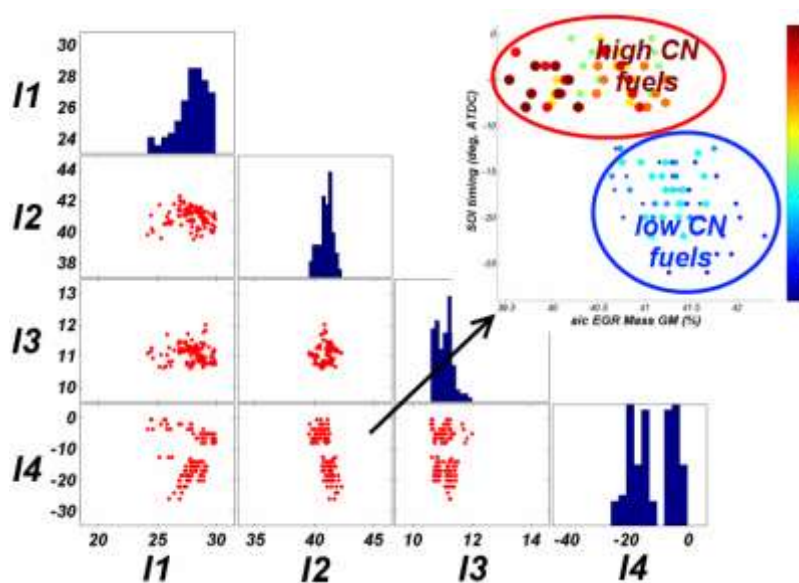


Figure 28 – Scatter representation of cross-correlation between control input engine parameters, where meanings of the labels (I*) are given in Figure 27. Histograms for each variable are located on the diagonal. The inset in the top-right corner magnifies the existing relations between *aic EGR Mass GM* parameter and start of injection (SOI) timing.

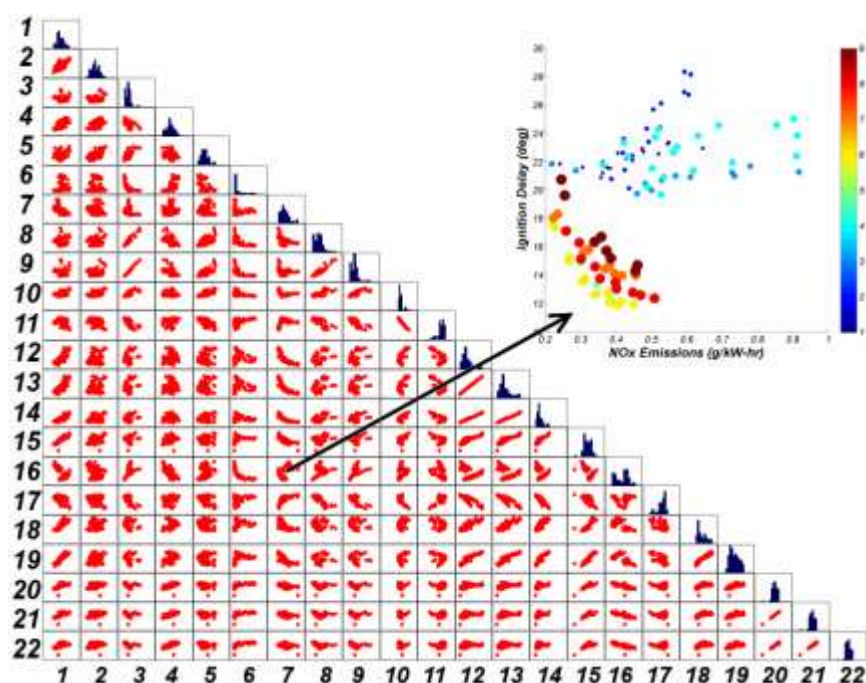


Figure 29 – Scatter representation of cross-correlation existing among output engine parameters obtained for HECC engine, where numbers indicate consecutive output parameters (O* - see Figure 27). The inset located in the top-right corner magnifies the existing relations between ignition delay and NOx emissions.

PCA was performed both on the input and output data provided to CanmetENERGY. The scores plot for the input data (Figure 30) shows clear separation of engine control parameters into two groups: low and high CN. This indicates that the engine control parameters were strongly influenced by fuel chemistry, in this case CN.

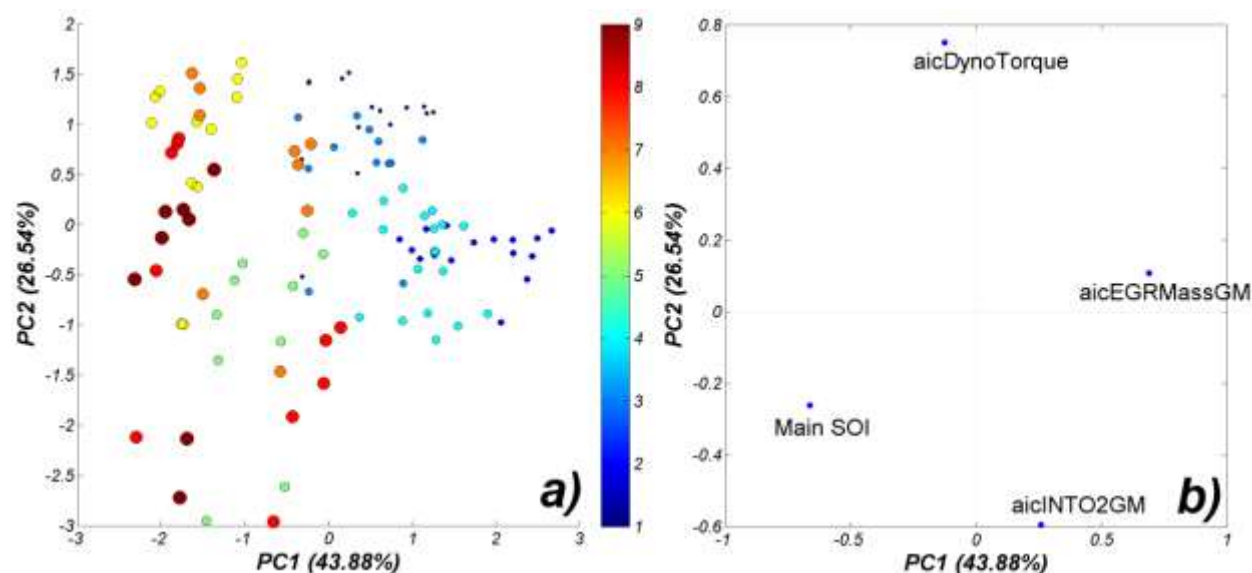


Figure 30 – PCA scores plots (a) and loadings plot (b) (PC1-PC2) obtained after running PCA for HECC ORNL engine input (control) parameters. On the scores plot, bubble colour and size are associated with respective fuels (i.e., smallest dark blue dots – FD1, largest brown bubbles - FD9).

Figure 31 presents results of PCA analyses calculated on all 22 engine output parameters obtained for all 111 engine event scenarios. Based on the observed results we conclude:

- In scores plots (Figure 31a) there is a noticeable separation of events derived from high- and low-CN fuels (pair 1 and 4 – low CN, pair 2 and 3 – high CN).
- Groups 1 and 2 can be treated as outliers in PCA output space because of high numerical output value of specific parameters (see next bullet(s)).
- The associated loadings plot describes the clustering of points in the group 1. Engine events located in group 1 produce high non-methane hydrocarbons, O₂, CO, and THC emissions. (see variables located in the corresponding to the group 1 position in loadings plot)

- Engine events found in group 2 are characterized by high exhaust rates and retarded combustion phasing characteristics.
- Events set in group 3 (high-CN fuels) produce high particulate matter and CO₂ emissions.
- Group 4 consists of low-CN fuel characterized by higher ignition delay and higher NO_x emissions.
- Both groups 3 and 4 consist of engine events with high brake thermal efficiency.
- We can observe some groups of variables in the loadings plot, which may indicate correlations among them. For example: NMHC/THC/CO/O₂, combustion phasing CA05/CA10/CA50, combustion noise and BTE.

A few examples of the aforementioned relations between variables (in the loadings plot) and engine events are shown in Figure 32.

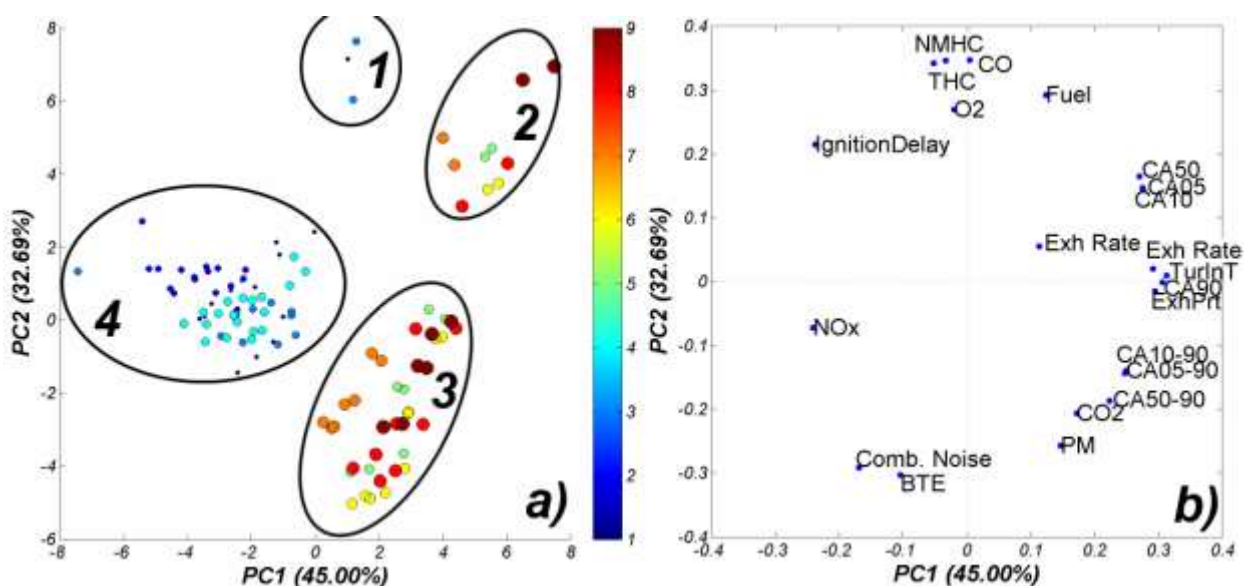


Figure 31 – PCA score (a) and loading (b) plots (PC1-PC2) obtained after running PCA for ORNL HCCI engine output parameters for all events. On the scores plot, bubble colour and size are associated with respective fuels (i.e., smallest dark blue dots – FD1, largest brown bubbles – FD9).

In the next step, the ‘outliers’ (events found in groups 1 and 2, Figure 3a) were removed and two PCA models were calculated: one for events from low-CN fuels (group 4) and the second for events from high-CN fuels (group 3). PCA results from this exercise are presented in Figure 33b and Figure 33c (only scores plots are shown here). On the scores plot (Figure 33b) we can observe that fuel FD2 is located in the lower part and FD3/FD4 in the upper part of the plot. FD1 is dispersed in FD2. FD2 events are characterized by high NMHC/THC/ignition delay and FD3 events have high CO₂ emissions (this conclusion was taken from the loadings plot, not shown here).

In Figure 33c we may also observe clear separation of events (distributed along PC2 axis) on three groups FD8/FD9, FD7, FD6/FD5. The observed separation of events in PCA space calculated from output engine parameters may be affected by physical and chemical properties of the fuels. We call the PC2 axis (Figure 33b and Figure 33c) an aromatic axis, as it separates events based on the aromatics content of the fuels explored in this part of the study.

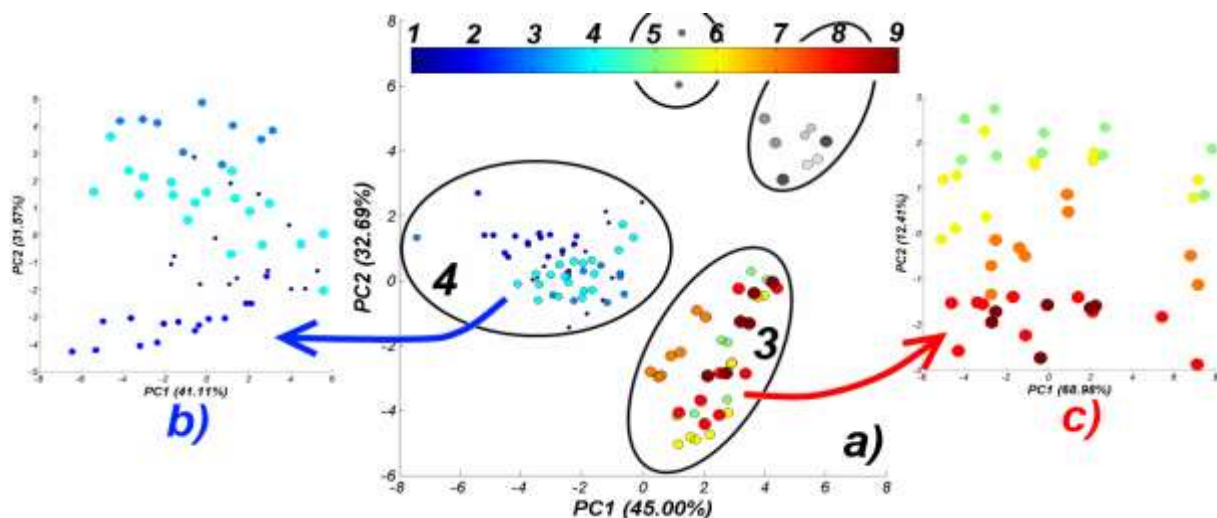


Figure 33 – After removing events located in groups 1 and 2 (two groups on the top – behind the colour bar) and calculating two PCA models for both groups 4 and 3, the resulting scores plots show a characteristic distributions of events. Points on the plot are coloured by fuel type (see respective colour bar).

7.2.3 MODELING APPROACH

Simple one-to-one correlations between engine control parameters and selected engine output parameters are presented in Figure 34.

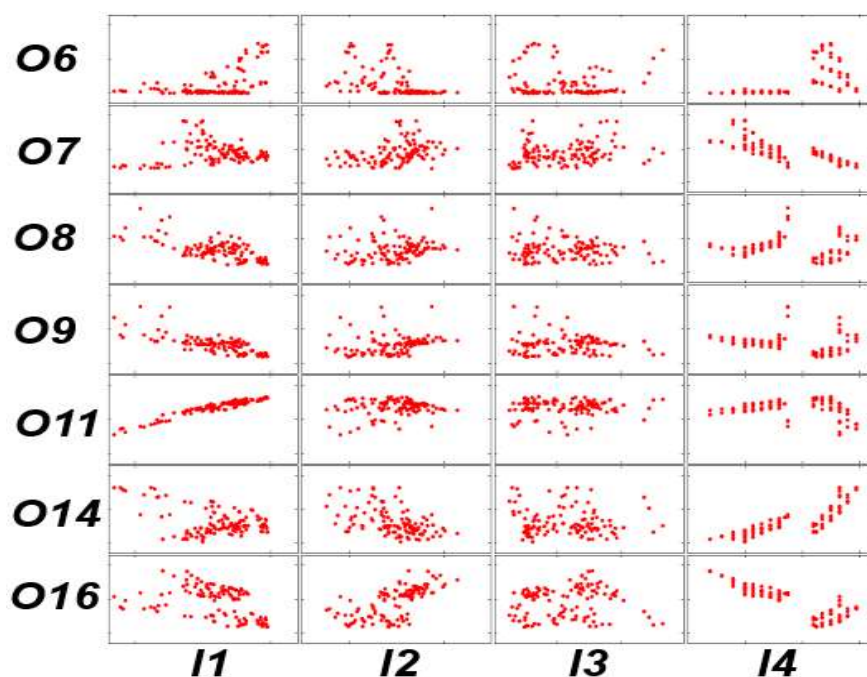


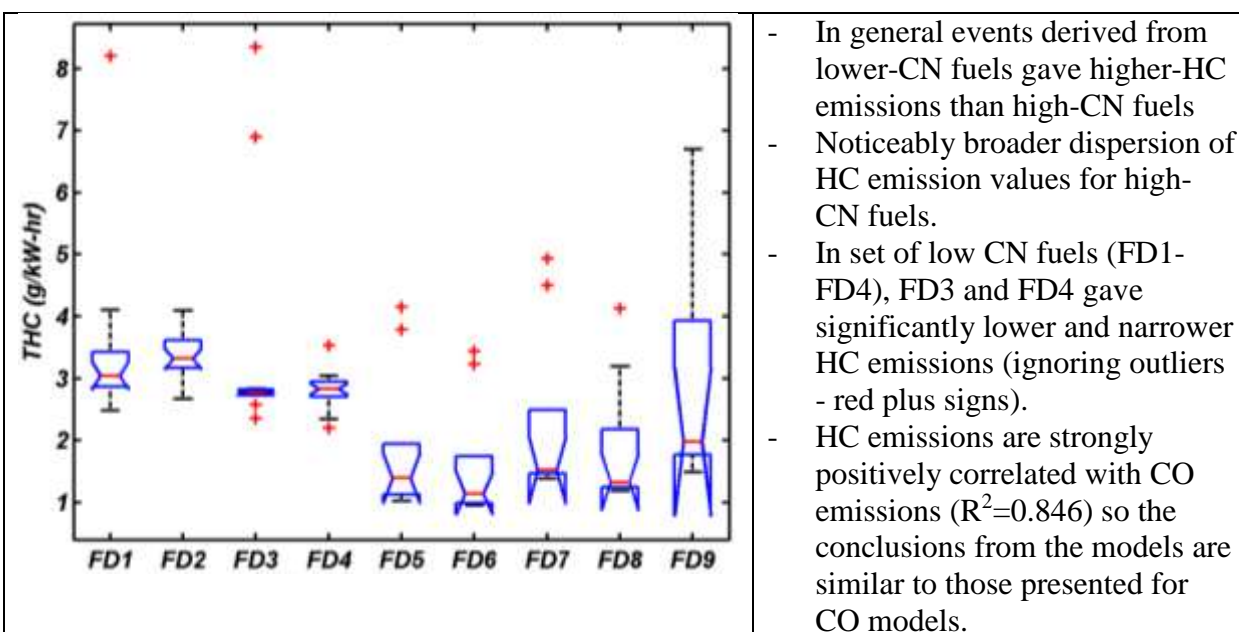
Figure 34 – Scatter representation of cross-correlations between output engine parameters (O^*) and engine control variables (I^*) obtained for HECC engine (for meaning of O^* and I^* see Figure 27). I_4 is a SOI timing control variable and plots similar to these presented may be found in the original paper (5). Note the strong correlation between O_{11} (BTE) vs I_1 (aicDynoTorque).

In this section the results of modeling of emission and efficiency parameters are presented. The engine was modeled using strategies described in detail in section 6:

- (strategy: 1a): aicDynoTorque (I_1), aic_EGR_Mass_GM (I_2), aic_INT_O2_GM (I_3), Main SOI (I_4),
- (strategy: 1b): strategy 1a + cross terms for all variables,
- (strategy: 2a): strategy 1a + FACE fuel CN/T90/aromatic content,
- (strategy: 2b): strategy 2a + cross terms for all variables,
- (strategy: 3a): strategy 1a + 40 FACE fuel descriptors (see section 6),
- (strategy: 3b): strategy 3a + cross terms for all variables.

PLS models were built for all events, and events performed only on high CN fuels and only on low CN fuels. For each modeled property, 18 PLS models were calculated and a summary of the results is presented in the following sections (7.2.4 – 7.2.8).

7.2.4 HYDROCARBON EMISSIONS



All HC models:

Similar to CO and NO_x emissions, HC emissions are relatively easy to model based on both control parameters and additional chemical and physical properties of FACE diesel fuels. The best models are achieved after including physical-chemical fuel descriptors (models 2a and 3a). For models calculated using engine events with high-CN fuels, the models seem to be outstanding after including fuel descriptors. All of the models were tested for statistical significance of their estimated predicting power (using Y-scrambling test as well as Monte Carlo random data model/test splitting tests).

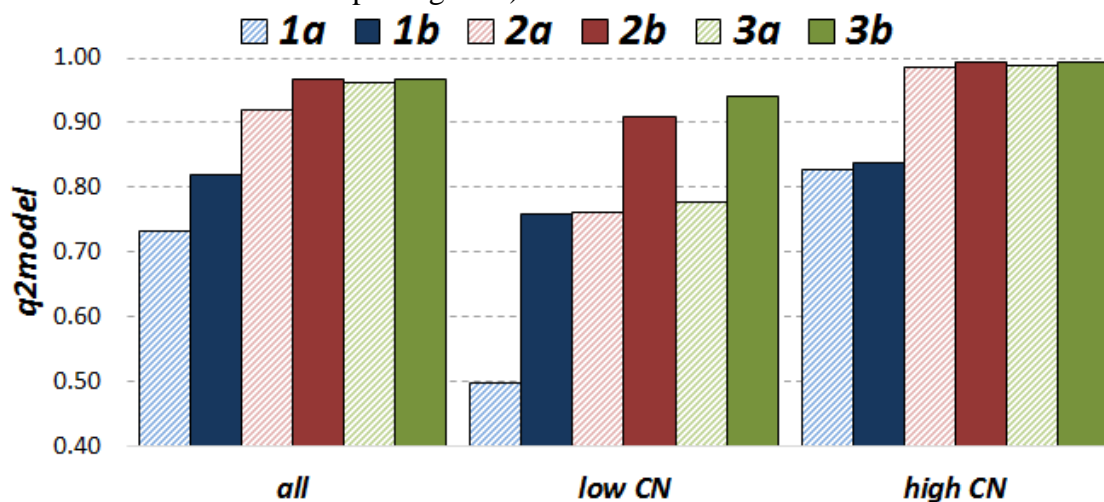


Figure: HC model performance

HC model 1a-all:

Similar to other emission profile models, the HC emissions for low-CN fuels are difficult to model using only control on engine parameters. Both aicDynoTorque and SOI engine control parameters are important and for both parameters, the sign of the regression coefficient is negative, which means that in order to obtain low HC emissions, aicDynoTorque and SOI timing values should be kept as low as possible – regardless of the fuel used.

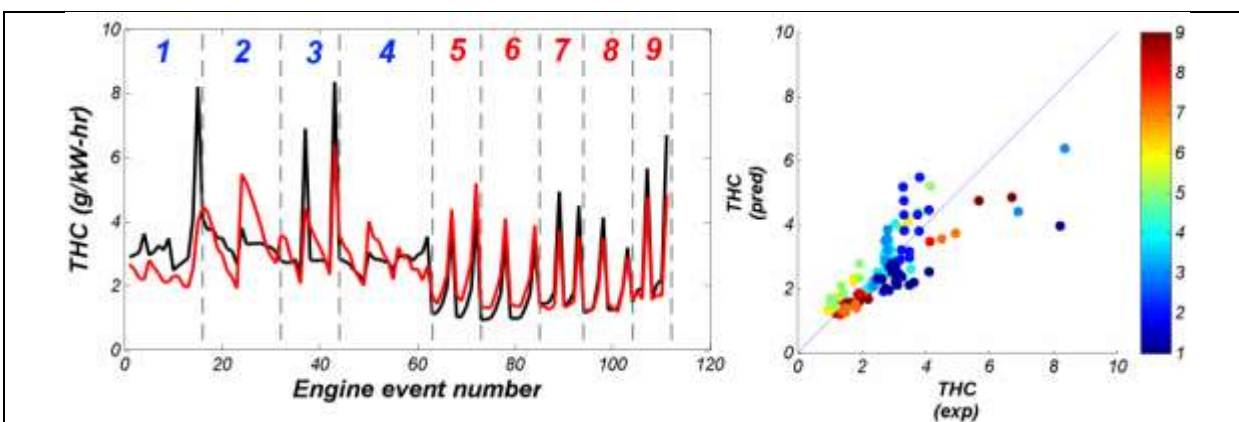


Figure: HC model 1a performance

HC model 2a-all:

Including three fuel descriptors (CN/T90/Aro) as independent variables significantly improved the modeling capability of the HC emissions PLS model. Improvement is clearly noticed for engine events run with low-CN fuels (compared with the previous model 1a).

In contrast to the previous model (model 1a-all), after inclusion of fuel descriptors, the sign of the regression coefficient for SOI is inverted, and here has positive sign. Among fuel descriptors, the most important is CN which seems to have an even larger impact (and negative PLS regression coefficient) than SOI in this model.

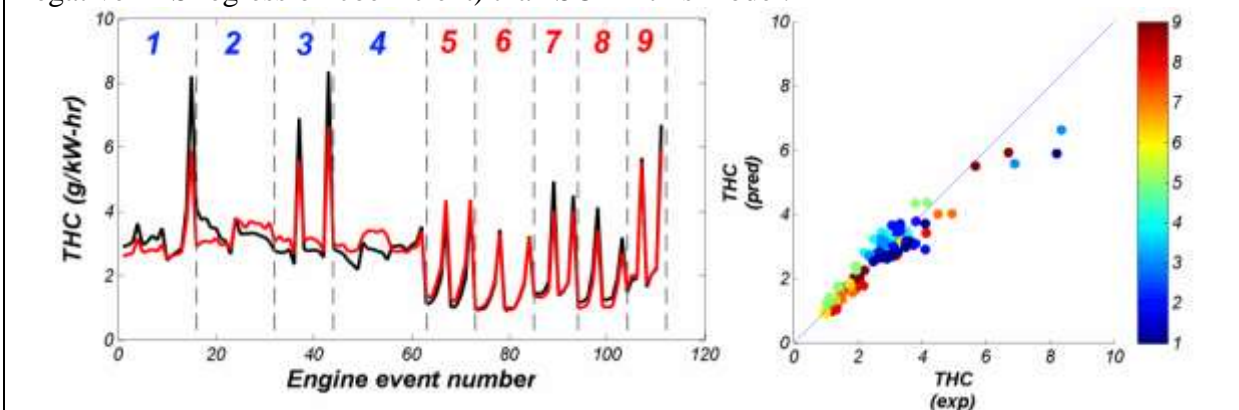


Figure: HC model 2a performance

HC model 3a - all:

Very similar (in performance q^2 model, RMS etc.) to model 2a-all. The effects of additional fuel descriptors does not seem to be so important. Fuel descriptors selected as important are **CN, T20-T50, CT2, CT8, alkylbenzenes, monoaromatic, and olefin content**. CN, distillation characteristics and CT2 are characterized by negative values of regression coefficients. The other four are positive.

HC models for low-CN fuels:

HC emissions for engine events run with low-CN fuels are more-or-less constant (not counting HC emission reading spikes at the end of the SOI timing test). This may create difficulties in modeling the HC parameter.

The HC model 1a is the poorest in the set of models built for low-CN fuels. Inclusion of cross terms, fuel descriptors, or both seems to improve HC models. However, removal of three 'spiky' readings leads to very weak PLS models (not shown here).

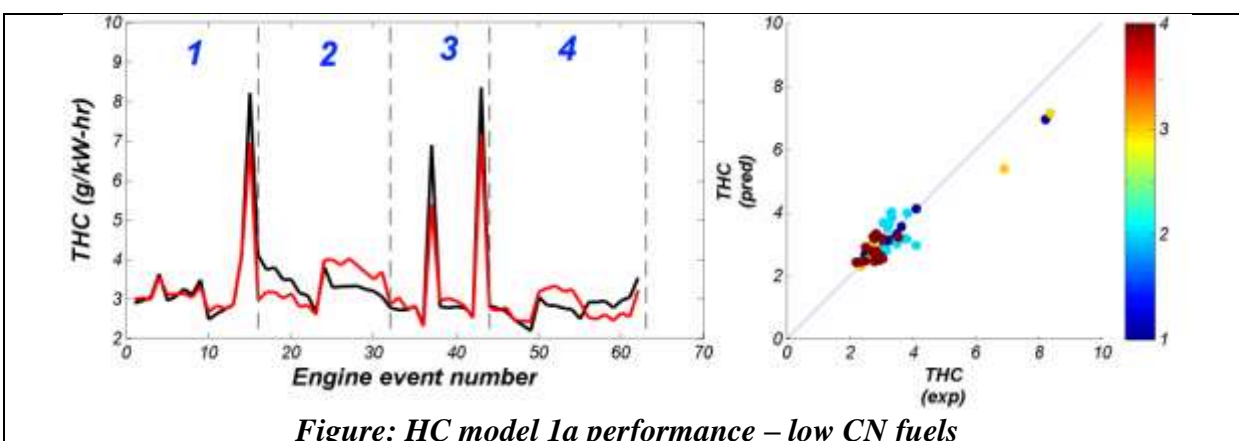


Figure: HC model 1a performance – low CN fuels

HC models for high-CN fuel:

HC models generated for high-CN fuels are characterized by high performance characteristics. Models 2a and 3a (without cross terms) provide the same level of performance as models including cross terms of all variables (models 2b and 3b). The PLS model 2a shows two important engine parameters and two important fuel parameters affecting the HC model performance: aicDynoTorque (-), SOI (+), CN (-), and aromatics (+). The signs after variable names indicate the sign of their model regression coefficient. Based on this we can conclude that, for high-CN fuels, high CN reduces and high aromatic content increases HC emissions.

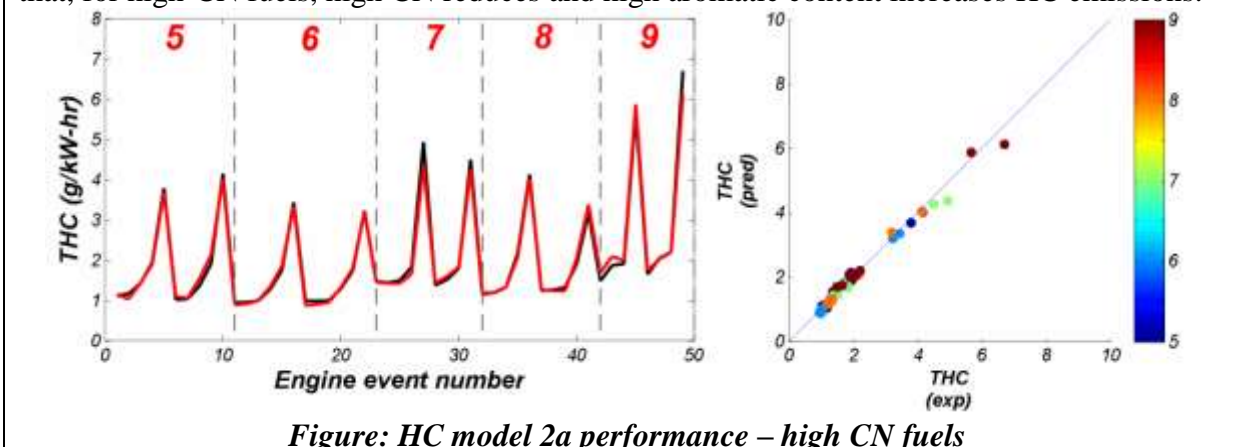
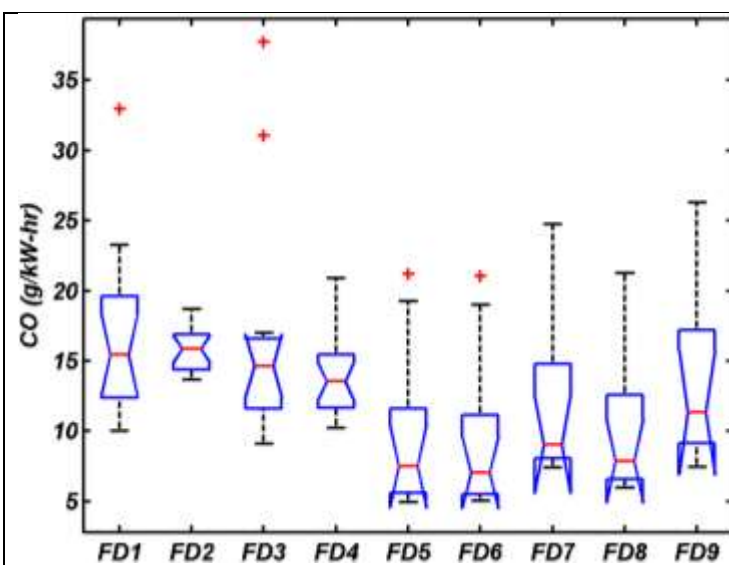


Figure: HC model 2a performance – high CN fuels

7.2.5 CARBON MONOXIDE EMISSIONS



- In general, low-CN fuels have slightly higher CO emissions.
- Two medium-CN fuels (FD7 and FD9) have higher CO than high-CN FACE fuels.
- High variance for high-CN fuels.

All CO models:

CO emissions are relatively easy to model based on both control parameters and additional chemical and physical properties of FACE diesel fuels. However, the best models may be achieved after including physical-chemical fuel descriptors (models 2 and 3). For models calculated on engine events performed only with low-CN fuels, the improvement is significant after including cross terms for all the variables (models 2b, 3b). Interestingly, model 3a based on engine control parameters and 40 fuel descriptors does not seem to be any better than models having only three fuel descriptors (model 2a).

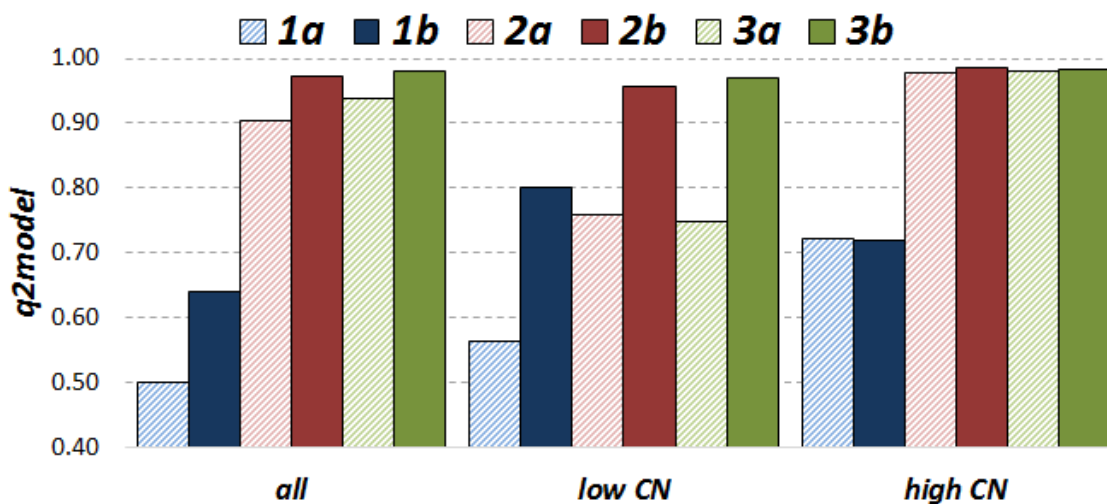


Figure: CO model performance

CO model 1a-all:

The PLS model based on engine control parameters does not work for low-CN fuels (does not predict the CO emissions profile). The behaviour of high-CN fuels is impeccably profiled. The control parameter *aicDynoTorque* (I1) is the most important variable, and its negative PLS regression coefficient value means that events having higher values of I1 will lower CO

emissions regardless of the fuel. Including cross terms for all control variables improves the model's performance.

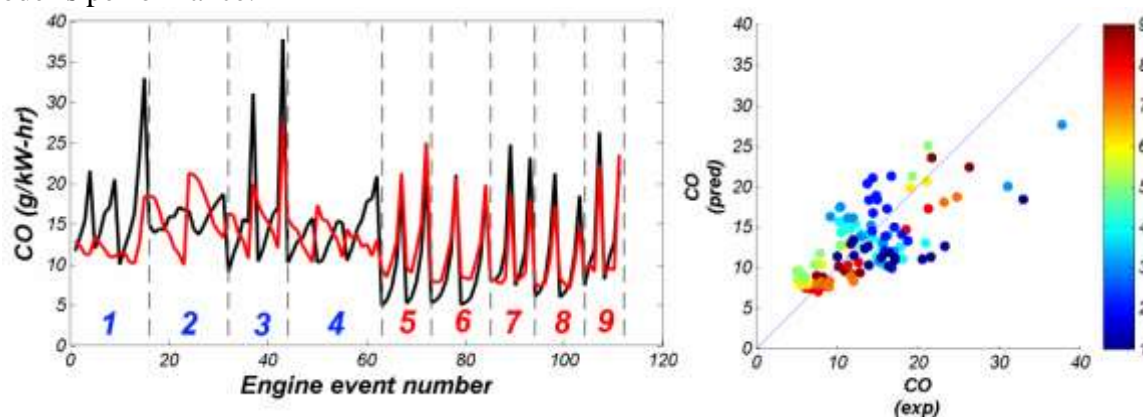


Figure: CO model 1a-all performance

CO model 2a-all:

Including three fuel descriptors (CN/T90/Aro) as independent variables in the PLS model significantly improves modeling of CO emissions, where CN is especially important. Improvement is clearly evident for engine events run with low-CN fuels (compared to the previous model 1a).

In contrast to the previous model (model 1a-all), two engine control parameters are the most important in the PLS equation, namely *aicDynoTorque* followed by *SOI timing*. Among fuel descriptors, the most important is *CN* which seems to have an even larger impact (and negative PLS regression coefficient) than *SOI* in this model, but not as large as *aicDynoTorque*.

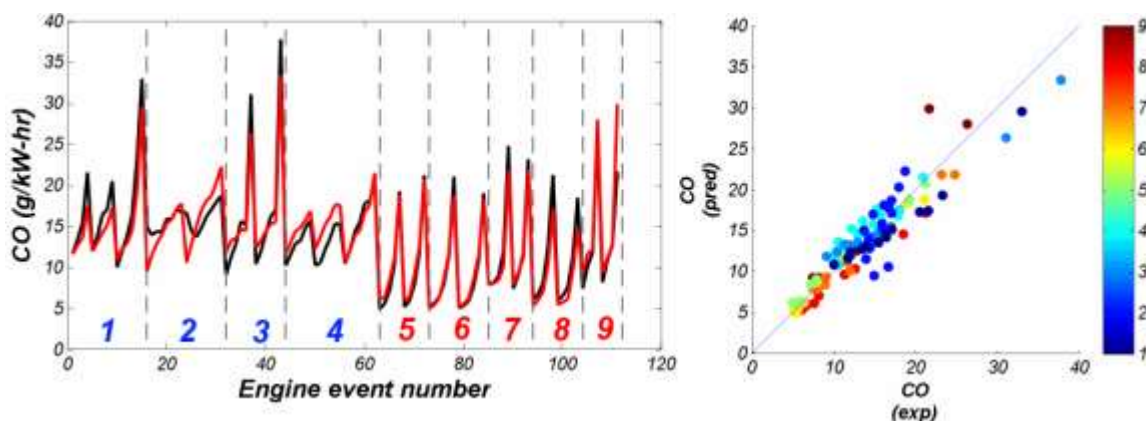


Figure: CO model 2a performance

CO model 3a - all:

Very similar in performance (*q2model*, *RMS* etc.) to model 2a-all. The effects of additional fuel descriptors do not seem to be as important as *CN*. However, fuel descriptors selected as important are *CN*, *CT2*, *CT8*, *olefins*, *alkylbenzenes content*.

CO models for low-CN fuel:

The CO model 1a is better than the model calculated for all events. Additional improvement is achieved after adding fuel descriptors (models 2a, 3a). In the case of PLS model 2a (see following figure) we can observe that the CO profiles for events derived from fuel FD2 are not well reproduced (on parity plot CO (pred) vs CO (exp) we see the non-linear positioning of FD2 points). This situation changes after introducing cross-terms for all variables (models 1b, 2b, and 3b) and significantly better models are thus obtained (not shown here).

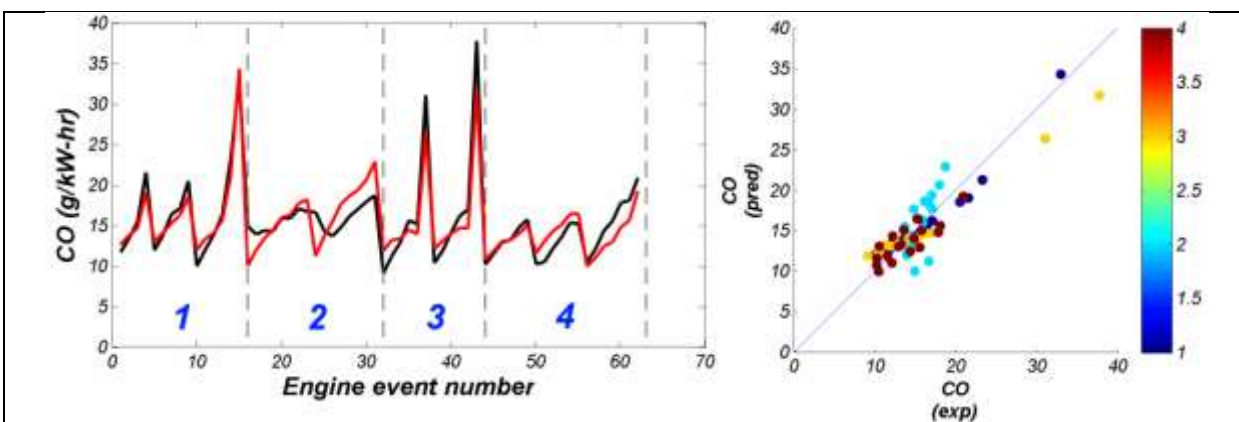


Figure: CO model 1a performance – low CN fuels

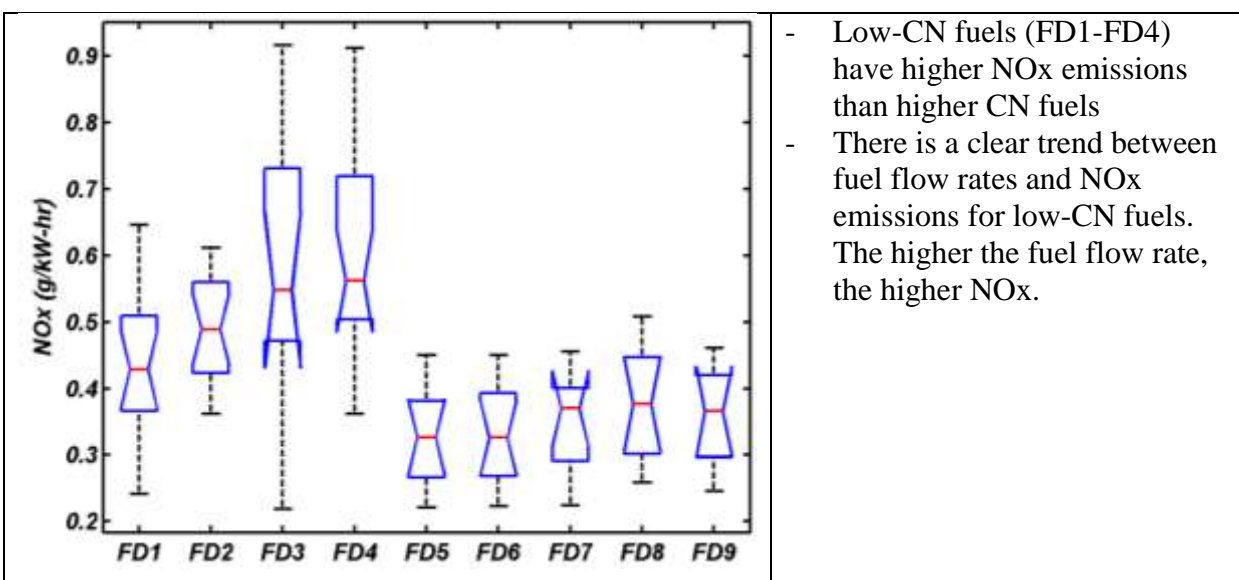
Adding fuel descriptors (*CN/T90/Aro*) is important in modeling CO emissions for low-CN fuels. However, the most important of the three physical-chemical variables is **aromatic** content (negative PLS regression coefficient in model 2a). **Low-CN** fuels with **higher** contents of aromatic compounds tend to reduce CO emissions (when engine control parameters are constant).

From the pool of 40 fuel descriptors included in models 3a and 3b, **cloud point** and **n-paraffin content** are important, among others.

CO models for high-CN fuel:

The CO models generated for high-CN fuels are characterized by high performance characteristics. Models 2a and 3a (without cross terms) provide the same level of performance as models including cross terms of all variables (models 2b and 3b). The models (like all others in this study) were checked by Y-scrambling and external validation to confirm model significance. The PLS model 2a shows two important engine parameters and two important fuel parameters affecting the CO model performance: *aicDynoTorque* (-), *SOI* (+), *CN* (-), and *Aro* (+). The signs after variable names indicate the sign of their model regression coefficient. Based on this, we can conclude that, for high-CN fuels, a decrease in CN or an increase in aromatic content increases CO emissions.

7.2.6 NITROGEN OXIDES EMISSIONS



All NOx models:

NOx emissions are relatively easy to model based on both control parameters and additional chemical and physical properties of FACE diesel fuels.

PLS models based only on control parameters are weaker than respective models with physical and chemical information added. Strong influence of cross-term parameters on model internal predictability expressed by the *q2model* parameter is observed.

PLS models built for high-CN fuels are stronger than respective models for low-CN fuels. Still, in these types of models, NOx emissions for high-CN fuels strongly depends on the engine control parameters, which is not the case for low-CN fuels.

In general, including all 40 fuel descriptors does not improve any of these models significantly.

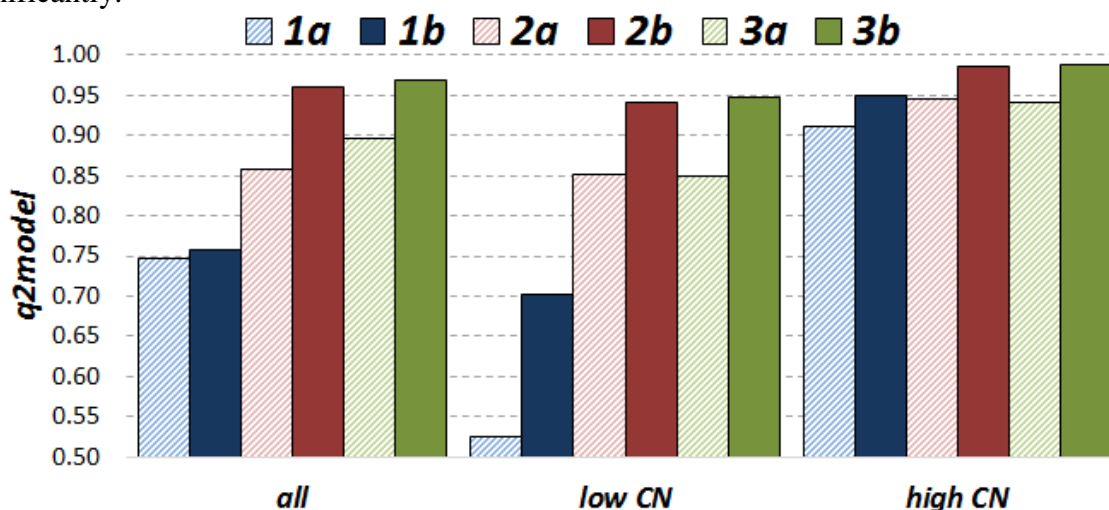


Figure: NOx models performance

NOx model 1a-all:

SOI timing is the most important engine control parameter. SOI timing alone reproduces the shape of the NOx profile. It can be observed that NOx emissions are much better reproduced for events derived from high-CN fuels (FD5-FD9) than low CN fuels. Without any additional

information about fuels - NO_x emissions are underpredicted for FD3 and FD4, and overpredicted for FD2.

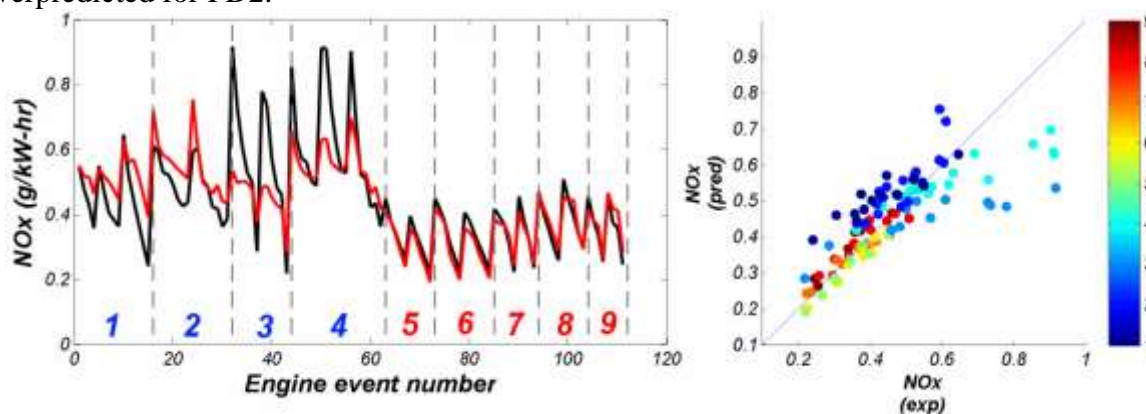


Figure: NO_x model 1a performance

NO_x Model 2a-all:

Including information about the fuels (CN/T90/Aromatic content) improves the modeling capability of the PLS models, of which CN is the most important added parameter, followed by aromatic content. Both terms have positive regression coefficients in PLS equation - positive influence on log(NO_x) emission. This means that high CN fuels will have higher log(NO_x) emissions parameter. After including cross-term interactions of the dependent variables the PLS model tends to improve.

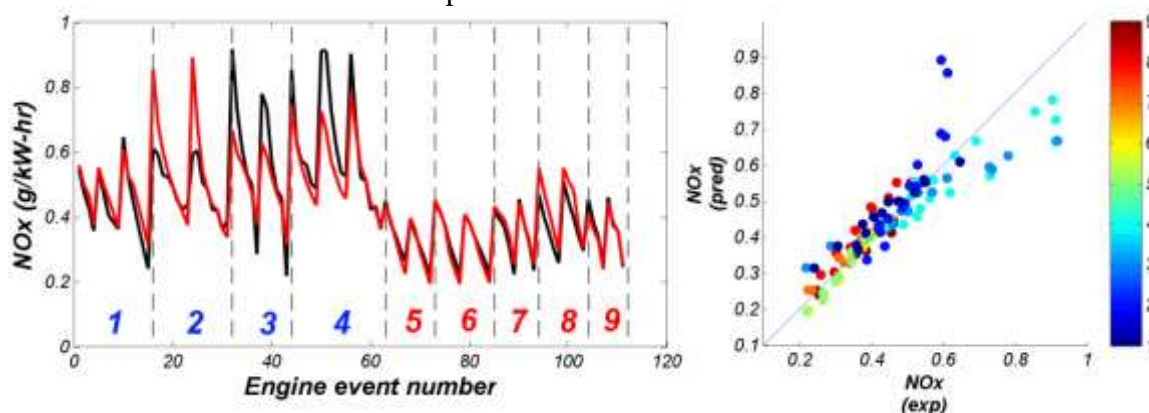
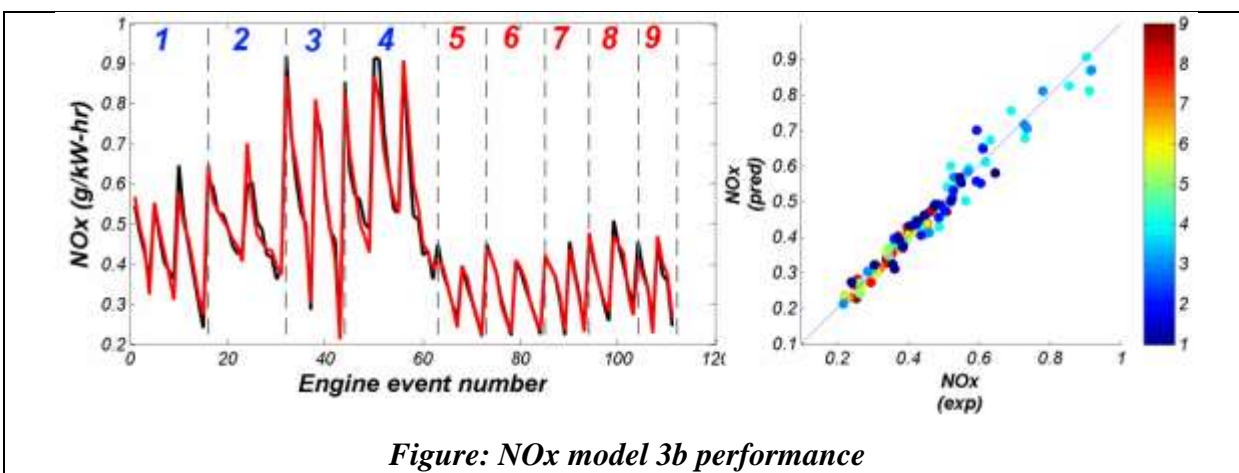


Figure: NO_x model 2a performance

NO_x model 3b-all

This PLS model includes all 40 physical and chemical fuel descriptors and all interaction terms (in total 1034 variables). The PLS model is characterized by high q^2 model, low RMS parameter values. Complexity (i.e. the number of PLS components) of the model is high ($A_{opt}=20$). Based on VIP or SR we may reduce the number of parameters to ~150 and still have a good PLS NO_x model. Most of the important terms have engine control parameters though. Important phys/chem fuel properties: CN, cloud point, CT2, CT8, n-paraffins.



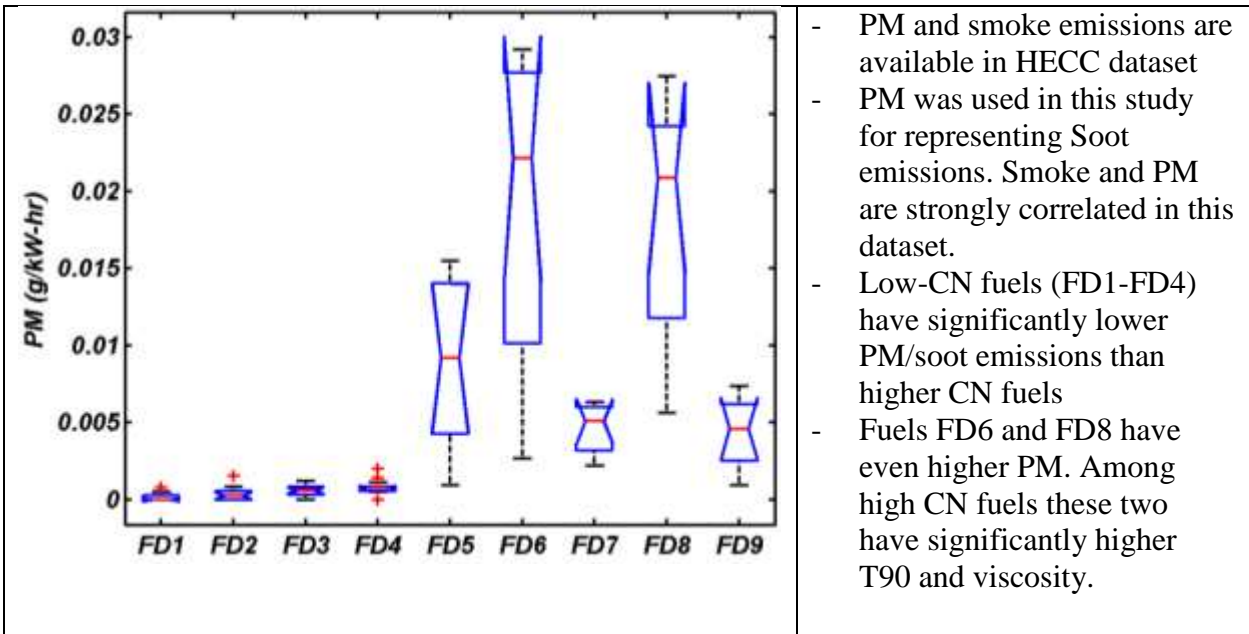
NOx models for low-CN fuels

Weak models. Including fuel descriptors significantly improves the PLS model performance for events using low-CN fuels. Interestingly, CN is not as important as **aromatic content**. Low-CN and low-aromatic fuels should have lower NOx emissions than high-CN and high-aromatic fuels. In fact, events for the low aromatics fuels (FD1 and FD2) have a lower NOx emissions than the high aromatics fuels (FD3 and FD4).

NOx models for high-CN fuels

Besides SOI, the other important term having high influence on modeling NOx is the *aicDynoTorque* control parameter, which has the opposite effect on NOx emissions to that of SOI timing.

7.2.7 SOOT EMISSIONS



All PM models:

A few engine events for low-CN fuels resulted in 'zero' values of PM. Hence we could not apply logarithmic transformation for PM data. Fourth-root transformation was applied here, instead. It is sufficiently close to the logarithm transformation but may be applied in cases where variables equal to zero.

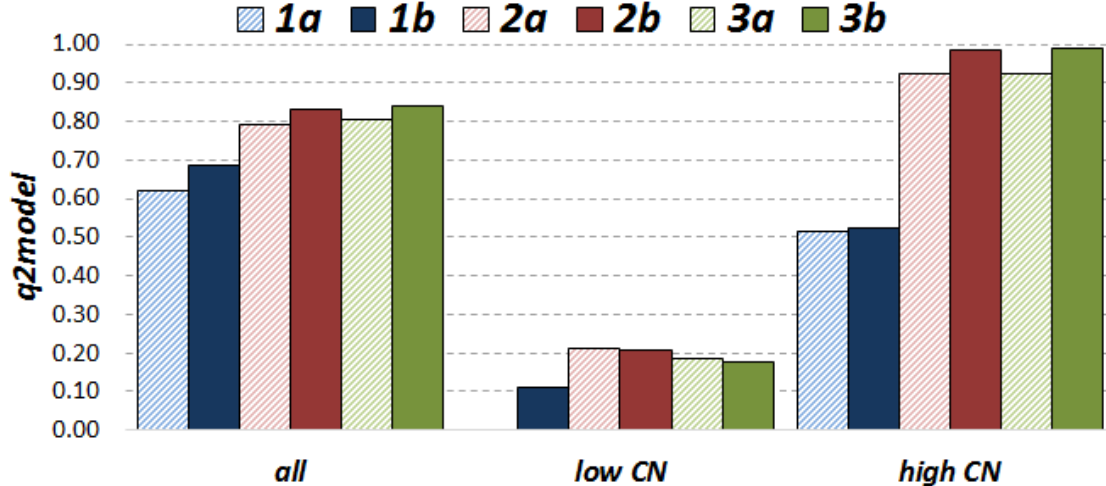


Figure: PM models performance

PM model 1a-all:

In general the model based on engine control parameters is weak. However, notice that the rough, 'jagsaw' trend for PM values for high-CN fuels is maintained (see the Figure below).

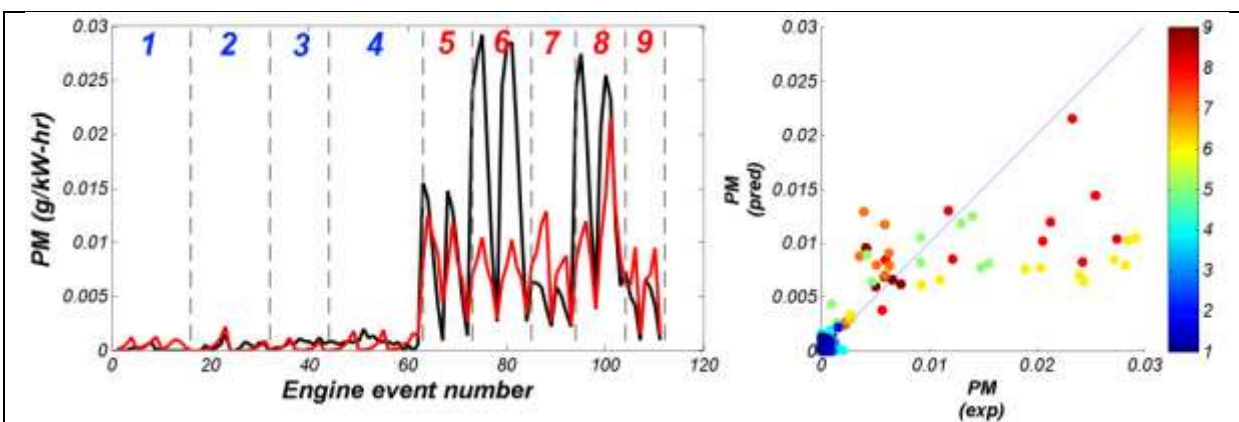


Figure: PM model 1a performance

PM model 2a-all:

Including fuel descriptors in the PM model improves the modeling capability. However, it is difficult to model engine events for higher PM values. Adding cross terms of all variables improve predictive capability of PM models.

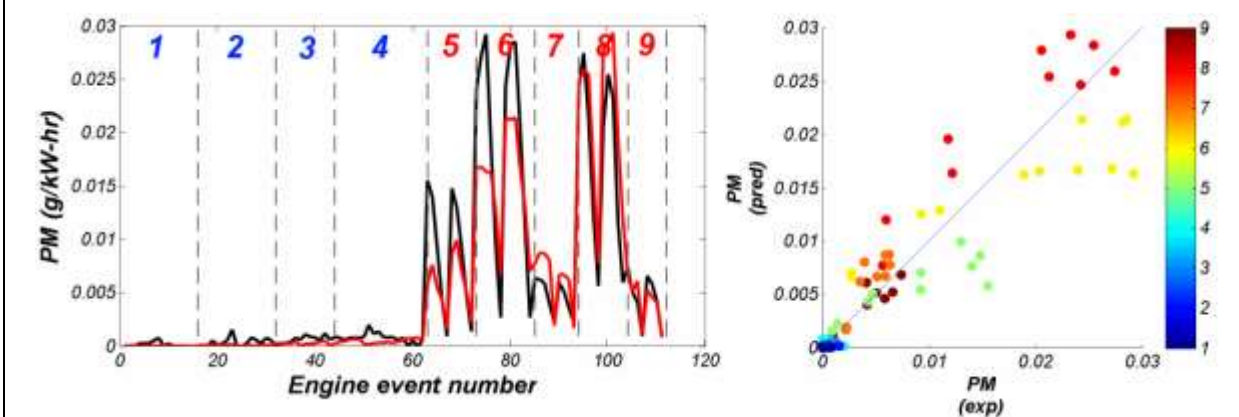


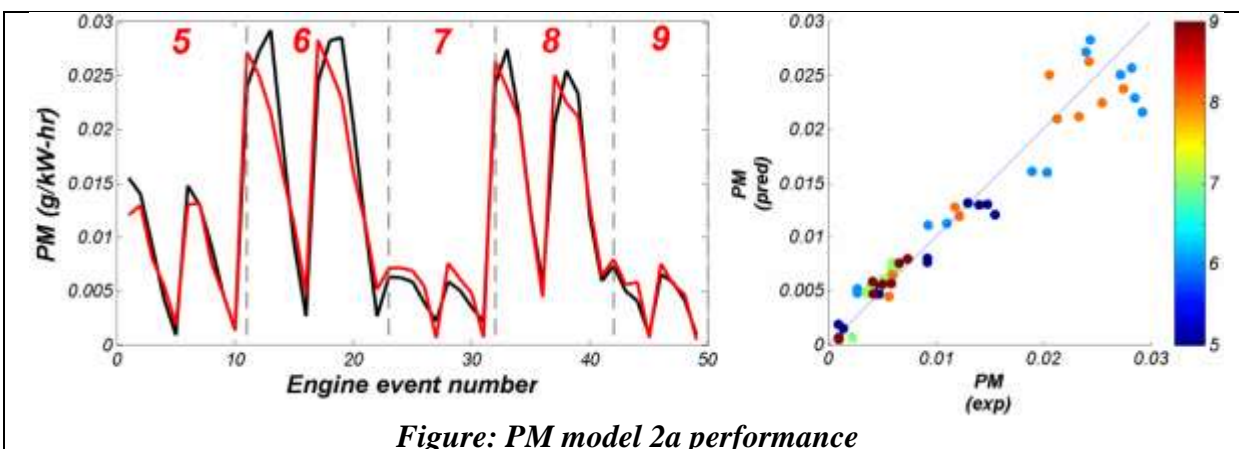
Figure: PM model 2a performance

PM models for low-CN fuels

Modeling PM for engine events derived from low-CN fuels is poor or even impossible due to the fact that PM content is extremely low and most likely on the 'edge' of smoke meter precision.

PM models for high-CN fuels

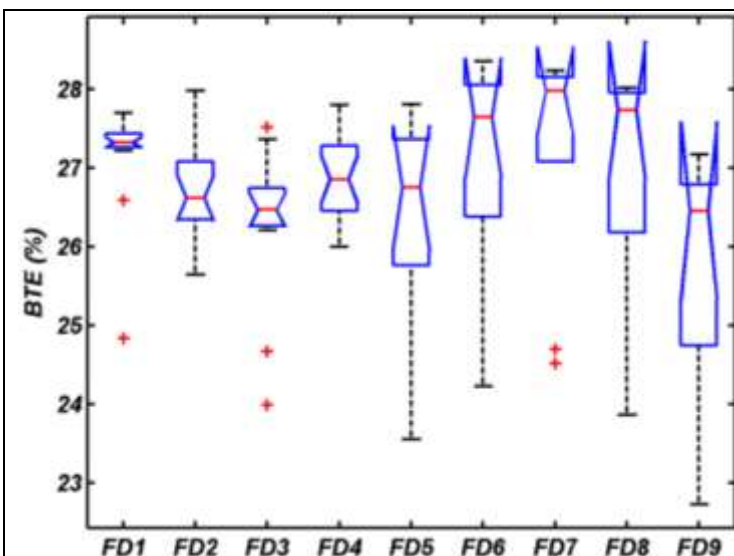
PM model 1a consisting of only engine control parameters is weak and is not practical for use in prediction. Including CN/T90/Aro descriptors significantly improved PM modeling capabilities (model 2a). The PLS model 2a shows two important engine and two important fuel parameters that affect PM model performance: *aicDynoTorque* (+), *SOI* (-), *CN*(+) and *T90* (+). The signs after variable names indicate the sign of their model regression coefficient. Based on that we can conclude that for high-CN fuels both high CN and high T90 will increase PM.



7.2.8 INDICATED SPECIFIC FUEL CONSUMPTION

ISFC was not reported in the HECC study performed by ORNL.

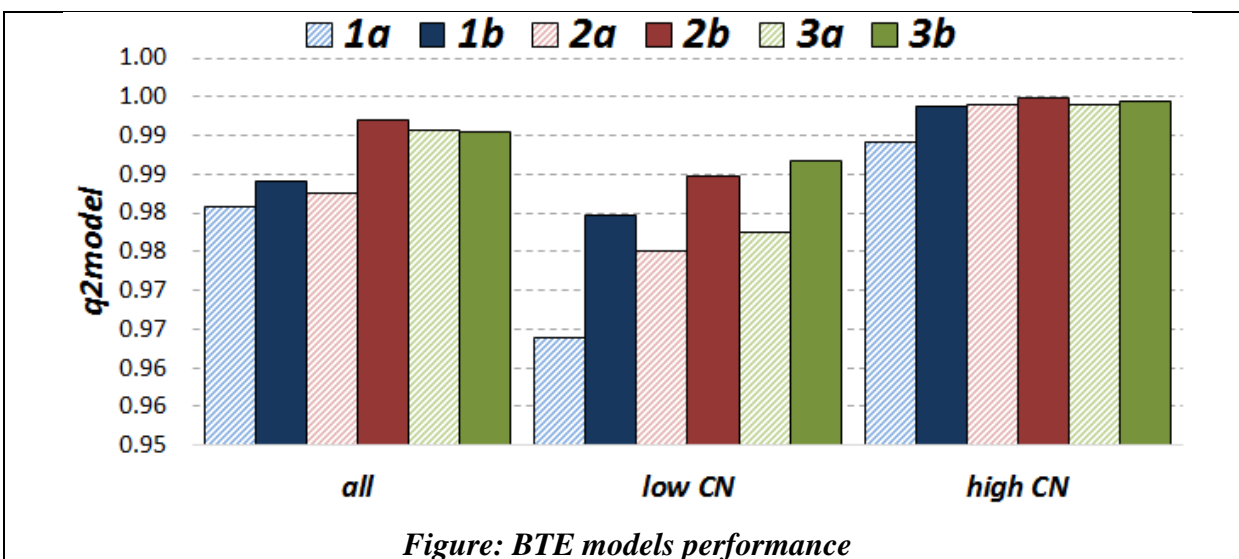
7.2.9 THERMAL EFFICIENCY



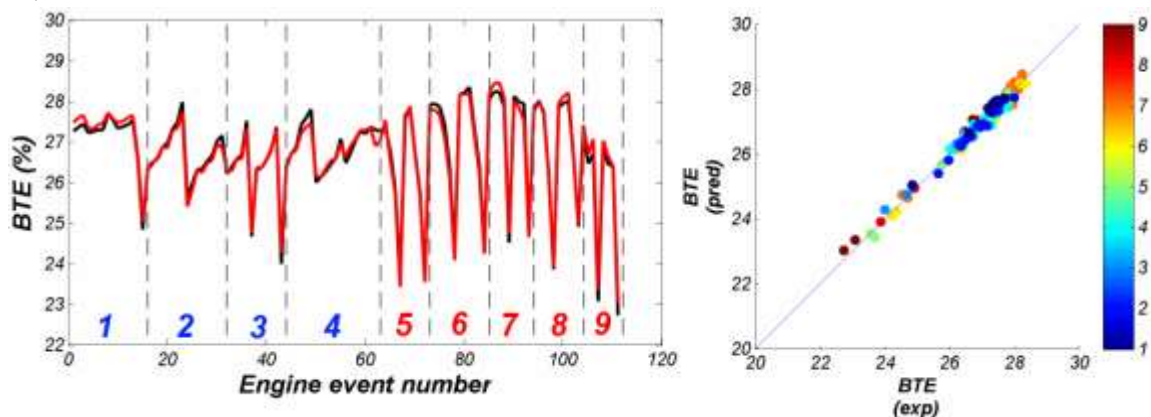
- There is no clear trend in brake thermal efficiency for HECC-ORNL dataset.

All models:

All models obtained for thermal efficiency (BTE) for the HECC engine are characterized by high regression performance. Chemistry and physical properties of the FACE fuels have minimal effect on the model of BTE parameter.



The aicDynoTorque is the most important control parameter in the PLS regression equation for model BTE values. Which makes sense since for a fixed fuel and fueling rate, the torque output of the engine will increase as the BTE increases. In fact correlation coefficient for aicDynoTorque and BTE is high and equal 0.9571. The influence of other variables (such as fuel descriptors in models 2a-3b) is hidden by described relation between aicDynoTorque and BTE.



Removing aicDynoTorque from the list of input parameters significantly decreased model predictability in all modeling strategies. Only models for BTE based on high-CN fuels present reasonable modeling options.

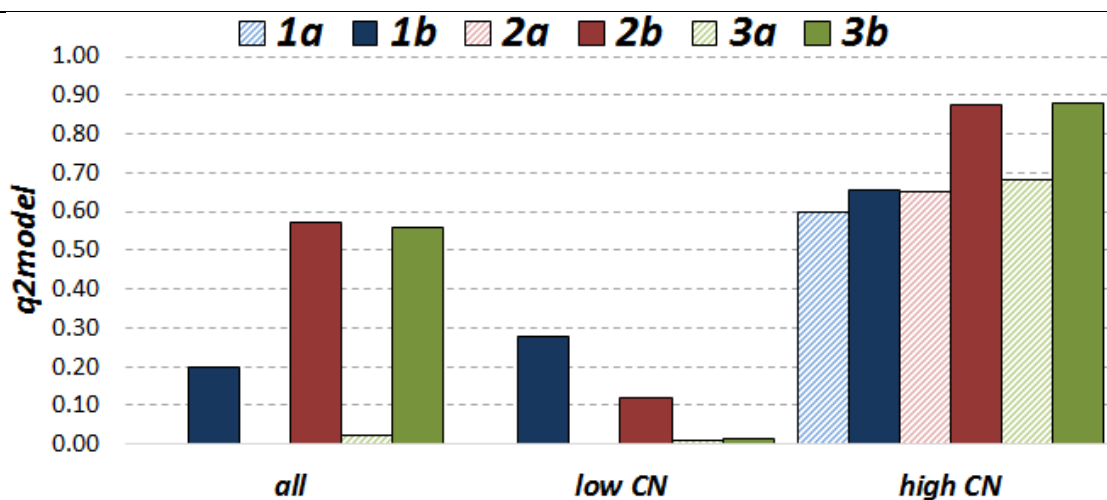


Figure: BTE models performance (after removing *aicDynoTorque* from the list of input parameters)

Following is an example of the Model 1a for all engine events

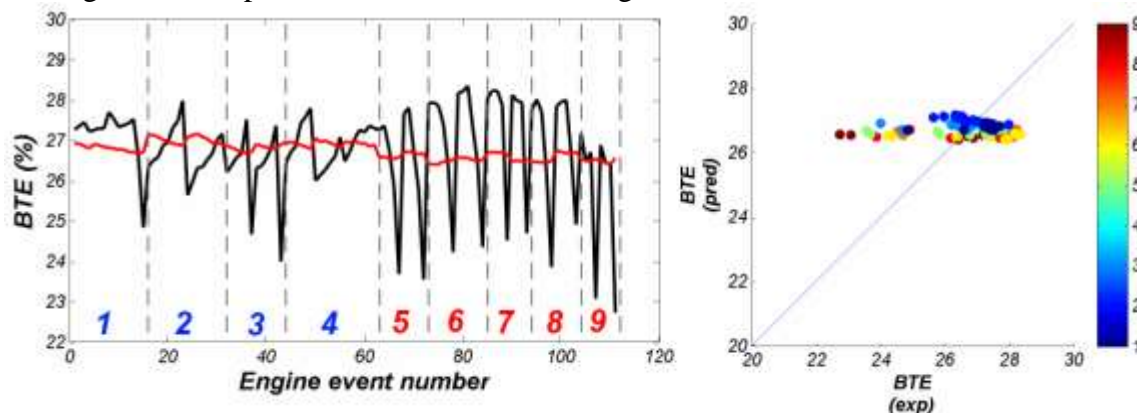


Figure: BTE Model 1a performance (after removing *aicDynoTorque* from the list of input parameters)

Model improved after combining both fuel descriptors and cross terms for all variables.

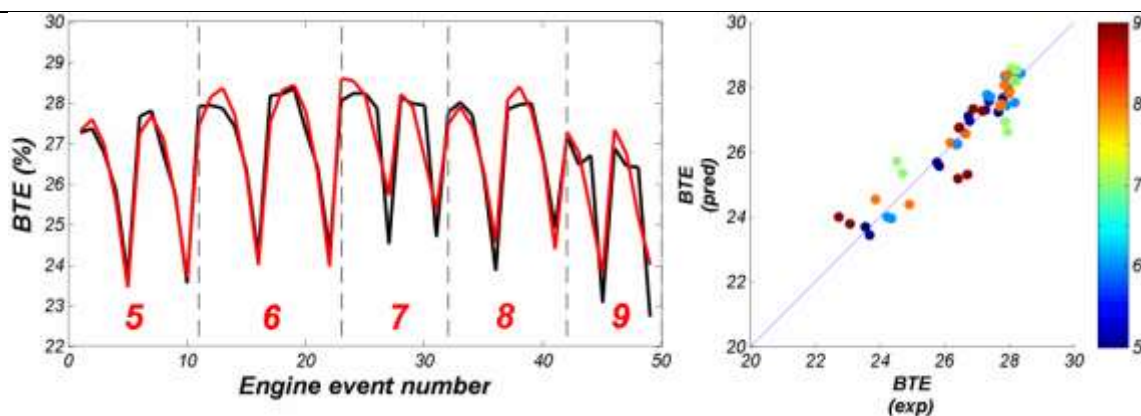


Figure: BTE Model 1a performance (after removing *aicDynoTorque* from the list of input parameters) build only for high-CN fuels.

7.2.10 SUMMARY OF PLS MODEL 2a RESULTS

A summary of the PLS models containing the 3 fuel design + engine control parameters (model 2a) for this engine is tabulated below. The filled circles (•) represent highly influential parameters and the open circles (◦) represent moderately influential parameters. Plus (+) symbols indicate parameters are directly correlated while negative (-) symbols indicate inverse correlations.

LTC MCE 1.9L Engine (HECC-ORNL):

Parameter	HC		CO		NO _x		PM		ITE
CN	•	(-)	•	(-)	•	(+)	•	(+)	(-)
T90		(-)		(-)		(+)		(+)	(+)
Aromatics		(-)		(-)		(+)		(+)	(-)
aicDynoTorque	•	(-)	•	(-)	◦	(+)		(+)	(-)
aicEGRMassGM	◦	(-)		(+)	•	(-)	•	(+)	•
aicINTO2GM		(-)		(-)		(+)		(+)	(+)
MainSOI(ATDC)	•	(+)	•	(+)	•	(-)	•	(+)	•

• - highly influential parameter

(+) - directly correlated

◦ - moderately influential parameter

(-) - inversely correlated

7.3 CDC – SCE 0.5L (CDC-ORNL)

An experimental study was performed at Oak Ridge National Laboratory using a 513-cc Hatz single-cylinder diesel engine equipped with advanced combustion analysis equipment. In the original study the experiments were performed using 16 fuels: nine FACE diesel fuels, one ultra-low sulfur diesel (ULSD) fuel, and six P20 blends of unprocessed plant oils (6). Only engine data related to FACE diesel fuels were extracted from the original work and evaluated here in order to understand fuel property effects on conventional diesel combustion mode.

7.3.1 DATA DESCRIPTION

The scheme in Figure 35 presents the list of input and output control parameters provided from the CDC engine experiments. For the nine fuels, a total of 90 experimental data points (aka engine events) were obtained.

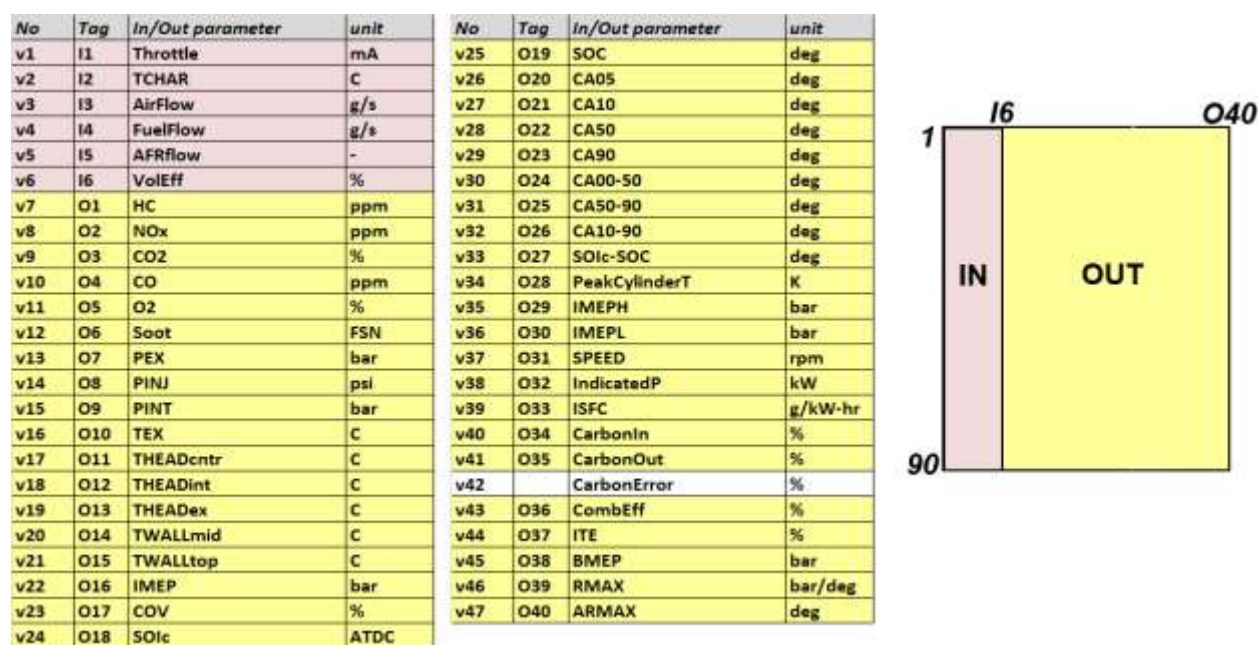


Figure 35 – Scheme presenting list of control input (I*) and output (O*) engine parameters for CDC- SCE 0.5L (as provided by ORNL researchers and located in FACE diesel engine database)

7.3.2 EXPLORATORY ANALYSIS

In the first step of analysis of the engine data, the correlation among input engine control and engine output parameters was performed and presented in scatter form. The results are presented in Figure 36 and Figure 37. In the case of a strong one-to-one relationship a plot of the two variables should yield a straight line (or curve in case of non-linear correlation – see relation between I5 (AFRflow) vs I4 (FuelFlow) in Figure 36).

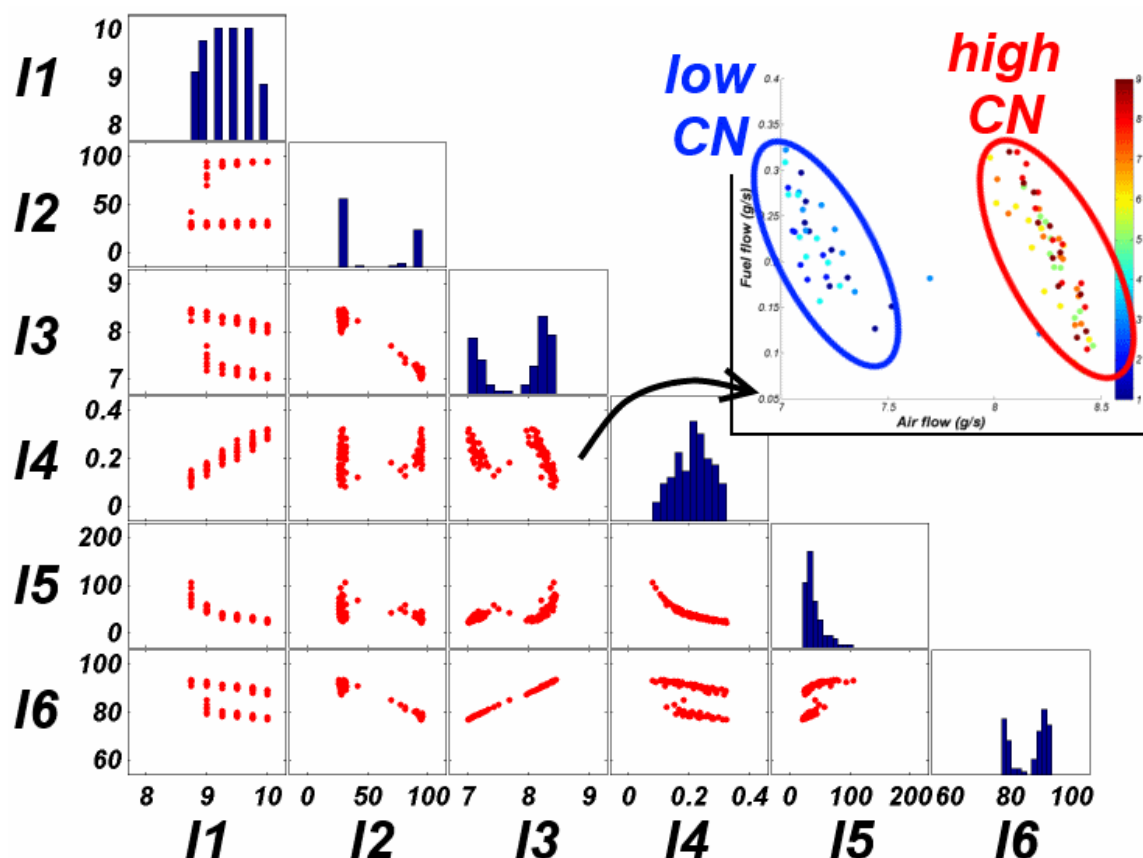


Figure 36 – Scatter representation of cross-correlation between control input engine parameters.

For meaning of labels (I*) on both x and y axes please refer to the Figure 35. Histograms for each variable are located on the diagonal. The inset in the top-right corner magnifies the existing relations between air flow and fuel flow.

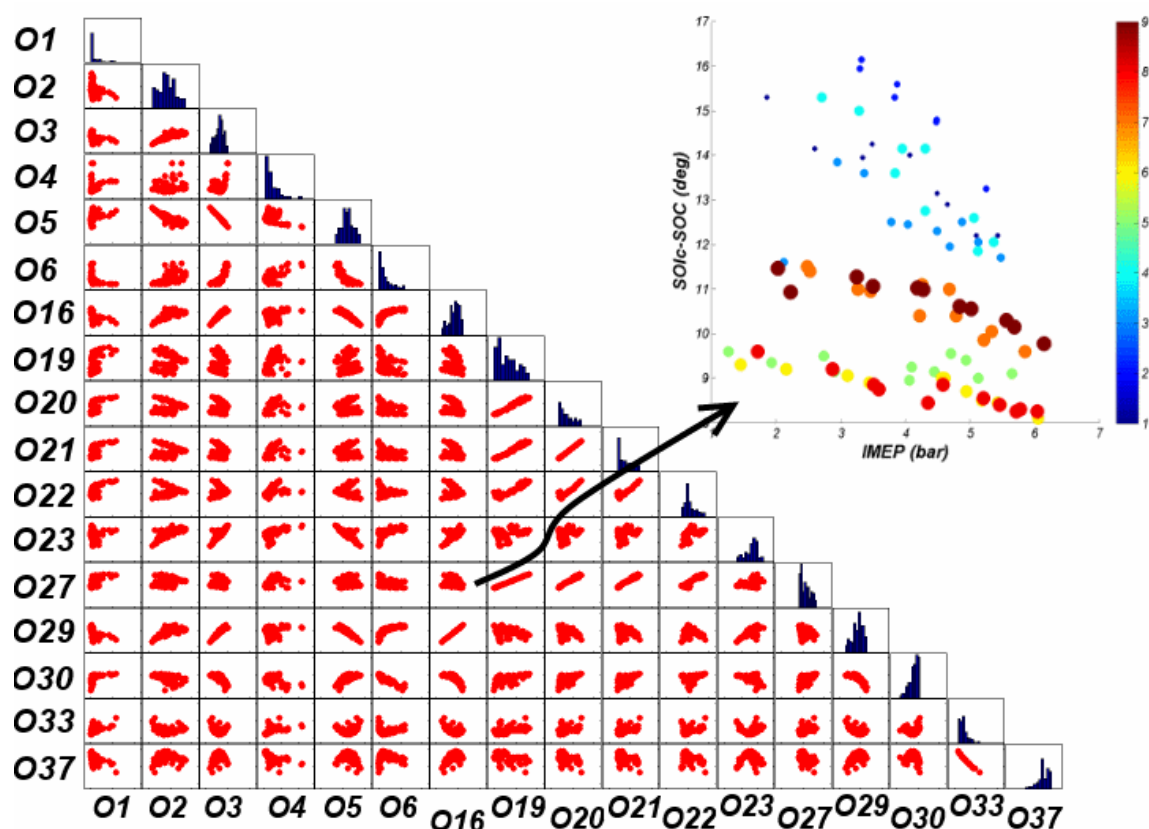


Figure 37 – The scatter representation of cross-correlation existing among selected output engine parameters obtained for CDC engine, where numbers indicate consecutive output parameters (O* - see Figure 35). The inset located in the top-right corner magnifies the existing relations between SOIc-SOC (aka ignition delay) and IMEP.

PCA was performed both on the input and output data provided to CanmetENERGY. The scores plot for input data (Figure 38) shows clear separation of control parameters into two groups (based on low- and high-CN fuels). Based on information given in the loadings plot, we observed that TCHAR and the pair VolEff /AirFlow are the control parameters mainly responsible for the existing separation of the engine/fuel input space. The diagonal spread of the points observed in the PCA scores plot is a result of the other parameters (i.e., Throttle, FuelFlow and AFR flow). Based on these results we conclude that the choice of engine control parameters was strongly influenced by fuel chemistry - in this case CN.

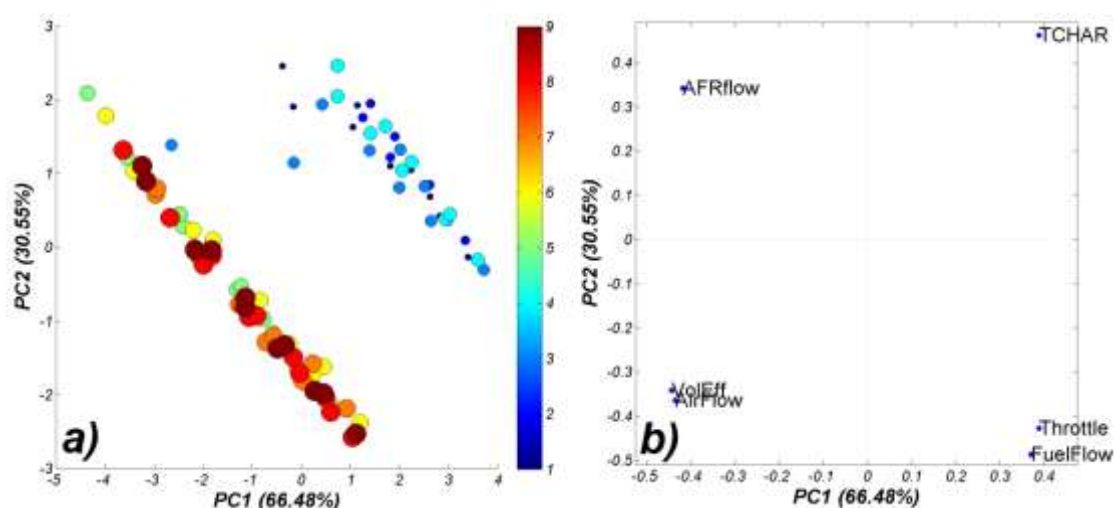


Figure 38 – The PCA scores (a) and loadings (b) plot (PC1-PC2) obtained after running PCA for HECC ORNL engine input (control) parameters. On scores plot bubbles colour and size are associated with respective fuels (i.e., smallest dark blue dots – FD1, largest brown bubbles – FD9).

Figure 39 presents results of the PCA analysis on 38 engine output parameters obtained for all 90 engine events scenarios. Two parameters were excluded from the calculations (i.e., SOIc (O18) and SPEED (O31)) because of low variance in those parameters. On the scores plot (Figure 39a) we notice the clear separation of events for low- and high-CN fuels, which is strongly associated and may be explained by the relative locations of output parameters in the loadings plot (Figure 39b).

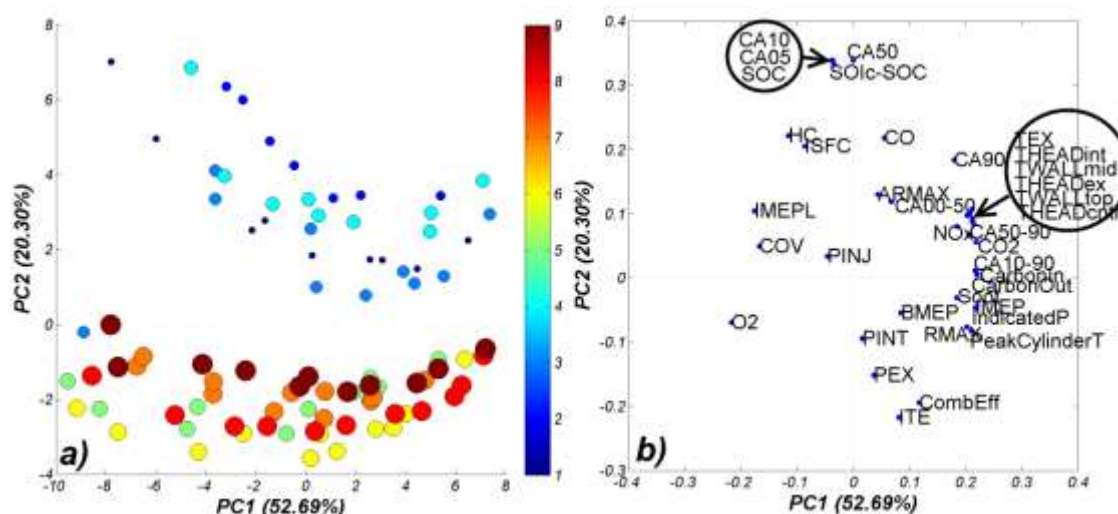


Figure 39 – The PCA score (a) and loading (b) plots (PC1-PC2) obtained after running PCA for CDC ORNL engine output parameters for all events. On scores plot bubble colour and size are associated with respective fuels (i.e., smallest dark blue dots – FD1, largest brown bubbles – FD9).

7.3.3 MODELING APPROACH

Simple one-to-one correlations between engine control parameters and selected engine output parameters are presented in Figure 40.

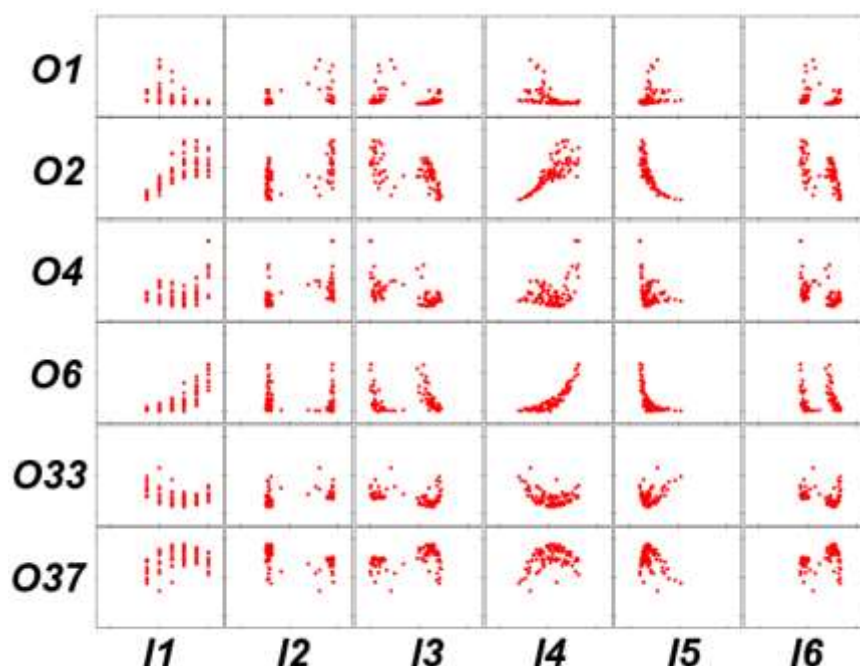


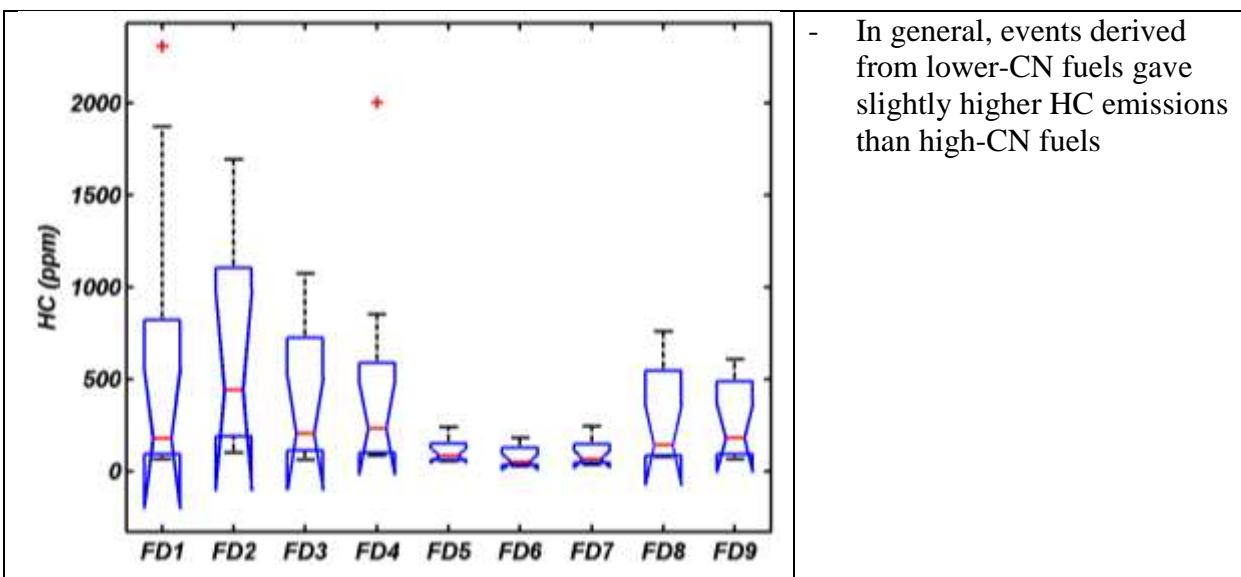
Figure 40 – Scatter representation of cross-correlation between output engine parameters obtained for CDC engine run where I* are control variables and O* indicate output engine parameters.

In the following section the results of modeling emissions and efficiency parameters are presented. The engine was modeled using strategies described in detail in section 6:

- (strategy: 1a): Throttle (I1), TCHAR (I2), AirFlow (I3), FuelFlow (I4), AFRflow (I5), VolEff (I6),
- (strategy: 1b): strategy 1a + cross terms for all variables,
- (strategy: 2a): strategy 1a + FACE fuel CN/T90/aromatic content,
- (strategy: 2b): strategy 2a + cross terms for all variables,
- (strategy: 3a): strategy 1a + 40 FACE fuel descriptors (see section 6),
- (strategy: 3b): strategy 3a + cross terms for all variables.

PLS models were built for all events, events with only high CN fuels, and events with only low CN fuels. For each modeled property 18 PLS models were calculated, a summary of the results is presented in the following sections (7.3.4 – 7.3.8)

7.3.4 HYDROCARBON EMISSIONS



All HC models:

Good models were calculated for engine events run with high-CN fuels. Inclusion of fuel descriptors without cross terms in models for low-CN fuels reduced their performance.

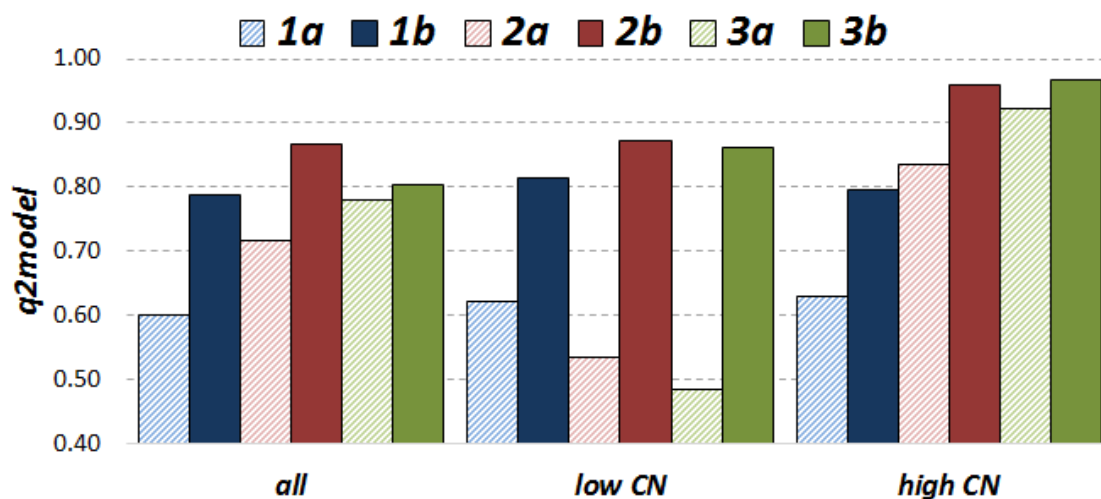


Figure: All HC models performance

HC model 1a-all:

In general the PLS model generated for HC emissions is weak. The following profiles of experimental (black) and predicted (red) HC emission shows the inability to properly model high HC emission events. A few underestimated values are observed, especially for low-CN fuels.

Important control parameters: Throttle (-) and Fuelflow (-).

Prediction of HC emission significantly improves after including cross terms in the modeling.

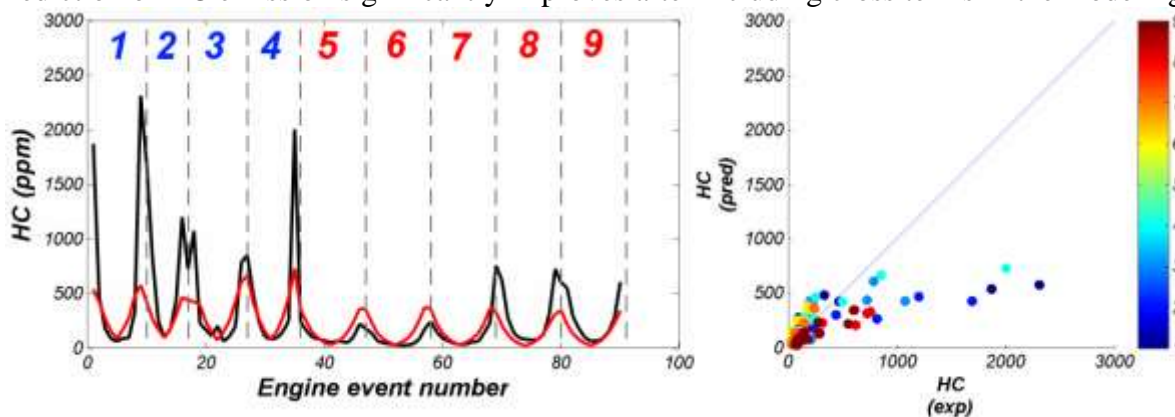


Figure: HC model 1a performance

HC model 2a-all:

Inclusion of three fuel descriptors in HC modeling improves modeling capability. However improvement is not significant.

Important control parameters: Throttle (-) and Fuelflow (-)

Important fuel parameters: CN (-)

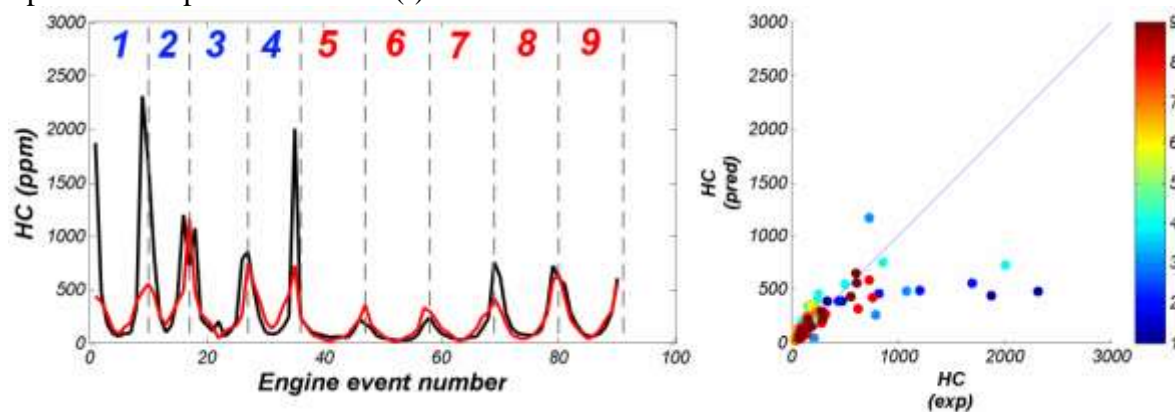


Figure: HC model 2a performance

HC model 3a-all:

Similar in performance to model 2b

Important control parameters: Throttle (-) and Fuelflow (-)

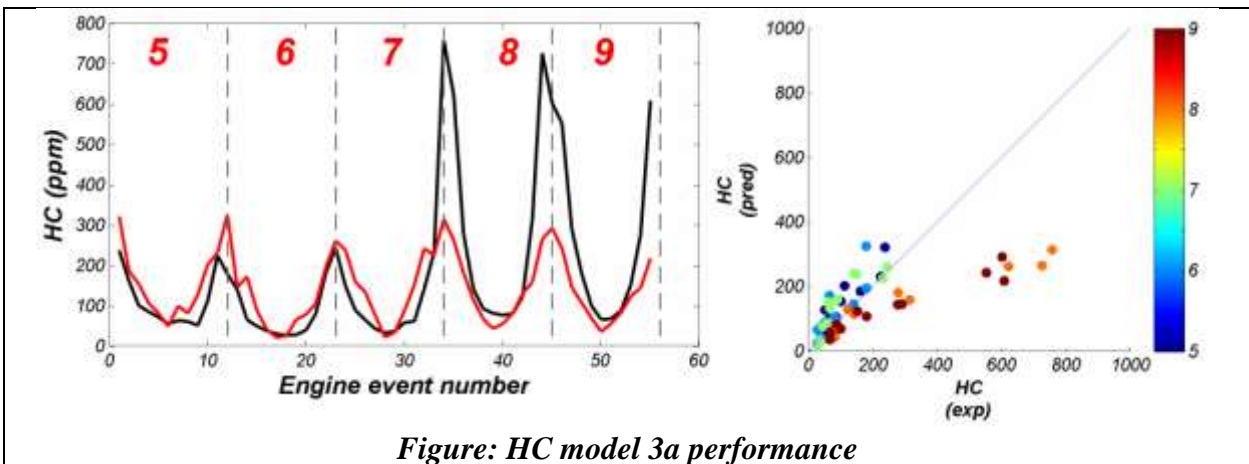
Important fuel parameters: CN (-), T30(-), T40(-), CT1(-), CT2(-), CT8(-), olefins(+)

HC models for low-CN fuels

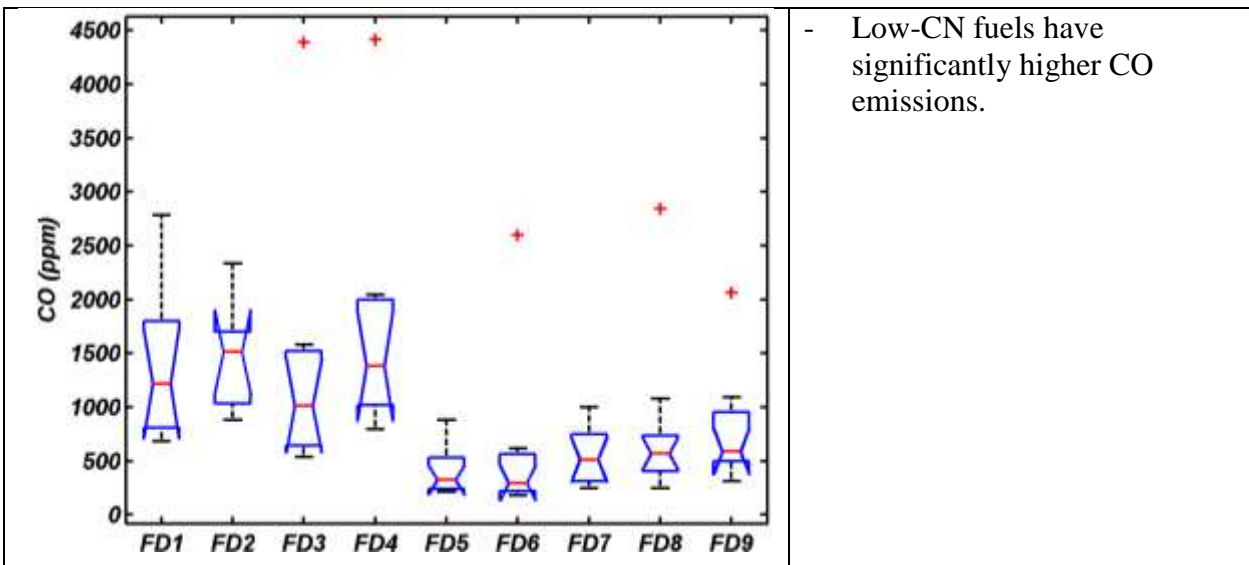
Poor models – better after adding cross terms.

HC models for high-CN fuels

In general HC models calculated for high-cetane fuels are characterized by better performance in comparison to HC models for low-cetane fuels, especially after including fuel descriptors in the models.



7.3.5 CARBON MONOXIDE EMISSIONS



All CO Models:

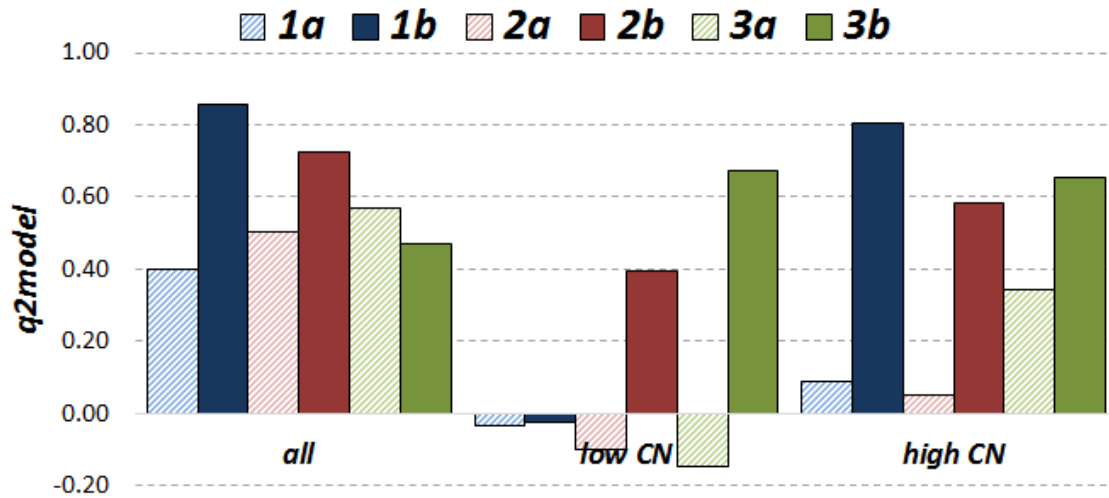


Figure: All CO model performance

CO model 1a-all:

In general the PLS model generated for CO emissions is weak. The following profiles of experimental (black) and predicted (red) HC emission shows the inability to properly model this property using only engine control parameters. We may observe a few underestimated values. Prediction of CO emissions significantly improves after including cross terms in the model (see performance for CO model 1b).

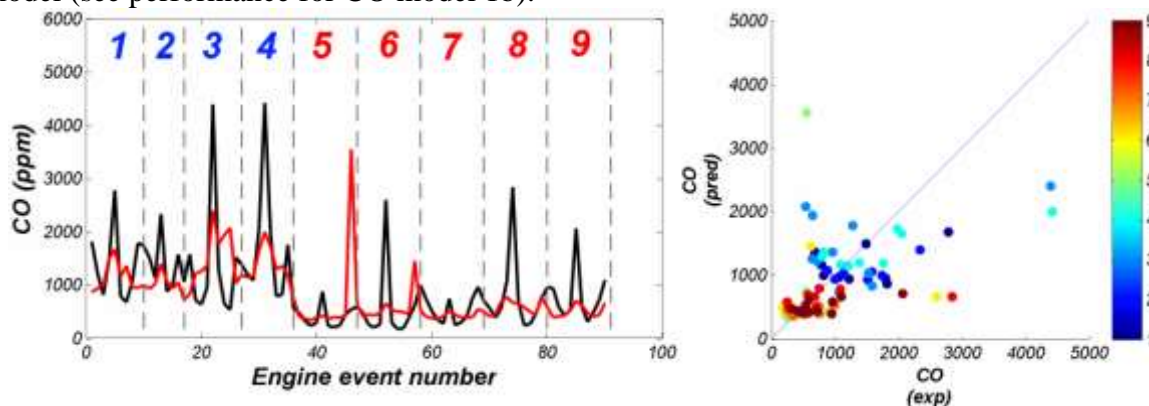


Figure: CO model 1a performance

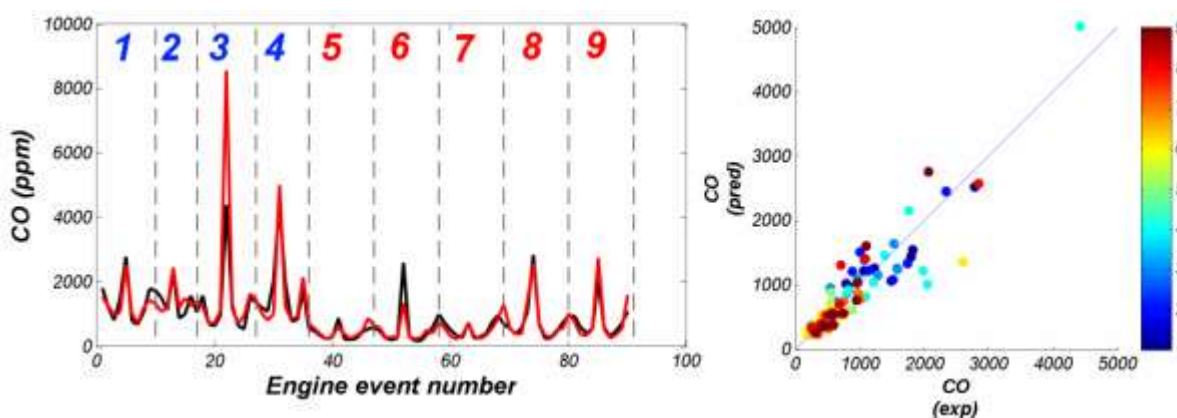
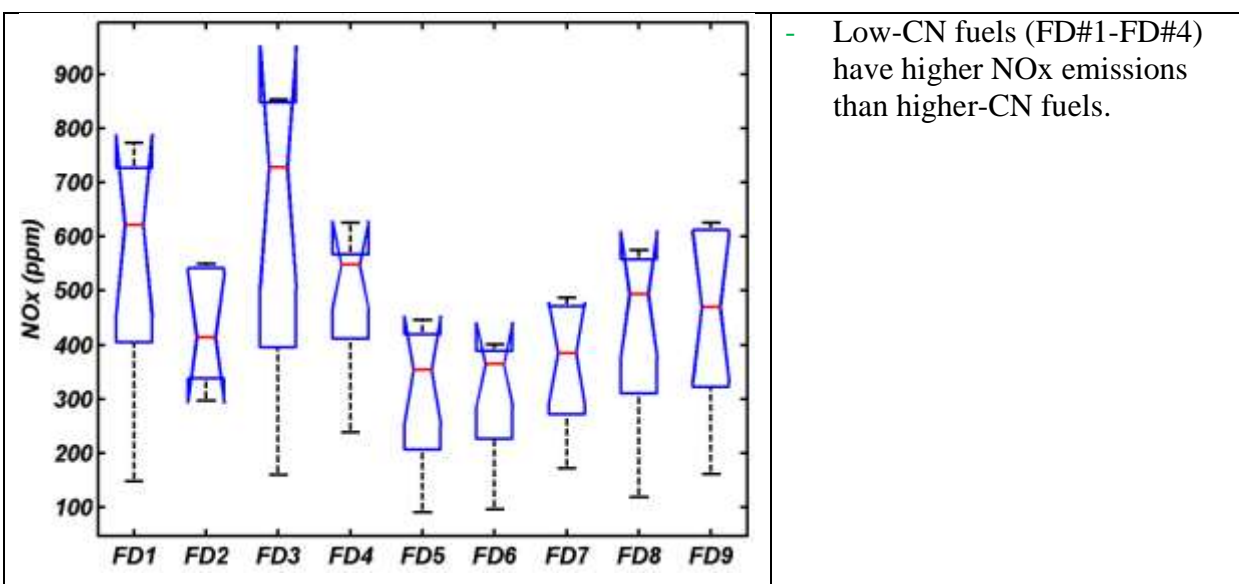


Figure: CO model 1b performance

7.3.6 NITROGEN OXIDES EMISSIONS



All NOx models:

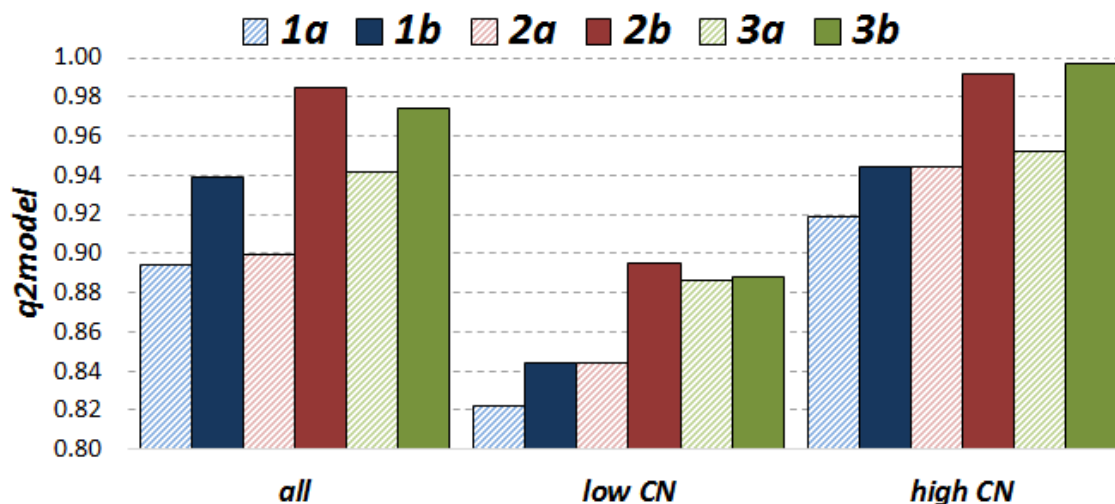


Figure: NOx models performance

NOx model 1a-all:

The PLS model built for NOx emissions shows good capability to model that property. Important control parameters: Throttle (-), Fuelflow (+), and AFRflow (-)

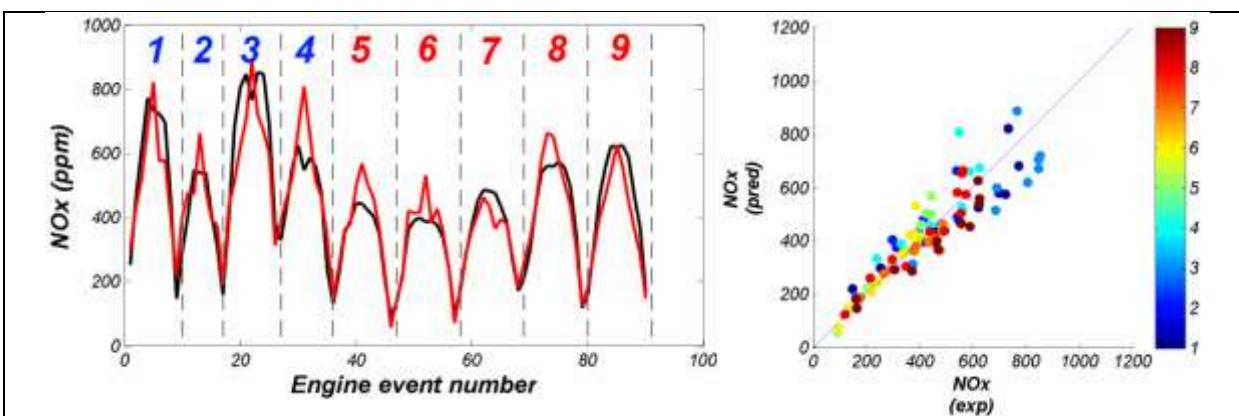


Figure: NOx model 1a performance

NOx model 2a-all:

The PLS model built for NOx emissions shows good capability for modeling that property. Information about fuel does not improve NOx models significantly.

Important control parameters: Throttle (-), Fuelflow (+), and AFRflow (-)

Important fuel parameters: CN (+), then aromatics (+)

NOx model 3a-all:

Important fuel parameters: CN (+), CT2 (+), CT8(-) and alkylbenzenes

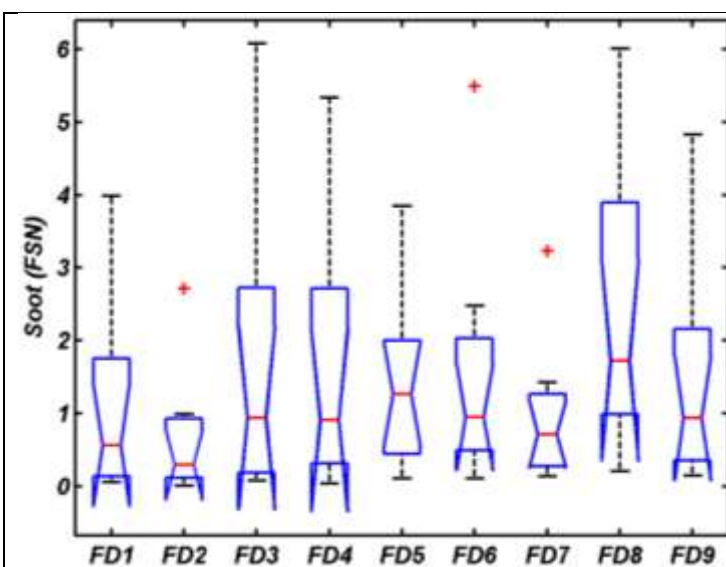
NOx models for low-CN fuels

Important fuel parameters: CN (+), CT2 (+), CT8(-) and alkylbenzenes

NOx Models for high-CN fuels

Important fuel parameters: CN (+), CT2 (+), CT8(-) and alkylbenzenes

7.3.7 SOOT EMISSIONS



- No clear trends

SOOT all models:

SOOT emissions are relatively easy to model based on both control parameters and additional chemical and physical properties of FACE diesel fuels. PLS models built for high-CN fuels are stronger than respective models for low-CN fuels. In these types of models, SOOT emissions strongly depends on the engine control parameters such as AFR_{flow}, Throttle, fuel flow. CN is the most important fuel parameter

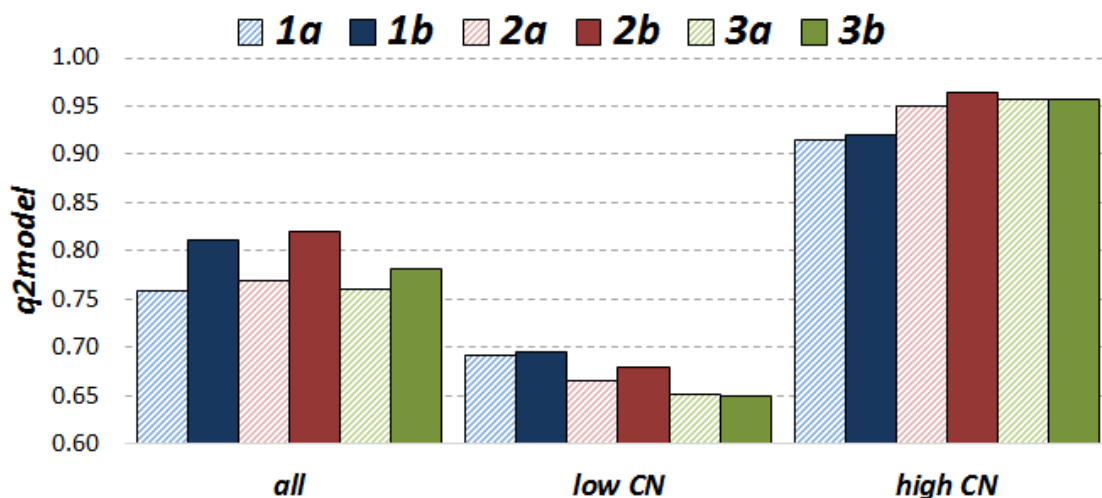


Figure: SOOT models performance

SOOT Model 1a-all:

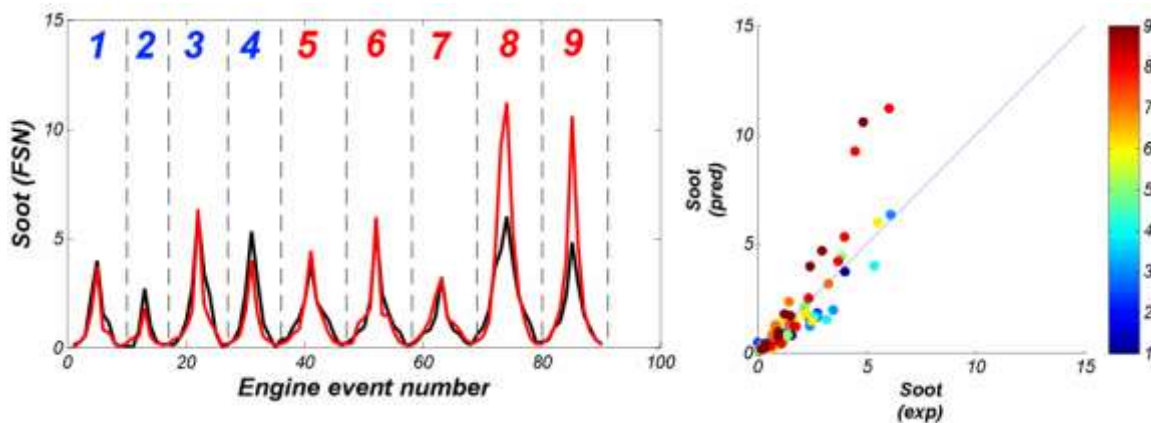
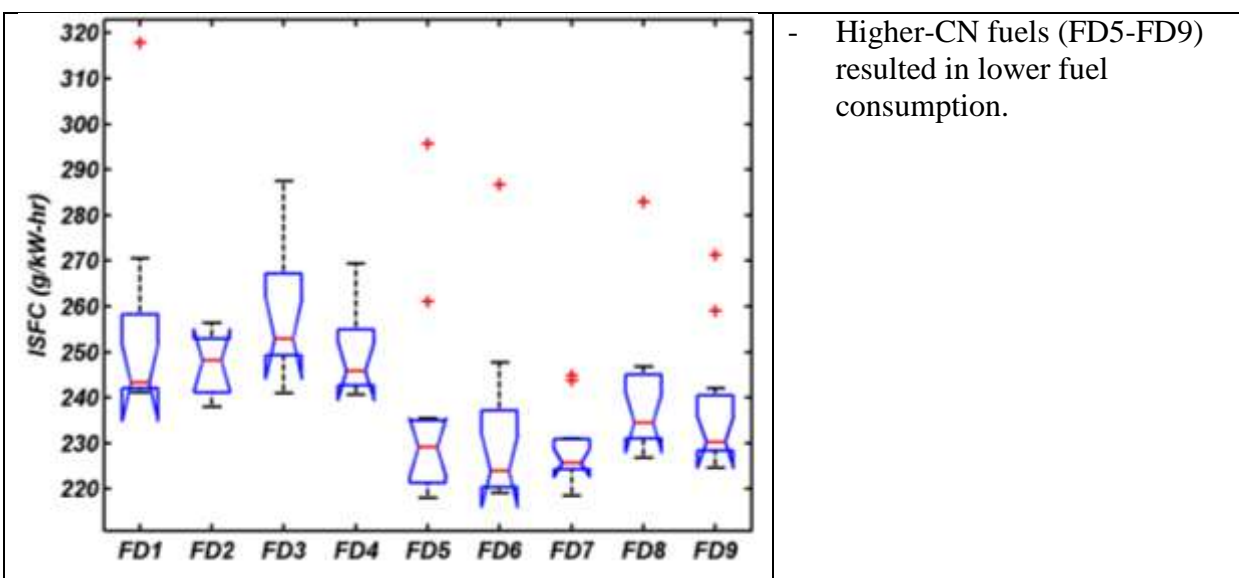


Figure: SOOT model 1a performance

7.3.8 INDICATED SPECIFIC FUEL CONSUMPTION



All ISFC models:

ISFC models built for only low CN fuels events has poor ISFC predictability. PLS models built for high-CN fuels are significantly stronger than respective models for low-CN fuels. In these types of models, ISFC strongly depends on the engine control parameters such as AFRflow, Throttle, fuel flow. CN and T90 are the most important fuel parameters.

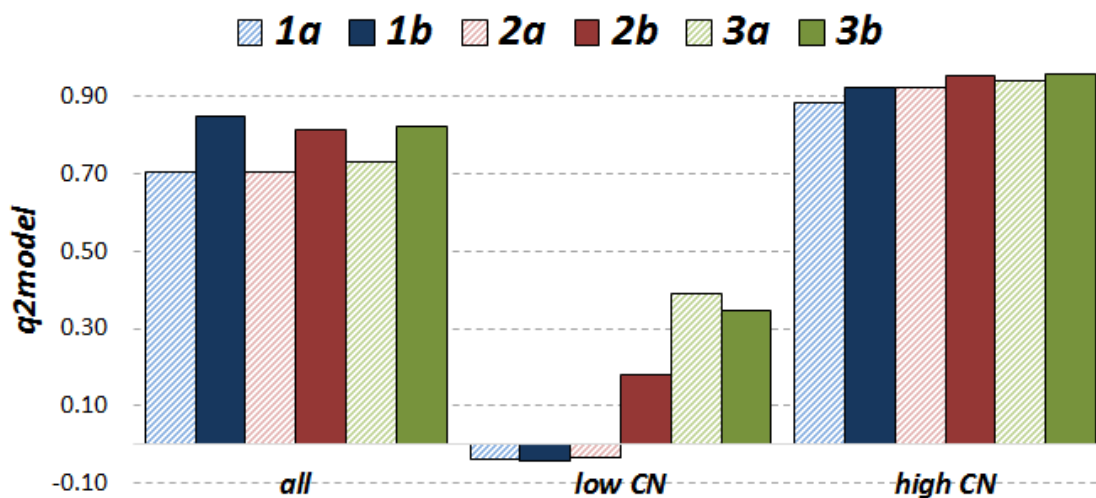
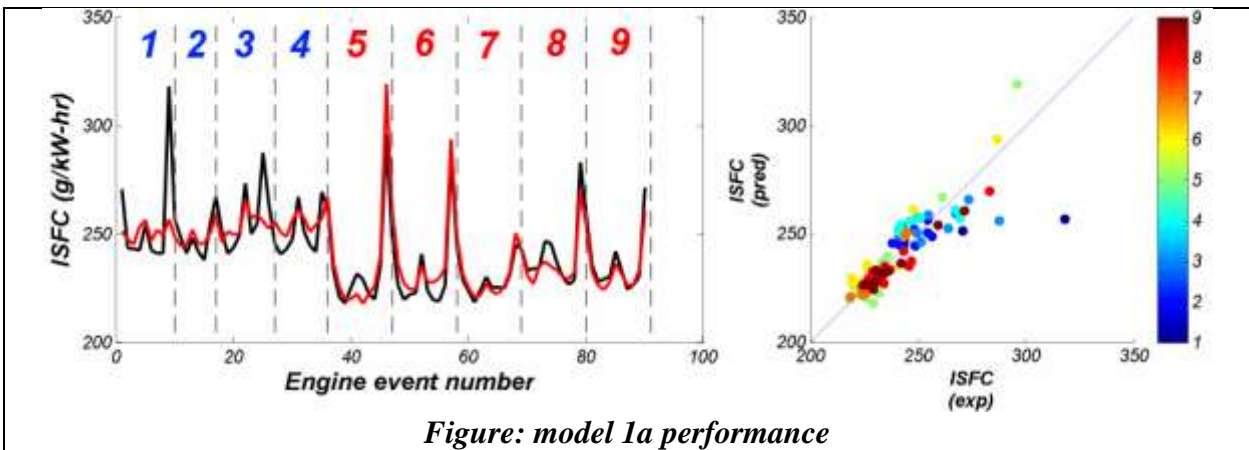
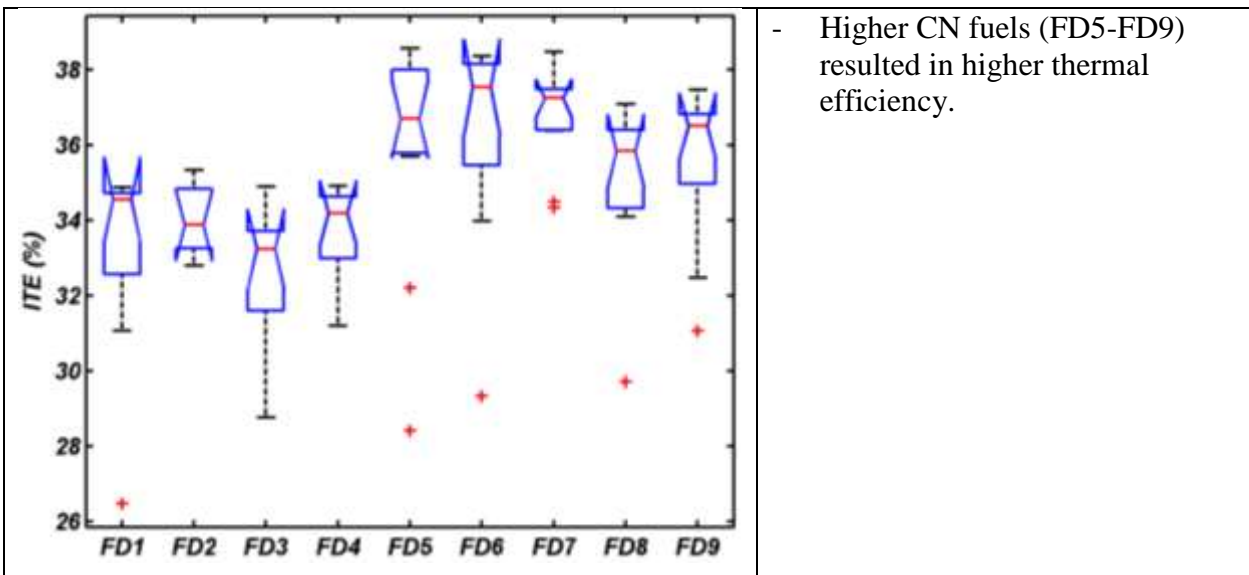


Figure: ISFC models performance

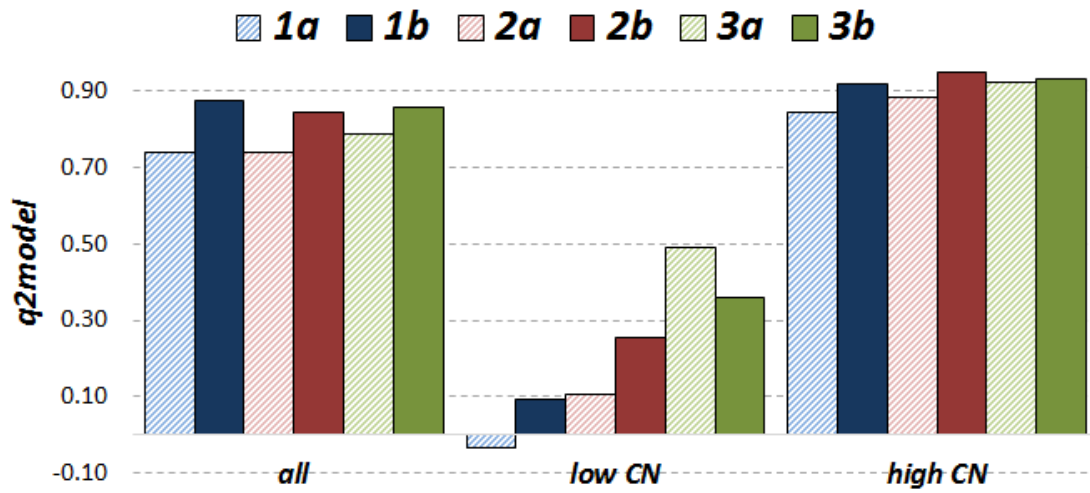
ISFC Model 1a-all:

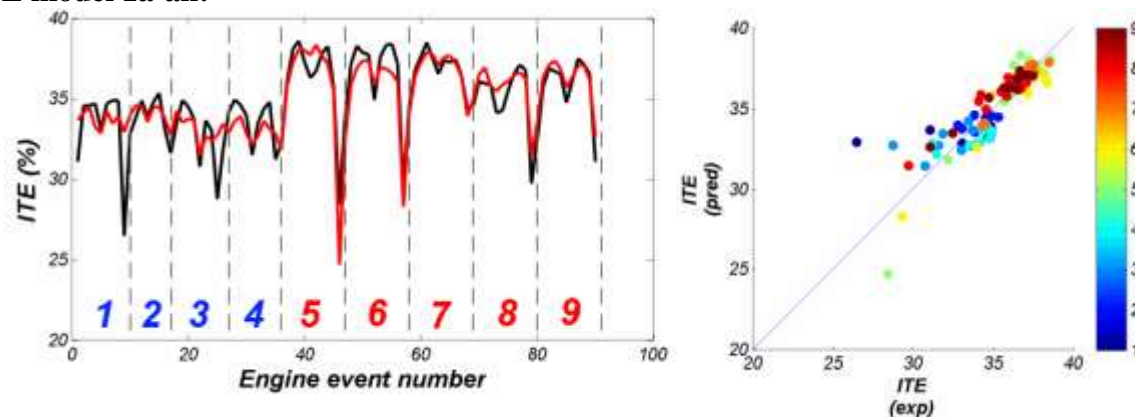


7.3.9 THERMAL EFFICIENCY



ITE all models:



ITE model 1a-all:*Figure: ITE model 1a performance*

7.3.10 SUMMARY OF PLS MODEL 2a RESULTS

A summary of the PLS models containing the 3 fuel design + engine control parameters (model 2a) for this engine is tabulated below. The filled circles (•) represent highly influential parameters and the open circles (◦) represent moderately influential parameters. Plus (+) symbols indicate parameters are directly correlated while negative (-) symbols indicate inverse correlations.

CDC-SCE 0.5L Engine (CDC-ORNL):

Parameter	HC		CO		NOx		Soot		ISFC		ITE	
CN	•	(-)	•	(-)		(+)		(+)	•	(-)	•	(+)
T90		(+)		(+)		(-)		(-)		(-)		(+)
Aromatics		(+)	◦	(+)		(+)		(+)		(+)		(-)
Throttle (mA)	•	(-)		(+)	•	(-)	•	(-)	•	(-)	•	(+)
TCHAR (°C)	◦	(+)	•	(+)		(+)		(+)	◦	(+)	•	(-)
Airflow		(+)	•	(-)	◦	(+)		(+)	◦	(-)	◦	(+)
Fuelflow	•	(-)		(+)	•	(+)	•	(+)	•	(+)	•	(-)
AFR Flow	•	(-)	◦	(+)	•	(-)	•	(+)	•	(+)	•	(-)
Vol.Effic.		(+)	•	(-)	◦	(+)		(+)	◦	(-)	◦	(+)

• - highly influential parameter

◦ - moderately influential parameter

(+) - directly correlated

(-) - inversely correlated

7.4 LTC – SPLIT INJECTION (LTC-AVFL-16)

An experimental study was performed in the Center for Alternative Fuels, Engines & Emissions (CAFEE) West Virginia University under CRC AVFL-16 project titled “Fuels to Enable Light-Duty Diesel Advanced Combustion Regimes” (10). The goal of that project was to investigate the effects of FACE diesel fuel properties on the combustion, emissions, and performance characteristics of light-duty production engines during advanced combustion operation. In that study not all FACE fuels were investigated. The FD2 fuel was replaced by an ultra-low sulfur diesel (ULSD) emissions certification fuel (aka CFA). The engine utilized in that study was a GM Z19DTH light-duty compression-ignition engine run with both a split-injection control strategy and a single-injection control strategy.

The modeling studies presented in this work are focused only on investigation of the effect of fuels properties on a split injection control strategy.

7.4.1 DATA DESCRIPTION

The scheme in Figure 41 presents the list of input and output control parameters provided from the light-duty compression-ignition engine experiment. The split injection control strategy comprised of varying the start of the pilot injection, the start of the main injection, and the fuel split ratio with a constant intake oxygen concentration of 16% and constant rail pressure of 1600 bar. For the nine fuels, a total of 396 experimental data points (aka engine events) were taken.

No	Tag	In/Out parameter	unit
v1		Engine Speed	RPM
v2	O1	BMEP	bar
v3	I1	Main SOI	BTDC
v4	I2	Pilot SOI	BTDC
v5	I3	Fuel Split	% Pilot
v6		Rail Pressure	bar
v7		Intake O2	%
v8	O2	HC	ppm
v9	O3	Nox	ppm
v10	O4	CO	ppm
v11	O5	Soot	mg/m3
v12		HC	g/kW-hr
v13		Nox	g/kW-hr
v14		CO	g/kW-hr
v15		Soot	mg/kW-hr
v16	O6	BTE	%
v17	O7	Max Pressure Rise Rate	bar/deg
v18	O8	CA50	deg

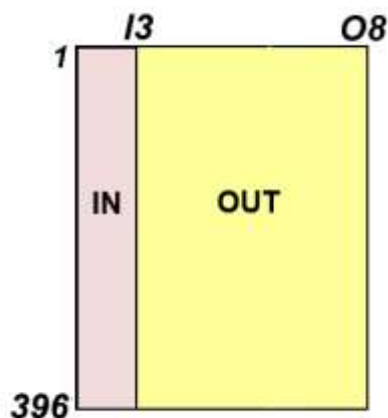


Figure 41 – Scheme presenting list of control input and output engine parameters for light-duty compression-ignition engine run in split injection control strategy

7.4.2 EXPLORATORY ANALYSIS

In the first step of the analysis of the engine data, the correlation between input engine control and engine output parameters was calculated and presented in scatter form in Figure 42a. The engine control parameters for split injection control strategy were selected in order to orthogonally map the space described by three parameters (Pilot SOI, Main SOI and Fuel split), therefore no correlation between parameters in such a design experiment are found.

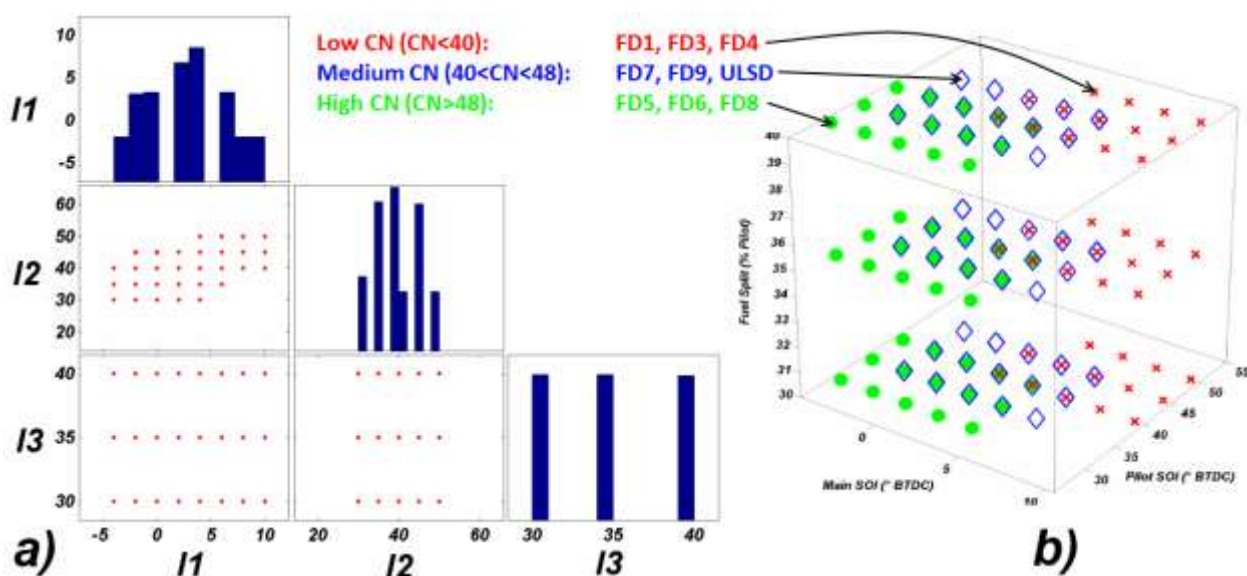


Figure 42 – (a) Scatter representation of cross-correlation between control input engine parameters. For meaning of labels (I^*) on both x and y axes please refer to the Figure 41. Histograms for each variable are located on the diagonal. (b) The design cube for split injection strategy, where points indicate specific conditions used for each engine run with a particular fuel (categorized as low-, medium-, and high-CN fuels)

Because of the orthogonal selection of control parameters for split injection, PCA was not performed for the input data. Instead the experimental design cube is plotted in Figure 42b where each point depicts an experimental condition used for running a particular fuel.

The correlation among the eight output engine parameters is presented in scatter form in Figure 43. In the inset presented in that figure, strong correlations between CO and HC emissions can be seen.

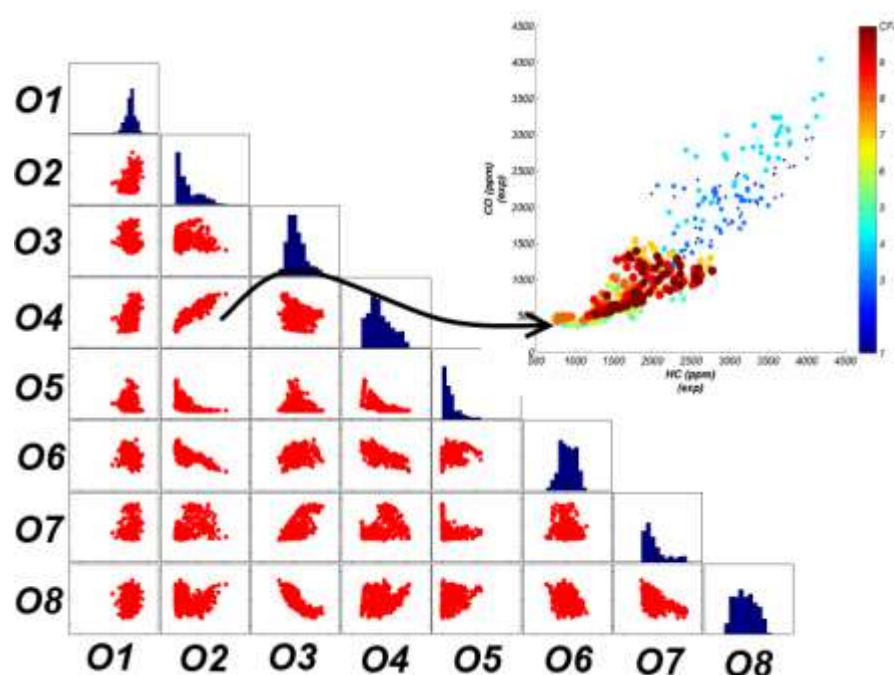


Figure 43 – Scatter representation of cross-correlation existing among output engine parameters obtained for CDC engine, where numbers indicate consecutive output parameters (O* - see Figure 41). Histograms for each variable are located on the diagonal. The inset in the top-right corner magnifies the high correlation between hydrocarbon and carbon monoxide emissions.

Figure 44a presents results of PCA analysis calculated on eight engine output parameters obtained for all 396 engine events scenarios. Notice the clear separation of the low-CN events from the group of medium- and high-CN fuels.

In the design cube for the split injection strategy, we see that some control input parameters are common for all fuels (see Figure 42b for points where three markers are overlaid). In the following step only output parameters for these points were selected for consecutive PCA, resulting in the scores plot presented in Figure 44b. Again, the obvious separation of events into three groups related to CN of the fuels is observed in the scores plot.

This simple example demonstrates that emissions and efficiencies in this type of engine are affected to a great extent by the FACE fuel properties.

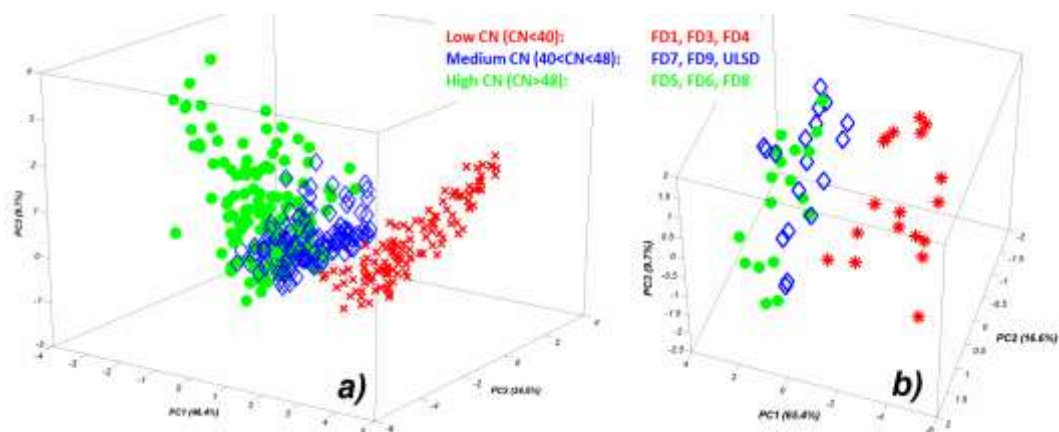


Figure 44 – PCA scores plot (PC1-PC2-PC3) obtained after running PCA for LTC-AVFL engine output parameters for all events (a), and events characterized by the same control input parameter values (b). On the scores plots, colour and marker type are associated with respective fuel set (low-, medium-, or high-CN fuel)

7.4.3 MODELING APPROACH

Simple one-to-one correlations between engine control parameters and selected engine output parameters are presented in Figure 45.

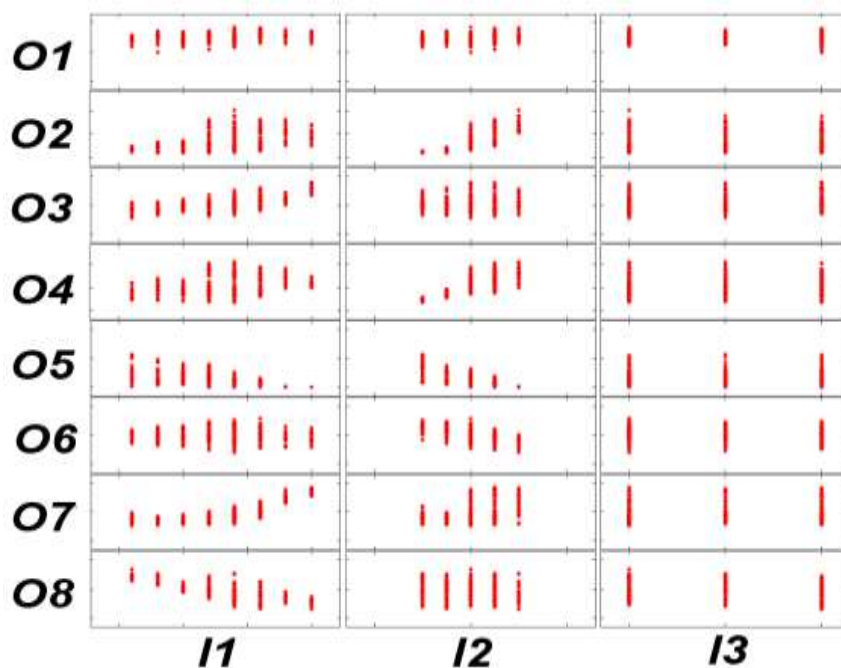


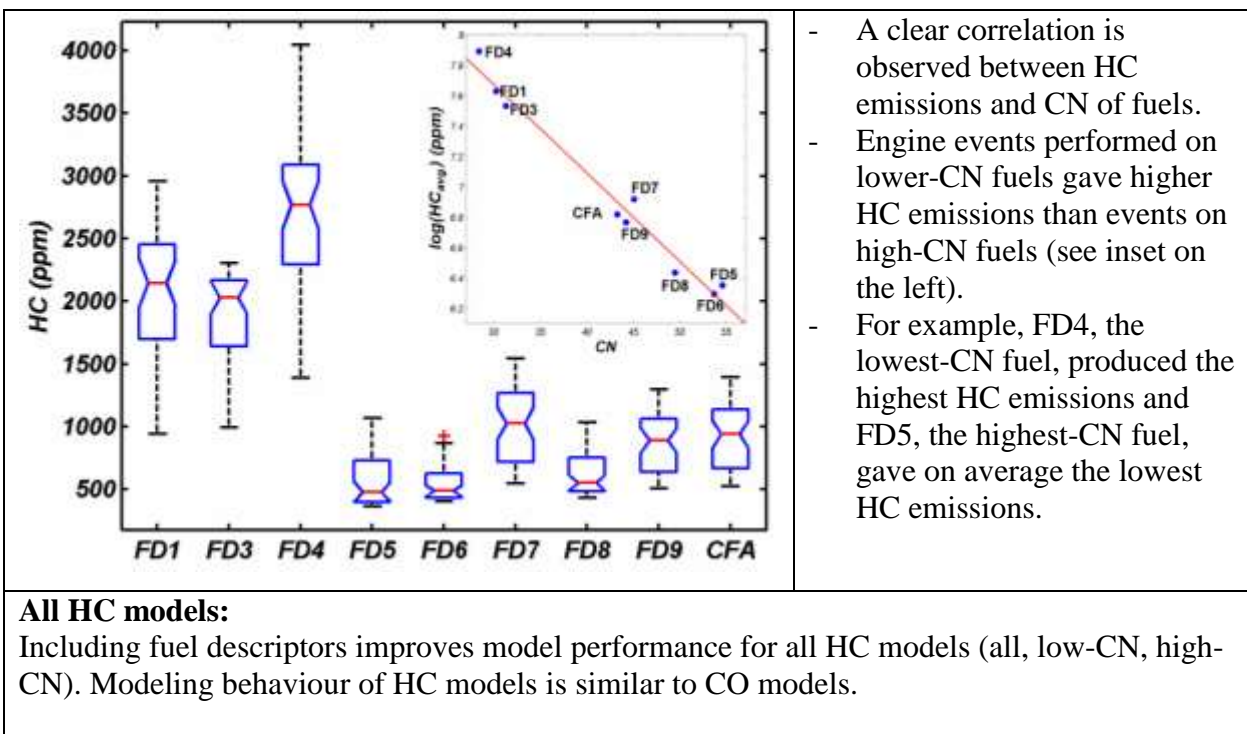
Figure 45 – Scatter representation of cross-correlation between output engine parameters obtained for LTC-AVFL engine runs, where I* are control variables and O* indicate output engine parameters (see Figure 41)

In the following section the results of modeling of emission and efficiency parameters will be presented. The engine was modeled using strategies described in detail in section 6:

- (strategy: 1a): Main SOI (I1), Pilot SOI (I2), Fuel split (I3),
- (strategy: 1b): strategy 1a + cross terms for all variables,
- (strategy: 2a): strategy 1a + FACE fuel CN/T90/aromatic content,
- (strategy: 2b): strategy 2a + cross terms for all variables,
- (strategy: 3a): strategy 1a + 40 FACE fuel descriptors (see section 6),
- (strategy: 3b): strategy 3a + cross terms for all variables.

PLS models were built for all events, events with only high CN fuels, and events with only low CN fuels. For each modeled property, 18 PLS models were calculated and a summary of the results is presented in following sections (7.4.4 – 7.4.8)

7.4.4 HYDROCARBON EMISSIONS



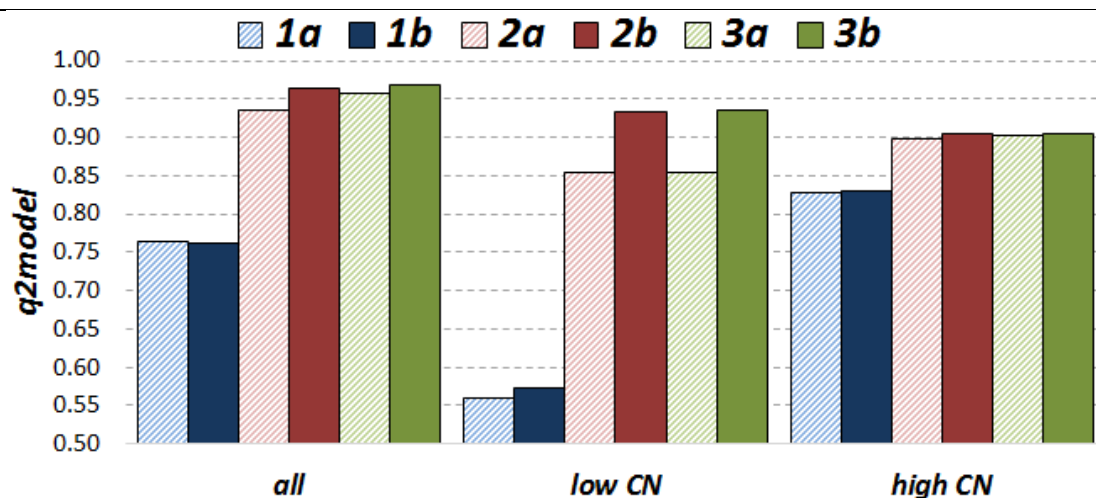


Figure: All HC models performance

HC model 1a-all:

In general the PLS model generated for HC emissions is weak. The following profiles of experimental (black) and predicted (red) HC emissions shows the inability to properly model events with low HC emission.

Important control parameters: Pilot SOI (+), Main SOI (+)

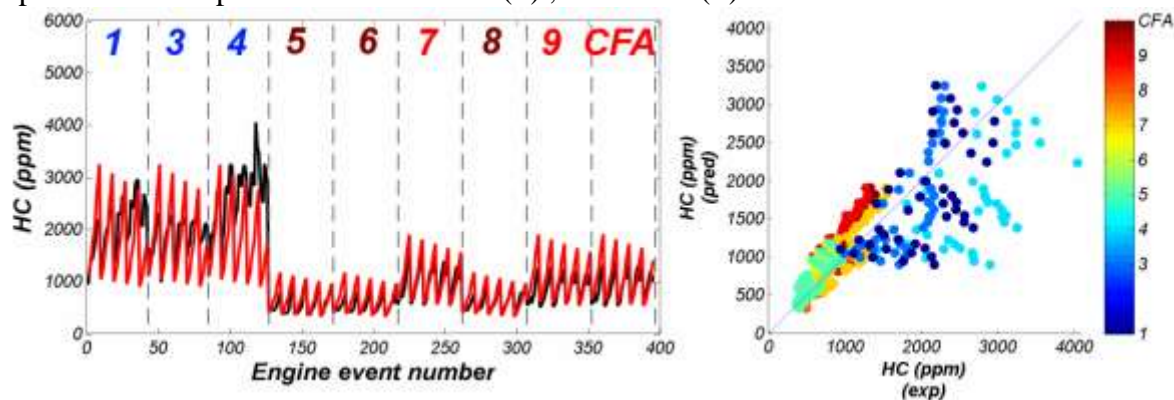


Figure: HC model 1a performance

HC model 2a-all:

Inclusion of the three fuel design descriptors in HC modeling improves modeling capability.

Important control parameters: Pilot SOI (+), Main SOI (-)

Important fuel parameters: CN (-)

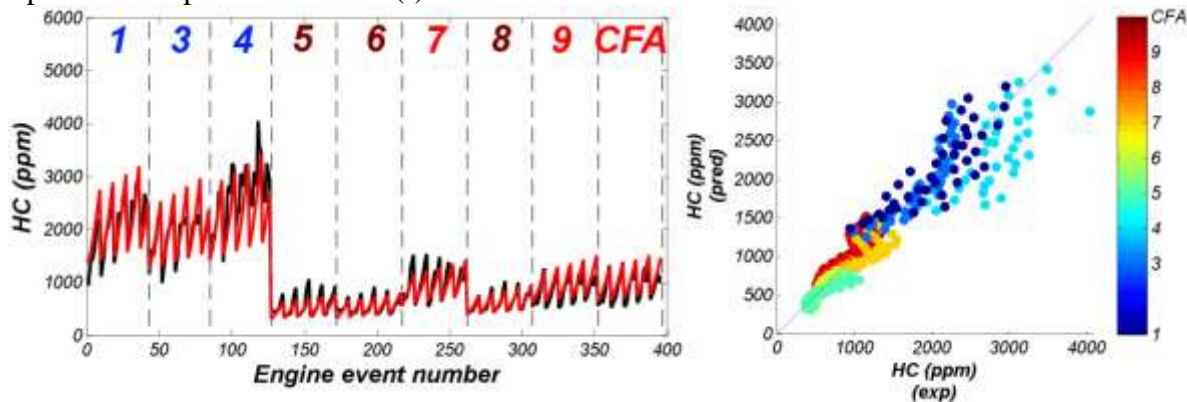


Figure: HC model 2a performance

HC model 3a-all:

Similar in performance to model 2b

Important control parameters: Pilot SOI (+), Main SOI (-)

Important fuel parameters: viscosity (+), CN (-), T10-T50(-), monoaromatics (-)

HC models for low-CN fuels

Models based just on engine control parameters are poor and improve significantly after adding fuel descriptors.

HC models for high-CN fuels

In general HC models calculated for high-CN fuels are characterized by good performance in comparison to HC models for low-CN fuels. Including fuel descriptors in the HC model improves predictability of the models.

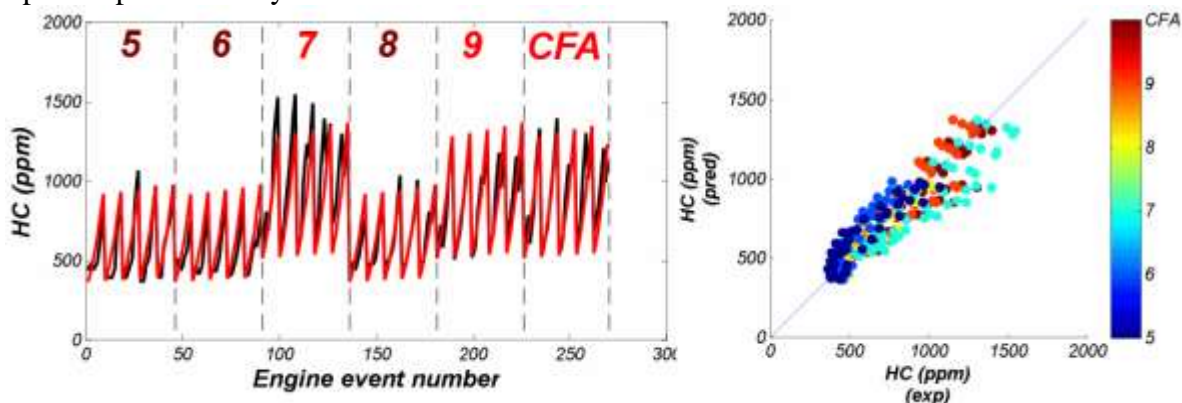
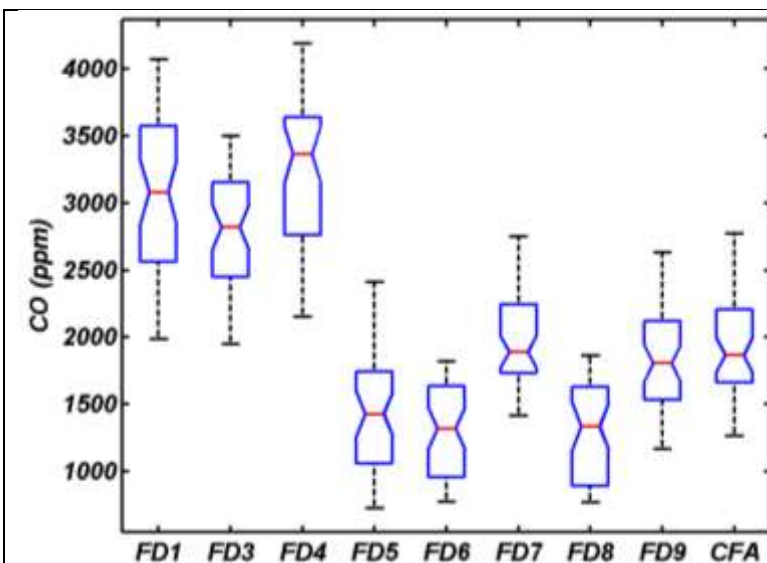


Figure: HC model a performance

7.4.5 CARBON MONOXIDE EMISSIONS



- Similar to HC emissions trend.
- Engine events performed on lower-CN fuels gave higher CO emissions than events for high-CN fuels.

All CO models:

Including fuel descriptors improves CO model performance for all models (all, low-CN, high-CN).

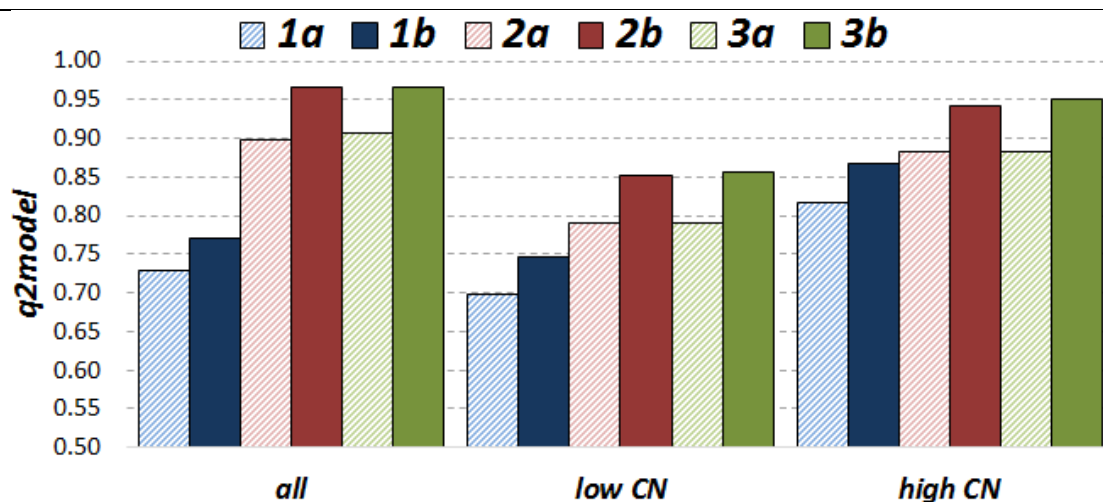


Figure: All CO Models performance

CO model 1a-all:

In general, the PLS model generated for CO emission works for engine events run on high-CN fuels.

Important control parameters: Pilot SOI (+) , Main SOI (-)

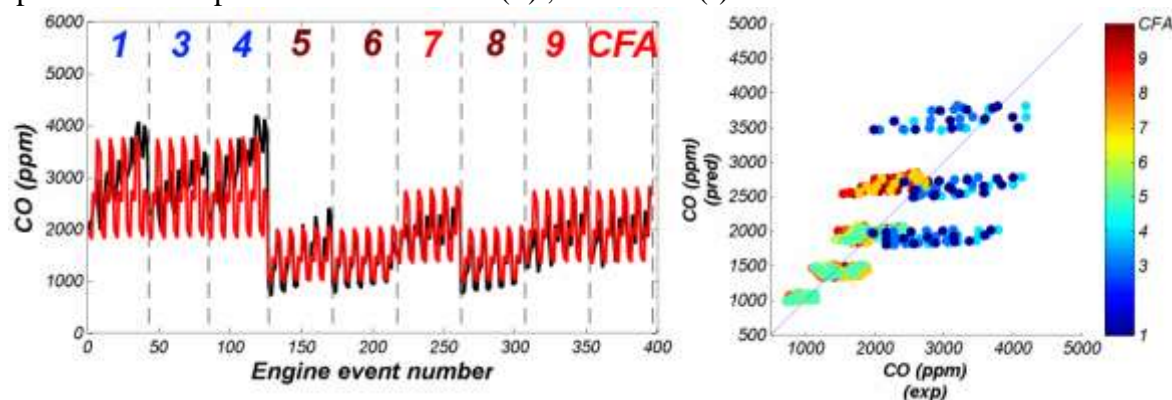


Figure: CO model 1a performance

CO model 2a-all:

Prediction of CO emissions significantly improves after including fuel information in the modeling.

Important control parameters: Pilot SOI (+) , Main SOI (-)

Important fuel parameters: CN (-)

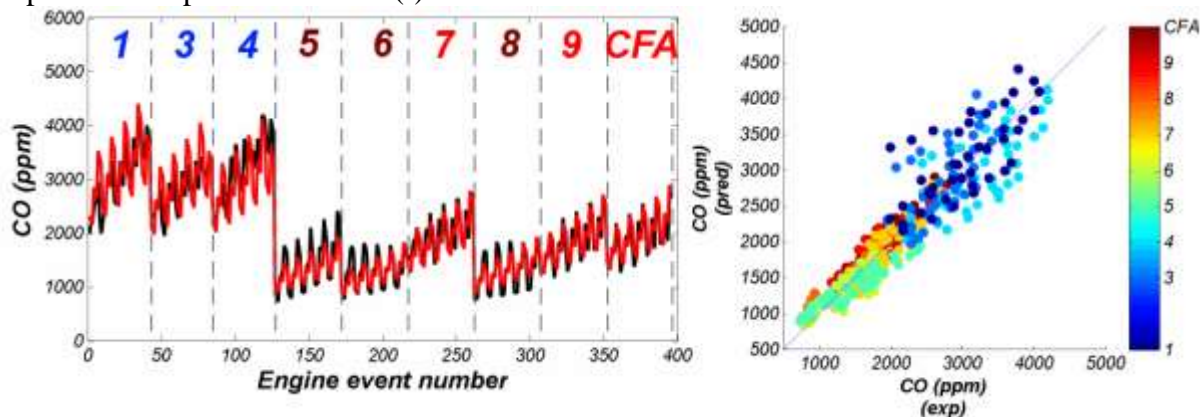


Figure: CO model 2a performance

CO models for low-CN fuels

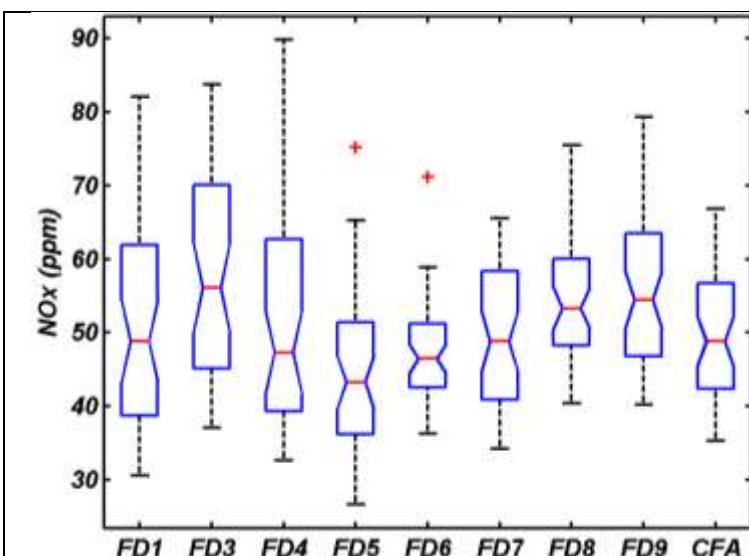
Predictions of CO emissions for events utilizing low-CN fuels is better than from models including events with all fuels.

Important control parameters: Main SOI (-)

CO models for high-CN fuels

Different important control engine parameters were used in modeling CO emission for low- and high-CN fuels. In the case of events run with low-CN fuels, the 'main SOI' is important, whereas in the case of events run with high-CN fuels, the 'Pilot SOI' is important.

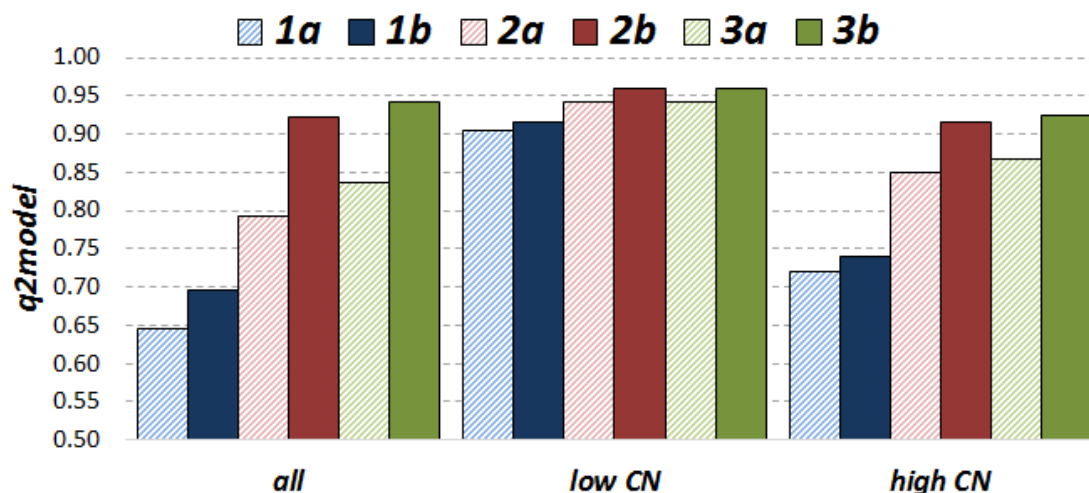
Important control parameters (model 2a): CN (-) and T90 (-), which may be translated as follows: high-CN and T90 fuels will produce lower CO emissions than other fuel types.

7.4.6 NITROGEN OXIDES EMISSIONS

- No obvious correlations for NOx emissions data presented in the box plot.
- Broader distribution of NOx emissions values for events run on low-CN fuels.

All NOx models:

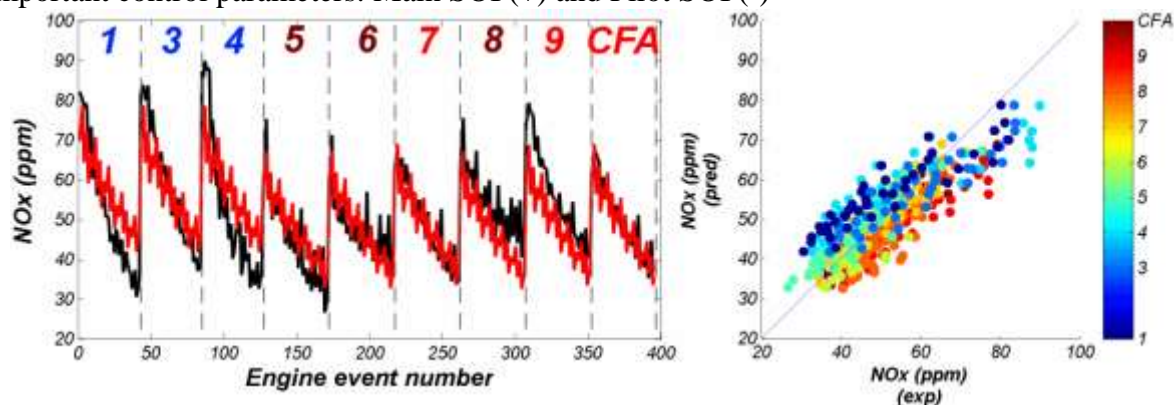
Best NOx models were obtained for engine events run on low-CN fuels.



*Figure: NOx models performance***NOx model 1a-all:**

The PLS model built for NOx emission shows good capability for model NOx behaviour.

Important control parameters: Main SOI (+) and Pilot SOI (-)

*Figure: NOx model 1a performance***NOx model 2a-all:**

Using the three fuel design descriptors improves NOx predictability of PLS models.

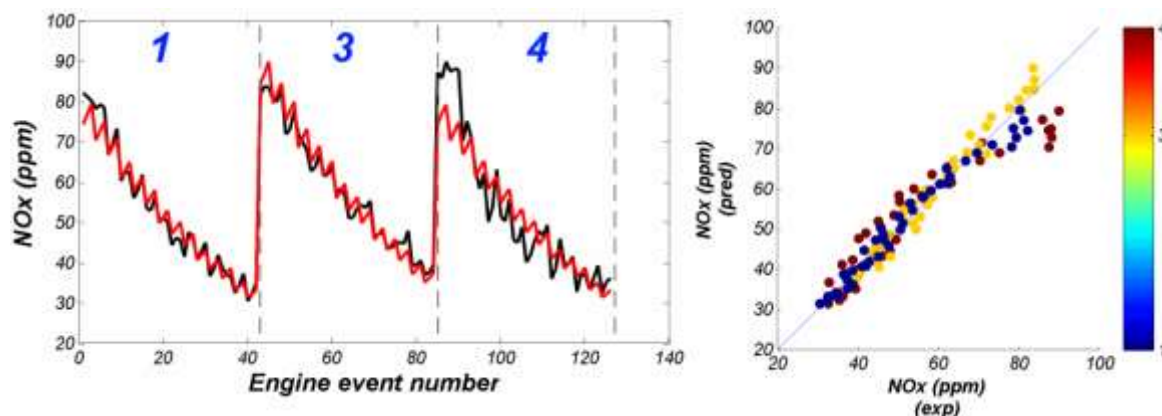
Important control parameters: Main SOI (+), Pilot SOI (-)

Important fuel parameters: CN (+), then aromatic content (+)

NOx models for low-CN fuels

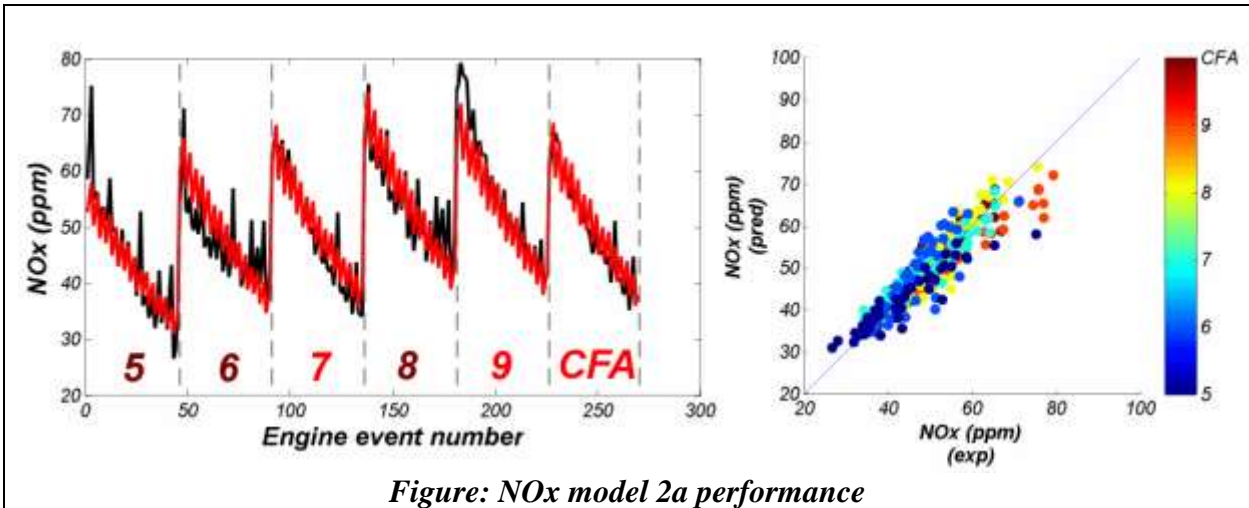
Good models for NOx emission prediction based only on engine control parameters. Does not improve significantly after combining with fuel physical and chemical information.

Important control parameters: Main SOI (+)

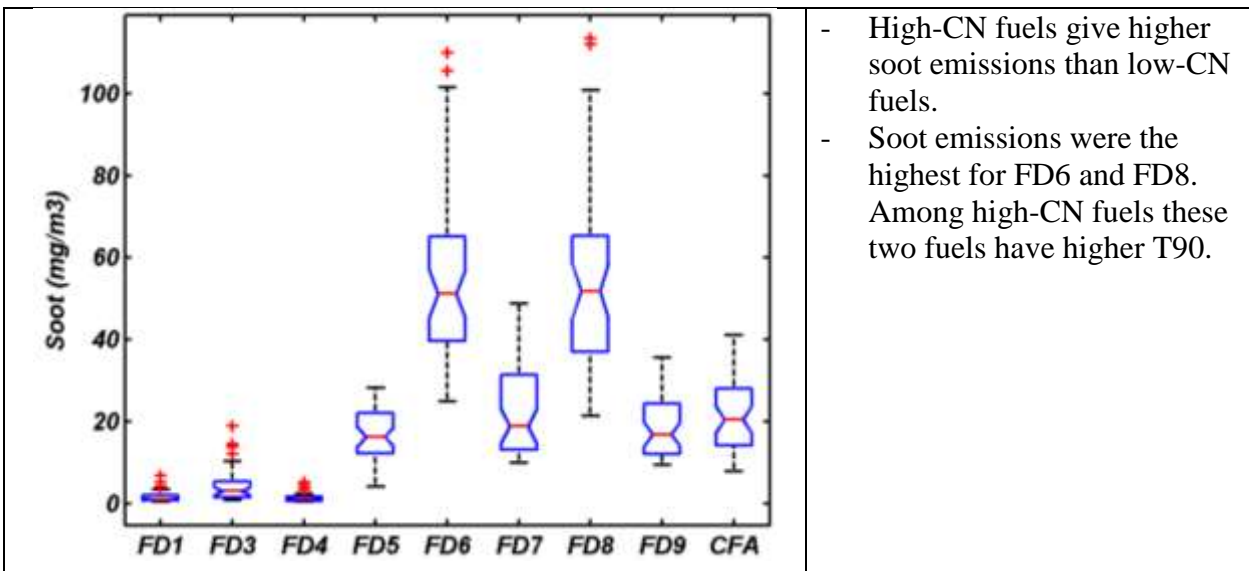
*Figure: NOx model 2a performance***NOx models for high-CN fuels**

Good models (1a) based on control engine parameters which may be enhanced by taking into consideration the three fuel design descriptors (CN/T90/Aro).

Important fuel parameters: T90 (+) and aromatic content (+). CN is less important for this set of engine events.



7.4.7 SOOT EMISSIONS



All soot models:

Models generated for engine events for all fuels are better than those generated using events with either low-CN or high-CN fuels.

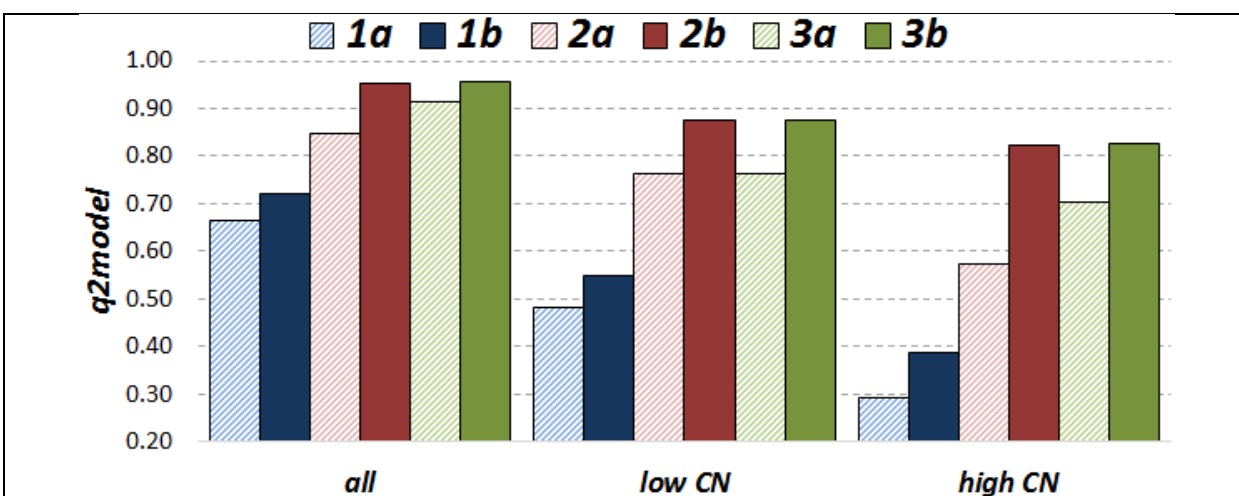


Figure: Soot models performance

Model 1a-all:

The PLS model overestimates soot content produced by engine events on FD5.

Important control parameters: Main SOI (-) and Pilot SOI (-).

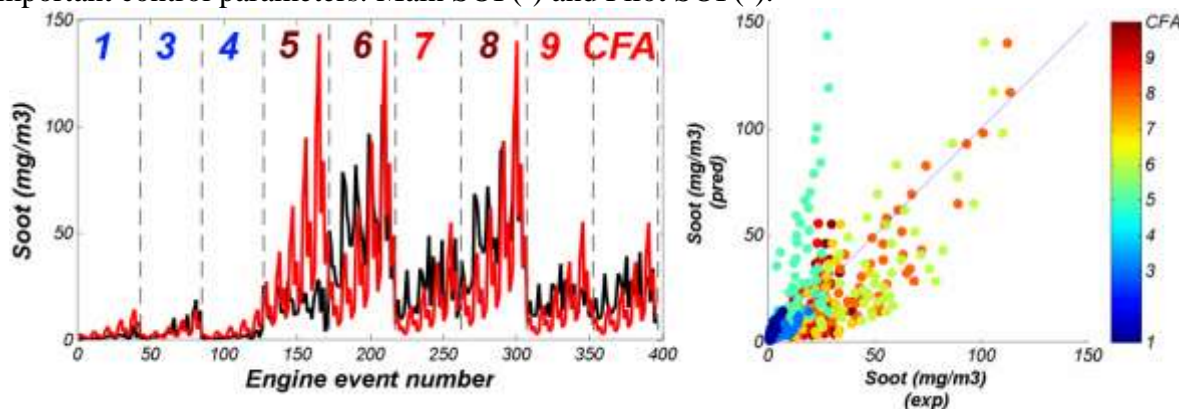


Figure: Model 1a performance

Model 2a-all:

After adding CN/T90/Aro into modeling, predictive capability of the soot model increased.

Model 2a does not produce a 'spike'-like predicted soot values, unlike model 1a. However, we can observe that some predicted soot values are overestimated (FD5) while others are underestimated (FD6).

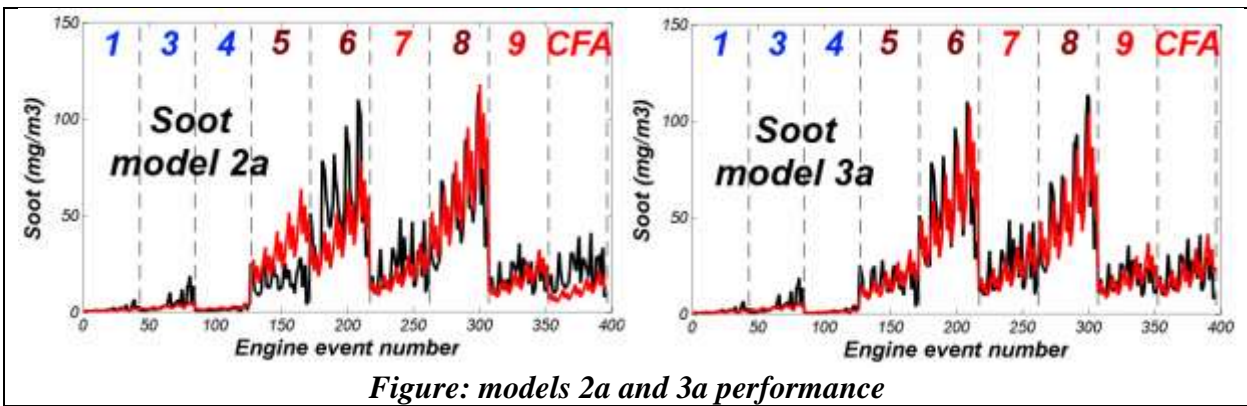
Important control parameters: Pilot SOI (-) and Main SOI (-).

Important fuel parameters: mainly CN (+), then aromatics (+).

Model 3a-all:

Including all 40 fuel descriptors further improves model 2a (see the following figure).

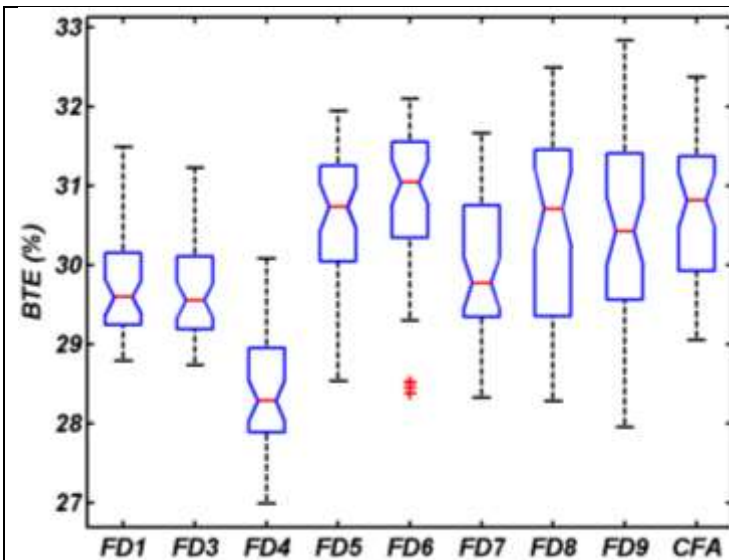
Important fuel parameters: cloud point (+), T10-T70, isoparaffins, n-paraffins.



7.4.8 INDICATED SPECIFIC FUEL CONSUMPTION

ISFC was not reported for this engine.

7.4.9 THERMAL EFFICIENCY



- Engine events on FD4 have the lowest BTE.

All BTE models:

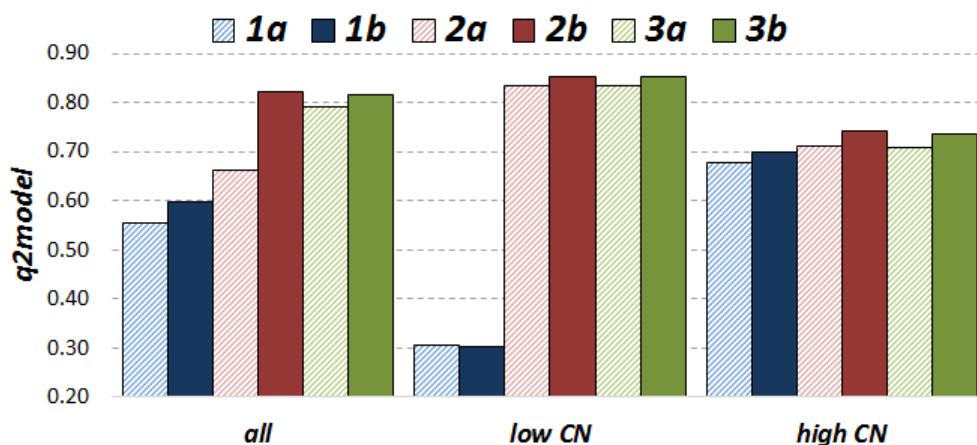


Figure: BTE models performance

BTE model 1a-all:

In general, the BTE model based on engine control parameters predicts thermal efficiency relatively well, with the exception of events run on FD4. For these runs the model predicts BTE higher than the experimental values.

Important control parameters: Pilot SOI (-) and Main SOI (+). Pilot SOI is the dominant parameter in the regression equation.

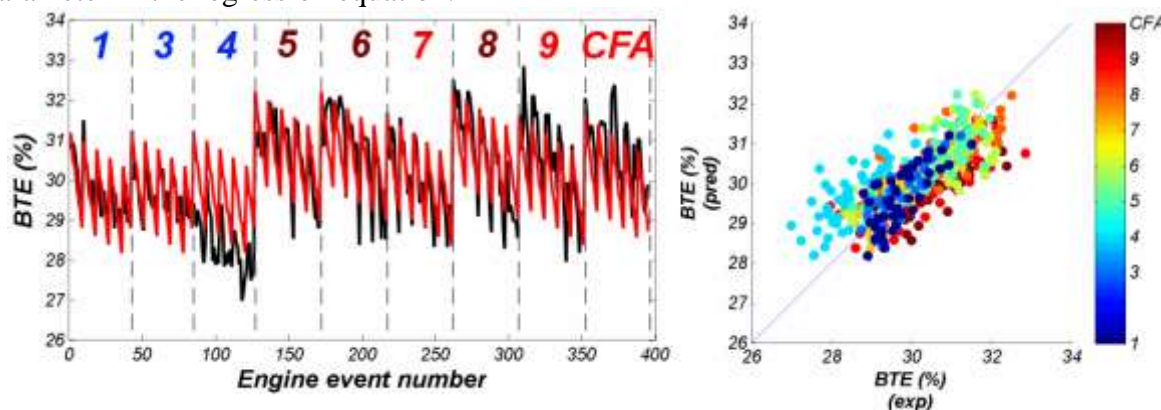


Figure: BTE model 1a performance

BTE model 2a-all:

After adding CN/T90/Aro to the BTE models, the capability of the model to predict events from FD4 were slightly improved.

Important control parameters: Pilot SOI (-) and Main SOI (+).

Important fuel parameters: CN (+) and T90 (-).

BTE model 3a-all:

By using all 40 fuel descriptors in the BTE modeling, the capability of the model to predict BTE for fuel FD4 was increased significantly. Interestingly, predictions for the other fuels did not improve.

Important control parameters: Pilot SOI (-) and Main SOI (+).

Important fuel parameters: cloud point (-), CN (+), T90 (-), CT2 (+), CT8 (-), indanes/tetralins (+).

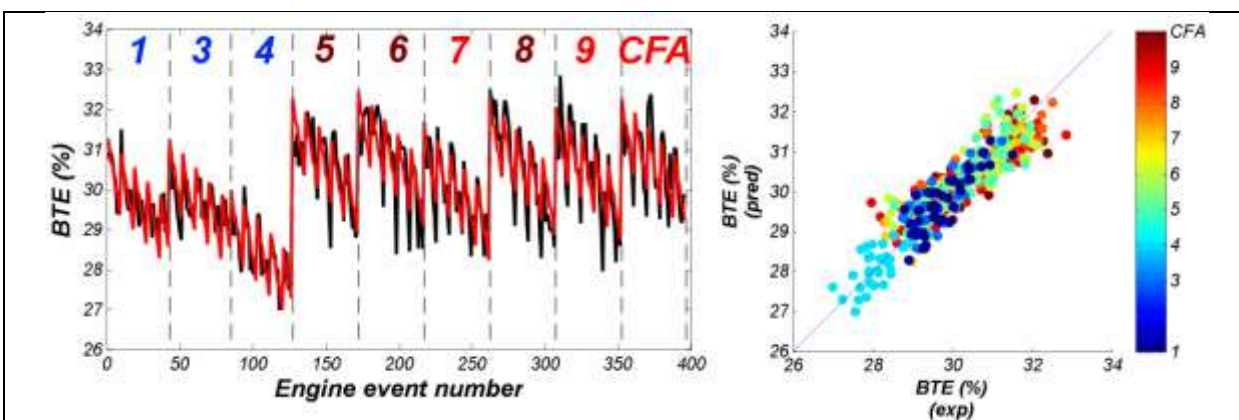


Figure: BTE model 3a performance

BTE models for low-CN fuels:

A significant improvements in the 2a/3a models performance were observed after including fuel descriptors in the modeling.

Important control parameters: Main SOI (+) and Pilot SOI (-).

Important fuel parameters: CN (+), T90 (-). Aromatics do not seem to have any significant effect on the BTE modeling.

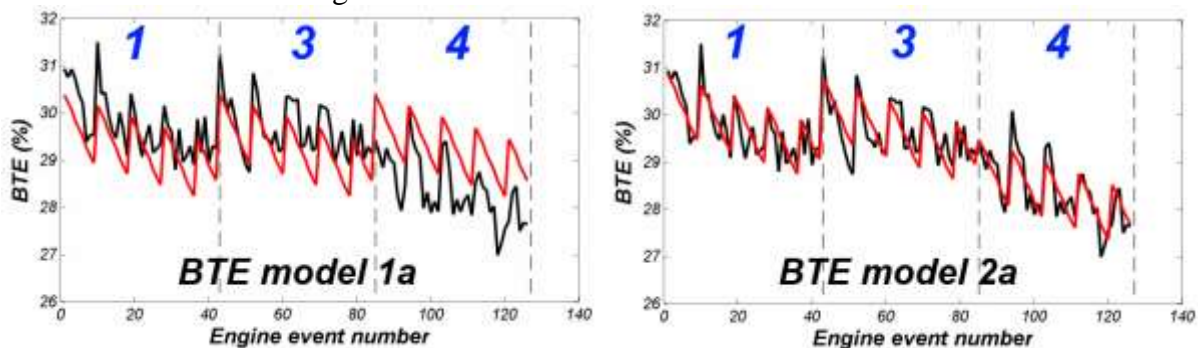


Figure: BTE models 1a and 2a performance

BTE models for high-CN fuels:

Including fuel descriptors (as well as cross terms) in BTE modeling does not improve predictive capability of resulting models. Weak effects of CN (-), T90 (+), and Aro (-).

7.4.10 SUMMARY OF PLS MODEL 2a RESULTS

A summary of the PLS models containing the 3 fuel design + engine control parameters (model 2a) for this engine is tabulated below. The filled circles (•) represent highly influential parameters and the open circles (◦) represent moderately influential parameters. Plus (+) symbols indicate parameters are directly correlated while negative (-) symbols indicate inverse correlations.

LTC Split Injection Engine (WVU-AVFL-16):

Parameter	HC		CO		NOx		Soot		BTE	
CN	●	(-)	●	(-)	●	(+)	●	(+)	●	(+)
T90		(+)		(-)		(+)		(+)		(-)
Aromatics		(-)		(-)		(+)		(+)		(-)
Main SOI (BTDC)	●	(-)	●	(-)	●	(+)	●	(-)	●	(+)
PilotSOI (BTDC)	●	(+)	●	(+)	●	(-)	●	(-)	●	(-)
Fuel Split(%Pilot)		(+)		(-)		(+)		(+)		(-)

● - highly influential parameter

○ - moderately influential parameter

(+) - directly correlated

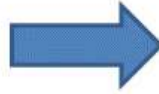
(-) - inversely correlated

7.5 SANDIA HEAVY-DUTY OPTICAL ENGINE

An experimental study was conducted at Sandia National Laboratories (3) to investigate the effects of fuel properties on mixing-controlled combustion. Four FACE diesel fuels (FD1, FD2, FD6, and FD8) and a ULSD#2 certification fuel were run in a single-cylinder heavy-duty optical compression-ignition (CI) engine. The engine experiments were performed using two injection pressures, three dilution levels, and constant start-of-combustion (SOC) timing (3). Considering that each fuel was run under two different injection pressures and three dilution levels, with at least three replicates at each condition, there were about 18 events/measurement points for each fuel. This sums up to 96 measurement points in total for all fuels in this dataset.

The engine dataset consists of 238 input and output parameters measured for all five diesel fuels. Identification of input/output parameters, detection of missing or flawed values, and removal of constant or semi-constant variables was the most time-consuming aspect of the preprocessing step. As a result, four sets of parameters in the Sandia engine dataset were identified: fuel descriptors (11 parameters), constant values (34 parameters), engine inputs (55 parameters), and engine outputs (138 parameters). The list of parameters is long and is not presented in this section. The complete list of parameters collected during running Sandia heavy-duty optical engine experiments is presented in Appendix I. Figure 46 shows examples of variables in each of the four aforementioned groups. In the final step both fuel descriptors and constant values were removed and not considered during exploratory and modeling/trending analysis.

The original spreadsheet contains 238 columns. We split these columns into four groups.



	Fuel descriptors
	Engine input variables
	Engine output variables
	Constant values

data acquisition date [YYMMDD]		indicated specific NOx emissions [g/kW-hr]	
text descriptor [-]		indicated specific HC emissions [g/kW-hr]	
numerical descriptor [-]		indicated specific CO emissions [g/kW-hr]	
process data (0=no, ~0=yes) [-]	total mass flow rate of charge-gas [g/s]	raw LII voltage (i.e., raw PMT signal) before firing [V]	
order number for plotting [-]	mass flow rate of air [g/s]	exhaust LII signal adjusted for ND and PMT gain [a.u.]	
run number [-]	setpoint for air pressure upstream of orifice [bar]	gross IMEP motored cycles (during compr'n and exp'n)	
fuel/calibration number [-]	actual air pressure upstream of orifice [bar]	coeff. of variation of g/IMEPm [%]	
cycle number [-]	air temperature upstream of orifice [°C]	gross indicated work per motored cycle [J]	
engine speed [rpm]	air pressure downstream of orifice [bar]	carbon density from avl smoke meter [mg/m³]	
0 = motored, 1 = fired [-]	mass flow rate of N2 [g/s]	emission index NOx emissions [g/kg_fuel]	
	setpoint for N2 pressure upstream of orifice [bar]		piston number [-]
	actual N2 pressure upstream of orifice [bar]		cylinder pxd calibration rev. [-]
	N2 temperature upstream of orifice [°C]		intake pxd calibration rev. [-]
	N2 pressure downstream of orifice [bar]		exhaust pxd calibration rev. [-]
	mass flow rate of CO2 [g/s]		cyl. pxd temperature (deg t)
			fourier filter start frequency [Hz]
			fourier filter end frequency [Hz]

Figure 46 – The Sandia engine dataset cleaning process results in classification of all 238 parameters into four groups of variables. Examples of parameters classified in each set of variables are presented in colour-coded boxes.

7.5.1 EXPLORATORY ANALYSIS

Engine Input Space

The engine experiments were designed to have only two operating conditions changed: injection pressure (80 MPa and 180 MPa) and intake oxygen mole fraction (16%, 18%, and 21%). The resulting engine design experimental space is presented in Figure 47, which represents an idealized case taking into consideration these two variables.

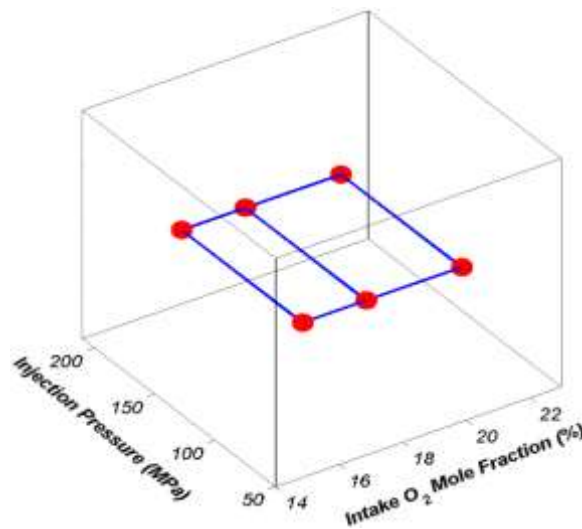


Figure 47 – Sandia experimental design. Six measurement points for each of five fuels resulted in 30 test conditions.

However, data provided from Sandia and parameters identified as input variables present a slightly different distribution of points in the multivariate space. The number of parameters exceeds three, so in order to visualize the engine input space, the PCA was performed to reduce the number of variables and visualize the high-dimensional data. Because the engine input parameters were measured in different units, the data were auto-scaled to unit variance before running PCA. The PCA results are presented in Figure 48 and Figure 49.

Engine events are distributed along PC1 according to parameters correlated with oxygen dilution level (see Figure 48c). PC2 splits the objects based on parameters related to injection pressure (see Figure 48c). In Figure 48d (along PC3), a very characteristic distribution of events according to fuel type is evident (looking from the bottom to the top: FD2 alone, then FD1/FD6 together, and finally pairs of CF/FD8).

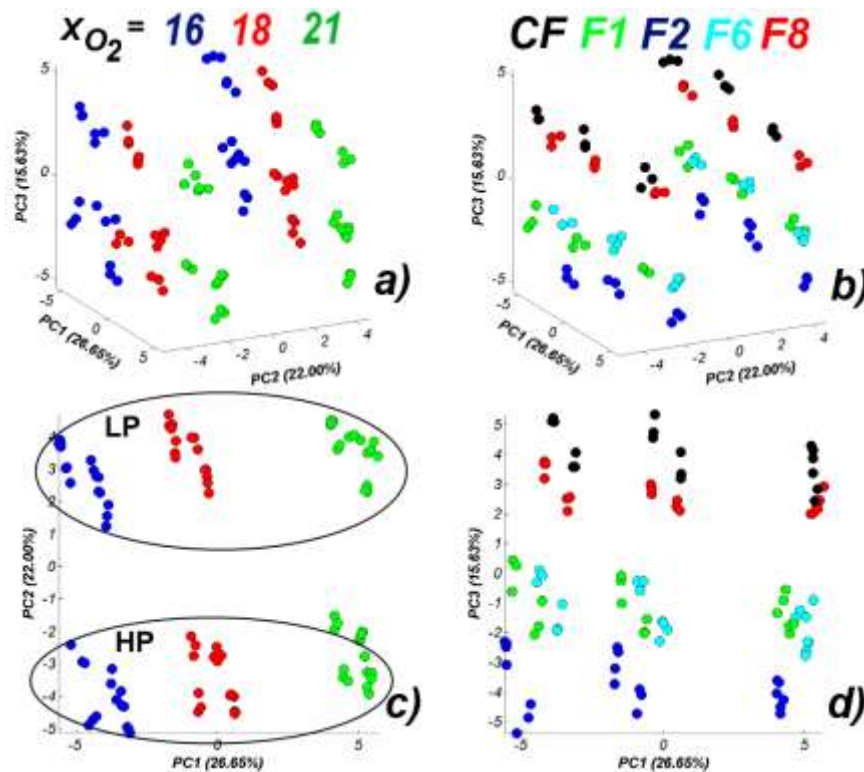


Figure 48 – The PCA scores plots (PC1-PC2-PC3) obtained after running PCA for Sandia engine input (control) parameters. The points on the dataset are coloured according to intake oxygen mole fraction (X_{O_2}) (a) and fuel type (b). The PCA scores plots (PC1-PC2) (c) present two main groups of points: low (LP) and high (HP) injection pressure. The PCA scores plots (PC1-PC3) (d) groups engine inputs based on fuel type and oxygen dilution.

We found that the first two principal components (PC1 and PC2) explain more than 45% of the data variance, which was sufficient to reproduce the target experimental engine conditions presented in Figure 47. The next two PCs (PC3-15.6% and PC4-11.5%) comprise an additional ~27% variance, which significantly improves the predictive ability of the PCA model. Figure 49 presents PCA scores plots of higher-order components (PC3-PC4), which groups the engine input events based on individual fuels. This means that each fuel was injected into the engine under slightly different conditions, which could affect combustion efficiency and engine emissions.

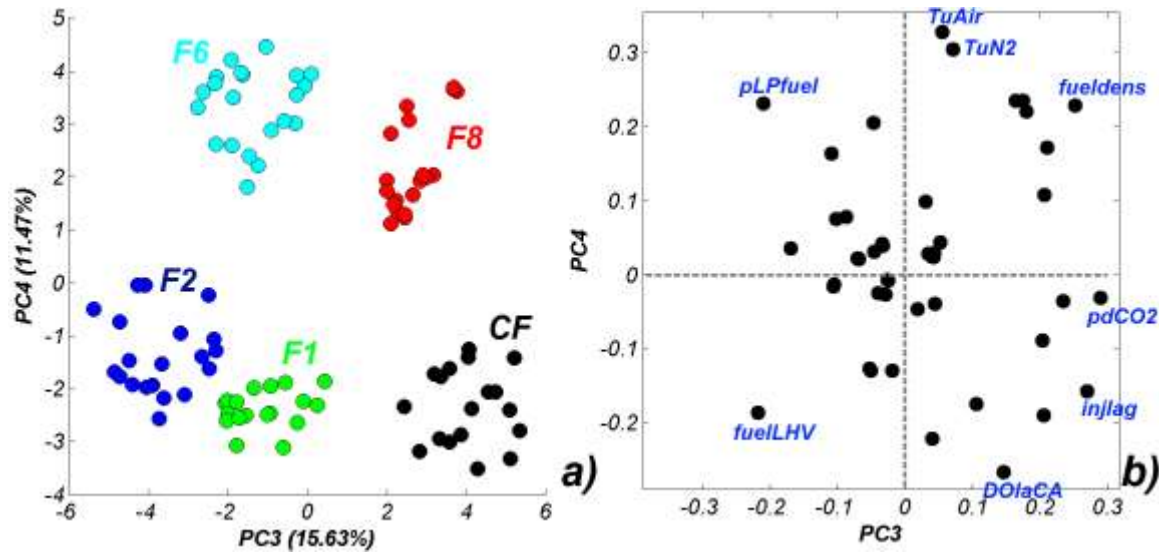


Figure 49 – PCA scores (a) and loadings plots (b) (PC3-PC4) of the Sandia engine input dataset, where pLPfuel=low-pressure fuel pressure [psig], TuAir/N2=air/nitrogen temperature upstream of orifice [°C], fueldens=fuel density [kg/m³], pdCO₂=CO₂ pressure downstream of orifice [bar], injlag=injector lag [s], DOIaCA=actual duration of injection [CAD], fuelLHV=lower heating value of fuel [MJ/kg].

The scores of both PC3 and PC4 are accompanied by the corresponding loadings. A scatter plot of the loadings of the fourth component vs the loadings of the third component (PC3-PC4) is shown in Figure 49b. This plot indicates the main engine parameters responsible for the grouping of the fuel events in the corresponding scores plots (Figure 49a).

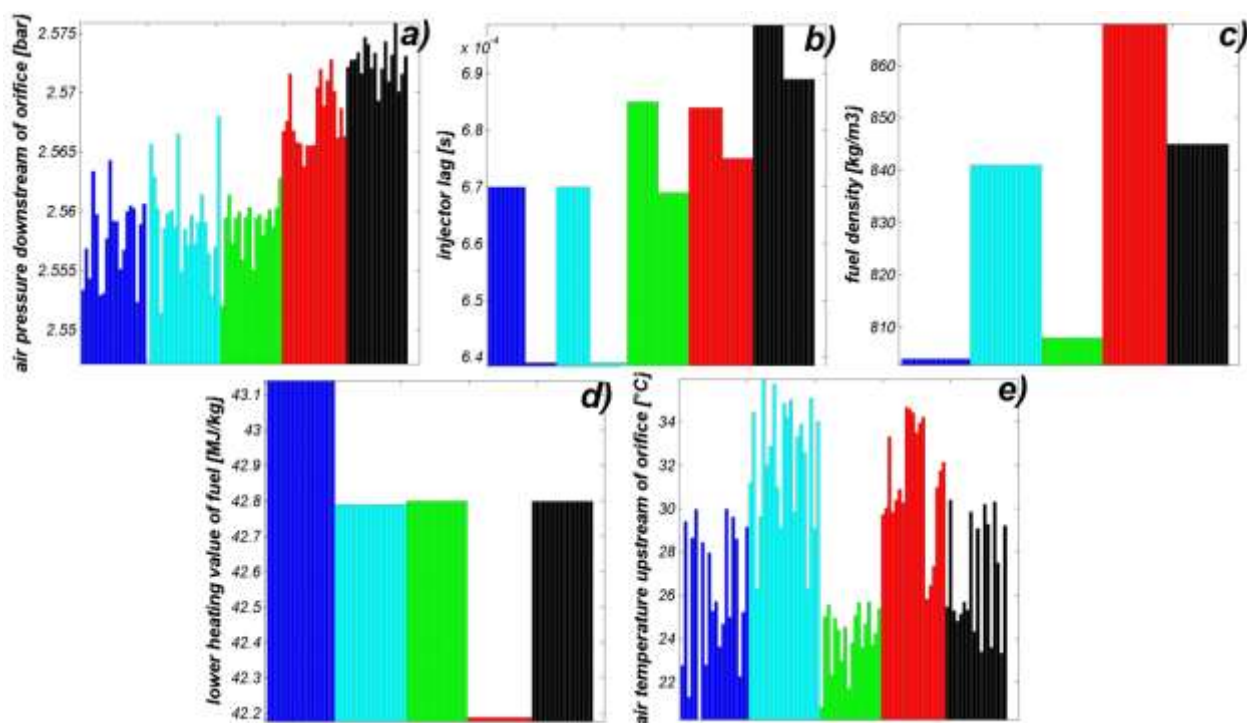


Figure 50 – Bar plots of parameters having the greatest influence on the observed grouping of fuels in the PCA scores plot (PC3-PC4).

Engine Output Space

The Sandia optical engine has the capability of providing a huge amount of data during a single experimental run. The set of variables selected as an engine output dataset consisted of 138 variables for which PCA projection resulted in the engine event distribution in the scores plots as presented in Figure 51.

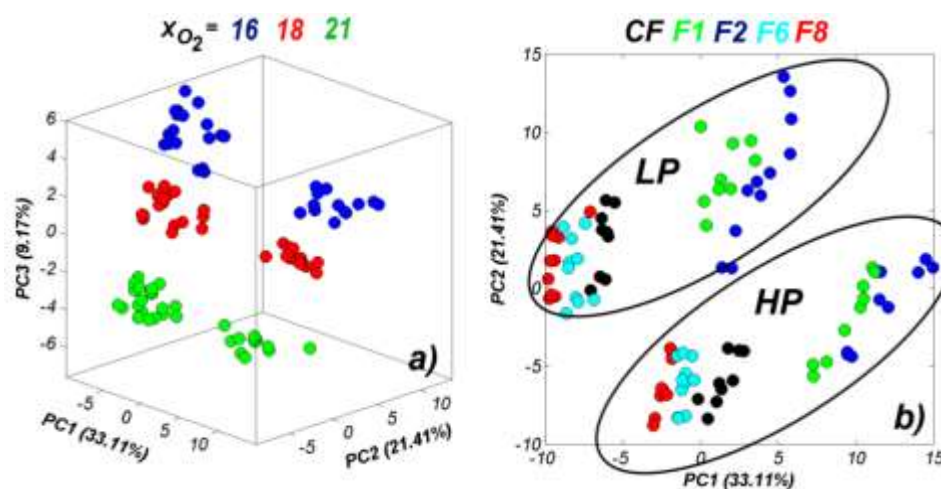


Figure 51 – PCA scores plots (PC1-PC2-PC3) obtained after running PCA for Sandia engine output parameters, coloured according to intake oxygen mole fraction (X_{O_2}) (a) and PCA scores plots (PC1-PC2), coloured according to fuel type (b). HP and LP denote high and low injector pressures, respectively.

7.5.2 MODELING APPROACH

Simple one-to-one correlations between engine control parameters and selected engine output parameters are presented in Figure 52.

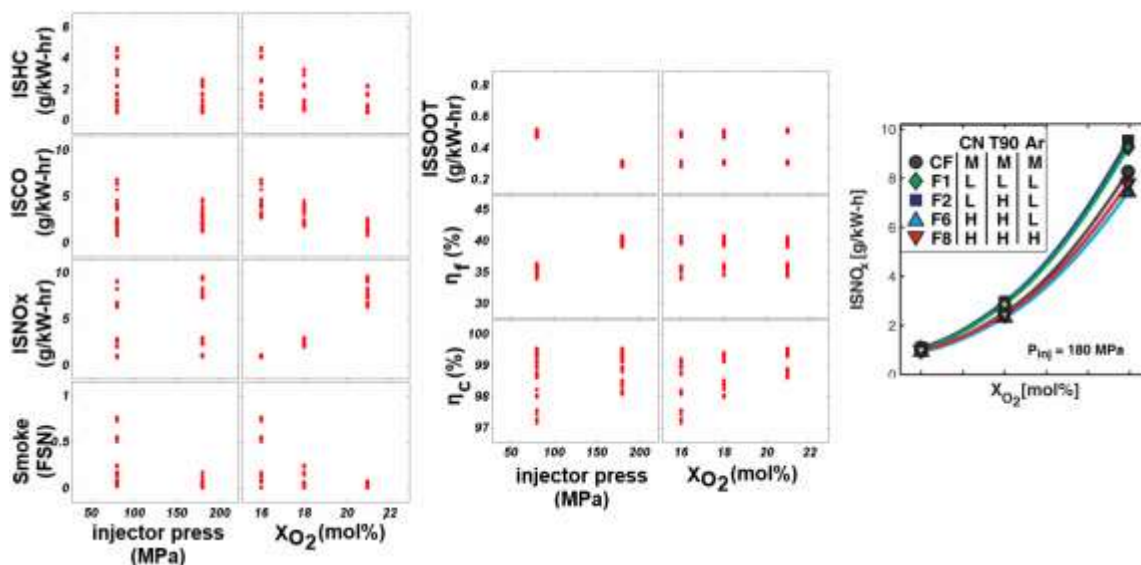


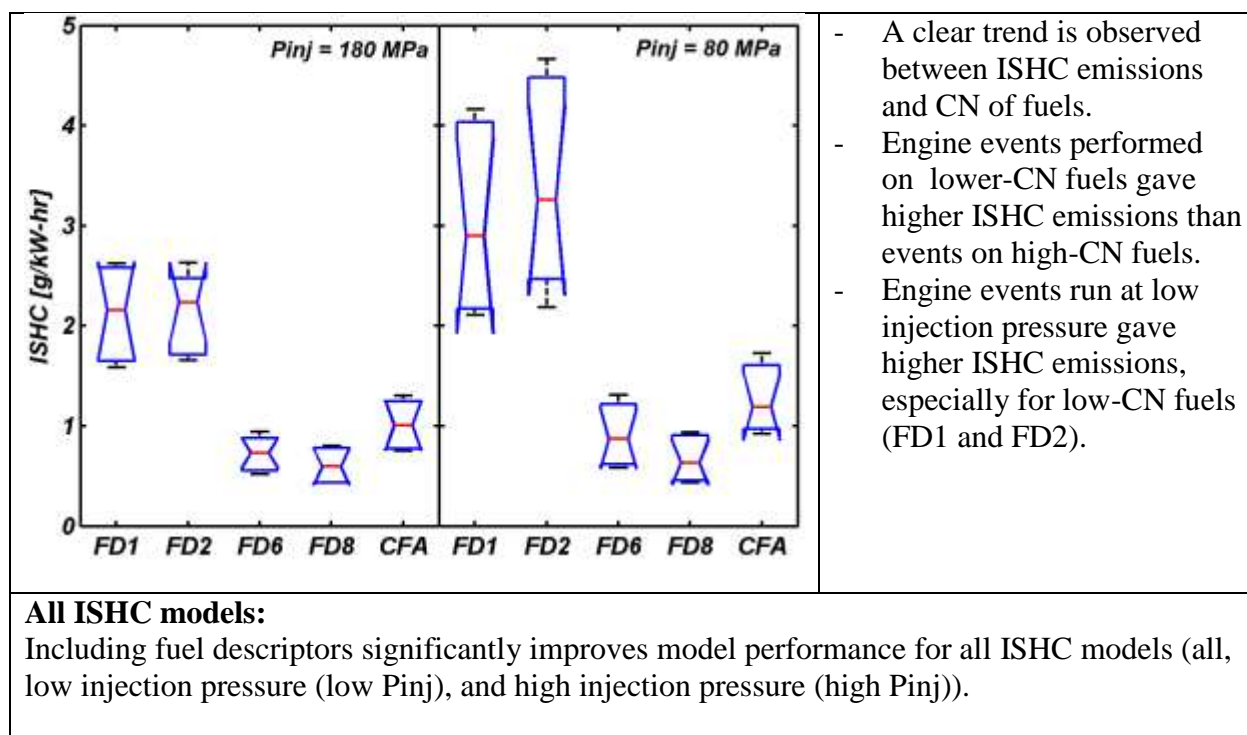
Figure 52 – Scatter representation of cross-correlation between output and experimentally designed engine input parameters obtained for Sandia engine runs. On the right is an example of the relationship between charge-dilution and indicated specific NOx emissions for engine events run with high injection-pressure.

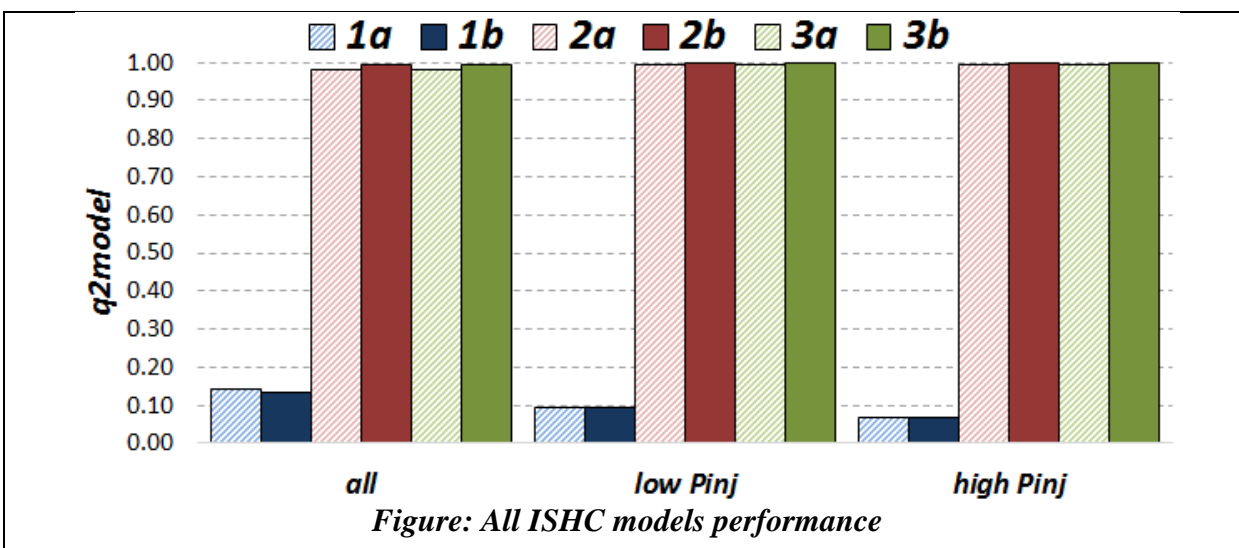
In the following section, the results of modeling emission and efficiency parameters will be presented. The engine was modeled using strategies described in detail in section 6:

- (strategy: 1a): Injector rail pressure and oxygen mole fraction,
- (strategy: 1b): strategy 1a + cross terms for all variables,
- (strategy: 2a): strategy 1a + FACE fuel CN/T90/aromatic content,
- (strategy: 2b): strategy 2a + cross terms for all variables,
- (strategy: 3a): strategy 1a + 40 FACE fuel descriptors (see section 6),
- (strategy: 3b): strategy 3a + cross terms for all variables.

PLS models were built for: 1) all events; 2) events run at low injector rail pressure; and finally 3) events performed at high injector rail pressure. In the studies presented in this section, only two input engine control parameters were used during modelling: the injector rail pressure and the oxygen mole fraction (a.k.a. charge-dilution), respectively. These two parameters were the primary design engine variables engaged during experiments performed at Sandia National Laboratories (3). For each modeled property, 18 PLS models were calculated and a summary of the results is presented in the following sections (7.5.3 – 7.5.9).

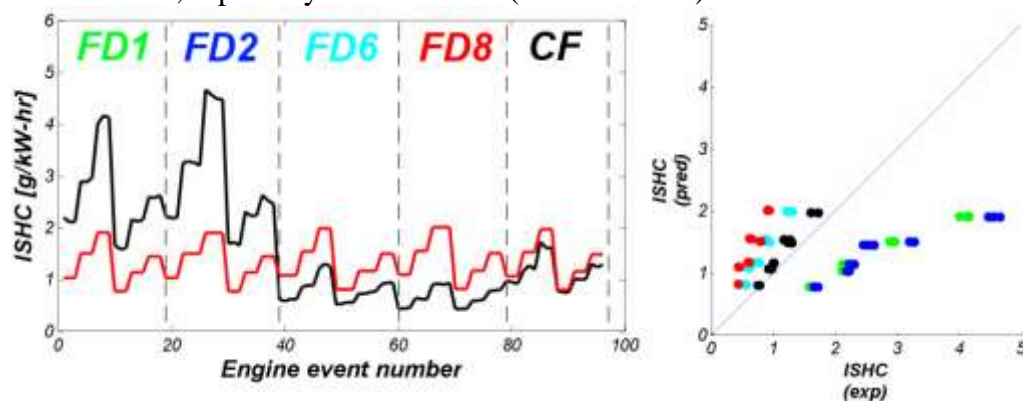
7.5.3 INDICATED SPECIFIC HYDROCARBON EMISSIONS (ISHC)





HC model 1a-all:

In general the PLS model generated for ISHC emissions is very weak. The following profiles of experimental (black) and predicted (red) ISHC emissions shows an inability to properly model HC emissions, especially low-CN fuels (FD1 and FD2).



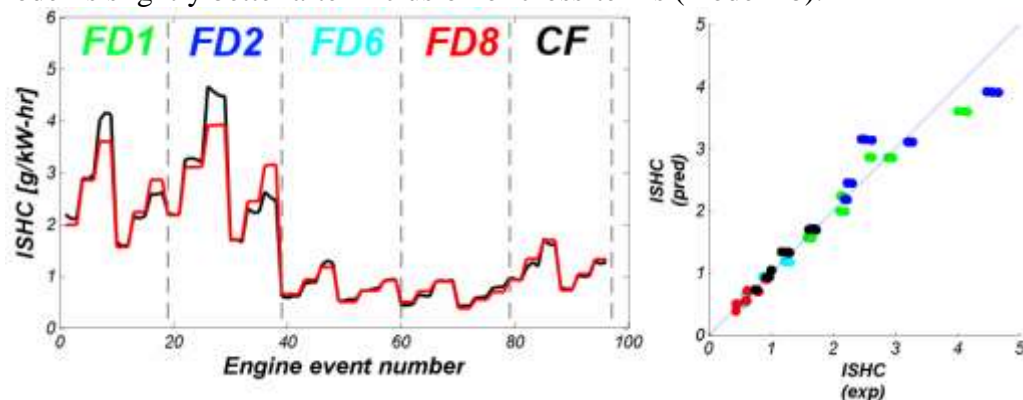
HC model 2a-all:

Inclusion of the three fuel design descriptors in ISHC modeling significantly improves modeling capability. Cetane number is the most important parameter out of the three used in modeling.

Important control parameters: O₂ mole fraction (-)

Important fuel parameters: CN (-) then Aro (-)

ISHC model is slightly better after inclusion of cross-terms (model 2b).



*Figure: ISHC model 2a performance***ISHC model 3a-all:**

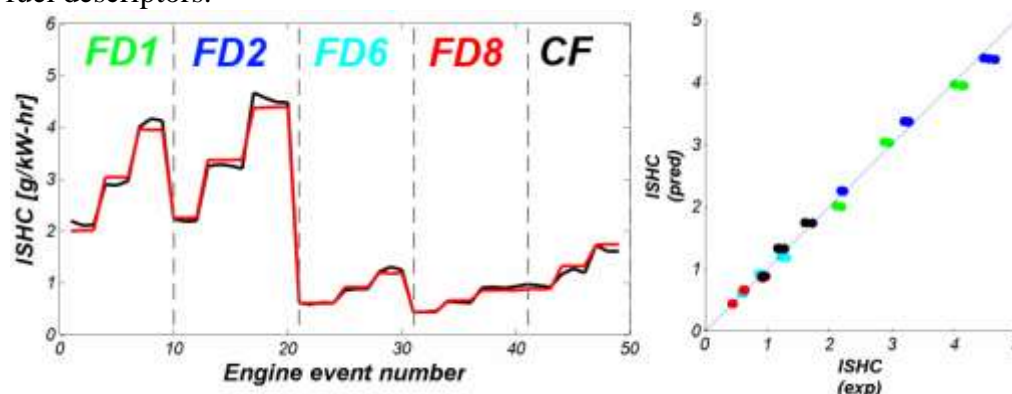
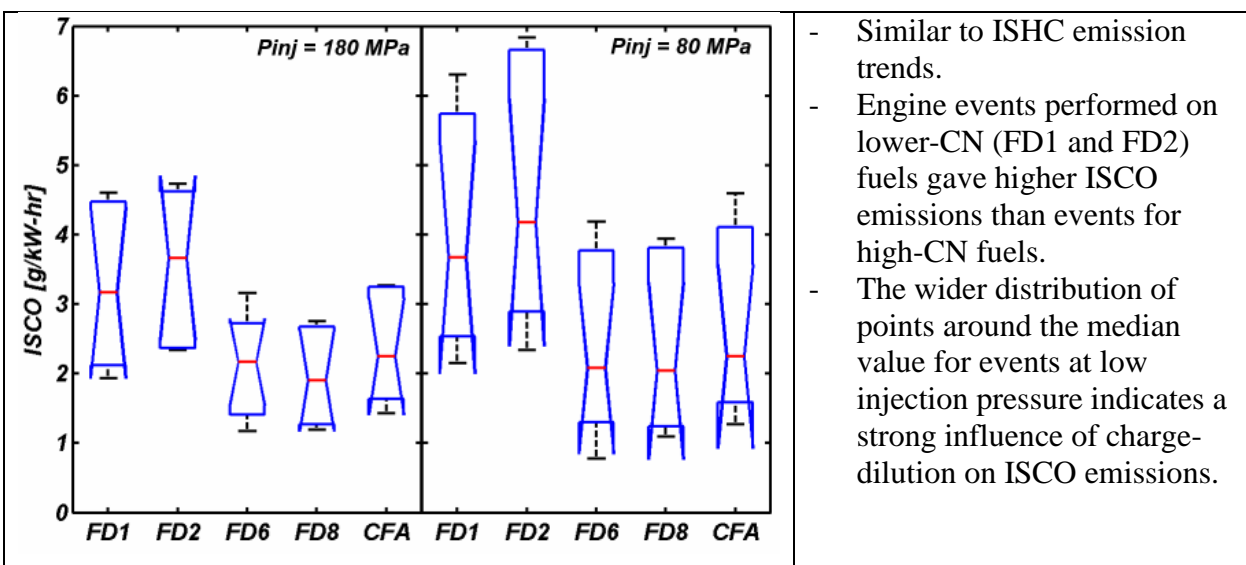
Similar in performance to model 2a

Important control parameters: O₂ mole fraction (-)

Important fuel parameters: CN (-), CT2 (-), T50-T80 (-), viscosity (-)

ISHC models for low/high pressure injection events

Models based on only engine control parameters are poor and improve significantly after adding fuel descriptors.

*Figure: ISHC model 2a performance for low pressure injection events***7.5.4 INDICATED SPECIFIC CARBON MONOXIDE EMISSIONS (ISCO)**

- Similar to ISHC emission trends.
- Engine events performed on lower-CN (FD1 and FD2) fuels gave higher ISCO emissions than events for high-CN fuels.
- The wider distribution of points around the median value for events at low injection pressure indicates a strong influence of charge-dilution on ISCO emissions.

All ISCO models:

Models based on only two engine parameters are relatively good. Including fuel descriptors significantly improves ISCO model performance for all models (all, low injection pressure (low Pinj), and high injection pressure (high Pinj)).

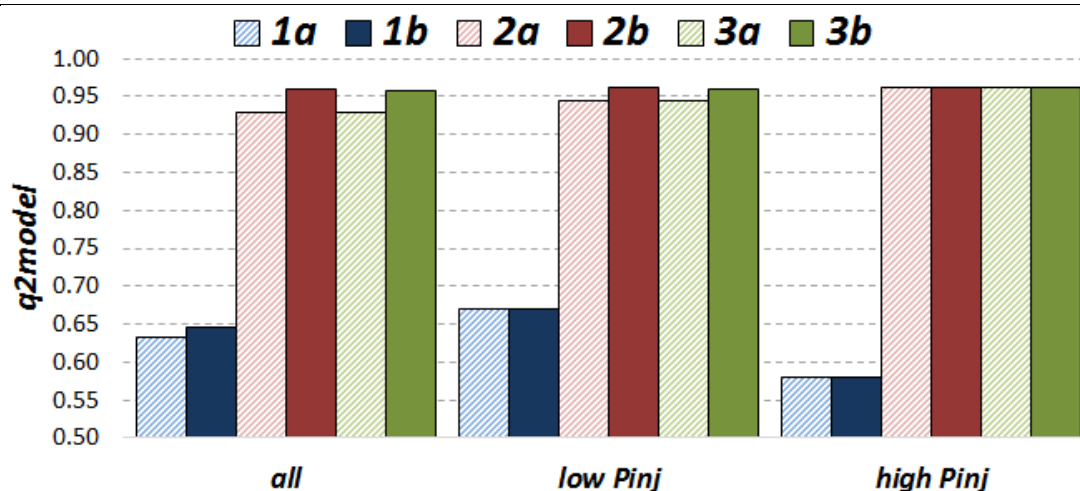


Figure: All ISCO Models performance

ISCO model 1a-all:

In general, the PLS model overestimates the ISCO values for high-CN fuels and underestimates the ISCO values for low-CN fuels.

Important control parameters: O₂ mole fraction (-)

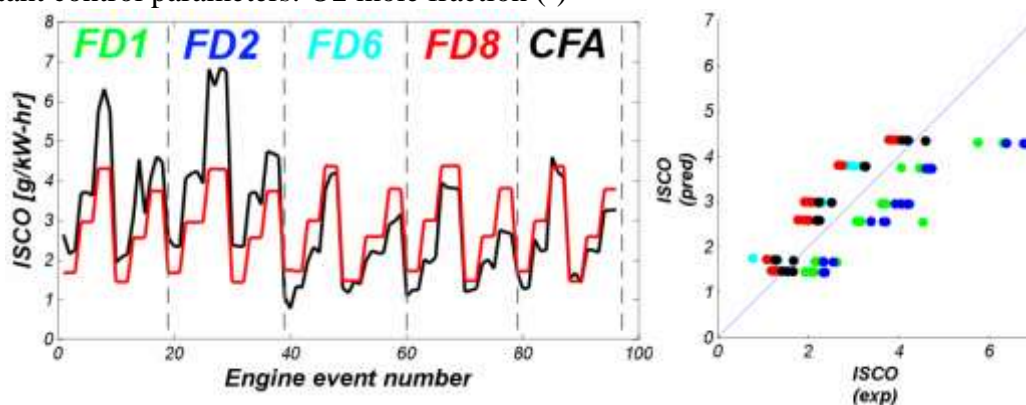


Figure: ISCO model 1a performance

ISCO model 2a-all:

Prediction of ISCO emissions improves after including fuel information in the modeling.

Important control parameters: O₂ mole fraction (-)

Important fuel parameters: CN (-)

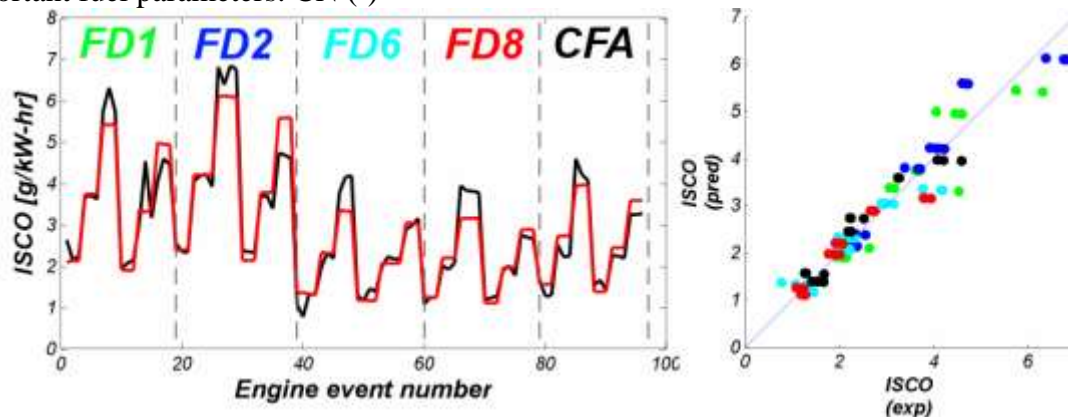


Figure: ISCO model 2a performance

ISCO model 3a-all:

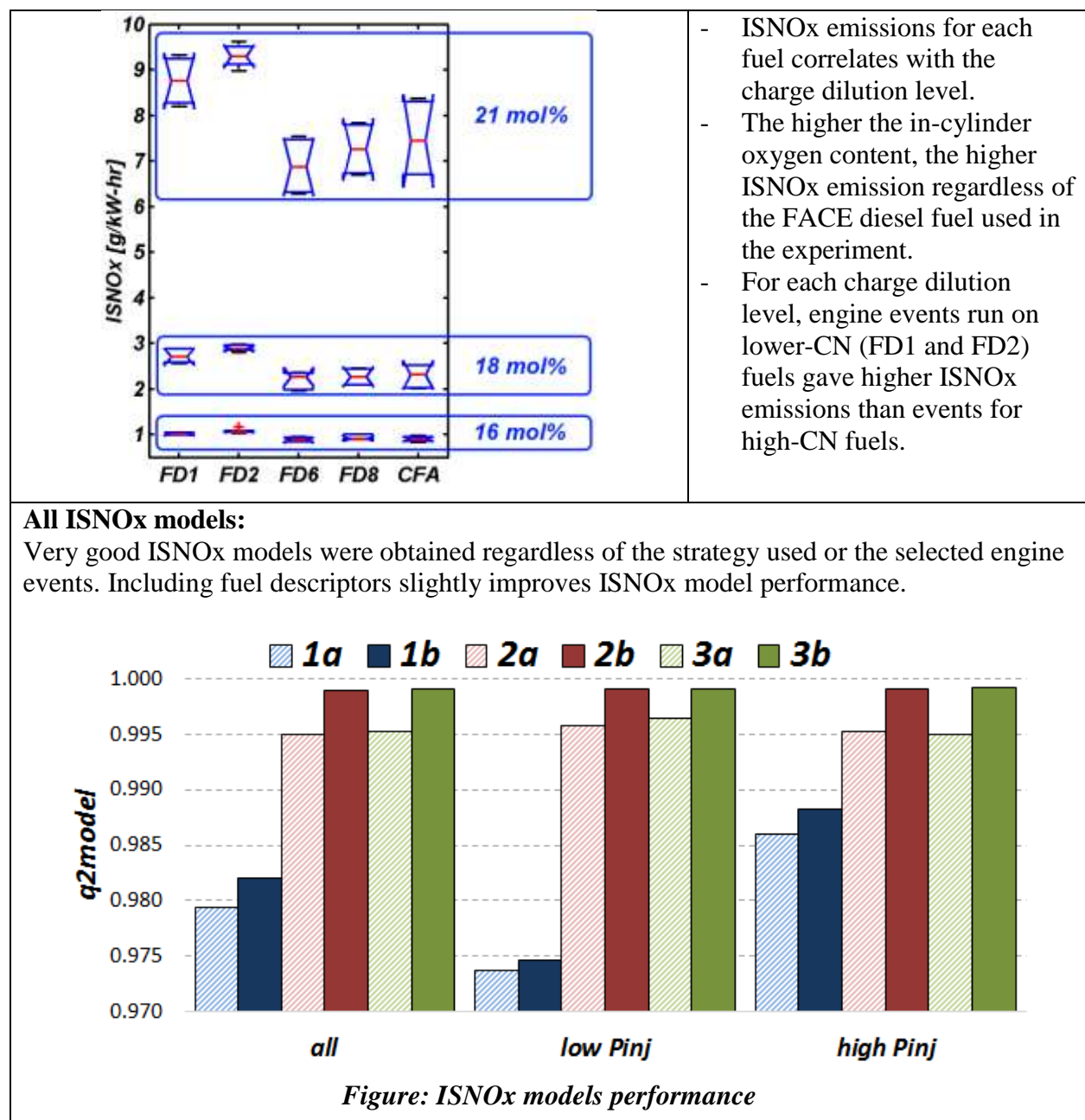
Similar in performance to model 2a

Important control parameters: O₂ mole fraction (-)

Important fuel parameters: CN (-), CT8 (+), T50-T80 (-), Alkylbenzenes (+)

ISCO models for low/high pressure injection events

Models based on only engine control parameters are poor and improve significantly after adding fuel descriptors (model 2a). For these models, CN is the most important parameter in ISCO emission modeling.

7.5.5 INDICATED SPECIFIC NITROGEN OXIDES EMISSIONS (ISNOX)

ISNOx model 1a-all:

The model built for ISNOx emissions shows excellent capabilities for model ISNOx behaviour. In fact, there is a strong correlation between O₂ mole fraction (-) and ISNOx emissions (see Figure 52).

Important control parameters: O₂ mole fraction (+)

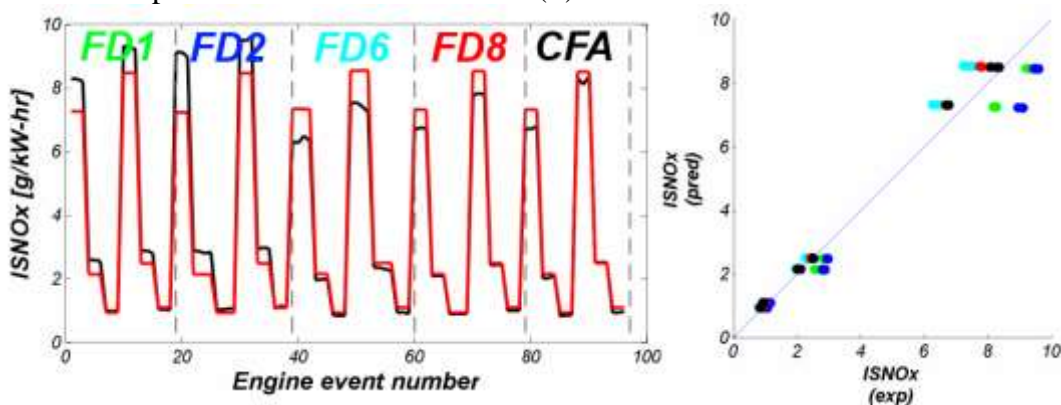


Figure: ISNOx model 1a performance

ISNOx model 2a-all:

The model built for ISNOx emission shows excellent capability for model NOx behaviour. CN/T90/Aro further refined predictability of ISNOx model 1a.

Important control parameters: O₂ mole fraction (+)

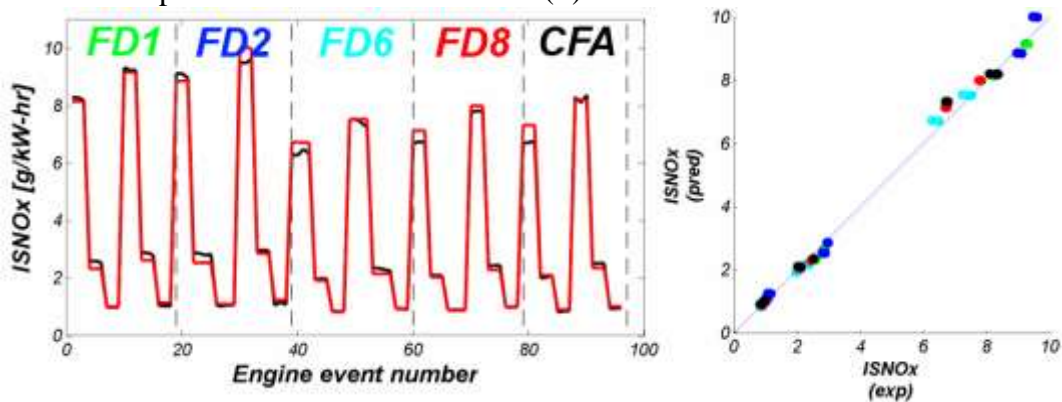
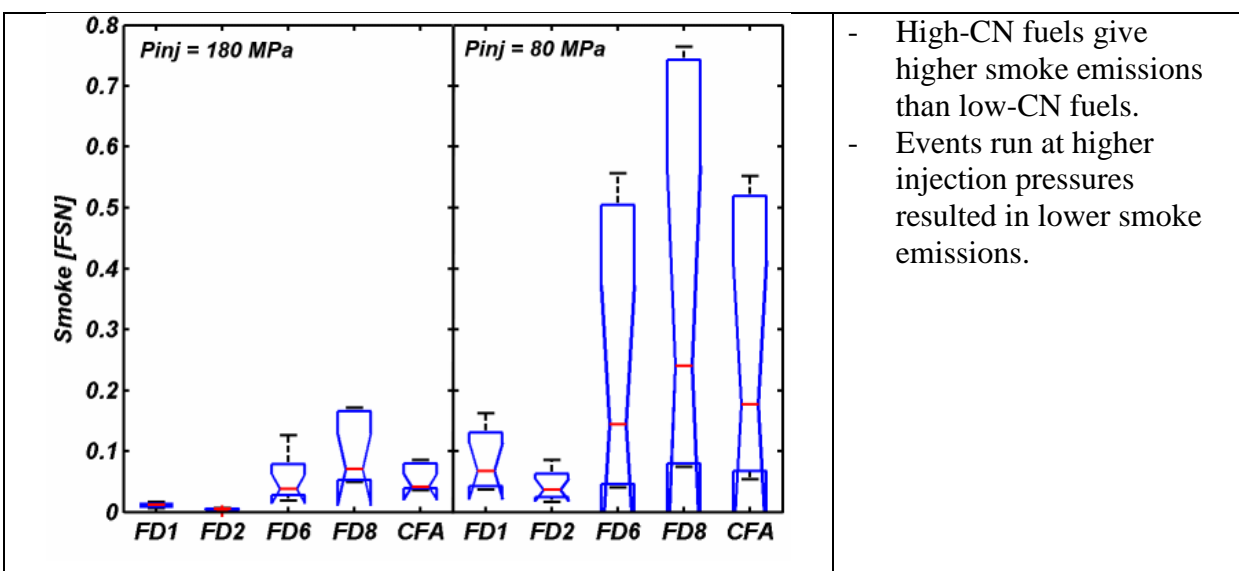


Figure: ISNOx model 2a performance

7.5.6 SMOKE EMISSIONS



All smoke models:

Including fuel descriptors significantly improves smoke model performance for all models, regardless of the choice of the events used in model development (i.e. all, low injection pressure (low Pinj), and high injection pressure (high Pinj)). On the contrary to smoke models (1a) developed on all events or the low Pinj events, events run at high injection pressure are useless.

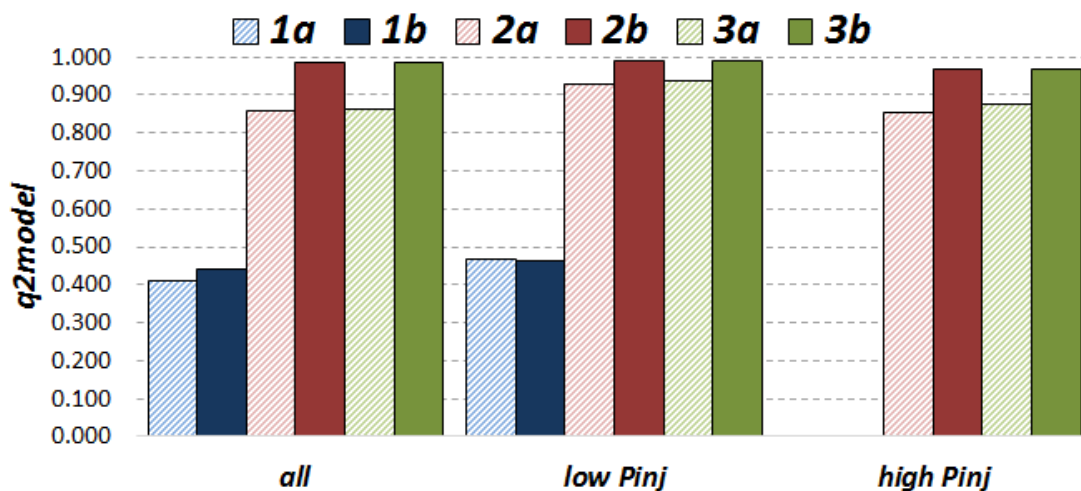
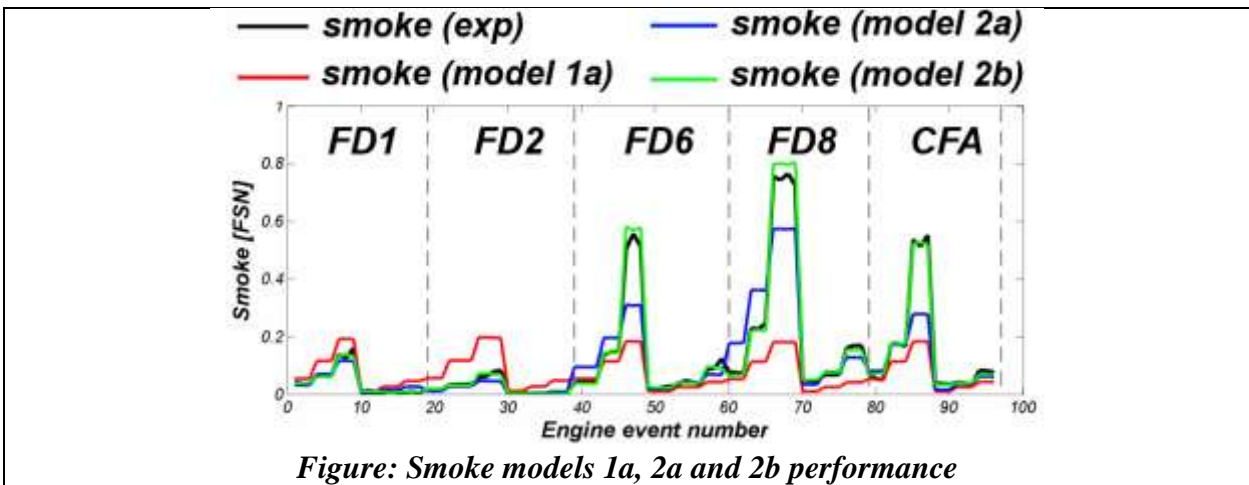


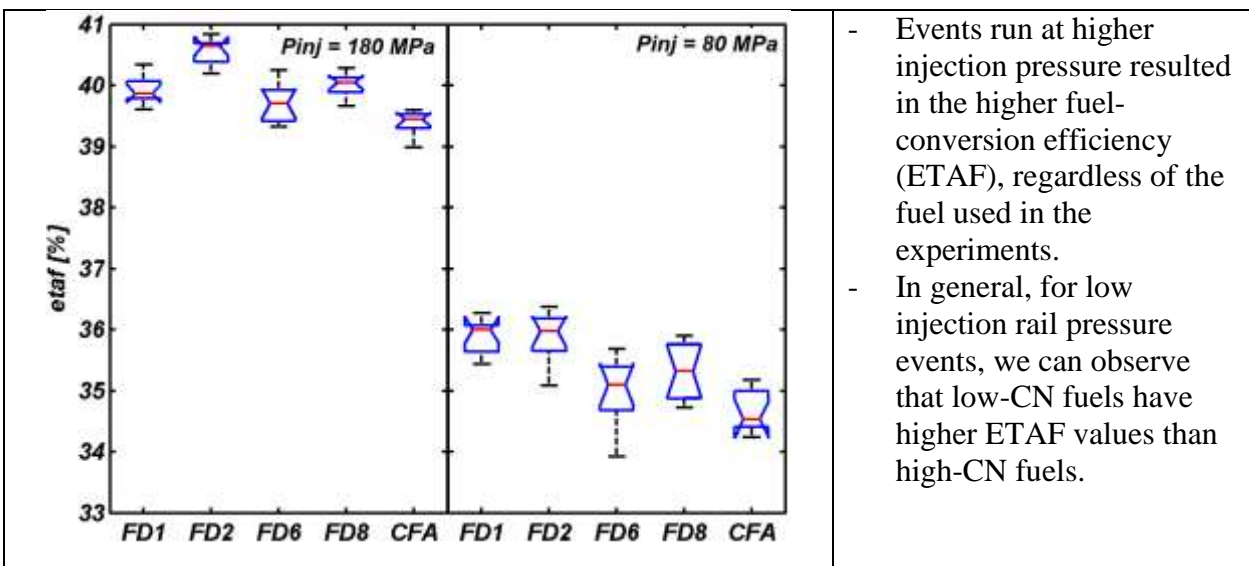
Figure: Smoke emissions models performance

Models all:

The PLS model 1a underestimates smoke emissions for high-CN fuels and overestimates smoke emissions for low-CN fuels. After inclusion of the three fuel design descriptors in smoke emission modeling (model 2a), we observed significant improvement in modeling capability. Model 3a further improves prediction ability for events with high smoke values. Important control parameters: Injector rail pressure (-) and oxygen mole fraction (-). Important fuel parameters: CN (+), aromatics (+).



7.5.7 INDICATED FUEL-CONVERSION EFFICIENCY (ETAF)



All ETAF models:

Using all 96 engine events in PLS model development resulted in the collection of the best ETAF models. ETAF models built for the other two cases (low or high injection rail pressure events) are characterized by lower values of the q^2_{model} parameter.

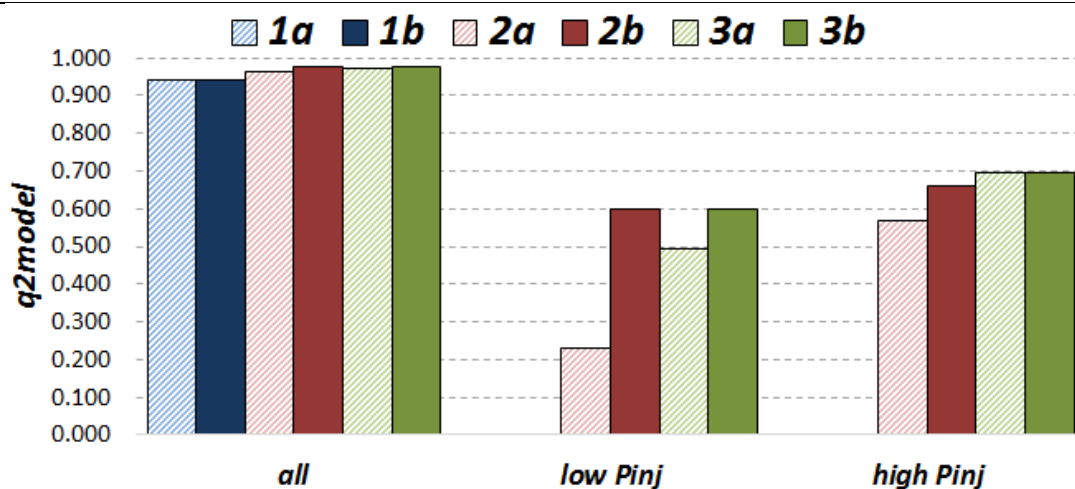


Figure: ETAF models performance

Models 1a-3b all:

The PLS model based on two engine control parameters showed a good capability for modeling indicated fuel-conversion efficiency. Models are characterized by high q^2_{model} parameter values, regardless of used strategy (1a – 3b). Including fuel descriptors in ETAF modeling allows for slight improvement of the modeling ability of specified parameters.

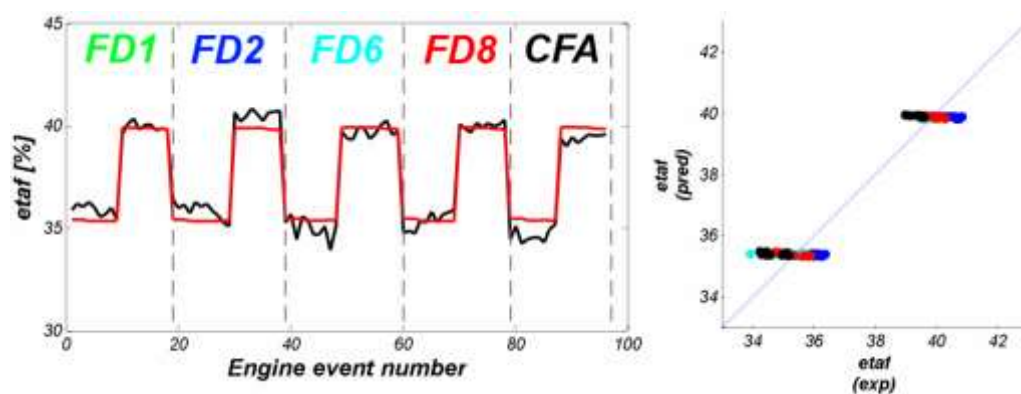


Figure: ETAF model 1a performance - fuel descriptors not included in modeling

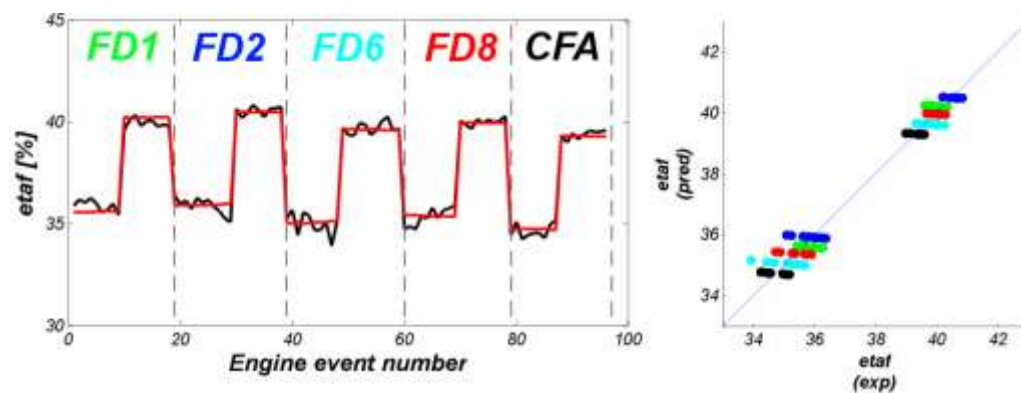


Figure: ETAF model 3a performance – All 40 fuel descriptors included in modeling

7.5.8 SUMMARY OF PLS MODEL 2a RESULTS

A summary of the PLS models containing the 3 fuel design + engine control parameters (model 2a) for this engine is tabulated below. The filled circles (●) represent highly influential parameters and the open circles (○) represent moderately influential parameters. Plus (+) symbols indicate parameters are directly correlated while negative (-) symbols indicate inverse correlations.

Sandia Heavy Duty Optical Engine:

Parameter	HC		CO		NOx		FSN		ETAF		ETAC	
CN	●	(-)	●	(-)		(-)	●	(+)		(-)	●	(+)
T90		(-)		(+)		(+)		(-)		(+)		(-)
Aromatics	○	(-)		(-)		(+)	○	(+)		(+)	○	(+)
inj rail P		(-)		(-)		(+)	●	(-)	●	(+)		(+)
O2 mole fraction		(-)	●	(-)	●	(+)	○	(-)		(-)	○	(+)

● - highly influential parameter

○ - moderately influential parameter

(+) - directly correlated

(-) - inversely correlated

8.0 SUMMARY AND CONCLUSIONS

The investigation described in this report had two main goals. The first goal was to perform ‘data mining’ and statistical analysis of the physical and chemical properties of the nine FACE diesel fuels. These fuels have previously been characterized by a considerable amount of both standard and advanced analytical approaches over the years by different research labs around the world. The following key points summarize what was done in the current study:

- Up to 28 datasets including both physical and compositional analyses were collected and summarized for this study, including spectroscopic and thermophysical data. All the available datasets for the FACE diesel fuels have been formatted and saved in Excel format and MATLAB structures. These files are ready for use in future projects.
- Each dataset was subjected to principal component analysis (PCA), which helped in the discovery of inner structures present in particular datasets, including similarities among fuels and correlations between variables (e.g., hydrocarbon content, physical and chemical properties, etc.).
- The RV coefficient was implemented in order to compare correlations among datasets.
- A new graphical representation of NMR-based carbon type fragments was developed, which can enhance future interpretation of this type of data.
- Correlations between physical and compositional information were evaluated using a cross-validation procedure. Generated CPRs (composition-property relationships) were validated using an external dataset of fuels. Some of the models (such as the model used to predict kinematic viscosity from SimDis data) are considered to be promising and need to be tested on larger and more representative fuel sets.

The second goal of this investigation was to identify relationships between engine output properties (i.e., NO_x, CO, HC, soot, fuel consumption, thermal efficiency) and engine control parameters, alone, as well as combined with physical and chemical properties of the FACE fuels. Based on the analyses performed, the following key points are noted:

- Principal component analysis (PCA) was used as an exploratory tool to emphasize variation and show any strong patterns existing in the engine datasets. PCA was performed on both engine control and output parameters.
- One of the major conclusions from the studies is that engine operating parameters have a considerable impact on engine output results. The selection of operating control parameters is dictated in all cases by the cetane number (CN) of the fuel. As a result, this fuel descriptor (CN) has a tremendous influence on emissions and fuel efficiency. Figure 53 underscores this finding.

IN: PCA space built on engine input parameters

OUT: PCA space built on engine output parameters (NO_x, CO, HC, Soot, ITE, ISFC, MFB50)

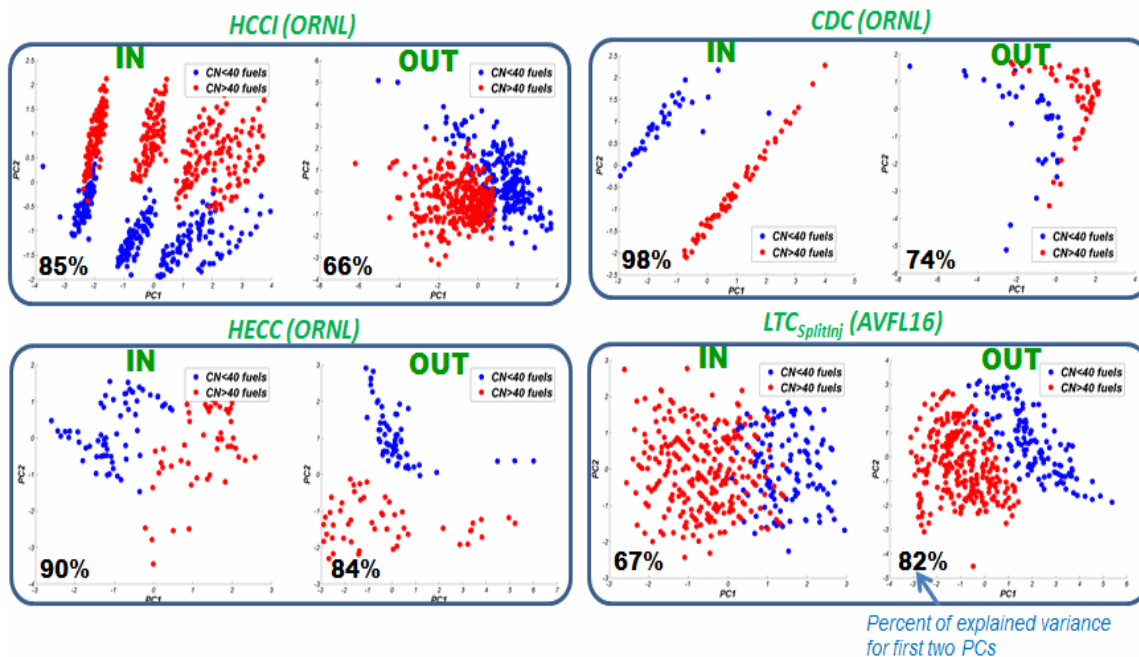


Figure 53 – PCA projection of input engine data (IN) and output engine data (OUT) into plane spanned by first two principal components. Engine events operating on low CN (CN<40) are plotted as blue points, and engine events operating on high CN (CN>40) are plotted as red points.

- Five engine datasets were evaluated and attempts to model emissions and the efficiency of these engines were undertaken. Partial least-squares (PLS) regression enhanced and extended the interpretation of the data.

- Six different PLS modeling strategies were implemented and used for emission or efficiency model development. In general, PLS models with cross terms provide the best emissions and efficiency prediction capabilities; however, they are difficult to interpret.
- The results from the statistical methods confirmed the existence of important relationships between fuel composition and engine output properties for a few cases. However, the impact of fuel chemical composition on engine performance varies from engine to engine depending on the combustion strategy employed and the engine operating conditions.
- A summary of the PLS models containing the 3 fuel design + engine control parameters (model 2a) for this engine is tabulated below. The filled circles (•) represent highly influential parameters and the open circles (◦) represent moderately influential parameters. Plus (+) symbols indicate parameters are directly correlated while negative (-) symbols indicate inverse correlations. Of the 3 FACE Diesel Fuel design parameters, CN appears to have the most influence on emissions and engine performance. More complex models containing 40 fuel parameters plus the engine control parameters (designated as models 3a and 3b) indicate that the number of secondary carbons in alkyl chains (determined by NMR) and aromatic ring carbons attached to alkyl groups also are influential parameters on emissions.

HCCI-SCE 0.5L Engine (ORNL):

Parameter	HC	CO	NOx	ISFC	ITE
CN	● (-)	● (-)	● (+)	● (+)	● (-)
T90	○ (+)	(+)	(-)	● (+)	● (-)
Aromatics	○ (-)	(-)	(-)	(+)	(-)
fuel rate	(+)	(+)	● (+)	(+)	(-)
airflow	○ (+)	(+)	(-)	● (+)	(-)
lambda	● (+)	● (+)	● (-)	(+)	(-)
intake T	● (+)	● (-)	(+)	● (+)	● (-)
intake P	● (-)	○ (-)	○ (+)	(-)	(+)

LTC MCE 1.9L Engine (HECC-ORNL):

Parameter	HC	CO	NOx	PM	ITE
CN	● (-)	● (-)	● (+)	● (+)	(-)
T90	(-)	(-)	(+)	(+)	(+)
Aromatics	(-)	(-)	(+)	(+)	(-)
aicDynoTorque	● (-)	● (-)	○ (+)	(+)	(-)
aicEGRMassGM	○ (-)	(+)	● (-)	● (+)	● (+)
aicINTO2GM	(-)	(-)	(+)	(+)	(+)
MainSOI(ATDC)	● (+)	● (+)	● (-)	● (+)	● (-)

CDC-SCE 0.5L Engine (CDC-ORNL):

Parameter	HC	CO	NOx	Soot	ISFC	ITE
CN	● (-)	● (-)	(+)	(+)	● (-)	● (+)
T90	(+)	(+)	(-)	(-)	(-)	(+)
Aromatics	(+)	○ (+)	(+)	(+)	(+)	(-)
Throttle (mA)	● (-)	(+)	● (-)	● (-)	● (-)	● (+)
TCHAR (°C)	○ (+)	● (+)	(+)	(+)	○ (+)	● (-)
Airflow	(+)	● (-)	○ (+)	(+)	○ (-)	○ (+)
Fuelflow	● (-)	(+)	● (+)	● (+)	● (+)	● (-)
AFR Flow	● (-)	○ (+)	● (-)	● (+)	● (+)	● (-)
Vol.Effic.	(+)	● (-)	○ (+)	(+)	○ (-)	○ (+)

LTC Split Injection Engine (WVU-AVFL-16):

Parameter	HC	CO	NOx	Soot	BTE
CN	● (-)	● (-)	● (+)	● (+)	● (+)
T90	(+)	(-)	(+)	(+)	(-)
Aromatics	(-)	(-)	(+)	(+)	(-)
Main SOI (BTDC)	● (-)	● (-)	● (+)	● (-)	● (+)
PilotSOI (BTDC)	● (+)	● (+)	● (-)	● (-)	● (-)
Fuel Split(%Pilot)	(+)	(-)	(+)	(+)	(-)

Sandia Heavy Duty Optical Engine:

Parameter	HC	CO	NOx	FSN	ETAF	ETAC
CN	● (-)	● (-)	(-)	● (+)	(-)	● (+)
T90	(-)	(+)	(+)	(-)	(+)	(-)
Aromatics	○ (-)	(-)	(+)	○ (+)	(+)	○ (+)
inj rail P	(-)	(-)	(+)	● (-)	● (+)	(+)
O2 mole fraction	(-)	● (-)	● (+)	○ (-)	(-)	○ (+)

● - highly influential parameter

○ - moderately influential parameter

(+) - directly correlated

(-) - inversely correlated

AVFL-23 group members need software to enhance visual interpretation of PCA models and support project progress and ongoing discussions. PCAVisTool software was written in MATLAB, an interpreted language that depends on the MATLAB virtual machine. It is easy to install on any user's Windows-based operating system. The set of PCA models generated and presented in this report is available upon request from the authors.

9.0 RECOMMENDATIONS

- Future studies similar to the one presented here would be helpful in order to obtain engine measurements on a significantly greater number of fuels with a broader range of properties (diversification of fuels).
- Implementation of nonlinear modelling such as neural network methods would be valuable in order to find better models of the effects of fuel properties on engine performance and emissions.

10.0 ACKNOWLEDGEMENTS

The authors thank the Coordinating Research council for co-funding this work and to the CRC AVFL-23 panel members – Tim Bays (PNNL), Bill Cannella (Chevron), Nick Killingsworth (LLNL), Chuck Mueller (Sandia), Johanna Paul (Ford), Kenneth Rose (ExxonMobil), and Scott Sluder (ORNL) – for their valuable discussion and helpful comments on the study presented in this report. Special thanks to Chuck Mueller, Scott Sluder, and Bill Cannella for providing spreadsheets of the engine data used in this study. The authors would like to acknowledge support from the Coordinating Research Council and the Government of Canada's ecoENERGY Innovation Initiative (eco-EII) program.

11.0 REFERENCES

1. Alnajjar, M., Cannella, W., Dettman, H., Fairbridge, C., Franz, J., Gallant, T., Gieleciak, R., Hager, D., Lay, C., Lewis, S., Ratcliff, M., Sluder, S., Storey, J., Yin, H., Zigler, B. "Chemical and Physical Properties of the Fuels for Advanced Combustion Engines (FACE) Research Diesel Fuels, FACE-1", Coordinating Research Council, 2010.
2. Gallant, T., Franz, J., Alnajjar, M., Storey, J., Lewis, S., Sluder, S., Cannella, W., Fairbridge, C., Hager, D., Dettman, H., Luecke, J., Ratcliff, M., Zigler, B. "Fuels for Advanced Combustion Engines Research Diesel Fuels: Analysis of Physical and Chemical Properties", *SAE Technical Paper 2009-01-2769*, 2009.
3. Dumitrescu, C.E, Polonowski, C.J., Fisher, B.T., Cheng, A.S., Lilik, G.K., Mueller, C.J. "An Experimental Study of Diesel-Fuel Property Effects on Mixing-Controlled Combustion in a Heavy-Duty Optical CI Engine", *SAE Int J Fuels Lubr* 7:65-81, 2014.
4. Bunting, B.G., Eaton, S.J., and Crawford, R.W. "Performance Evaluation and Optimization of Diesel Fuel Properties and Chemistry in an HCCI Engine," *SAE Technical Paper 2009-01-2645*, 2009.
5. Cho, K., Han, M., Sluder, S.C., Wagner, R., Lilik, G. "Experimental Investigation of the Effects of Fuel Characteristics on High Efficiency Clean Combustion in a Light-Duty Diesel Engine", *SAE Technical Paper 2009-01-2669*, 2009.
6. Bunce, M., Bunting D.G., Wang, J., Crawford, R. "Experimental and Statistical Comparison of Engine Responses as a Function of Fuel Chemistry and Properties in CI and HCCI Engines", *SAE Technical Paper 2012-01-0857*, 2012.
7. Hosseini, V., Neill, W., Guo, H., Dumitrescu, C., Chippior, W., Fairbridge, C., Mitchell, K. "Effects of Cetane Number, Aromatic Content and 90% Distillation Temperature on HCCI Combustion of Diesel Fuels", *SAE Technical Paper 2010-01-2168*, 2010.
8. Dumitrescu, C. E., Neill, W. S., Guo, H., Hosseini, V., Chippior, W. L. "Fuel Property Effects on PCCI Combustion in a Heavy-Duty Diesel Engine", *ASME Internal Combustion Engine Fall Technical Conference*, ICEF2010-35194, 2010.
9. De Ojeda, W., Bulicz, T., Han, X., Zheng, M., Cornforth, F. "Impact of Fuel Properties on Diesel Low Temperature Combustion," *SAE Int J Engines* 4:188-201, 2011.
10. Carder, D., Ryskamp, R., Nuskowski, J. Li, H., Clark, N., Thompson, G., Gautam, M., Wayne, S. "Fuels to Enable Light-Duty Diesel Advanced Combustion Regimes", Coordinating Research Council, CRC Report No. AVFL-16, 2012.
11. Weiss, G.M., Davison, B.D., *Data Mining, Handbook of Technology Management*, H. Bidgoli (Ed.), John Wiley and Sons, 2010.
12. Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. "From Data Mining to Knowledge Discovery in Databases", *AI Magazine* 17:37-54, 1996.
13. MATLAB 2013b; The MathWorks, Inc.: Natick, MA

14. Brereton, R. G., *Applied Chemometrics for Scientist*, John Wiley & Sons, 2007.
15. Martinez, W.L., Martinez, A.R., Solka, J.L., *Exploratory Data Analysis with MATLAB*, 2nd Edition, CRC Press, Taylor & Francis Group, 2011.
16. Wold, S., *Chemometric Methods in Molecular Design*, VCH, Weinheim, 1995.
17. World, S., Eriksson, L., Trygg J., Kettanneh, N. "The PLS Method - Partial Least Squares Projections to Latent Structures - and its Applications in Industrial RDP (Research, Development, and Production)". In Tech rep, Umea University, 2004.
18. Martens, H., Naes, T., *Multivariate Calibration*, John Wiley & Sons, New York, 1989.
19. Hoskuldsson, A. "PLS regression methods", *J Chemom* 2:211–228, 1998.
20. Zhang, M.H., Xu, Q.S., Massart, D.L. "Boosting Partial Least Squares", *Anal Chem* 77:1423-1431, 2005.
21. Van der Voet, H. "Comparing the Predictive Accuracy of Models Using a Simple Randomization Test", *Chemom Intell Lab Syst* 25:313-323, 1994.
22. Eriksson, L., Byrne, T., Johansson, E., Trygg, J., and Vikstrom, C., *Multi-and Megavariate Data Analysis. Basic Principles and Applications*, 3rd Edition, MKS Umetrics AB, Malmo, Sweden, 2013.
23. Rajalahti, T., Arneberg, R., Kroksveen, A.C., Berle, M., Myhr, K-M., and Kvalheim, O.M. "Discriminating Variable Test and Selectivity Ratio Plot: Quantitative Tools for Interpretation and Variable (Biomarker) Selection in Complex Spectral or Chromatographic Profiles", *Anal Chem* 81:2581-2590, 2009.
24. Mahmood, T., Liland, K.H., Snipen, L., Sæbø, S. "A Review of Variable Selection Methods in Partial Least Squares Regression", *Chemom Intell Lab Syst* 118:62-69, 2012.
25. Gieleciak, R., and Fairbridge, C. "Detailed Hydrocarbon Analysis of FACE Diesel Fuels Using Comprehensive Two-Dimensional Gas Chromatography". Division Report Devon 2013-2065-RT, CanmetENERGY, Natural Resources Canada, 2013.
26. Mueller, Ch. J., Cannella, W.J., Bruno, T.J., Bunting, B., Dettman, H.D., Franz, J. A., Huber, M. L., Natarajan, M., Pitz, W. J., Ratcliff, M. A. and Wright, K. "Methodology for Formulating Diesel Surrogate Fuels with Accurate Compositional, Ignition-Quality, and Volatility Characteristics", *Energy Fuels* 26:3284-3303, 2012.
27. Wen, Q., Arboleda, P., Michaelian, K.H., Fairbridge, C., Astrath, N. G. C, Shen, J. Raman, Infrared, and Ultraviolet Absorption Spectroscopy of FACE Diesel Fuels. Division Report Devon 2010-1027-IN, CanmetENERGY, Natural Resources Canada, 2010.
28. Astrath, N.G.C., Astrath, F.B.G., Shen, J., Zhou, J., Michaelian, K.H., Fairbridge, C., Malacarne, L.C., Pedreira, P.R., Medina, A.N., Baesso M.L. "Thermal-lens Study of Photochemical Reaction Kinetics", *Opt Lett* 34:3460-3462, 2009.
29. Steimacher, A. Medina, A.N., Bento, A.C., Rohling, J.H., Baesso, M.L., Reynoso, V.C.S., Lima, S.M., Petrovitch, M.N., Hewak, D.W. "The Temperature

- Coefficient of the Optical Path Length as a Function of the Temperature in Different Optical Glasses”, *J Non-Cryst Solids* 348:240-244, 2004.
30. Wen, Q., Shen, J., Gieleciak, R., Michaelian, K.H., Rohling, J.H., Astrath, N.G.C., Baesso, M.L. “Temperature Coefficients of the Refractive Index for Complex Hydrocarbon Mixtures”, *Int J Thermophys* 35: 930-41, 2014.
 31. Smilde, A.K., Kiers, H.A.L., Bijlsma, S., Rubingh, C.M., van Erk, M.J. “Matrix Correlations for High-dimensional Data: the Modified RV-coefficient”, *Bioinformatics* 25:401-405, 2009.
 32. Abdi, H. “RV Coefficient and Congruence Coefficient”, In: Salkind, N. (editor), *Encyclopedia of Measurement and Statistics*, Thousand Oaks, Sage, 2007.
 33. Cannella, W., Fairbridge, C., Arboleda, P., Bays, T., Dettman, H., Foster, M., Gieleciak, R., Gunter, G., Hager, D., Lay, C., Lewis, S., Luecke, J., Sluder, S., Zigler, B. Detailed Characterization of the Physical and Chemical Properties of the Reformulated FACE Diesel Fuels: FD2B, FD4B, and FD7B. CRC Report No. AVFL-19-1, 2013.
 34. Cannella, W., Fairbridge, Gieleciak, R., C., Arboleda, P., Bays, T., Dettman, H., Foster, M., Gunter, G., Hager, D., King, D., Lay, C., Lewis, S., Luecke, J., Sluder, S., Zigler, B., Natarajan, M. Advanced Alternative and Renewable Diesel Fuels: Detailed Characterization of Physical and Chemical Properties. CRC Report No. AVFL-19-2, 2013.
 35. Kim, J., Sluder, C.S., Wagner, R., “Combustion Studies with FACE Diesel Fuels: A Literature Review”, *SAE Technical Paper 2012-01-1688*, 2012.
 36. Aitchison, J., *The Statistical Analysis of Compositional Data*, Chapman & Hall, London, 1986.

APPENDIX A: PCA INTRODUCTION AND EXAMPLE

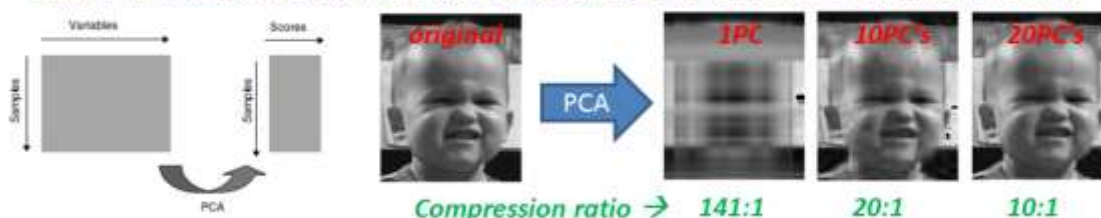
Principal Component Analysis - introduction

Rafal Gieleciak

1

Principal Component Analysis (PCA) – as seen by chemometricians

Principal component analysis (PCA) is a method of data reduction or data compression.

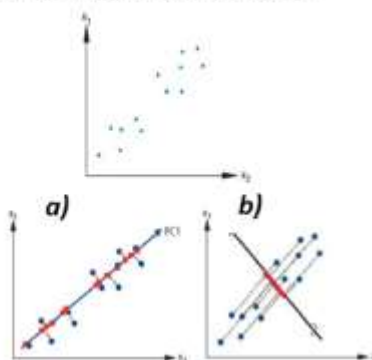


Applying PCA to a data matrix it constructs new variables, known as principal components (PCs), such that:

- the new variables are linear combinations of the original ones, with the weight vectors that define the combinations being of unit length and orthogonal (at right angles) to each other;

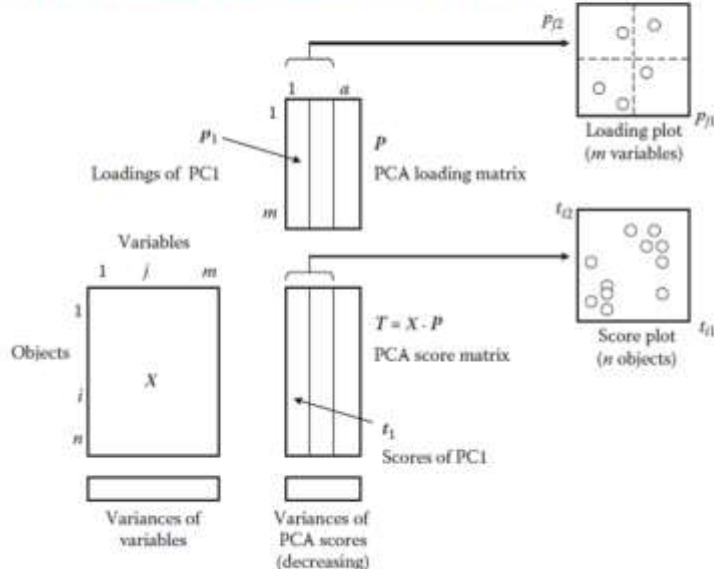
$$t_{i1} = x_{i1} p_1 + x_{i2} p_2 + \dots + x_{im} p_m = \mathbf{x}_i^T \cdot \mathbf{p}_1$$

- the new variables are uncorrelated;
- the first new variable captures as much as possible of the variability in all the original ones, having been constructed to have maximum variance amongst all such linear combinations
- each successive new variable accounts for as much of the remaining variability as possible.



Principal Component Analysis (PCA) – as seen by chemometricians

Matrix scheme for PCA. Since the aim of PCA is dimension reduction and the variables are often highly correlated a $a < \min(n, m)$ PCs are used



A **loading plot** shows similarities, differences between variables and their influence on the behaviour of the objects. Variables contributing similar information are grouped together (they are correlated). Moreover the further away from the plot origin a variable lies, the stronger impact that variable has on the model

A **score plot** shows similarities and differences between objects. a scatterplot of the first two new variables often being highly informative, but more importantly the first few new variables can be used as input to methods such as multiple regression or discriminant analysis that will not cope with highly multivariate data

3

Example of PCA Analysis Tim Hortons nutrition facts



A Canadian multinational fast casual restaurant known for its coffee and doughnuts. It is also Canada's largest fast food service; at the end of 2013, it had 3,588 restaurants in Canada, 859 in the United States and 38 in the Persian Gulf region.

Source: wikipedia



Aside from its coffee, tea, hot chocolate, and doughnuts, the Tim Hortons menu now contains a number of other baked goods, such as doughnut holes (branded as Timbits), muffins, croissants, tea biscuits, cookies, rolls, danishes, and more recently bagels.

Source: wikipedia

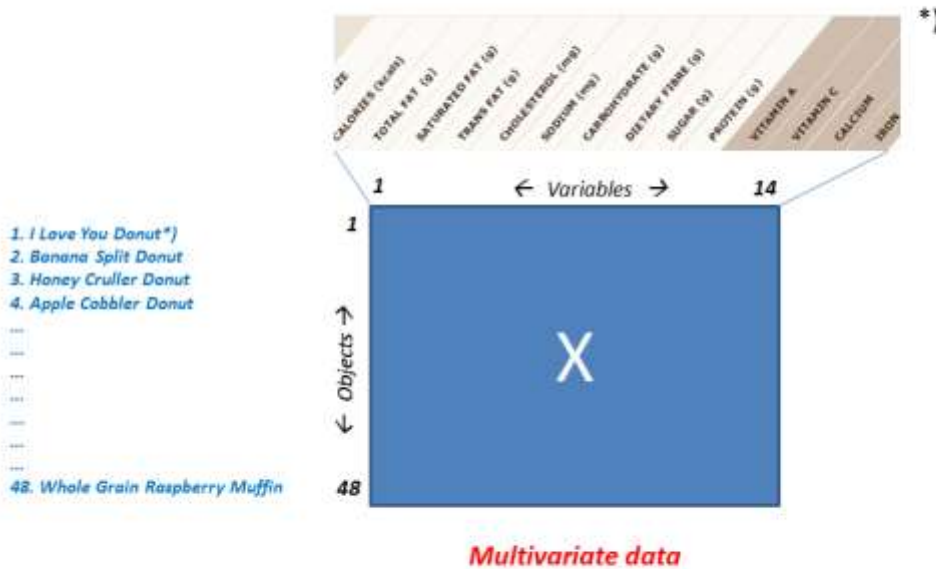


Each bakery good is associated with nutrition facts table containing such information as: Total fat, Calories, Proteins etc.

Custom Project: use Principal Component Analysis to explore some of the Tim Hortons bakery goods according to their nutrition details

4

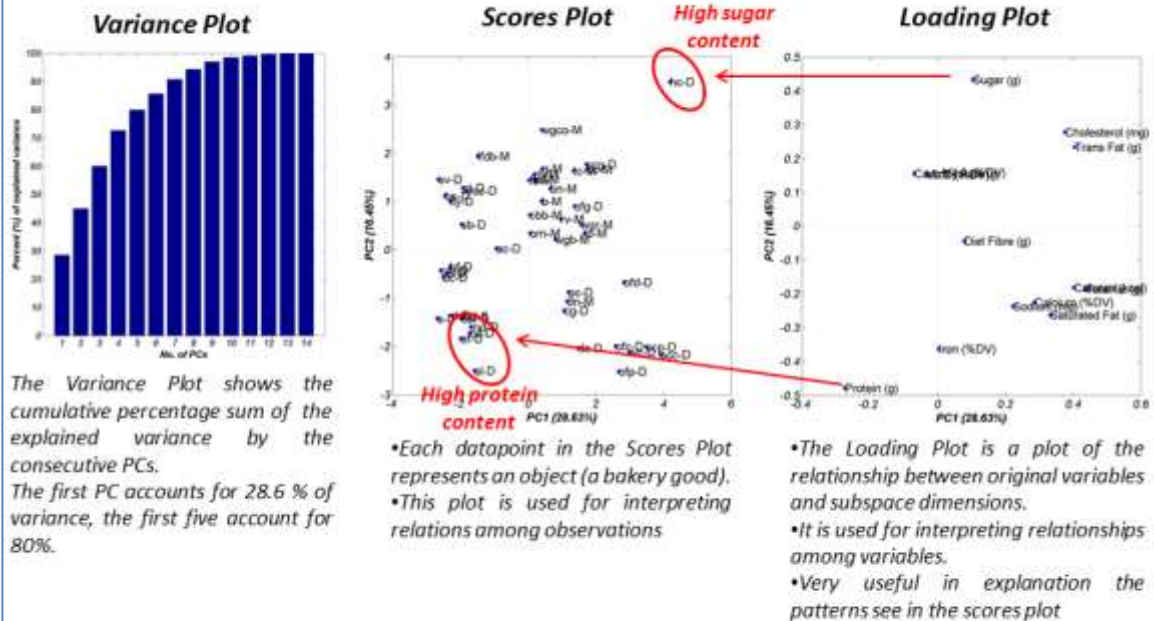
Tim Hortons Dataset



*) data taken from <http://www.timhortons.com/ca/en/menu/nutrition-calculator.html>

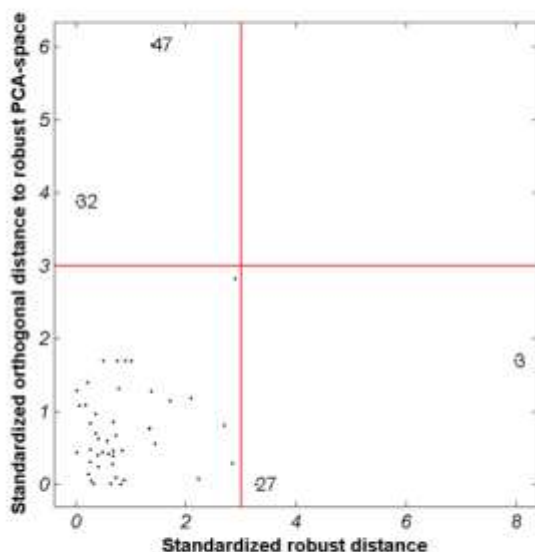
5

PCA-first approach



6

PCA - diagnostic tool



Based on this plot at least 2 objects are potential outliers in PC space:

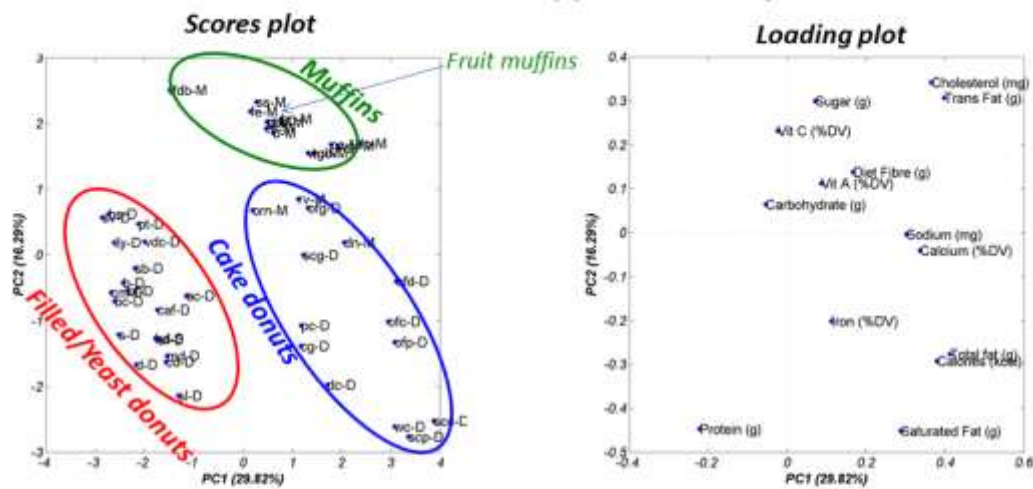
(47) *Whole Grain Carrot Orange Muffin* – High Vitamin A content

(3) *Honey Cruller Donut* – High Cholesterol content/low Iron

We will remove these objects in the next step

7

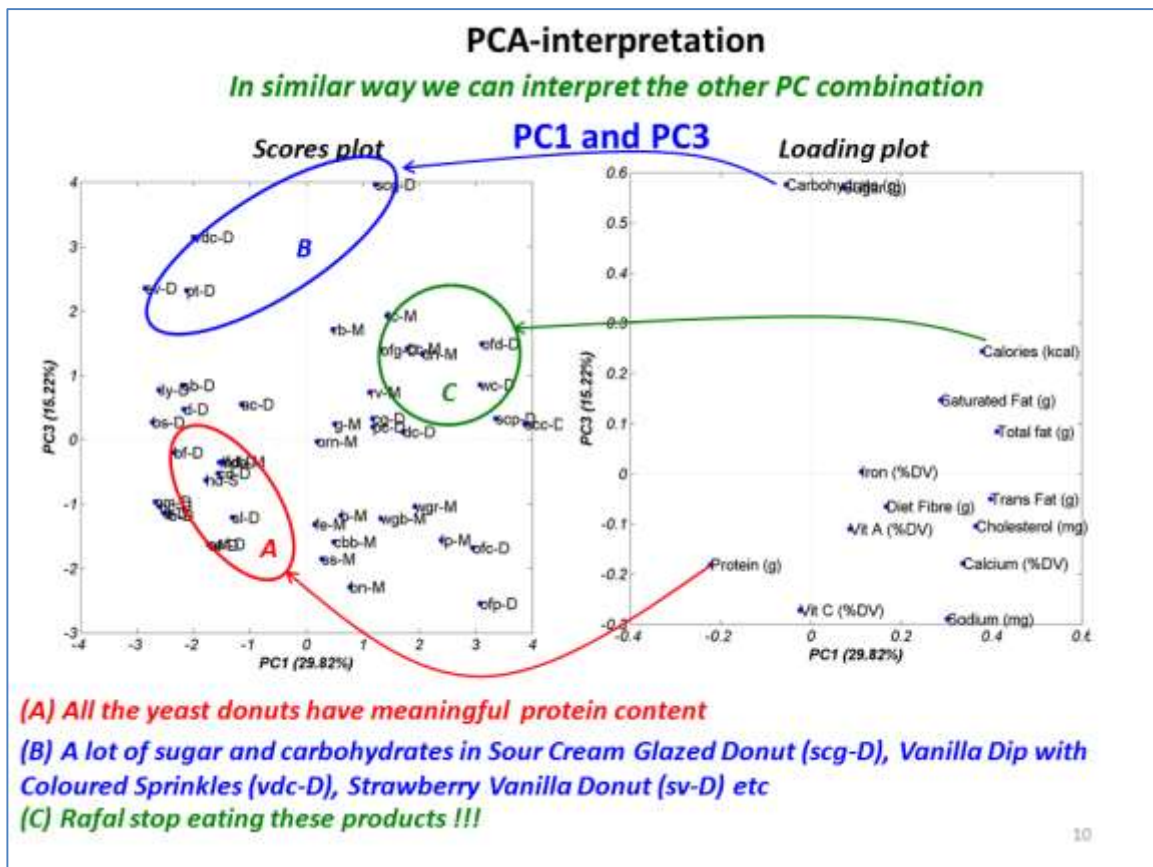
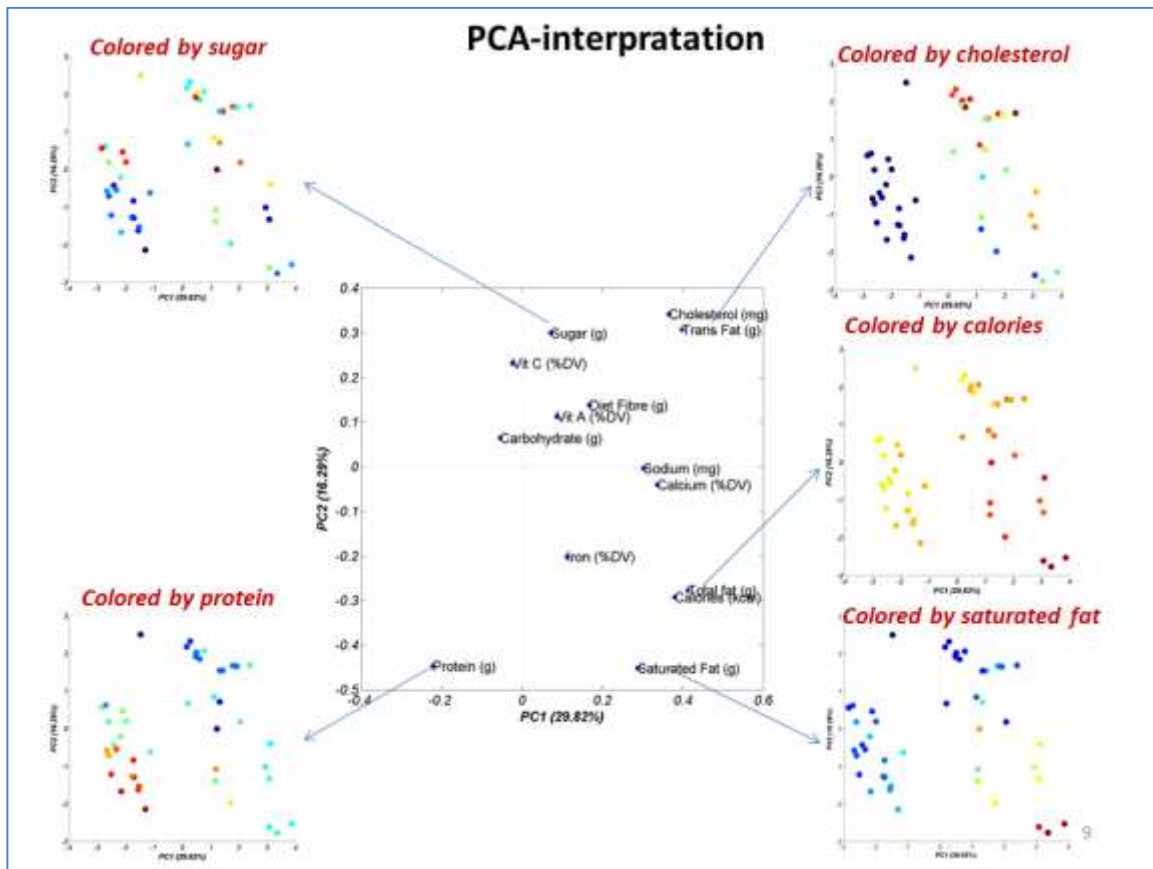
PCA-Second Approach - Projection



At least three groups of bakery goods based on PC1/PC2 score plot

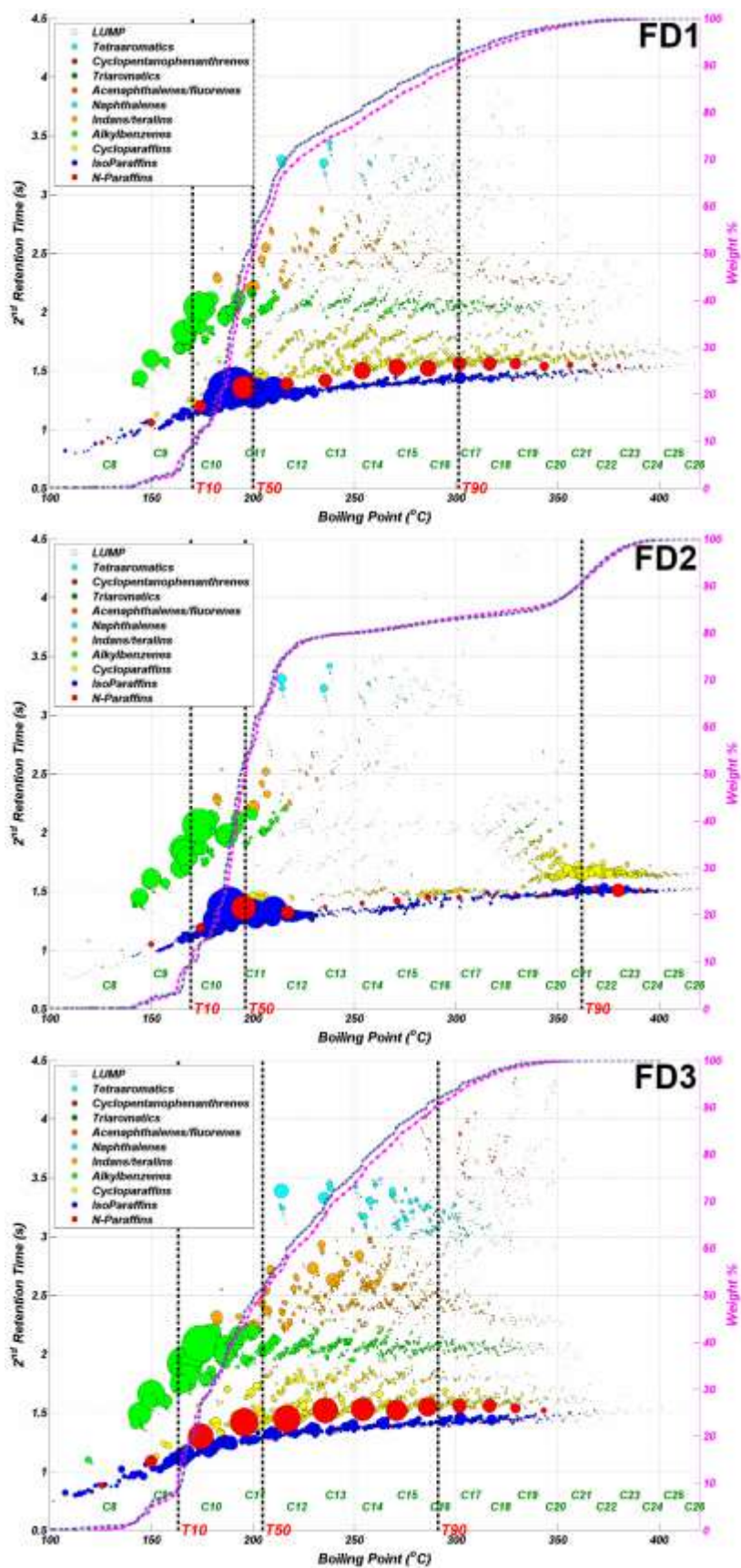
- Strongly correlated pairs of variables:
Cholesterol/trans fat
Total fat/Calories
- Carbohydrate seems to be not important variable (close to the center – low loading values)
- Fruit muffins have a lot of sugar and Vitamin C
-

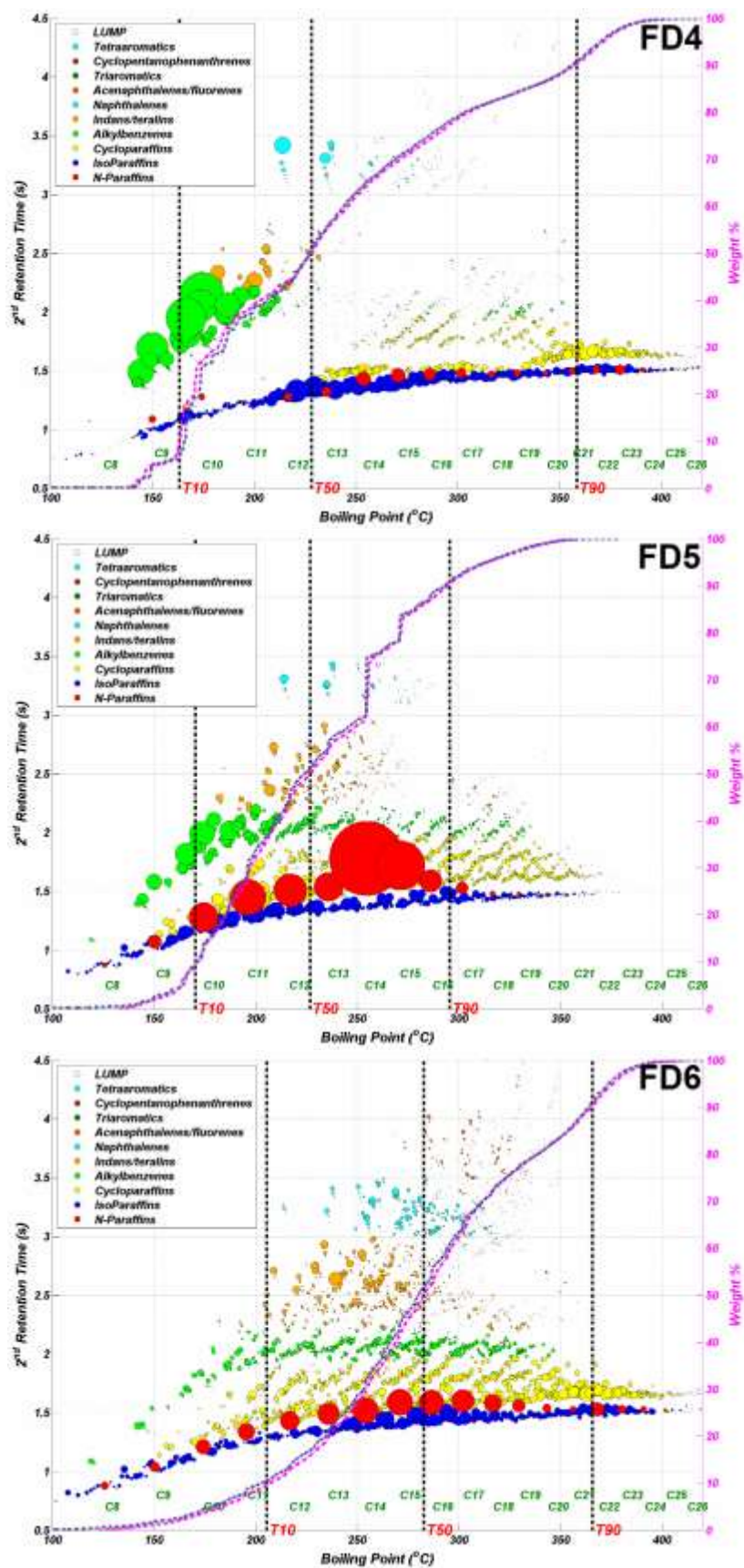
Visual Interpretation on the next slide → 8

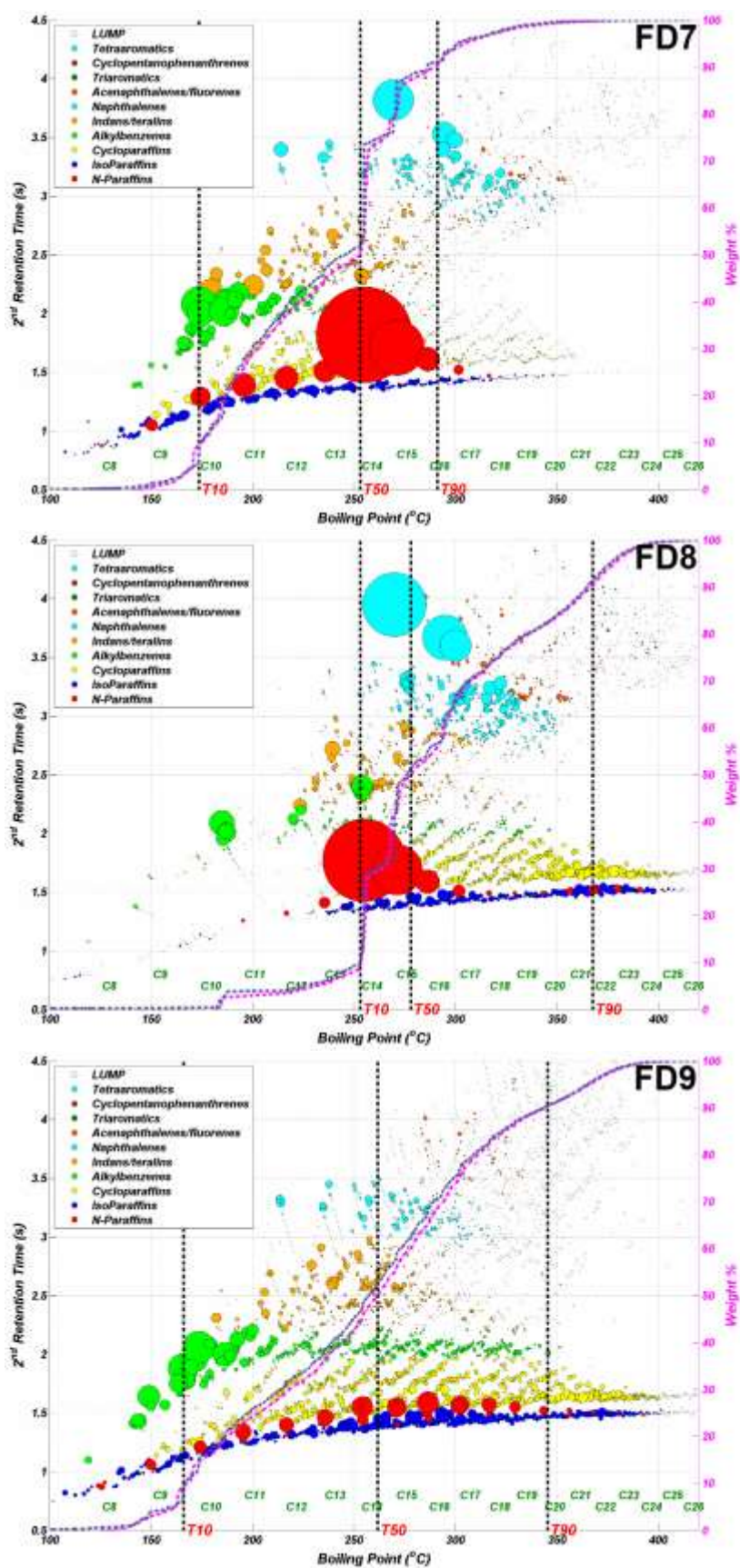


APPENDIX B: GC×GC ANALYSIS OF FACE DIESEL FUELS (TAKEN FROM REF 25)

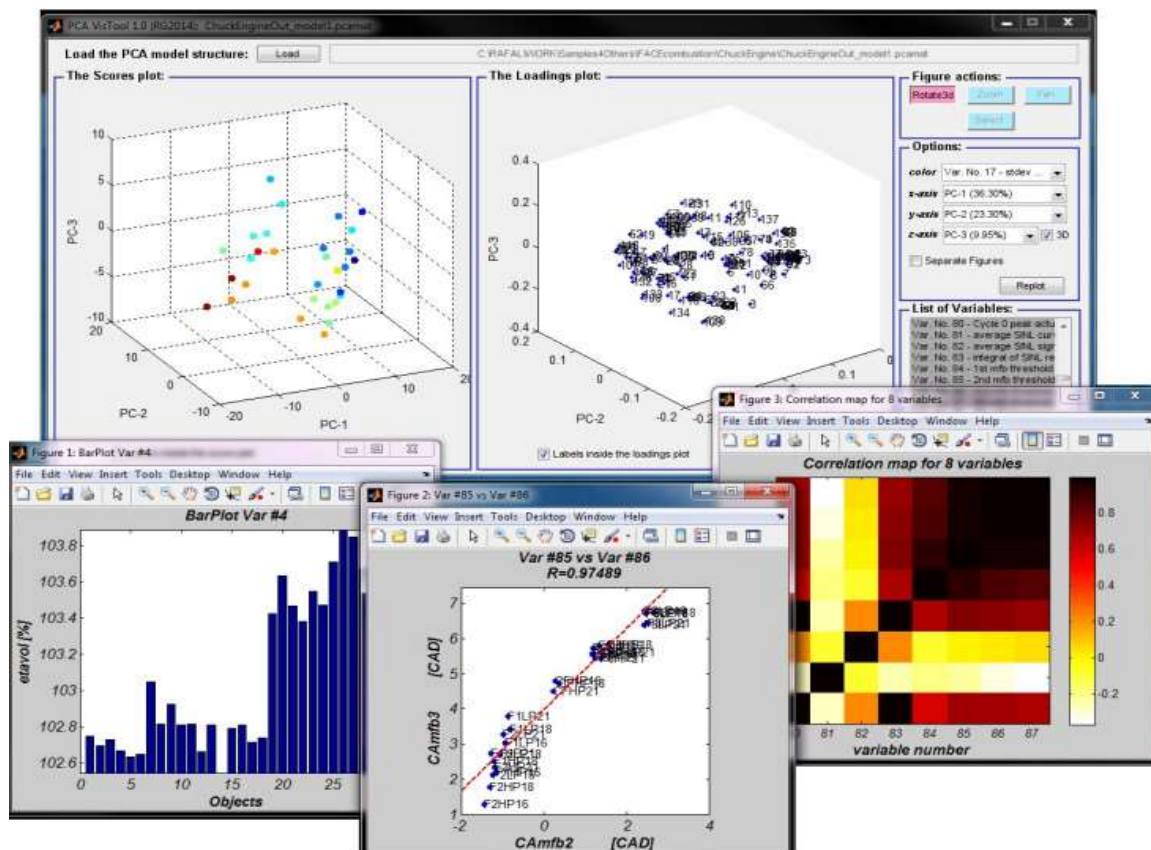
In the following figures, GC×GC-FID chromatograms (nonpolar x polar – ‘normal’ column configuration) are presented in the bubble plot format. The colour and the size of the bubbles are related to hydrocarbon type and compound concentration, respectively. The SimDis curve obtained from ASTM D2887 analysis was superimposed on each chromatogram as a dashed magenta line. The dashed blue line following the magenta SimDis line represents the GC×GC SimDis profile. Black dotted lines on the pictures show the SimDis regions T10, T50, and T90, respectively. On each of the bubble plots, just above the x-axis, there is indication of n-paraffin carbon number labeled in green.







APPENDIX C: PCAVISTOOL TUTORIAL



PCAVisTool was written at CanmetENERGY, Canada to support CRC project (AVFL-23) and can be used/extended to other projects.

C1. INTRODUCTION

In response to the need identified by AVFL-23 group members, a software package to enhance visual interpretation of PCA models and support the project progress discussions was developed. PCAVisTool is written in MATLAB, an interpreted language that depends on the MATLAB virtual machine and is easy to install on any user's Windows-based operating system. PCAVisTool is universal software and can be used to depict PCA models.

C2. INSTALLATION

This version of PCAVisTool runs on any 32-bit version of Windows operating system (i.e., XP/Vista/7). The source code of PCAVisTool was written in the MATLAB software environment. The standalone version of the PCAVisTool was created using MATLAB Compiler, so it can be deployed to new users without MATLAB installed on their machines (i.e., it does not require a MATLAB license). However, users do have to have MATLAB Compiler Runtime (MCR) installed on their own computers. The current version of PCAVisTool requires MCR version 8.0 (R2012b) run under 32-bit Windows operating system. The MCRs (various versions) can be downloaded (free of charge) from the MathWorks Website (<http://www.mathworks.com/products/compiler/mcr/>) and follow the instructions in the installation wizard.

To download and install the MCR:

1. Click the version and platform that corresponds to the application or component you are using.

Note: you can find this information in the `readme.txt` file that accompanies the application or component.

Release	Windows	Linux	Mac
R2014a (8.3)	32-bit / 64-bit	64-bit	Intel 64-bit
R2013b (8.2)	32-bit / 64-bit	64-bit	Intel 64-bit
R2013a (8.1)	32-bit / 64-bit	64-bit	Intel 64-bit
R2012b (8.0)	32-bit / 64-bit	64-bit	Intel 64-bit
R2012a (7.17)	32-bit / 64-bit	32-bit / 64-bit	Intel 64-bit

2. Save the MCR installer file on the computer on which you plan to run the application or component.
3. Double click the installer and follow the instructions in the installation wizard.

Figure C1 – Screen shot of the Mathworks website showing how to download and install the Matlab Compiler Runtime (MCR).

Once the installation of MCR is completed, reboot the computer and double click on the PCAVisTool.exe file to start the software. It may take a few minutes before the execution actually starts and software screen appears like in Figure C2.

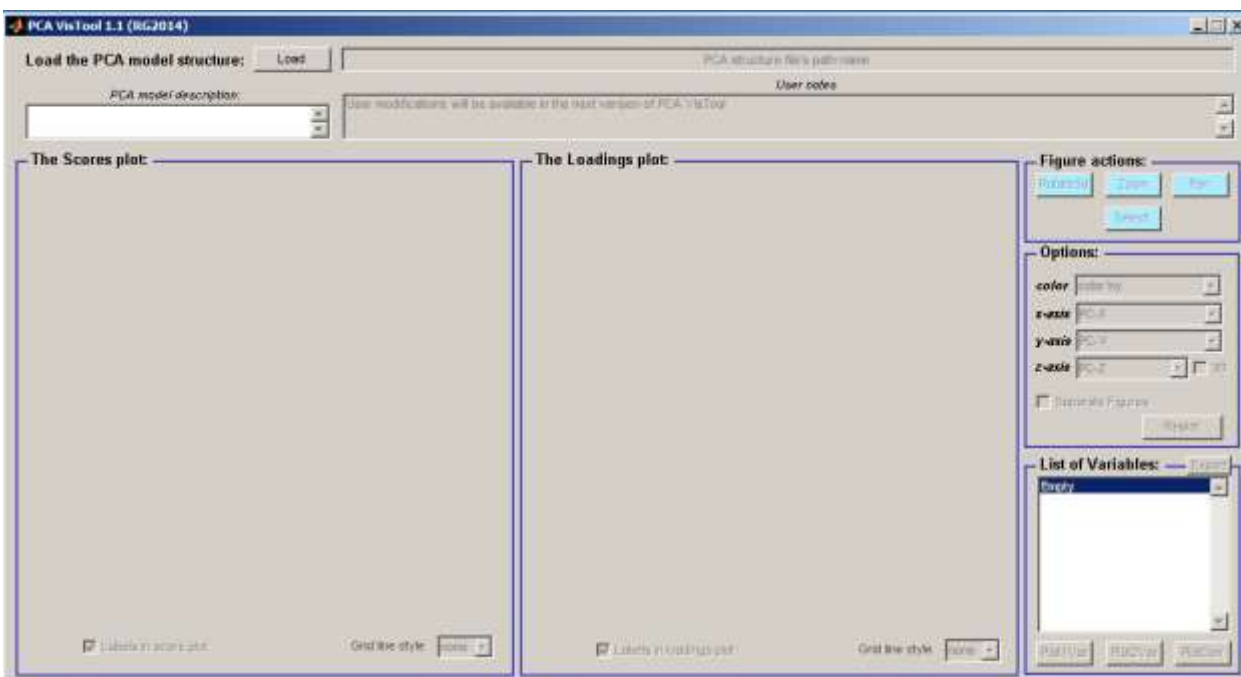


Figure C2 – The PCAVisTool desktop

C3. WORKING WITH PCAVISTOOL

C3.1. PCA MODELS CONTAINER

PCAVisTool software version 1.1 (RG2014) support only one Matlab file format with *.pcmat file extension. The *.pcmat file consists of a Matlab structure array which includes results after PCA analysis performed on a selected set of data. This information is necessary to run PCAVisTool properly. Each *.pcmat file has a few predefined ‘data containers’ called fields (and each field can also be a structure). An example of *.pcmat file structure is presented in Figure C3. The structure of *.pcmat file is created in such way that each Matlab user can export, form, save and use such files in the future.

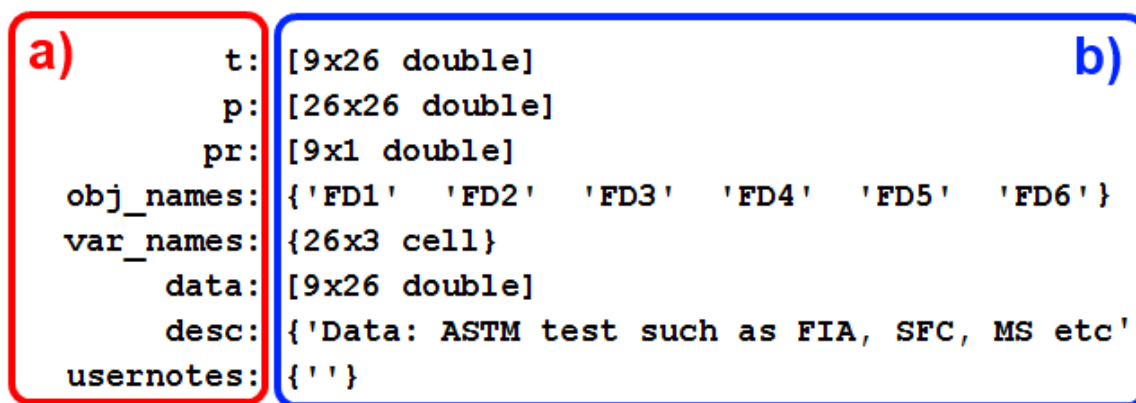


Figure C3 – The view into *.pcmat file structure with the predefined field names (a) and field's content (b). Short description of the meaning of the fields is presented in Table C1.

Table C1 – Description of fields of the *.pcmat structure

Field	Content	Format
t	PCA scores	Matrix (m x n)
p	PCA loadings	Matrix (n x n)
pr	Explained variance	Vector (m x 1)
obj_names	Object's tag names	String array (m x 1)
var_names	Variable's tag name	Cell array (n x 3)
data	Original dataset used in PCA	Matrix (m x n)
desc	Description of original dataset	String array
usernotes	Empty (default) used for additional notes made by user during running the PCAVisTool	String

C3.2. LOADING PCA MODELS

Before loading a PCA model all GUI components are inactive or invisible (see Figure C2). To load PCA model into PCAVisTool, the user has to press the 'Load' button and browse for a PCA model file (*.pcmat). After loading a selected PCA model into the PCAVisTool workspace, most of the GUI components will be activated and updated (shown in Figure C4) such as PCA model name and file path name, dataset description, scores and loadings plots, list of variables etc.

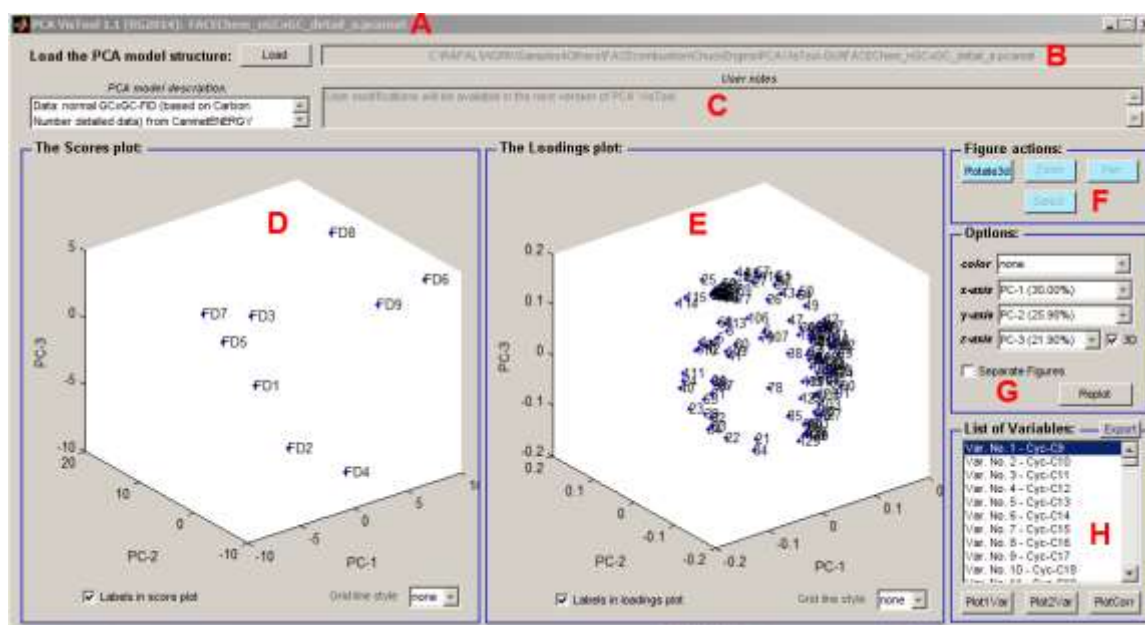


Figure C4 – The PCAVisTool desktop after loading a PCA model most of the GUI components are activated. The name (A) and the file path (B) of the current PCA model and the dataset description (C). The PCA scores (D) and loadings (E) plots. Activated: figure actions (F), plotting options (G) and interactive list of variables (H).

C3.3. PLOTTING OPTIONS

After loading PCA model into the PCAVisTool workspace, the scores and loadings plots appears in respective windows (see Figure C4 (d) and (e)). The default mode of presenting the PCA results is a three-dimensional representation of principal components (blue labeled points, no grids on the plots). The default plotting mode may be changed after selecting proper options followed by pressing the ‘Replot’ button (see Figure C5).

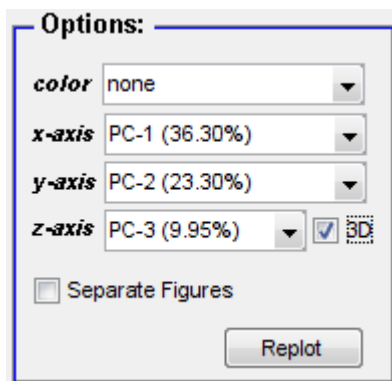


Figure C5 – The plotting ‘Options’ container. Remember; every time you make any changes in the ‘Options’ box press the ‘Replot’ button to update the scores and loadings plots.

Colour

The default colour for points in scores plots is blue ('none' option). However, in many cases it is very helpful to colour objects according to a selected variable. In Figure C6 one of the variables is used to colour the objects in the scores plot.

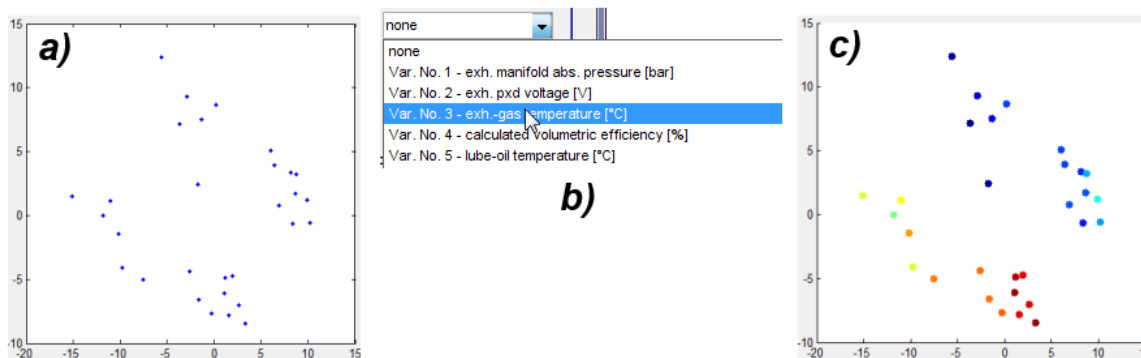


Figure C6 – The user can choose variable (b) which will be used to colour objects in the scores plot. In this example two scores plots with option 'none' (a) and coloured using selected variable (c) are shown.

x-axis, y-axis, z-axis

Both scores and loadings plots are scatter plots. The user can specify which principal components (PC) are used to plot PCA results. The first three PC's are used in default mode (after loading the PCA model). To change that, the user has to click on PC in x-, y- and/or z-axis drop-down list and select respective PC (see Figure C7) followed by pressing the 'Replot' button.

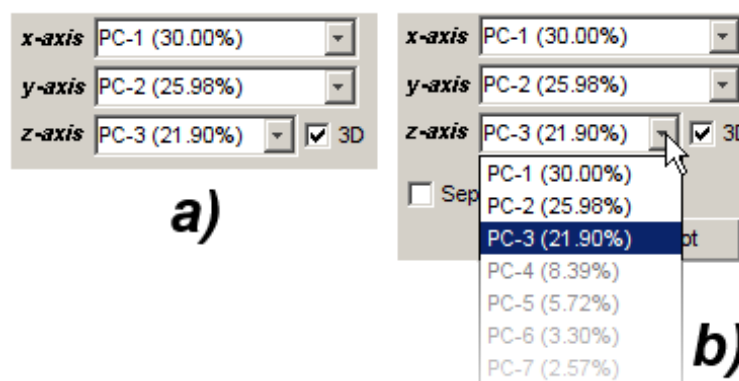


Figure C7 – The user can specify which PCs are plotted on x-, y-, and/or z-axis by selecting respective PCs (a) from the drop-down lists (b). Values in brackets in drop-down list (i.e. PC-1 (30.00%)) report amount of explained total variance for selected principal component. Notice that each consecutive PC has lower value of explained variance, hence is less informative.

3D mode

The default mode for visualization of PCA results is in three dimensional space (3D XYZ). 3D XYZ graphs summarize the relationships between three PC scores or loadings. The user can change both loading and scores plots to 2D XY. Usually 2D XY plots are more convenient as they allow the user to apply more interactive manipulations inside the figures such as zooming or selecting.

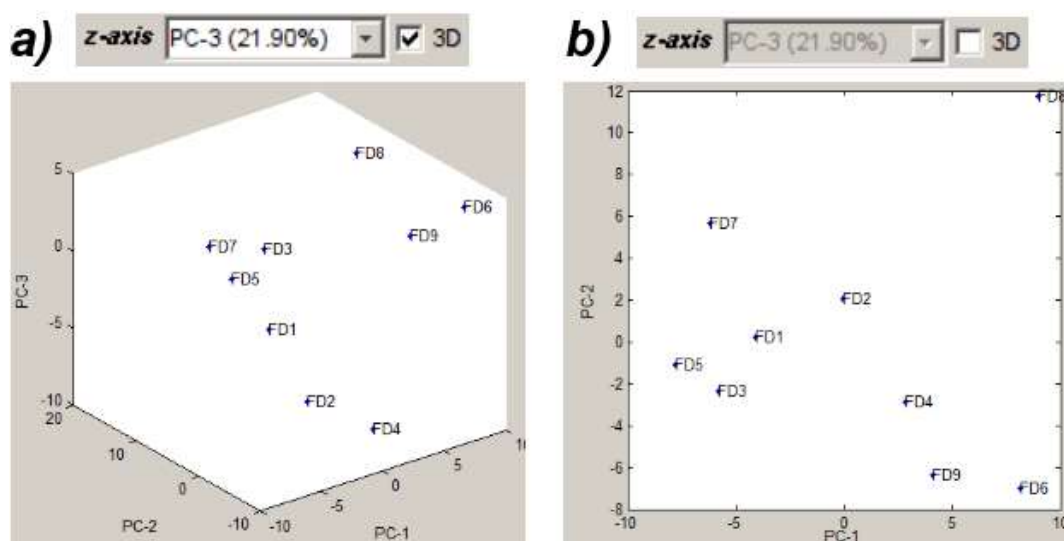


Figure C8 – Switching between 3D (a) and 2D representation of PCA models (b) (here in scores plots). Notice that after checking off the 3D option the ‘z-axis’ drop-down list will be inactive.

Labels

Each object in the scores plot and each variable in the loadings plot is displayed with the respective label (according to the content of the obj_names field saved in the *.pcmat structure). Variables in the loadings plot are labeled by variable number rather than variable name. User may remove labels from the plots by checking off the respective checkboxes located under the scores and loadings plot figures followed by pressing the ‘Replot’ button.

Grid lines

Grid lines on both Scores and Loadings plots are optional (the default is to display without grid lines). To control the presence and appearance of grid lines on a plot user has to select one of the four options from the grid line style drop-down list located under the scores and loadings plot figures (Figure C9) followed by pressing the ‘Replot’ button.

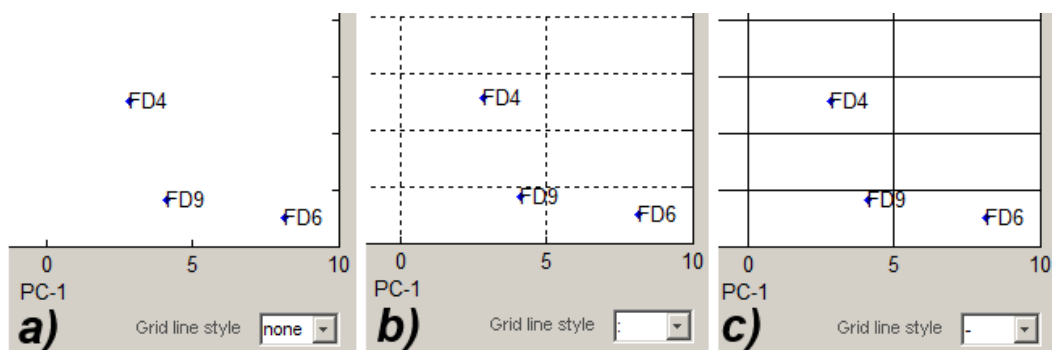


Figure C9 – The ‘grid line style’ drop-down list box used for controlling the presence and appearance of grid lines. On the plot three examples: no grid lines (a), dotted grid line (b) and solid grid line (c) style.

Separate figures

The scores and loadings plots are docked in PCAVisTool by default. However, the user can replot these graphs in separate figures after checking ‘On’ in the ‘Separate Figures’ CheckBox and pressing the ‘Replot’ button. As a result of this operation, user will notice the two separate windows containing the scores and loadings plot figures, respectively (Figure C10).

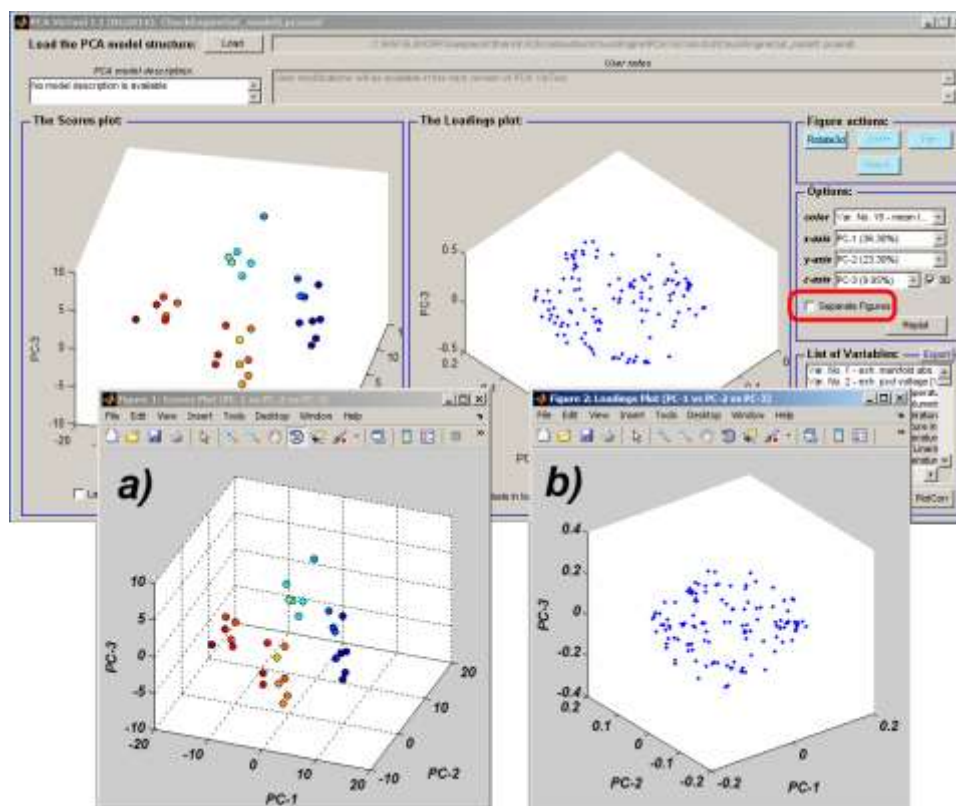


Figure C10 – The ‘Separate Figures’ function allows for reploting the scores and loadings plots in separate figure windows.

C3.4. FIGURE ACTIONS

To simplify work within PCAVisTool software the user can perform a few actions inside both scores and loadings plots. These figure actions (see Figure C11) consist of rotating axis in 3D space, zooming, panning, and selecting variables in the loadings plots. Active and inactive appearance of buttons in the ‘Figure actions’ container depends on the selected display mode (3D turn on/3D turn off).

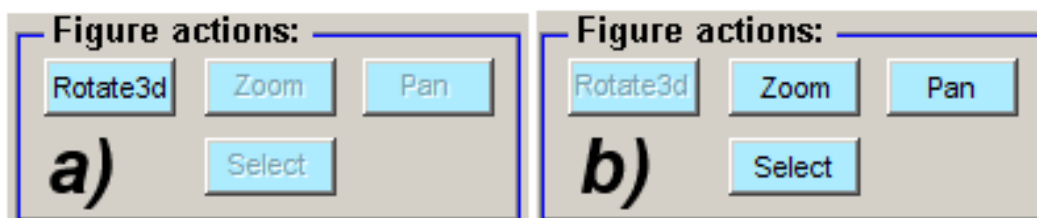


Figure C11 – The ‘Figure actions’ container consists of four functional buttons used for performing additional routines such as rotating axis, zooming, panning and selecting points in Scores and loadings plots. Figure (a) presents the buttons’ state (Rotate3d active and Zoom/Pan/Select inactive) in the case where plots are displayed in 3D space. Changing the display mode from 3D to 2D activates the Zoom/Pan/Select buttons and inactivates the Rotate3D action.

Rotate3d (Works only in 3D display mode)

The user can easily rotate axis in score or loadings plots after pressing the ‘Rotate3d’ button. Pink colour of the button indicates that ‘rotate3d’ mode is in action.

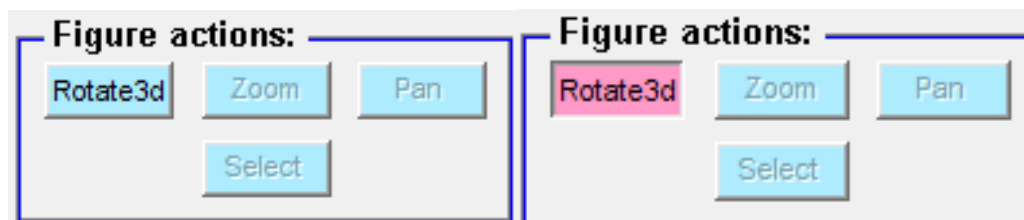


Figure C12 – The ‘Rotate3d’ action button

In plotting area use the left mouse button to rotate the graphs. After pressing the right mouse button in the plot area the user gets access to a few other options (see Figure C13) such as selection of predefined views and rotation style for complex graphs. Pressing the ‘Rotate3d’ button again switches off the axis rotation mode.

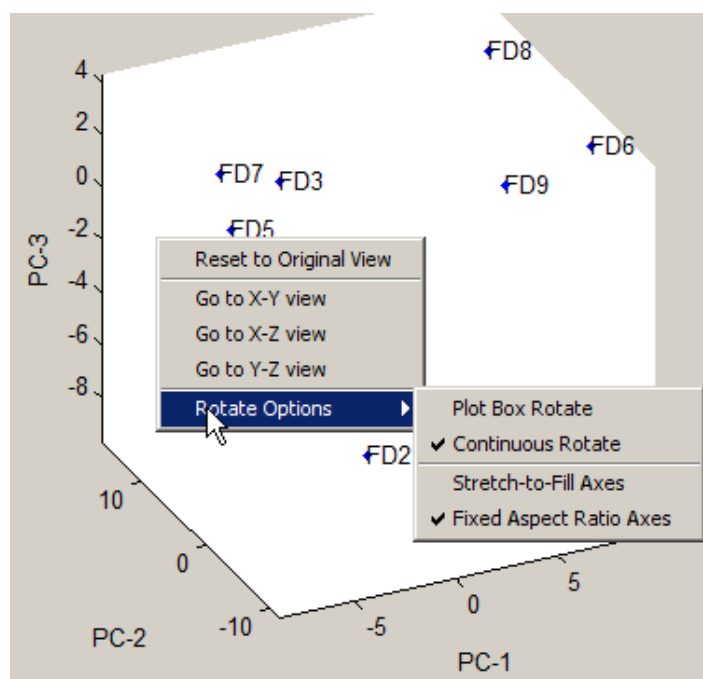


Figure C13 – After activating rotate3D mode, user can control various rotation options from right-click context menu

Zoom (*Works only in 2D display mode*)

The user has access to interactive zooming in/out action when the ‘zoom’ button is pressed. The zoom mode will be indicated by button color change (See Figure C14). The Zoom feature enables a more detailed look at the data while reviewing PCA models.

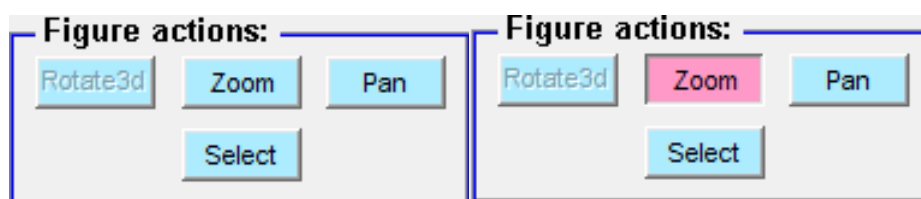


Figure C14 – The ‘Zoom’ action button

When interactive zooming is enabled, pressing the left mouse button while the cursor is within an axes zooms into the point or out from the point beneath the mouse. The user has three modes of zooming (Figure C15): unconstrained (in both dimensions), horizontal and vertical. In order to return to the original view the user should press the right button and then choose ‘Reset to Original View’. The same effect can be achieved by double-clicking over the axes.

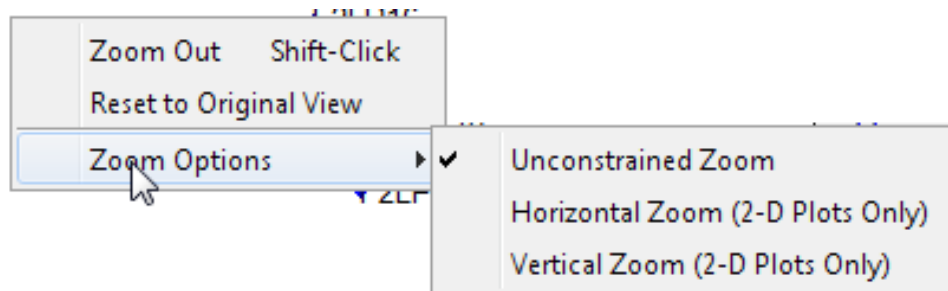


Figure C15 – After activating zoom mode, user can control various rotation options from right-click context menu.

Pan (Works only in 2D display mode)

The user can move the view of a graph both horizontally and vertically after turning on the ‘Pan’ action. The pan mode will be indicated by button color change (See Figure C16) and a hand icon in the plots. Panning is a very practical tool especially when coupled with zoom action.

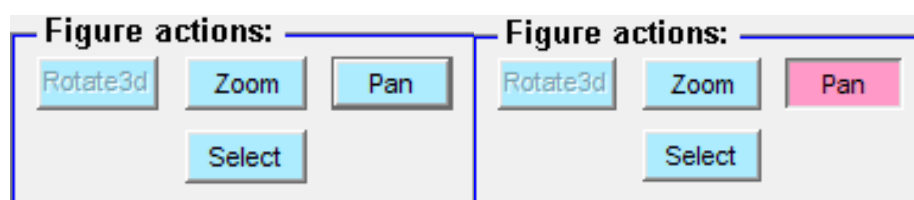


Figure C16 – The ‘Pan’ action button

Additional pan tool options are available by right-clicking in the plotting area when pan mode is in action (see Figure C17). In order to return to the original view the user should press the right button and then choose ‘Reset to Original View’. The same effect can be achieved by double-clicking over the axes.

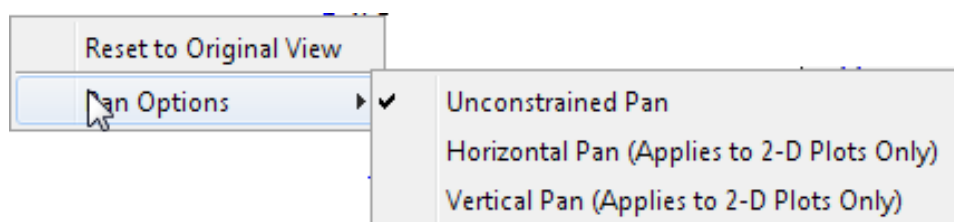


Figure C17 – After activating pan mode, user can control various rotation options from right-click context menu.

Select (Works only in 2D display mode inside the Loadings plot)

The user can enhance exploration of PCA models by selecting (or extracting) respective points (variables) in the loadings plot by using the select mode. By default, the points selector function

is off. To initialize the select mode press the ‘Select’ button. The select mode will be indicated by button colour change (See Figure C18). In the loadings plot area the user will notice a cross icon. It lets create the user a polygon which surrounds the point on the graph. Left-click to add more points to the polygon. Double-click to add a final point and complete the polygon selection. Selected points will change color from blue to red. In the meantime, variables associated with the selected points in the Loadings plot will be highlighted in the ‘List of Variables’ container. Figure C19 shows the entire process of selecting a group of points located at the upper part of the Loadings plot.

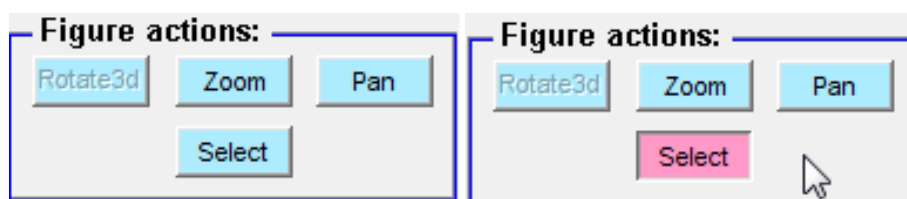


Figure C18 – The ‘Select’ action button

Notice:

- Right-click to exit ‘Select’ mode without selecting any points. This operation can be performed at any stage of creating a polygon.
- Pressing Enter or Return ends the polygon creation without adding a final point.
- Pressing Backspace or Delete removes the last point in polygon.

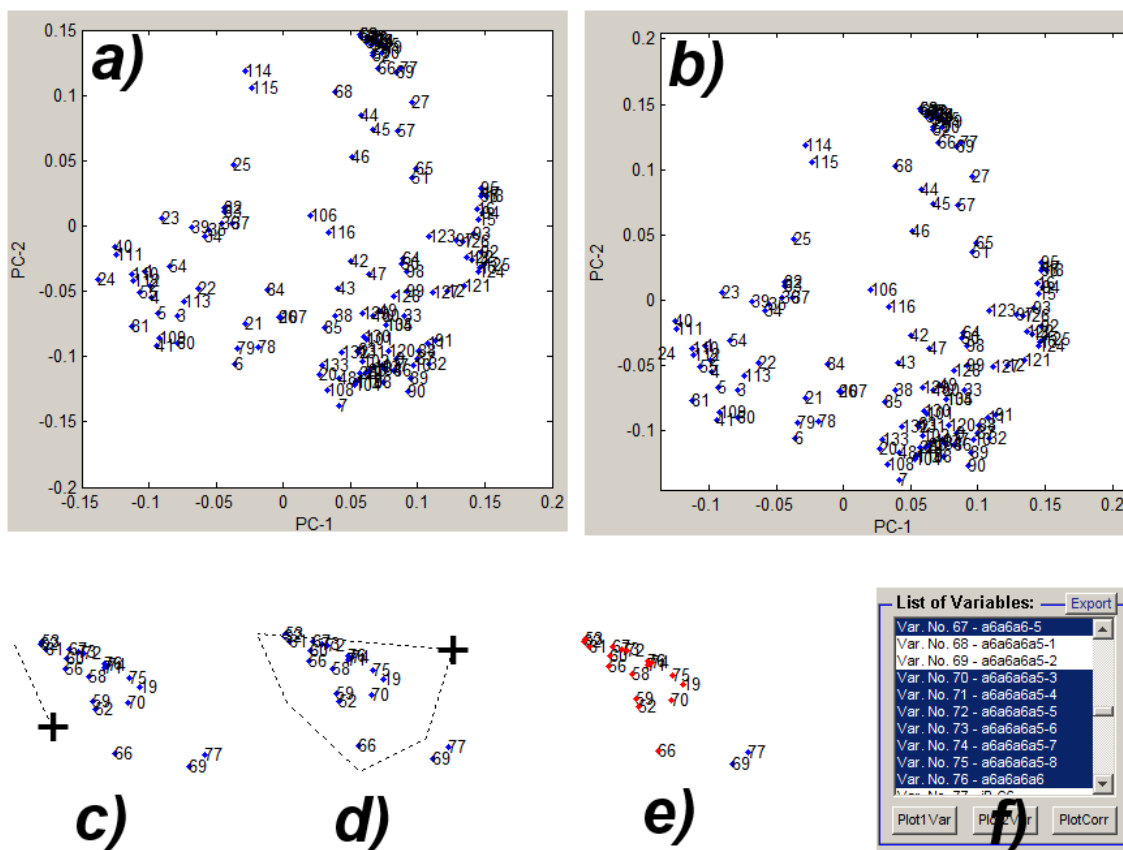


Figure C19 – The points selection procedure. Original plot (a), panning a plot to access the group of points in the most upper part of the figure (b), start the selection tool and drag the polygon line (c), expand the number of polygon points by mouse left-click (d), double-click to add the final polygon point and complete selection process, selected variables with the detailed description will be highlighted in the List of Variables container (f).

C3.5. LIST OF VARIABLES

The 'List of Variables' field contains information about all original variables used in the imported PCA model. This list is very useful during interpretation of relations among objects in the scores plot, examination of the variable profiles and analysis of correlations between variables (Figure C20).

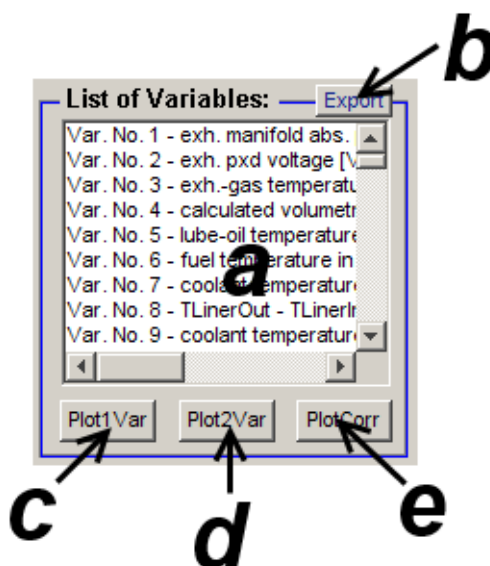


Figure C20 – The ‘List of variables’ container. Variables are arranged according to the order found in the original dataset and are listed in the listbox (a). Selected variables can be exported to the excel format (*.xls) (b). Buttons ‘Plot1Var’, ‘Plot2Var’, and ‘PlotCorr’ (c-e) used for visualization of one or more selected variables.

Selection of variables may be performed inside the listbox or by interactive selection inside the Loadings plot (see section C3.4). The standard way to select a single variable in a ListBox is by clicking on it (all others variables will be deselected, however). Pressing Shift key and left-clicking allows user select a range of variables. To select all of the variables in the list, the user can use ‘Ctrl+A’ hotkey.

Variables (at least one) may be visualized using three modes: the bar plot of one variable (the Plot1Var button), the correlation parity plot (the Plot2Var button), and the correlation maps (the plotCorr button).

Plot1Var

The user can review profile (as a barplot) of selected variables after selecting one variable in the list and pressing the ‘Plot1Var’ button.

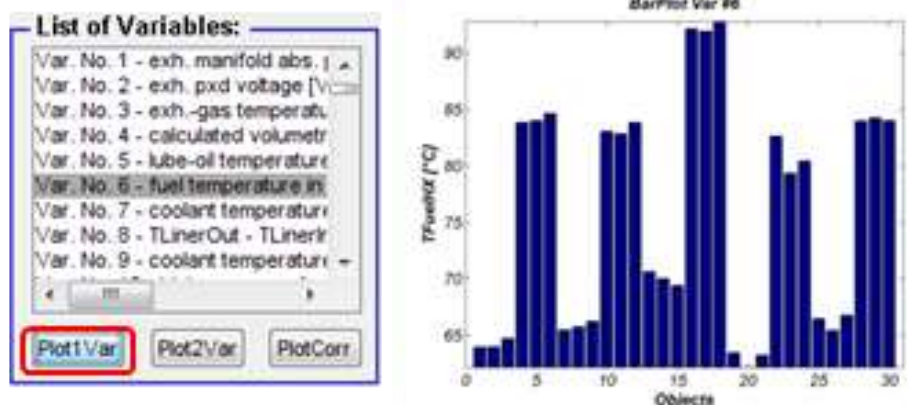


Figure C21 – The result of ‘Plot1Var’ action

Plot2Var

The user can review correlations between two selected variables after selecting two variables in the list and pressing the ‘Plot2Var’ button. The correlation coefficient (R) is reported in the figure title.

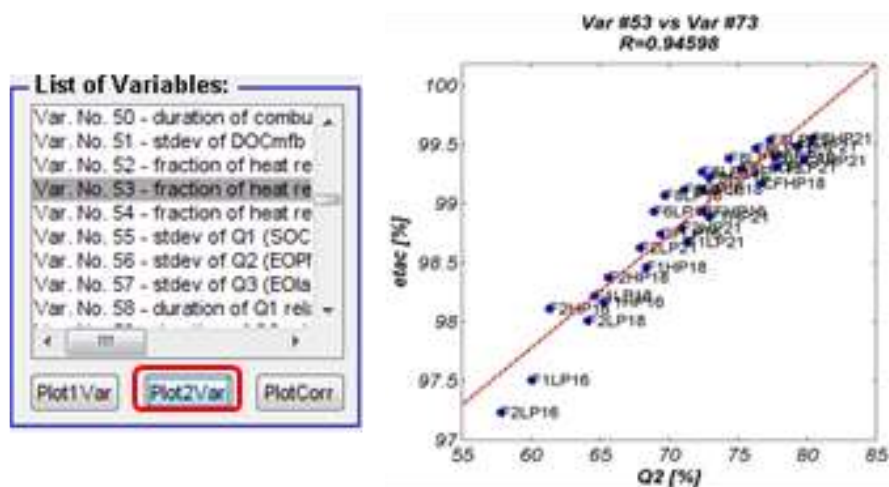


Figure C22 – The result of ‘Plot2Var’ action.

PlotCorr

The user can review correlations among all selected variables after selecting more than one variable in the list and pressing the ‘PlotCorr’ button. As a result the correlation map is displayed with respective colorbar on the right side which explains the color meaning in regards to correlation coefficient between two variables. The diagonal of a correlation matrix always consists of ones. A correlation matrix is always a symmetric matrix.

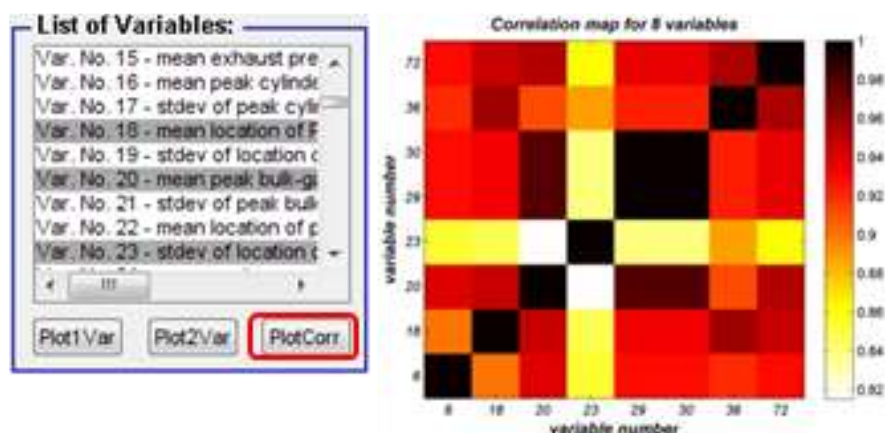


Figure C23 – The result of ‘PlotCorr’ action.

Finally selected variables may be exported to the excel file (*.xls) and submitted for further user selected analysis (see Figure C20b).

C3.6. SAVING IMAGES

The user can generate separate figures such as the scores and loadings plots (C3.3), bar plots, parity plots, and correlation plots (see C3.5). Inside these figures, the user has access to all the functional MATLAB features in the figure such as zooming, panning, printing, editing, etc. (see Figure C24. Description of these tools can be found in Matlab documentation available online (www.mathworks.com)).



Figure C24 – The meaning of the action icons inside the figure window bar.

It is worth mentioning that the ‘save’ button (a disc icon) was reprogrammed in order to export graphs or images to high-quality graphic formats such as the Joint Photographic Experts Group (jpeg), Portable Network Graphics (png) etc. To save a figure the user has to press the ‘disc’ icon, type a file name, and select a file format from the list (as in example in Figure C25).



Figure C25 – The list of all available graphic file formats used to store graphs created in PCAVistool.

C4. SOFTWARE SUPPORT

The collection of functions and algorithms are provided as MATLAB source files (m-files), with no requirements for any add-ins beyond the standard MATLAB installation. The toolbox has been developed under MATLAB 7.9., (MathWorks Inc.), but the functions have been designed for optimal MATLAB 4.2c compatibility.

Please contact Rafal Gieleciak (rgieleci@nrcan.gc.ca) for PCAVisTool support and to report any bugs.

APPENDIX D: BIPLOTS

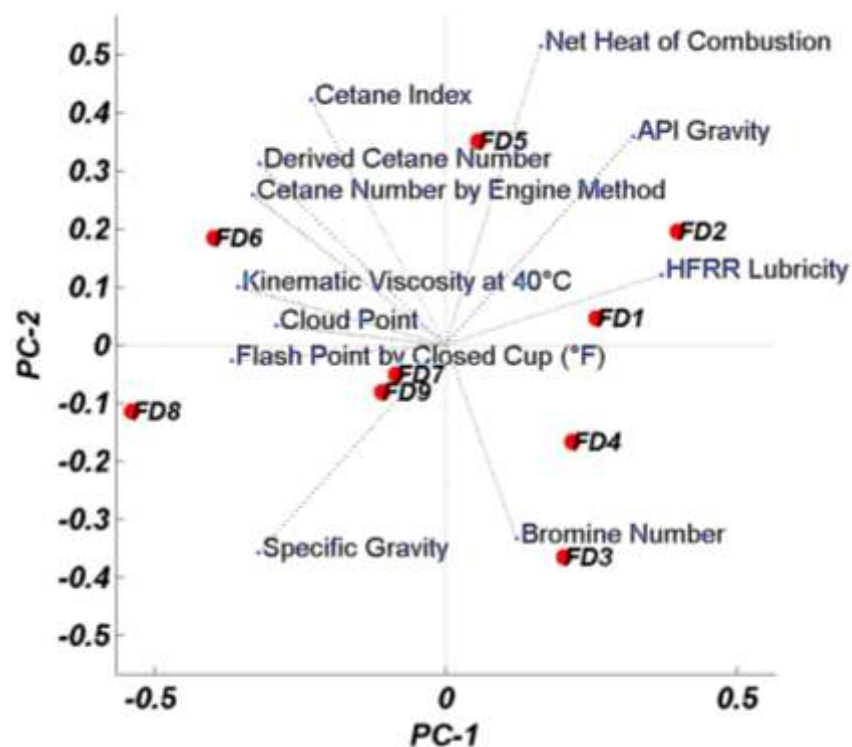


Figure D1 – A sample biplot.

Biplots are usually used in the context of principal component analysis (PCA) as a tool for visualization of the distances between observations (aka. samples, objects) as well as a graphical interpretation of variances and correlations existing among the variables. They are helpful tools that can be used for the explanation of clusters, multicollinearity and outliers seen in the dataset after PCA analysis.

The two axes of biplot represents the principal components (usually the first few PC's corresponding to the largest eigenvalues are chosen). Biplot (see Figure D1) consists of a set of lines and points. Lines represent the variables of the dataset, and points are used to show the objects. The length of the lines approximates the variances of the variables (longer lines represents more important variables). In case of Figure D1 the 'Net Heat of Combustion' has higher variance than 'cloud point' variable. The angle (the cosine of the angle) between the dotted lines approximates the correlation between respective variables. An angle of 0 or 180 degrees represent a variables with a total positive (1) or negative (-1) correlation, respectively. Orthogonal variables are located at angles close to 90 or 270 degree. In Figure D1 a strong relationship is observed between 'Cetane number by Engine Method' and 'Derived Cetane Number', and a weak relationship between 'Cetane Number by Engine Method' and 'API Gravity'. The correlation between 'API Gravity' and 'Specific Gravity' is negative (close to -1).

The distance between two points (red points in Figure D1) approximates the Euclidean distance between two samples in the multidimensional space. Samples located close each other are similar, and vice versa. For instance, in the example biplot fuels FD7 and FD9 are almost identical in this variable space represented by physical properties of FACE diesel fuels. In fact, the interpretation of points in a biplot is the same as in a scores plot.

APPENDIX E: MATRIX CORRELATION (RV-COEFFICIENT)

The simplest way of determining whether two blocks are similar is by inspection of correlation between them; the higher the correlation, the better. Among others there are two most popular: Mantel test and RV-coefficient. In this report the later method was used and will be described here.

The RV-coefficient is a generalization of the squared Pearson correlation coefficient for multidimensional data. Simply speaking, the RV-coefficient measures correlation between two matrices (**A** and **B**) and is defined as:

$$RV(A,B) = \frac{trace(AA'BB')}{\sqrt{trace(AA'AA')trace(BB'BB')}}}$$

where ‘*trace*’ means the sum of matrix diagonal elements and prime (‘) sign transposition of a given matrix. The RV-coefficients values span the [0,1] range, where RV=1 means perfect correlation. The RV values are invariant to rotation and scaling, hence the RV-coefficient can be used on scores obtained after PCA projection.

The significance of the RV-coefficient is determined using Monte Carlo permutations. The order of the vectors in one of the matrices (**A**) is permuted while the other remains unchanged (**B**) and for such scrambled case the RV-coefficient is calculated. This process should be repeated many times (~1000) and the resulting RV values are used to form empirical null distribution. Significance of the RV-coefficient may be estimated by calculating the proportion of the times the Monte Carlo simulation results in a RV-correlation greater than the observed value of RV for the not-scrambled case. These approximate probabilities are denoted by the *p-values*. For example, the RV-coefficient between two sets of data is 0.75 and is greater than 990 (out of 1000) of the random permutations. This means that the empirical significance that these two matrices are correlated is 99% which is reported as a *p-value* of 0.01. The following Figure E1 presents calculated significance of all RV-coefficients calculated for 25 FACE datasets.

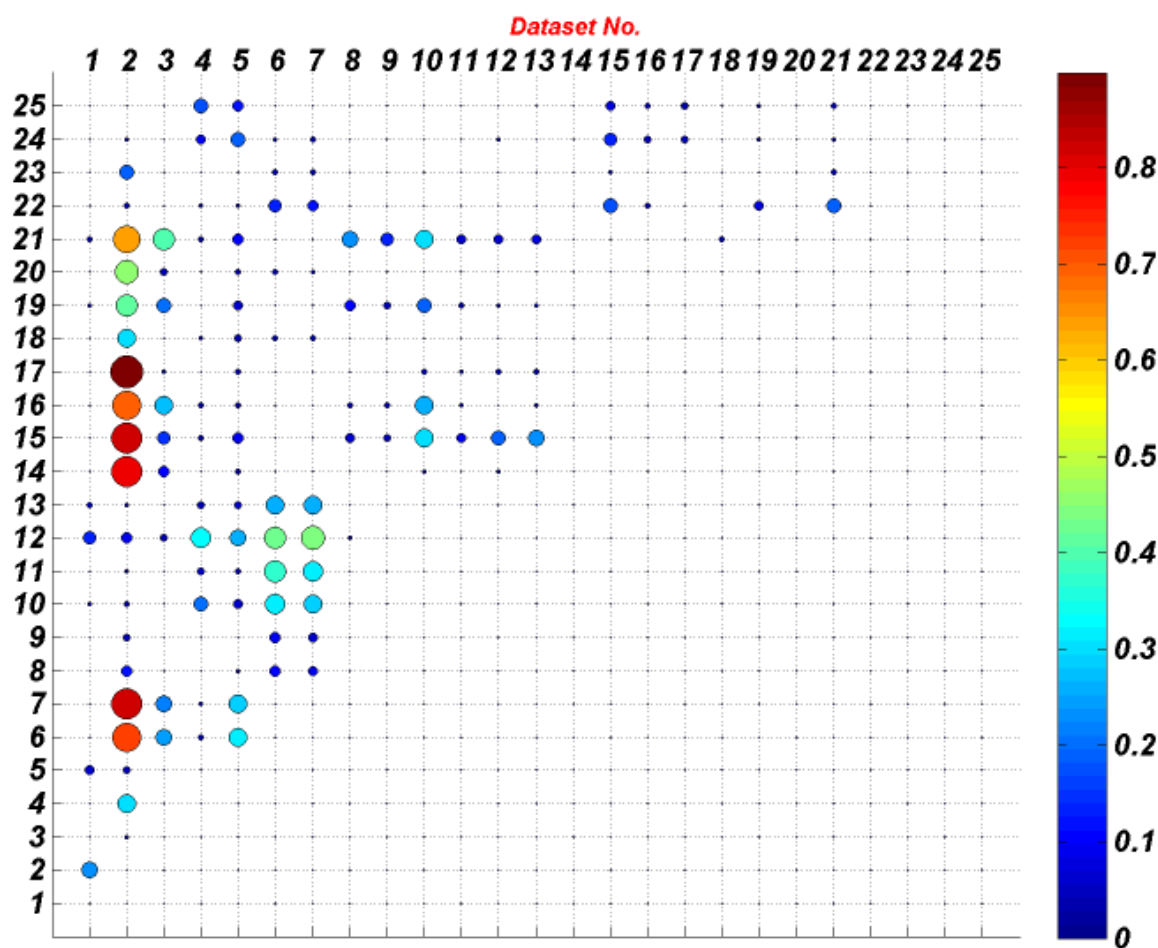


Figure E1 – The estimated significance of RV-coefficient expressed as p-value (see text) calculated for all RVs shown in main text in section 4.0.

The calculations of RV-coefficient were performed in MATLAB using the *rvcoef.m* matlab function available from Multi-block Toolbox for MATLAB available free-of-charge on www.models.life.ku.dk/MBToolbox.

APPENDIX F: PREPROCESSING METHODS

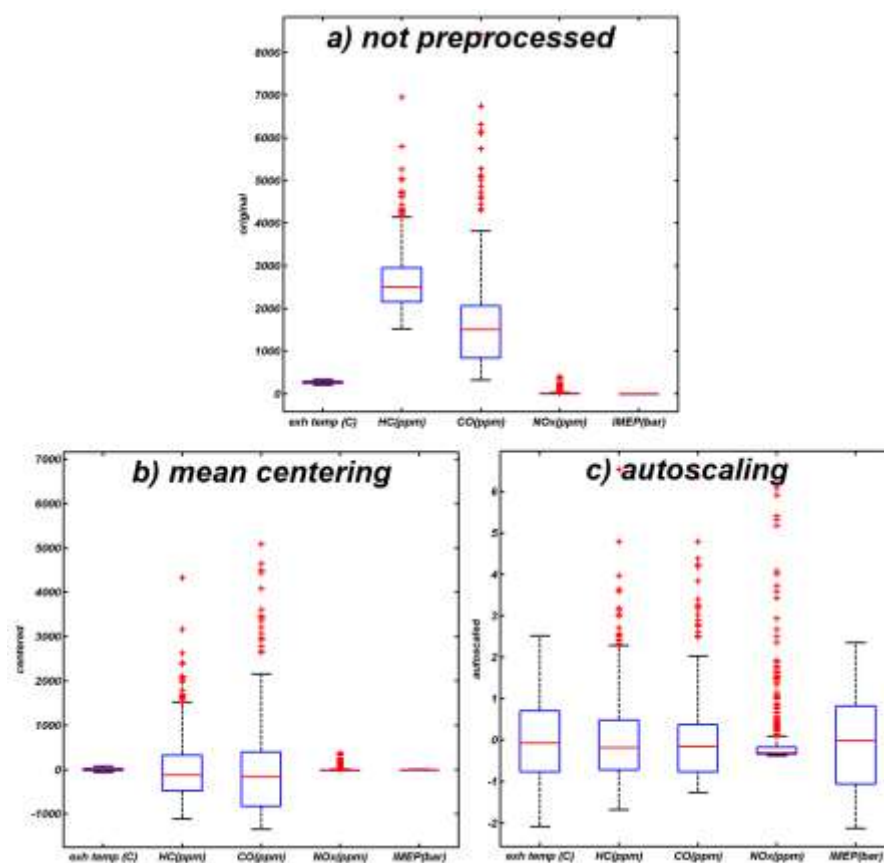


Figure F1 – Effect of mean centering (b) and autoscaling (c) on original set of variables (a). Five variables are shown reported in different units. Carbon monoxide (CO) and Nitrogen oxides (NOx) are reported in the same units (ppm) but they have significantly different value ranges.

Preprocessing is any mathematical operation performed on the original data (i.e. raw data) and refers to the preparation of data for consecutive statistical analysis (such as calibration or pattern recognition). In general, a data preprocessing step is used to correct undesired measurement effects such as data offset or different variable units, in order to increase the quality and interpretability of the results.

In this study we are considered only three data pretreatment methods: centering, autoscaling and logarithmic transformation.

Centering (aka. column mean centering)

Centering is one of the standard steps in most statistical analyses. Mean centering removes the offset from the data and leaves only the relevant variation (variation between

samples). This translates the variance of the data set to be centered about the ordinate axis. In centering, the mean of each column is subtracted from all the values in that column (see Equation F1) resulting in a centered data (c_{ij}) where the average value of each variable is zero after preprocessing.

$$\bar{x}_i = \frac{1}{J} \sum_{j=1}^J x_{ij}$$

$$c_{ij} = x_{ij} - \bar{x}_i \quad \text{Eq F1}$$

Autoscaling (aka. standardization, z-transformation or autoscaling to unit variance)

Autoscaling removes any inadvertent weighting that arises when variables of the data are in different ranges and units. Autoscaling consists of data mean centering followed by standardization (scaling) when data values are divided by standard deviation s_{ij} (see Equation F2). As a result autoscaling gives equal weighting to all measured data variables (z_{ij}) and makes them equally important which can also be a problem when the same weight is given both to “the noisy” variables and to the informative ones. The variance of the autoscaled variables is equal to 1 and its mean value is equal 0.

$$s_i = \sqrt{\frac{\sum_{j=1}^J (x_{ij} - \bar{x}_i)^2}{J-1}}$$

$$z_{ij} = \frac{c_{ij}}{s_i} \quad \text{Eq F2}$$

Logarithmic transformation

The log transformation is applied to the data that do not have a normal distribution. A log transformation will reduce positive skewness in the variable (see Figure F2a). Also log transformation is used to improve linear regression models (especially in the case when a dependent variable is strongly nonlinear).

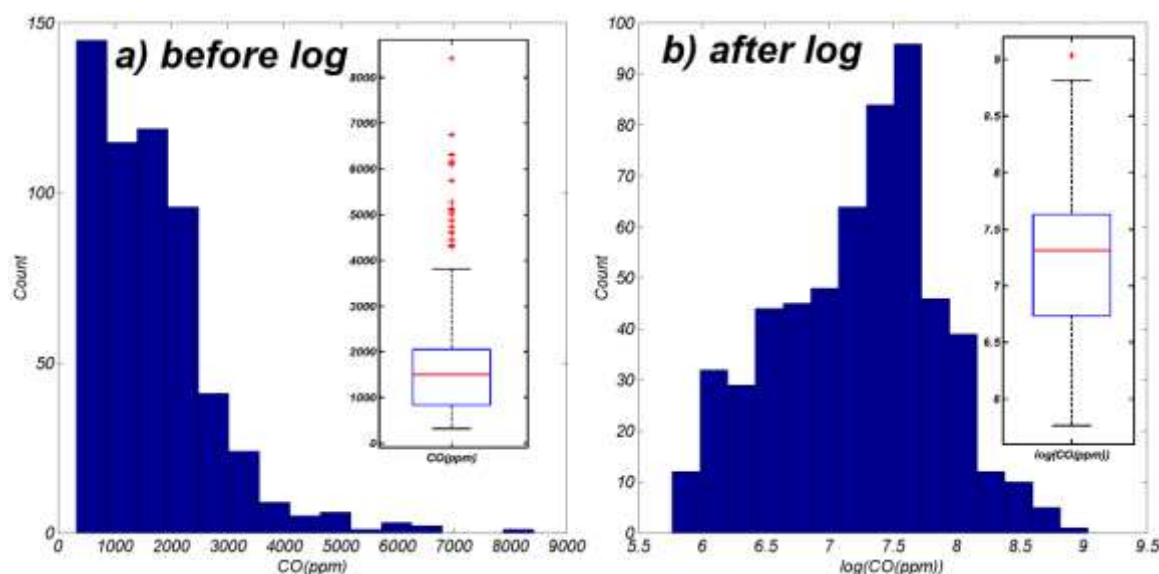


Figure F2 – Effect of logarithmic transformation (b) on original variable (a). Notice a change in variable distribution: after log transformation CO concentration has more or less a normal type distribution.

Transformation for compositional data

Most of the compositional dataset reported in this work were originally provided in normalized form which means that the values of a sample sum up to 100 % which relates to relative concentrations of species (or hydrocarbon types) in the fuel mixture. Such data are called compositional data or closed data and may give artifacts in some methods of data analysis (i.e. PCA) and should be submitted for appropriate data treatment or transformation (36)

APPENDIX G: BOX-PLOTS

In this report we used the box-plot technique for comparison and examination of the univariate distribution of selected engine output parameters in regards to FACE diesel fuel. A boxplot is a way of summarizing a set of data measured on an interval scale and it is often used in exploratory data analysis.

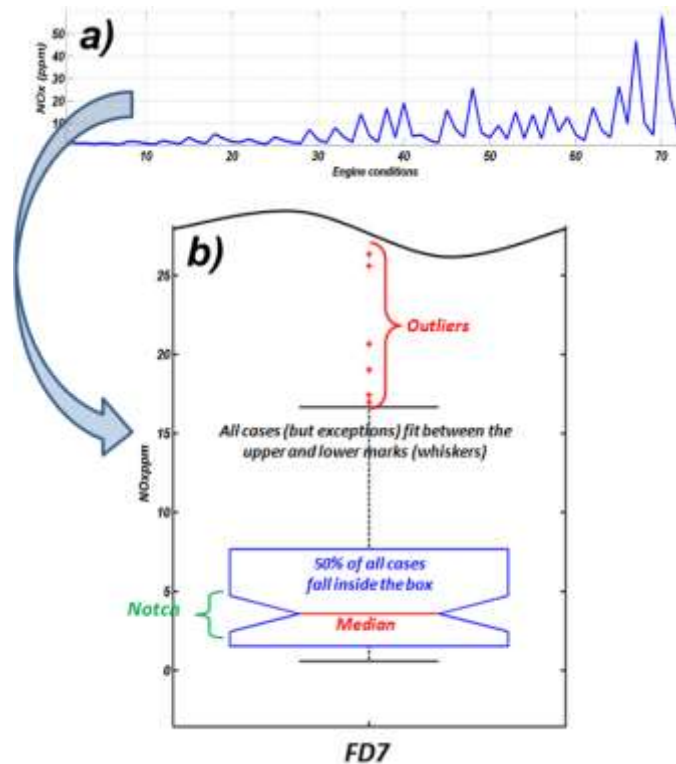


Figure G1 – The result presented in (a) as a profile plot is converted to box-plot (b). Details in the text.

The boxplot presents five simple distribution statistics: the minimum, the lower quartile (q1), the median (q2), the upper quartile (q3) and the maximum in a visual form. Figure G1 shows an example boxplot calculated based on data extracted from one of the engine datasets presented in this study. The tops and the bottoms of the blue rectangle are the first and the third quartiles of the data distribution. The horizontal red line inside box denotes the data median. If the median is not centered in the box, it shows distribution skewness. The dashed lines (whiskers) above and below the rectangle, have a length equal to 1.5 times the interquartile range. Red plus signs on the plot denote the outliers. To emphasize the relative locations, boxplots can be drawn with notches. Notches display the variability (aka. confidence interval) of the median between observations and represent a robust estimate of the uncertainty about the medians for box to box comparison. If two boxes' notches do not overlap there is evidence (95%

confidence) that their medians differ. The advantage of using box-plots is that the quartiles are practically insensitive to outliers.

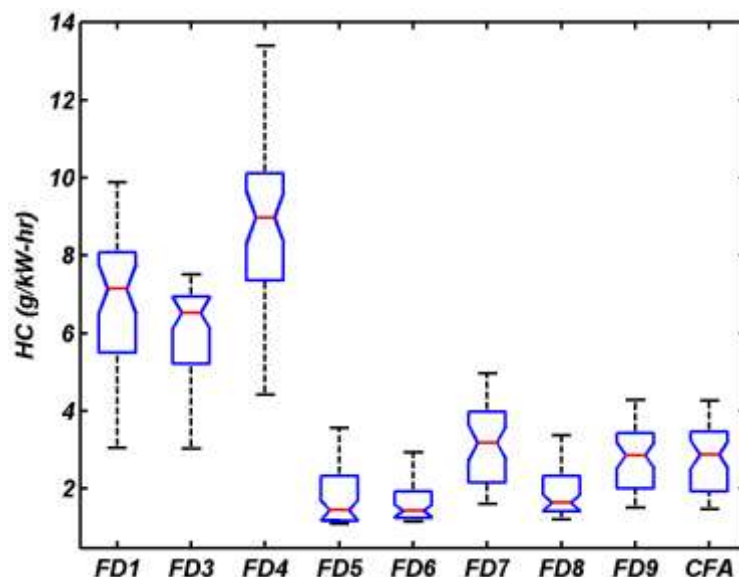


Figure G2 – Comparison of medians across the dataset can be informally performed using the box-plot technique

A careful reader will notice that sometimes box-plots look rather strange (see Figure G3) where notches edges are located outside the box. This is because the length of the notch is bigger than the interquartile range.

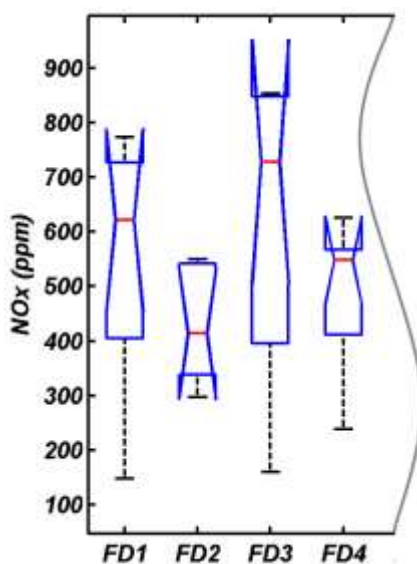


Figure G3 – Example of box-plots with the length of the notches bigger than the interquartile range

APPENDIX H: PLS MODELS DETAILS

Results in this Appendix are presented for strategies 1a and 2a (see section 6.1), consisting of a low number of parameters. The models built for other strategies will not be presented in this report due to their complexity.

Variables (x) important for the modeling of Y (an output parameter) may be identified by large absolute values of PLS regression coefficients, b . Note that variables may also be important for the modeling of X (input parameters), which is characterized by large absolute values of PLS loadings. A summary of the importance of a j th variable for both Y and X is given by the VIP (variable importance for the projection) score calculated from the relation:

$$VIP_j = \sqrt{\frac{\sum_{a=1}^A SS_a \times (w_{aj}/\|w_a\|)^2 \times M}{\sum_{a=1}^A SS_a}}$$

where SS_a is the sum of squares explained by the a th component, A is the number of retained PLS components, w_{aj} is the PLS weight of the j th variable for the a th component, and M is the number of variables. By convention, a variable is considered to be highly influential on the PLS model when its VIP score is above 1.0, moderately influential if the VIP score is within the range of 0.8 to 1.0. In the case where VIP is below 0.8, the variable has a small influence on the PLS model (22).

The following set of tables (Table H1 – Table H27) present the overall statistical description of the models generated (for selected strategies) in this study. Each table provides information about PLS model performance ($q2model$, $rfit$, $RMSECV$, etc. defined in section 2.0), a collection of numerical values of regression coefficients of variables participating in the final PLS equation, as well as variable importance as described by the VIP parameter. We used different colors to differentiate negative (blue) and positive (red) PLS regression coefficients. Influential variables (these with $VIP > 1$) are indicated by highlighting (a red and a bold font) the respective VIP scores values in the tables.

Table H1 – HC models - the engine dataset: HCCI-SCE 0.5L (ORNL). See section 7.1.4

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_6</i>	<i>2a_6</i>	<i>1a_8</i>	<i>2a_8</i>	<i>1a_10</i>	<i>2a_10</i>	<i>1a_12</i>	<i>2a_12</i>
<i>q2model</i>	0.200	0.236	0.027	0.443	0.005	0.159	0.485	0.482	0.598	0.764
<i>rfit</i>	0.464	0.511	0.235	0.705	0.202	0.479	0.723	0.739	0.802	0.895
<i>RMSCV</i>	0.206	0.201	0.242	0.183	0.172	0.158	0.151	0.151	0.172	0.132
<i>RMS</i>	0.204	0.198	0.238	0.174	0.169	0.151	0.145	0.142	0.163	0.121
<i>Aopt</i>	4.0	7.0	1.0	5.0	1.0	6.0	4.0	7.0	4.0	6.0
<i>Input</i>	<i>PLS regression coefficients</i>									
fuel rate (g/min)	0.386	0.343	-0.001	0.031	-0.007	-0.011	-0.002	0.018	-0.029	0.014
airflow (g/min)	0.293	0.313	-0.021	0.064	0.009	0.057	-0.075	-0.425	-0.002	-0.051
lambda (corrected)	0.234	0.247	-0.019	0.051	0.010	0.038	-0.356	-0.180	-0.501	-0.043
int temp actual (°C)	0.295	0.178	0.021	-0.152	-0.015	-0.103	-0.466	-0.625	-0.602	-0.017
intake press (psi gage)	-0.411	-0.438	-0.001	-0.135	-0.011	-0.082	0.321	0.482	0.398	0.022
Cetane Number		-0.124		-0.375		-0.187		0.000		0.268
T90 (°C)		0.021		-0.005		0.017		-0.001		-0.007
Aromatics (wt%)		-0.013		-0.003		0.008		-0.021		0.011
Intercept	7.848	7.848	7.906	7.906	7.803	7.803	7.785	7.785	7.922	7.922
<i>Input</i>	<i>Variable importance in the projection (VIP)</i>									
fuel rate (g/min)	0.68	0.62	0.05	0.85	0.63	1.15	0.89	0.90	0.49	0.52
airflow (g/min)	0.88	0.91	1.33	0.79	0.84	0.79	0.09	0.35	0.23	0.39
lambda (corrected)	1.09	1.11	1.18	0.74	0.96	0.79	0.24	0.29	0.50	0.42
int temp actual (°C)	1.00	1.05	1.35	0.94	1.41	0.92	1.47	1.48	1.67	1.50
intake press (psi gage)	1.26	1.32	0.07	0.89	0.99	0.90	1.42	1.40	1.29	1.15
Cetane Number		1.03		1.50		1.27		1.58		1.80
T90 (°C)		0.81		0.95		0.84		0.40		0.59
Aromatics (wt%)		1.00		1.12		1.20		0.42		0.47

^{a)} Model label (XX_YY) consists of two parts, where XX - means calculation strategy (1a or 2a) and YY – particular engine events used for modelling the given output engine parameter (i.e., *all* - all events; 6, 8, 10, and 12 – only events run for specified fuel flow rate: ~6.3 g/min, ~8.4 g/min, ~10.5 g/min, and ~12.5 g/min, respectively)

Table H2 – CO models - the engine dataset: HCCI-SCE 0.5L (ORNL). See section 7.1.5

	<i>1a_all^{a)}</i>	<i>2a_all</i>	<i>1a_6</i>	<i>2a_6</i>	<i>1a_8</i>	<i>2a_8</i>	<i>1a_10</i>	<i>2a_10</i>	<i>1a_12</i>	<i>2a_12</i>
<i>q2model</i>	0.653	0.724	0.482	0.741	0.654	0.690	0.750	0.803	0.695	0.913
<i>rfit</i>	0.812	0.856	0.710	0.874	0.818	0.846	0.876	0.905	0.848	0.962
<i>RMSCV</i>	0.372	0.332	0.353	0.250	0.352	0.332	0.342	0.303	0.380	0.203
<i>RMS</i>	0.368	0.326	0.346	0.238	0.343	0.318	0.329	0.290	0.364	0.187
<i>Aopt</i>	4.0	7.0	2.0	7.0	2.0	6.0	4.0	5.0	3.0	6.0
<i>Input</i>	<i>PLS regression coefficients</i>									
fuel rate (g/min)	0.330	0.260	0.025	-0.010	-0.047	-0.003	0.035	0.074	0.094	0.029
airflow (g/min)	0.280	0.489	0.107	0.494	0.144	0.224	0.459	0.123	0.093	0.157
lambda (corrected)	0.503	0.523	0.098	-0.067	0.145	0.164	-0.917	0.111	0.084	0.087
int temp actual (°C)	-0.089	-0.132	-0.163	-0.473	-0.209	-0.272	-0.966	-0.331	-0.408	-0.437
intake press (psi gage)	-0.471	-0.614	-0.085	-0.298	-0.115	-0.188	0.471	-0.148	-0.255	-0.286
Cetane Number		-0.125		-0.584		-0.141		0.088		0.094
T90 (°C)		0.179		0.135		0.135		0.180		0.338
Aromatics (wt%)		-0.032		-0.031		-0.048		-0.028		0.038
Intercept	7.225	7.225	7.465	7.465	7.165	7.166	7.000	6.999	7.231	7.232
<i>Input</i>	<i>Variable importance in the projection (VIP)</i>									
fuel rate (g/min)	0.60	0.54	0.20	0.31	0.22	0.39	0.72	0.75	0.45	0.48
airflow (g/min)	0.46	0.41	1.21	1.16	1.19	1.27	0.93	0.95	0.66	0.66
lambda (corrected)	1.28	1.20	1.15	1.08	1.20	1.28	0.88	0.85	0.70	0.69
int temp actual (°C)	1.42	1.30	1.39	1.32	1.36	1.46	1.55	1.61	1.72	1.57
intake press (psi gage)	0.89	0.88	0.51	0.77	0.50	0.47	0.67	0.60	0.96	0.89
Cetane Number		1.71		1.33		1.36		1.60		1.69
T90 (°C)		0.63		0.96		0.54		0.47		0.81
Aromatics (wt%)		0.56		0.62		0.34		0.28		0.30

^{a)} Model labelled (XX_YY) consists of two parts where XX - means calculation strategy (1a or 2a) and YY – particular engine events used for modelling the given output engine parameter (i.e., *all* - all events; 6, 8, 10, and 12 – only events run for specified fuel flow rate: ~6.3 g/min, ~8.4 g/min, ~10.5 g/min, and ~12.5 g/min, respectively)

Table H3 – NO_x models - the engine dataset: HCCI-SCE 0.5L (ORNL). See section 7.1.6

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_6</i>	<i>2a_6</i>	<i>1a_8</i>	<i>2a_8</i>	<i>1a_10</i>	<i>2a_10</i>	<i>1a_12</i>	<i>2a_12</i>	<i>1a_lCN</i>	<i>2a_lCN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.745	0.784	0.485	0.718	0.774	0.793	0.827	0.838	0.706	0.889	0.916	0.918	0.553	0.729
<i>rfit</i>	0.867	0.889	0.723	0.872	0.884	0.902	0.916	0.925	0.858	0.953	0.959	0.961	0.753	0.862
<i>RMSCV</i>	0.729	0.671	0.516	0.382	0.556	0.532	0.593	0.574	0.799	0.491	0.457	0.452	0.601	0.467
<i>RMS</i>	0.719	0.659	0.496	0.352	0.547	0.503	0.571	0.541	0.755	0.447	0.445	0.435	0.591	0.455
<i>Aopt</i>	4.0	7.0	4.0	7.0	3.0	7.0	4.0	7.0	4.0	7.0	4.0	7.0	4.0	7.0
Input	PLS regression coefficients													
fuel rate (g/min)	0.887	1.315	-0.296	0.222	0.230	0.245	0.018	0.030	0.110	0.275	2.957	3.077	-0.353	-0.614
airflow (g/min)	-1.281	-1.588	1.502	-0.386	0.791	0.412	0.113	1.373	0.251	-0.395	-0.538	-0.977	0.124	-0.441
lambda (corrected)	-0.653	-0.449	-0.716	0.705	0.874	1.211	2.037	1.465	2.358	2.111	0.032	0.082	-0.918	-1.441
int temp actual (°C)	0.006	0.101	1.257	1.428	2.824	3.308	3.227	3.997	3.632	2.802	1.050	0.993	0.021	0.251
intake press (psi)	0.589	0.653	-0.650	-0.272	-1.647	-1.598	-2.708	-3.336	-3.031	-2.210	-0.518	-0.209	0.035	0.697
Cetane Number		0.051		0.651		0.520		0.108		-0.382		0.448		0.200
T90 (°C)		-0.319		-0.306		-0.212		-0.140		-0.610		0.356		-0.566
Aromatics (wt%)		-0.093		-0.147		0.039		-0.006		-0.307		-0.024		0.088
Intercept	1.795	1.795	0.650	0.650	1.849	1.848	2.632	2.631	2.621	2.618	2.618	2.618	1.144	1.144
Input	Variable importance in the projection (VIP)													
fuel rate (g/min)	1.02	1.22	0.52	0.67	0.23	0.26	0.63	0.60	0.36	0.48	1.07	1.30	1.04	1.18
airflow (g/min)	0.77	0.63	1.29	1.26	1.31	1.45	1.23	1.40	1.11	1.13	0.78	0.97	0.86	1.01
lambda (corrected)	1.48	1.79	1.16	1.13	1.32	1.46	1.18	1.33	1.16	1.16	1.37	1.70	1.37	1.55
int temp actual (°C)	0.83	0.66	1.27	1.27	1.21	1.34	1.23	1.36	1.40	1.38	0.87	1.06	0.81	1.06
intake press (psi)	0.71	0.91	0.32	0.73	0.16	0.48	0.43	0.37	0.58	0.20	0.78	0.94	0.80	1.03
Cetane Number		1.19		1.15		1.17		1.30		1.50		0.35		0.55
T90 (°C)		0.38		0.96		0.47		0.33		0.91		0.38		0.59
Aromatics (wt%)		0.29		0.54		0.27		0.33		0.39		0.46		0.59

^{a)} Model label (XX_YY) consists of two parts where XX - means calculation strategy (1a or 2a) and YY – particular engine events used for modelling the given output engine parameter (i.e., *all* - all events; 6, 8, 10, and 12 – only events run for specified fuel flow rate: ~6.3 g/min, ~8.4 g/min, ~10.5 g/min, and ~12.5 g/min, respectively; *lCN*, *hCN* – engine events run on low-CN or high-CN fuels, respectively)

Table H4 – ISFC models - the engine dataset: HCCI-SCE 0.5L (ORNL). See section 7.1.8

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_6</i>	<i>2a_6</i>	<i>1a_8</i>	<i>2a_8</i>	<i>1a_10</i>	<i>2a_10</i>	<i>1a_12</i>	<i>2a_12</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.184	0.542	0.149	0.214	0.238	0.708	0.493	0.853	0.106	0.832	0.589	0.637	0.526	0.759
<i>rfit</i>	0.448	0.747	0.440	0.521	0.506	0.863	0.728	0.932	0.405	0.929	0.778	0.814	0.737	0.879
<i>RMSCV</i>	25.068	18.768	11.595	11.141	17.575	10.890	24.491	13.179	37.516	16.281	9.940	9.338	20.225	14.415
<i>RMS</i>	24.809	18.450	11.288	10.724	17.369	10.172	23.582	12.506	36.273	14.716	9.732	8.998	19.867	13.989
<i>Aopt</i>	4.0	7.0	2.0	3.0	3.0	7.0	4.0	6.0	2.0	7.0	4.0	7.0	4.0	7.0
Input	PLS regression coefficients													
fuel rate (g/min)	56.183	50.224	4.144	4.625	4.441	1.359	11.422	5.236	9.682	9.384	53.220	53.146	68.470	53.371
airflow (g/min)	0.697	24.081	0.623	1.034	2.972	39.838	9.362	-6.031	0.613	-21.975	-11.441	-15.222	-50.950	-27.334
lambda (corrected)	33.702	24.165	-0.199	-0.007	-14.186	-22.869	-58.022	-8.717	0.406	48.633	36.218	35.696	19.159	17.192
int temp actual (°C)	13.808	45.671	1.201	0.290	-18.819	34.808	-53.975	16.526	-10.680	59.104	12.027	11.799	9.237	8.805
intake press (psi)	-25.038	-32.618	2.875	2.020	5.980	-9.523	40.548	12.060	-7.705	-28.737	-16.996	-13.466	19.988	7.536
Cetane Number		25.009		-1.046		29.546		50.674		60.353		-17.286		15.609
T90 (°C)		11.244		3.579		9.244		11.452		11.526		-16.244		12.725
Aromatics (wt%)		6.789		-0.666		7.476		12.565		11.227		4.593		11.476
Intercept	269.267	269.270	269.887	269.897	260.502	260.467	268.850	268.738	286.253	286.321	254.470	254.484	279.829	279.767
Input	Variable importance in the projection (VIP)													
fuel rate (g/min)	0.88	0.79	1.70	1.71	1.28	1.04	1.50	1.35	1.01	0.68	1.16	1.17	0.94	1.05
airflow (g/min)	0.80	0.85	0.37	0.48	0.98	0.94	0.37	0.51	0.22	0.29	0.79	1.01	0.94	0.83
lambda (corrected)	1.20	0.64	0.15	0.31	0.94	0.90	0.48	0.45	0.20	0.38	0.90	1.31	0.99	1.10
int temp actual (°C)	1.21	1.18	0.48	0.62	1.14	1.02	1.28	1.12	1.46	0.97	1.23	1.26	1.34	0.83
intake press (psi)	0.82	0.77	1.31	1.35	0.44	0.45	0.86	0.81	1.32	0.86	0.84	1.06	0.68	0.87
Cetane Number		1.39		0.72		1.30		1.52		1.73		0.30		0.83
T90 (°C)		1.33		1.35		1.36		1.12		1.52		0.51		1.63
Aromatics (wt%)		0.74		0.45		0.68		0.49		0.56		0.92		0.47

^{a)} Model label (XX_YY) consists of two parts where XX - means calculation strategy (1a or 2a) and YY – particular engine events used for modelling the given output engine parameter (i.e., *all* - all events; 6, 8, 10, and 12 – only events run for specified fuel flow rate: ~6.3 g/min, ~8.4 g/min, ~10.5 g/min, and ~12.5 g/min, respectively; *ICN*, *hCN* – engine events run on low-CN or high-CN fuels, respectively)

Table H5 – ITE models - the engine dataset: HCCI-SCE 0.5L (ORNL). See section 7.1.9

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_6</i>	<i>2a_6</i>	<i>1a_8</i>	<i>2a_8</i>	<i>1a_10</i>	<i>2a_10</i>	<i>1a_12</i>	<i>2a_12</i>	<i>1a_1CN</i>	<i>2a_1CN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.180	0.541	0.107	0.184	0.172	0.726	0.477	0.847	0.125	0.844	0.542	0.569	0.495	0.750
<i>rfit</i>	0.443	0.746	0.412	0.495	0.474	0.869	0.718	0.932	0.415	0.935	0.749	0.773	0.716	0.874
<i>RMSCV</i>	2.612	1.954	1.312	1.254	2.127	1.225	2.611	1.412	3.522	1.488	1.338	1.298	2.041	1.437
<i>RMS</i>	2.586	1.922	1.265	1.207	2.059	1.158	2.515	1.306	3.425	1.336	1.310	1.253	2.006	1.396
<i>Aopt</i>	4.0	7.0	4.0	3.0	4.0	7.0	4.0	7.0	2.0	7.0	4.0	6.0	4.0	7.0
Input	PLS regression coefficients													
fuel rate (g/min)	-5.366	-5.135	0.081	-0.389	0.124	0.527	-0.760	0.043	-0.717	-0.650	-6.102	-5.534	-5.772	-4.584
airflow (g/min)	-0.684	-3.347	-1.399	-0.126	-1.934	-6.464	-2.531	-5.508	-0.076	0.549	1.270	1.168	4.648	2.211
lambda (corrected)	-3.317	-2.589	2.109	0.032	4.632	5.606	7.916	3.774	-0.030	-3.923	-3.961	-3.671	-1.360	-1.333
int temp actual (°C)	-1.646	-5.212	0.590	-0.036	3.632	-2.985	6.017	-4.708	1.084	-6.121	-1.779	-1.613	-0.883	-0.992
intake press (psi)	2.925	4.015	-0.783	-0.244	-1.800	0.396	-4.437	1.781	0.792	3.591	1.914	1.767	-2.084	-0.705
Cetane Number		-2.533		0.117		-3.512		-5.495		-5.667		2.688		-1.478
T90 (°C)		-1.174		-0.414		-1.090		-1.329		-1.127		2.389		-1.251
Aromatics (wt%)		-0.477		0.229		-0.953		-1.395		-0.870		-0.505		-0.860
Intercept	31.690	31.690	31.375	31.374	32.661	32.663	31.919	31.934	30.051	30.047	33.236	33.235	30.535	30.541
Input	Variable importance in the projection (VIP)													
fuel rate (g/min)	0.82	0.67	1.37	1.17	1.09	0.99	1.37	1.24	0.76	0.63	1.33	1.15	0.89	1.05
airflow (g/min)	0.80	0.77	0.40	0.50	0.95	0.97	0.38	0.54	0.12	0.28	0.93	1.08	0.93	0.82
lambda (corrected)	1.12	0.63	0.57	0.37	1.00	0.95	0.58	0.46	0.06	0.35	0.80	1.18	0.97	1.04
int temp actual (°C)	1.33	1.18	0.65	0.73	1.24	0.97	1.36	1.16	1.59	1.03	0.90	1.15	1.39	0.83
intake press (psi)	0.82	0.78	1.50	1.45	0.61	0.52	0.90	0.87	1.37	0.88	0.96	1.04	0.68	0.86
Cetane Number		1.51		0.77		1.35		1.56		1.75		0.67		0.85
T90 (°C)		1.34		1.63		1.35		1.14		1.46		0.60		1.66
Aromatics (wt%)		0.69		0.60		0.56		0.37		0.60		0.94		0.46

^{a)} Model label (XX_YY) consists of two parts where XX - means calculation strategy (1a or 2a) and YY – particular engine events used for modelling the given output engine parameter (i.e., *all* - all events; *6*, *8*, *10*, and *12* – only events run for specified fuel flow rate: ~6.3 g/min, ~8.4 g/min, ~10.5 g/min, and ~12.5 g/min, respectively; *1CN*, *hCN* – engine events run on low-CN or high-CN fuels, respectively)

Table H6 – HC models - the engine dataset: LTC-MCE 1.9L (HECC-ORNL). See section 7.2.4

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.732	0.920	0.498	0.763	0.826	0.986
<i>rfit</i>	0.867	0.965	0.773	0.912	0.924	0.995
<i>RMSCV</i>	0.247	0.135	0.173	0.119	0.212	0.061
<i>RMS</i>	0.238	0.125	0.155	0.100	0.194	0.052
<i>Aopt</i>	2.0	6.0	3.0	6.0	3.0	6.0
Input	PLS regression coefficients					
aicDynoTorque (ft-lb)	-0.364	-0.315	-0.191	-0.250	-0.562	-0.431
aic_EGR_Mass_GM (%)	0.072	-0.014	0.030	0.016	0.000	0.024
aic_INT_O2_GM (%)	-0.079	-0.041	-0.071	-0.012	-0.022	-0.061
Main SOI (ATDC)	-0.136	0.228	0.090	0.111	-0.116	0.072
Cetane Number		-0.499		-0.198		-0.184
T90 (°C)		-0.069		-0.295		0.000
Aromatics (wt%)		-0.055		-0.120		0.055
Intercept	0.885	0.887	1.139	1.138	0.567	0.567
Input	Variable importance in the projection (VIP)					
aicDynoTorque (ft-lb)	1.55	1.67	1.77	2.09	1.53	1.95
aic_EGR_Mass_GM (%)	0.82	0.92	0.44	0.59	0.56	0.52
aic_INT_O2_GM (%)	0.32	0.41	0.55	0.62	0.16	0.24
Main SOI (ATDC)	0.90	1.07	0.61	0.80	1.15	1.37
Cetane Number		1.32		0.52		0.76
T90 (°C)		0.43		0.45		0.32
Aromatics (wt%)		0.37		0.89		0.55

^{a)} Model label (XX_YY) consists of two parts where XX - means calculation strategy (1a or 2a) and YY – particular engine events used for modelling the given output engine parameter (i.e., *all* - all events; *ICN*, *hCN* – engine events run on low-CN or high-CN fuels, respectively)

Table H7 – CO models - the engine dataset: LTC-MCE 1.9L (HECC-ORNL). See section 7.2.5

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.499	0.904	0.564	0.758	0.721	0.977
<i>rfit</i>	0.730	0.960	0.808	0.903	0.878	0.991
<i>RMSCV</i>	0.309	0.135	0.178	0.132	0.255	0.074
<i>RMS</i>	0.298	0.123	0.158	0.116	0.231	0.063
<i>Aopt</i>	2.0	6.0	3.0	4.0	3.0	4.0
Input	PLS regression coefficients					
aicDynoTorque (ft-lb)	-0.294	-0.224	-0.122	-0.162	-0.417	-0.268
aic_EGR_Mass_GM (%)	0.071	0.041	0.057	0.016	-0.003	0.033
aic_INT_O2_GM (%)	-0.076	-0.067	-0.031	-0.040	-0.056	-0.095
Main SOI (ATDC)	-0.055	0.560	0.230	0.254	0.010	0.232
Cetane Number		-0.707		0.004		-0.172
T90 (°C)		-0.021		-0.007		0.010
Aromatics (wt%)		-0.054		-0.130		0.115
Intercept	2.517	2.519	2.710	2.708	2.275	2.275
Input	Variable importance in the projection (VIP)					
aicDynoTorque (ft-lb)	1.68	1.56	0.77	0.92	1.50	1.86
aic_EGR_Mass_GM (%)	0.76	0.77	0.58	0.60	0.37	0.44
aic_INT_O2_GM (%)	0.37	0.45	0.42	0.69	0.20	0.32
Main SOI (ATDC)	0.68	1.05	1.70	1.87	1.25	1.52
Cetane Number		1.35		0.77		0.75
T90 (°C)		0.70		0.78		0.24
Aromatics (wt%)		0.60		0.78		0.58

^{a)} Model label (XX_YY) consists of two parts where XX - means calculation strategy (1a or 2a) and YY – particular engine events used for modelling the given output engine parameter (i.e., *all* - all events; *ICN*, *hCN* – engine events run on low-CN or high-CN fuels, respectively)

Table H8 – NO_x models - the engine dataset: LTC-MCE 1.9L (HECC-ORNL). See section 7.2.6

	<i>1a_all^{a)}</i>	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.748	0.857	0.526	0.851	0.911	0.946
<i>rfit</i>	0.877	0.938	0.788	0.942	0.963	0.981
<i>RMSCV</i>	0.169	0.127	0.206	0.115	0.067	0.052
<i>RMS</i>	0.162	0.116	0.184	0.100	0.061	0.044
<i>Aopt</i>	3.0	6.0	3.0	5.0	2.0	5.0
<i>Input</i>	<i>PLS regression coefficients</i>					
aicDynoTorque (ft-lb)	0.082	0.061	0.043	0.105	0.097	0.081
aic_EGR_Mass_GM (%)	-0.115	-0.076	-0.107	-0.037	-0.046	-0.058
aic_INT_O2_GM (%)	0.083	0.060	0.064	0.121	0.063	0.069
Main SOI (ATDC)	-0.340	-0.556	-0.251	-0.282	-0.125	-0.154
Cetane Number		0.266		0.029		0.032
T90 (°C)		0.008		-0.001		0.000
Aromatics (wt%)		0.086		0.184		-0.019
Intercept	-0.856	-0.857	-0.680	-0.677	-1.081	-1.081
<i>Input</i>	<i>Variable importance in the projection (VIP)</i>					
aicDynoTorque (ft-lb)	0.70	0.83	0.60	0.84	1.29	1.59
aic_EGR_Mass_GM (%)	1.05	1.01	0.80	0.67	0.31	0.36
aic_INT_O2_GM (%)	0.40	0.56	0.64	0.81	0.46	0.59
Main SOI (ATDC)	1.50	1.59	1.61	1.66	1.43	1.77
Cetane Number		1.25		0.81		0.61
T90 (°C)		0.65		0.84		0.29
Aromatics (wt%)		0.68		1.03		0.63

^{a)} Model label (XX_YY) consists of two parts where XX - means calculation strategy (1a or 2a) and YY – particular engine events used for modelling the given output engine parameter (i.e., *all* - all events; *ICN*, *hCN* – engine events run on low-CN or high-CN fuels, respectively)

Table H9 – PM models - the engine dataset: LTC-MCE 1.9L (HECC-ORNL). See section 7.2.7

	<i>1a_all^{a)}</i>	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.622	0.793	-0.024	0.213	0.517	0.922
<i>rfit</i>	0.805	0.906	0.369	0.592	0.776	0.972
<i>RMSCV</i>	0.074	0.055	0.075	0.066	0.049	0.020
<i>RMS</i>	0.071	0.051	0.069	0.060	0.044	0.017
<i>Aopt</i>	3.0	6.0	2.0	3.0	3.0	5.0
<i>Input</i>	<i>PLS regression coefficients</i>					
aicDynoTorque (ft-lb)	0.040	0.030	-0.009	0.003	0.076	0.041
aic_EGR_Mass_GM (%)	-0.009	0.011	0.002	0.016	-0.011	-0.002
aic_INT_O2_GM (%)	0.020	0.002	0.019	0.015	0.013	0.005
Main SOI (ATDC)	0.082	0.002	0.024	0.019	0.027	-0.014
Cetane Number		0.108		-0.002		0.047
T90 (°C)		0.035		0.009		0.028
Aromatics (wt%)		0.034		0.041		0.017
Intercept	0.199	0.198	0.113	0.114	0.306	0.306
<i>Input</i>	<i>Variable importance in the projection (VIP)</i>					
aicDynoTorque (ft-lb)	0.69	0.77	0.53	0.50	1.53	1.69
aic_EGR_Mass_GM (%)	1.12	1.13	0.09	0.59	0.29	0.21
aic_INT_O2_GM (%)	0.45	0.51	1.33	0.99	0.20	0.39
Main SOI (ATDC)	1.44	1.40	1.39	1.01	1.24	1.13
Cetane Number		1.56		0.47		1.03
T90 (°C)		0.58		0.68		1.16
Aromatics (wt%)		0.37		1.93		0.54

^{a)} Model label (XX_YY) consists of two parts where XX - means calculation strategy (1a or 2a) and YY – particular engine events used for modelling the given output engine parameter (i.e., *all* - all events; *ICN*, *hCN* – engine events run on low-CN or high-CN fuels, respectively)

Table H10 – BTE models (the aicDynoTorque parameter is not taking into consideration in these models) - the engine dataset: LTC-MCE 1.9L (HECC-ORNL). See section 7.2.9

	<i>1a_all^{a)}</i>	<i>2a_all</i>	<i>1a_LCN</i>	<i>2a_LCN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	-0.029	-0.047	-0.130	-0.087	0.598	0.651
<i>rfit</i>	0.167	0.156	0.223	0.386	0.808	0.854
<i>RMSCV</i>	1.187	1.198	0.775	0.760	0.988	0.919
<i>RMS</i>	1.154	1.156	0.710	0.672	0.918	0.811
<i>Aopt</i>	1.0	1.0	1.0	2.0	2.0	2.0
<i>Input</i>	<i>PLS regression coefficients</i>					
aic_EGR_Mass_GM (%)	0.099	0.080	-0.050	-0.094	0.490	0.521
aic_INT_O2_GM (%)	0.046	0.038	0.071	0.036	-0.122	-0.027
Main SOI (ATDC)	-0.104	-0.084	0.140	0.200	-1.220	-1.351
Cetane Number		-0.023		-0.071		0.336
T90 (°C)		0.018		-0.007		0.148
Aromatics (wt%)		-0.021		-0.212		-0.079
Intercept	26.776	26.776	26.787	26.787	26.733	26.721
<i>Input</i>	<i>Variable importance in the projection (VIP)</i>					
aic_EGR_Mass_GM (%)	1.14	1.54	0.52	0.62	0.58	0.77
aic_INT_O2_GM (%)	0.53	0.72	0.74	0.78	0.30	0.35
Main SOI (ATDC)	1.19	1.62	1.48	1.23	1.60	2.13
Cetane Number		0.45		0.91		0.64
T90 (°C)		0.34		0.79		0.28
Aromatics (wt%)		0.41		1.43		0.53

^{a)} Model label (XX_YY) consists of two parts where XX - means calculation strategy (1a or 2a) and YY – particular engine events used for modelling the given output engine parameter (i.e., *all* - all events; *LCN*, *hCN* – engine events run on low-CN or high-CN fuels, respectively)

Table H11 – HC models - the engine dataset: CDC-SCE 0.5L (CDC-ORNL). See section 7.3.4

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.600	0.716	0.621	0.535	0.629	0.836
<i>rfit</i>	0.798	0.877	0.864	0.861	0.805	0.933
<i>RMSCV</i>	0.667	0.562	0.662	0.733	0.532	0.354
<i>RMS</i>	0.636	0.507	0.541	0.546	0.518	0.314
<i>Aopt</i>	3.0	8.0	2.0	2.0	3.0	4.0
<i>Input</i>	<i>PLS regression coefficients</i>					
Throttle (mA)	-0.663	-0.170	-0.370	-0.264	-0.257	-0.128
TCHAR (°C)	0.282	1.557	0.132	-0.032	0.061	0.278
AirFlow (g/s)	-0.131	0.168	-0.075	0.066	0.361	0.257
FuelFlow (g/s)	-0.314	-0.279	-0.450	-0.292	0.270	-0.068
AFRflow (-)	-0.174	-0.072	0.313	0.240	0.011	0.099
VolEff (%)	-0.108	1.717	-0.043	0.082	0.351	0.249
Cetane Number		-0.769		-0.144		-0.203
T90 (°C)		0.297		0.153		0.376
Aromatics (wt%)		0.036		-0.115		0.250
Intercept	5.087	5.092	5.635	5.627	4.721	4.733
<i>Input</i>	<i>Variable importance in the projection (VIP)</i>					
Throttle (mA)	1.37	1.47	1.09	1.33	1.08	1.18
TCHAR (°C)	0.81	0.92	0.84	0.86	0.23	0.75
AirFlow (g/s)	0.57	0.62	0.87	0.95	1.16	1.22
FuelFlow (g/s)	1.34	1.48	1.21	1.42	1.07	1.09
AFRflow (-)	1.04	1.07	1.06	1.32	0.99	1.02
VolEff (%)	0.52	0.57	0.87	0.98	1.15	1.23
Cetane Number		1.13		0.57		0.72
T90 (°C)		0.64		0.60		0.82
Aromatics (wt%)		0.49		0.43		0.79

^{a)} Model label – see Table H10

Table H12 – CO models - the engine dataset: CDC-SCE 0.5L (CDC-ORNL). See section 7.3.5

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.399	0.502	-0.037	-0.103	0.087	0.049
<i>rfit</i>	0.714	0.764	0.237	0.307	0.579	0.553
<i>RMSCV</i>	0.575	0.524	0.509	0.525	0.597	0.610
<i>RMS</i>	0.520	0.478	0.486	0.476	0.510	0.521
<i>Aopt</i>	4.0	4.0	1.0	1.0	5.0	4.0
<i>Input</i>	<i>PLS regression coefficients</i>					
Throttle (mA)	-0.430	0.321	0.046	0.046	-0.547	0.119
TCHAR (°C)	0.211	0.044	-0.005	-0.005	-0.080	0.108
AirFlow (g/s)	-0.273	-0.068	-0.014	-0.014	-0.086	-0.104
FuelFlow (g/s)	1.027	0.276	0.043	0.043	1.702	0.121
AFRflow (-)	0.794	0.697	-0.015	-0.015	1.026	0.488
VolEff (%)	-0.196	-0.047	-0.015	-0.014	0.314	-0.034
Cetane Number		-0.456		-0.038		-0.236
T90 (°C)		0.131		0.036		0.123
Aromatics (wt%)		0.103		-0.019		0.145
Intercept	6.570	6.573	7.143	7.145	6.195	6.203
<i>Input</i>	<i>Variable importance in the projection (VIP)</i>					
Throttle (mA)	0.84	0.75	1.66	1.57	1.04	0.65
TCHAR (°C)	1.14	1.19	0.17	0.16	1.08	0.91
AirFlow (g/s)	1.12	1.16	0.50	0.47	0.87	0.59
FuelFlow (g/s)	0.88	0.70	1.55	1.46	1.16	0.66
AFRflow (-)	0.87	0.94	0.54	0.50	1.05	1.52
VolEff (%)	1.10	1.15	0.52	0.49	0.74	0.63
Cetane Number		1.33		1.31		1.35
T90 (°C)		0.74		1.23		0.91
Aromatics (wt%)		0.81		0.65		1.27

^{a)} Model label – see Table H10

Table H13 – NO_x models - the engine dataset: CDC-SCE 0.5L (CDC-ORNL). See section 7.3.6

	<i>1a_all^{a)}</i>	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.894	0.899	0.822	0.844	0.919	0.944
<i>rfit</i>	0.956	0.960	0.937	0.975	0.967	0.981
<i>RMSCV</i>	0.166	0.162	0.187	0.176	0.140	0.117
<i>RMS</i>	0.150	0.142	0.156	0.099	0.127	0.096
<i>Aopt</i>	5.0	8.0	3.0	5.0	3.0	5.0
<i>Input</i>	<i>PLS regression coefficients</i>					
Throttle (mA)	-0.076	-0.001	-0.216	-0.028	0.058	0.048
TCHAR (°C)	0.809	0.696	0.166	0.137	0.018	0.064
AirFlow (g/s)	0.404	0.350	0.165	0.154	0.110	0.088
FuelFlow (g/s)	0.320	0.207	0.196	-0.090	0.193	0.146
AFRflow (-)	-0.347	-0.361	-0.533	-0.648	-0.378	-0.378
VolEff (%)	0.361	0.277	0.130	0.109	0.057	0.035
Cetane Number		0.029		0.070		-0.082
T90 (°C)		-0.023		-0.058		0.030
Aromatics (wt%)		0.053		-0.010		0.010
Intercept	5.963	5.963	6.204	6.199	5.815	5.818
<i>Input</i>	<i>Variable importance in the projection (VIP)</i>					
Throttle (mA)	1.13	1.26	1.08	1.11	0.99	1.20
TCHAR (°C)	0.70	0.79	0.91	0.94	0.29	0.51
AirFlow (g/s)	0.76	0.91	0.85	0.98	1.00	1.11
FuelFlow (g/s)	1.16	1.35	1.06	1.24	1.10	1.31
AFRflow (-)	1.30	1.50	1.21	1.46	1.32	1.46
VolEff (%)	0.79	0.93	0.85	1.00	0.99	1.14
Cetane Number		0.76		0.70		0.62
T90 (°C)		0.33		0.68		0.32
Aromatics (wt%)		0.57		0.58		0.64

^{a)} Model label– see Table H10

Table H14 – SOOT models - the engine dataset: CDC-SCE 0.5L (CDC-ORNL). See section 7.3.7

	<i>1a_all^{a)}</i>	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.759	0.769	0.691	0.666	0.915	0.950
<i>rfit</i>	0.887	0.902	0.867	0.886	0.965	0.980
<i>RMSCV</i>	0.642	0.629	0.878	0.912	0.306	0.235
<i>RMS</i>	0.603	0.564	0.786	0.732	0.275	0.206
<i>Aopt</i>	5.0	7.0	2.0	3.0	5.0	4.0
<i>Input</i>	<i>PLS regression coefficients</i>					
Throttle (mA)	-0.054	-0.127	0.565	0.594	-0.385	0.222
TCHAR (°C)	1.215	1.099	-0.303	-0.250	0.225	0.251
AirFlow (g/s)	0.825	0.401	0.135	0.039	-0.041	0.002
FuelFlow (g/s)	1.867	1.698	0.757	0.644	1.203	0.343
AFRflow (-)	0.448	0.376	-0.426	-0.340	-0.072	-0.362
VolEff (%)	0.737	0.379	0.088	-0.002	-0.041	-0.044
Cetane Number		0.691		0.042		0.210
T90 (°C)		-0.002		-0.086		0.181
Aromatics (wt%)		0.124		0.270		0.217
Intercept	-0.270	-0.272	-0.598	-0.591	-0.048	-0.051
<i>Input</i>	<i>Variable importance in the projection (VIP)</i>					
Throttle (mA)	1.32	1.58	1.07	1.31	1.09	1.30
TCHAR (°C)	0.59	0.53	0.90	0.88	0.47	0.54
AirFlow (g/s)	0.50	0.43	0.85	0.90	1.01	1.22
FuelFlow (g/s)	1.43	1.72	1.26	1.44	1.16	1.36
AFRflow (-)	1.17	1.39	1.00	1.23	1.08	1.28
VolEff (%)	0.49	0.43	0.85	0.92	1.03	1.25
Cetane Number		0.73		0.66		0.39
T90 (°C)		0.41		0.62		0.44
Aromatics (wt%)		0.51		0.65		0.39

^{a)} Model label – see Table H10

Table H15 – ISFC models - the engine dataset: CDC-SCE 0.5L (CDC-ORNL). See section 7.3.8

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.707	0.705	-0.039	-0.035	0.886	0.922
<i>rfit</i>	0.868	0.877	0.339	0.381	0.960	0.978
<i>RMSCV</i>	10.143	10.176	16.287	16.253	5.696	4.700
<i>RMS</i>	9.293	8.992	15.029	14.775	4.728	3.555
<i>Aopt</i>	4.0	7.0	1.0	1.0	5.0	8.0
Input	PLS regression coefficients					
Throttle (mA)	-11.803	-9.891	-0.968	-0.947	-8.874	1.780
TCHAR (°C)	4.150	1.601	-1.008	-0.986	0.511	2.317
AirFlow (g/s)	-7.787	-6.710	0.904	0.884	-1.065	-2.291
FuelFlow (g/s)	30.333	29.826	-0.783	-0.766	38.878	22.916
AFRflow (-)	33.997	34.669	1.377	1.347	40.082	36.295
VolEff (%)	-6.515	-4.900	0.943	0.922	3.152	3.291
Cetane Number		-5.499		0.840		-0.930
T90 (°C)		-0.615		-0.841		3.285
Aromatics (wt%)		0.275		0.518		3.585
Intercept	241.888	241.914	252.685	252.641	234.677	234.756
Input	Variable importance in the projection (VIP)					
Throttle (mA)	1.08	1.24	0.96	1.03	0.90	0.97
TCHAR (°C)	0.98	1.00	0.99	1.07	0.88	0.77
AirFlow (g/s)	0.89	0.90	0.89	0.96	0.91	1.01
FuelFlow (g/s)	1.15	1.24	0.77	0.83	1.01	0.92
AFRflow (-)	1.02	1.30	1.36	1.47	1.37	1.63
VolEff (%)	0.85	0.86	0.93	1.00	0.82	0.93
Cetane Number		1.01		0.92		0.58
T90 (°C)		0.52		0.92		1.14
Aromatics (wt%)		0.64		0.56		0.66

^{a)} Model label – see Table H10

Table H16 – ITE models - the engine dataset: CDC-SCE 0.5L (CDC-ORNL). See section 7.3.9

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.739	0.741	-0.035	0.105	0.847	0.886
<i>rfit</i>	0.886	0.896	0.322	0.760	0.949	0.968
<i>RMSCV</i>	1.274	1.269	1.930	1.795	0.893	0.771
<i>RMS</i>	1.158	1.109	1.795	1.233	0.722	0.571
<i>Aopt</i>	4.0	7.0	1.0	4.0	5.0	8.0
Input	PLS regression coefficients					
Throttle (mA)	1.807	1.428	0.111	-0.488	1.447	0.032
TCHAR (°C)	-0.703	-0.528	0.119	1.241	-0.065	-0.323
AirFlow (g/s)	1.030	0.776	-0.104	2.049	0.169	0.421
FuelFlow (g/s)	-4.328	-4.163	0.081	-1.871	-5.652	-3.503
AFRflow (-)	-4.548	-4.621	-0.152	-5.021	-5.450	-4.942
VolEff (%)	0.848	0.516	-0.108	1.623	-0.547	-0.632
Cetane Number		0.773		-0.366		0.145
T90 (°C)		0.037		0.316		-0.432
Aromatics (wt%)		-0.070		-0.605		-0.498
Intercept	34.951	34.947	33.385	33.339	35.993	35.982
Input	Variable importance in the projection (VIP)					
Throttle (mA)	1.11	1.15	0.97	1.13	0.92	0.98
TCHAR (°C)	0.99	1.03	1.04	1.21	0.85	0.73
AirFlow (g/s)	0.89	0.92	0.91	0.76	0.91	1.00
FuelFlow (g/s)	1.18	1.17	0.71	1.22	1.04	0.92
AFRflow (-)	0.94	1.26	1.33	1.36	1.36	1.61
VolEff (%)	0.85	0.88	0.95	0.72	0.82	0.93
Cetane Number		1.06		0.84		0.50
T90 (°C)		0.54		0.82		1.25
Aromatics (wt%)		0.78		0.65		0.62

^{a)} Model label – see Table H10

Table H17 – HC models - the engine dataset: LTC-Split Injection (AVFL-16). See section 7.4.4

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.763	0.936	0.561	0.854	0.827	0.898
<i>rfit</i>	0.876	0.968	0.767	0.932	0.912	0.950
<i>RMSCV</i>	0.302	0.158	0.190	0.109	0.156	0.120
<i>RMS</i>	0.299	0.155	0.183	0.104	0.154	0.117
<i>Aopt</i>	2.0	5.0	2.0	3.0	2.0	3.0
Input	PLS regression coefficients					
Main SOI (BTDC)	0.092	-0.098	-0.133	-0.133	-0.024	-0.053
Pilot SOI (BTDC)	0.486	0.305	0.193	0.192	0.336	0.284
Fuel Split (% Pilot)	0.087	0.068	0.032	0.032	0.087	0.083
Cetane Number		-0.416		-0.075		-0.092
T90 (°C)		0.002		0.078		-0.037
Aromatics (wt%)		-0.013		0.004		0.029
Intercept	6.912	6.912	7.660	7.662	6.561	6.561
Input	Variable importance in the projection (VIP)					
Main SOI (BTDC)	0.94	1.03	0.89	0.90	0.35	0.56
Pilot SOI (BTDC)	1.44	1.44	1.46	1.42	1.65	1.73
Fuel Split (% Pilot)	0.20	0.32	0.27	0.26	0.41	0.51
Cetane Number		1.47		1.23		1.21
T90 (°C)		0.40		1.23		0.48
Aromatics (wt%)		0.67		0.24		0.87

^{a)} Model label – see Table H10

Table H18 – CO models - the engine dataset: LTC-Split Injection (AVFL-16). See section 7.4.5

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.728	0.898	0.698	0.790	0.816	0.883
<i>rfit</i>	0.857	0.950	0.847	0.900	0.907	0.943
<i>RMSCV</i>	0.216	0.132	0.103	0.086	0.135	0.108
<i>RMS</i>	0.213	0.129	0.099	0.081	0.133	0.105
<i>Aopt</i>	2.0	4.0	2.0	3.0	2.0	3.0
Input	PLS regression coefficients					
Main SOI (BTDC)	-0.015	-0.142	-0.148	-0.148	-0.061	-0.083
Pilot SOI (BTDC)	0.360	0.236	0.077	0.077	0.290	0.251
Fuel Split (% Pilot)	-0.034	-0.029	-0.025	-0.025	-0.027	-0.025
Cetane Number		-0.275		-0.028		-0.081
T90 (°C)		-0.042		0.025		-0.058
Aromatics (wt%)		-0.029		-0.017		-0.025
Intercept	7.555	7.555	7.999	7.999	7.347	7.348
Input	Variable importance in the projection (VIP)					
Main SOI (BTDC)	0.86	1.06	1.58	1.88	0.32	0.65
Pilot SOI (BTDC)	1.50	1.51	0.66	0.81	1.69	1.81
Fuel Split (% Pilot)	0.12	0.24	0.28	0.33	0.17	0.20
Cetane Number		1.38		0.91		1.12
T90 (°C)		0.42		0.86		0.64
Aromatics (wt%)		0.68		0.36		0.78

^{a)} Model label – see Table H10

Table H19 – NO_x models - the engine dataset: LTC-Split Injection (AVFL-16). See section 7.4.6

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.646	0.793	0.905	0.941	0.721	0.849
<i>rfit</i>	0.809	0.895	0.955	0.973	0.855	0.926
<i>RMSCV</i>	0.138	0.106	0.088	0.070	0.105	0.077
<i>RMS</i>	0.137	0.104	0.086	0.066	0.103	0.075
<i>Aopt</i>	2.0	4.0	2.0	3.0	2.0	4.0
Input	PLS regression coefficients					
Main SOI (BTDC)	0.208	0.262	0.277	0.277	0.166	0.167
Pilot SOI (BTDC)	-0.096	-0.049	-0.046	-0.046	-0.046	-0.045
Fuel Split (% Pilot)	0.047	0.038	0.025	0.025	0.043	0.042
Cetane Number		0.125		0.020		0.008
T90 (°C)		0.037		-0.012		0.059
Aromatics (wt%)		0.064		0.040		0.049
Intercept	3.915	3.914	3.955	3.955	3.896	3.895
Input	Variable importance in the projection (VIP)					
Main SOI (BTDC)	1.49	1.64	1.71	2.32	1.64	1.94
Pilot SOI (BTDC)	0.82	1.05	0.25	0.29	0.34	0.58
Fuel Split (% Pilot)	0.34	0.52	0.15	0.21	0.45	0.54
Cetane Number		1.14		0.45		0.83
T90 (°C)		0.55		0.39		0.62
Aromatics (wt%)		0.61		0.39		0.73

^{a)} Model label – see Table H10

Table H20 – SOOT models - the engine dataset: LTC-Split Injection (AVFL-16). See section 7.4.7

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.665	0.847	0.483	0.762	0.292	0.573
<i>rfit</i>	0.819	0.923	0.717	0.885	0.559	0.772
<i>RMSCV</i>	0.843	0.570	0.613	0.416	0.538	0.418
<i>RMS</i>	0.834	0.560	0.595	0.397	0.531	0.407
<i>Aopt</i>	2.0	4.0	1.0	2.0	1.0	3.0
Input	PLS regression coefficients					
Main SOI (BTDC)	-0.777	-0.397	-0.535	-0.553	-0.203	-0.155
Pilot SOI (BTDC)	-0.593	-0.161	-0.175	-0.146	-0.249	-0.198
Fuel Split (% Pilot)	0.145	0.130	0.180	0.190	0.084	0.096
Cetane Number		0.984		0.170		0.124
T90 (°C)		0.172		-0.111		0.315
Aromatics (wt%)		0.357		0.288		0.181
Intercept	2.373	2.373	0.520	0.517	3.242	3.242
Input	Variable importance in the projection (VIP)					
Main SOI (BTDC)	1.28	1.28	1.57	1.62	1.06	0.95
Pilot SOI (BTDC)	1.15	1.16	0.51	0.51	1.30	1.16
Fuel Split (% Pilot)	0.17	0.25	0.53	0.55	0.44	0.42
Cetane Number		1.49		1.06		0.88
T90 (°C)		0.44		0.91		1.51
Aromatics (wt%)		0.73		0.94		0.73

^{a)} Model label – see Table H10

Table H21 – BTE models - the engine dataset: LTC-Split Injection (AVFL-16). See section 7.4.9

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_ICN</i>	<i>2a_ICN</i>	<i>1a_hCN</i>	<i>2a_hCN</i>
<i>q2model</i>	0.556	0.661	0.306	0.833	0.676	0.710
<i>rfit</i>	0.751	0.820	0.596	0.921	0.828	0.852
<i>RMSCV</i>	0.764	0.667	0.737	0.362	0.592	0.560
<i>RMS</i>	0.756	0.656	0.710	0.344	0.583	0.545
<i>Aopt</i>	2.0	4.0	1.0	2.0	2.0	4.0
<i>Input</i>	<i>PLS regression coefficients</i>					
Main SOI (BTDC)	0.403	0.655	0.317	0.325	0.605	0.597
Pilot SOI (BTDC)	-0.928	-0.714	-0.450	-0.438	-0.667	-0.674
Fuel Split (% Pilot)	-0.294	-0.238	-0.132	-0.122	-0.289	-0.291
Cetane Number		0.539		0.287		-0.149
T90 (°C)		-0.084		-0.341		0.111
Aromatics (wt%)		-0.038		-0.198		-0.210
Intercept	30.089	30.089	29.264	29.257	30.479	30.480
<i>Input</i>	<i>Variable importance in the projection (VIP)</i>					
Main SOI (BTDC)	0.81	1.27	0.97	0.76	1.06	1.39
Pilot SOI (BTDC)	1.47	1.44	1.38	1.05	1.22	1.54
Fuel Split (% Pilot)	0.45	0.62	0.40	0.30	0.62	0.78
Cetane Number		1.14		1.38		0.66
T90 (°C)		0.44		1.45		0.50
Aromatics (wt%)		0.65		0.46		0.65

^{a)} Model label – see Table H10

Table H22 – ISHC models - the engine dataset: Sandia heavy-duty optical engine. See section 7.5.3

	<i>1a_all^{a)}</i>	<i>2a_all</i>	<i>1a_lp</i>	<i>2a_lp</i>	<i>1a_hp</i>	<i>2a_hp</i>
<i>q2model</i>	0.142	0.983	0.093	0.994	0.068	0.995
<i>rfit</i>	0.440	0.993	0.407	0.998	0.382	0.998
<i>RMSCV</i>	0.617	0.087	0.683	0.055	0.553	0.043
<i>RMS</i>	0.598	0.081	0.655	0.051	0.530	0.038
<i>Aopt</i>	1.0	3.0	1.0	3.0	1.0	3.0
Input	PLS regression coefficients					
inj rail press (MPa)	-0.140	-0.123	0.000	0.000	0.000	0.000
O2 mole fract (mol%)	-0.256	-0.245	-0.295	-0.279	-0.221	-0.210
Cetane Number		-0.510		-0.555		-0.464
T90 (°C)		-0.032		-0.032		-0.039
Aromatics (wt%)		-0.172		-0.199		-0.141
Intercept	0.248	0.247	0.391	0.387	0.104	0.101
Input	Variable importance in the projection (VIP)					
inj rail press (MPa)	0.68	0.43	0.00	0.00	0.00	0.00
O2 mole fract (mol%)	1.24	0.78	1.41	0.82	1.41	0.77
Cetane Number		1.68		1.69		1.71
T90 (°C)		0.76		0.70		0.88
Aromatics (wt%)		0.90		0.99		0.85

^{a)} Model label (XX_YY) consists of two parts where XX - means calculation strategy (1a or 2a) and YY – particular engine events used for modelling the given output engine parameter (i.e., *all* - all events; *lp*, *hp* – engine events run at low (80MPa) and high (180MPa) injection pressure, respectively)

Table H23 – ISCO models - the engine dataset: Sandia heavy-duty optical engine. See section 7.5.4

	<i>1a_all^{a)}</i>	<i>2a_all</i>	<i>1a_lp</i>	<i>2a_lp</i>	<i>1a_hp</i>	<i>2a_hp</i>
<i>q2model</i>	0.632	0.928	0.671	0.945	0.581	0.962
<i>rfit</i>	0.809	0.968	0.836	0.977	0.784	0.984
<i>RMSCV</i>	0.299	0.132	0.319	0.131	0.261	0.079
<i>RMS</i>	0.289	0.124	0.305	0.118	0.250	0.071
<i>Aopt</i>	1.0	3.0	1.0	3.0	1.0	3.0
Input	PLS regression coefficients					
inj rail press (MPa)	-0.072	-0.054	0.000	0.000	0.000	0.000
O2 mole fract (mol%)	-0.392	-0.387	-0.469	-0.458	-0.319	-0.314
Cetane Number		-0.246		-0.274		-0.219
T90 (°C)		0.017		0.020		0.010
Aromatics (wt%)		-0.057		-0.043		-0.070
Intercept	0.960	0.960	1.032	1.030	0.887	0.885
Input	Variable importance in the projection (VIP)					
inj rail press (MPa)	0.25	0.31	0.00	0.00	0.00	0.00
O2 mole fract (mol%)	1.39	1.72	1.41	1.80	1.41	1.64
Cetane Number		1.20		1.17		1.24
T90 (°C)		0.48		0.39		0.61
Aromatics (wt%)		0.52		0.47		0.62

^{a)} Model label – see Table H22

Table H24 – ISNOx models - the engine dataset: Sandia heavy-duty optical engine. See section 7.5.5

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_lp</i>	<i>2a_lp</i>	<i>1a_hp</i>	<i>2a_hp</i>
<i>q2model</i>	0.979	0.995	0.974	0.996	0.986	0.995
<i>rfit</i>	0.990	0.998	0.988	0.998	0.994	0.998
<i>RMSCV</i>	0.124	0.061	0.138	0.056	0.102	0.060
<i>RMS</i>	0.120	0.057	0.133	0.051	0.098	0.054
<i>Aopt</i>	1.0	4.0	1.0	3.0	1.0	3.0
Input	PLS regression coefficients					
inj rail press (MPa)	0.075	0.059	0.000	0.000	0.000	0.000
O2 mole fract (mol%)	0.855	0.861	0.852	0.860	0.871	0.873
Cetane Number		-0.111		-0.132		-0.088
T90 (°C)		0.025		0.032		0.008
Aromatics (wt%)		0.004		0.004		0.008
Intercept	0.973	0.972	0.901	0.900	1.050	1.049
Input	Variable importance in the projection (VIP)					
inj rail press (MPa)	0.12	0.20	0.00	0.00	0.00	0.00
O2 mole fract (mol%)	1.41	2.21	1.41	2.20	1.41	2.23
Cetane Number		0.19		0.24		0.15
T90 (°C)		0.11		0.16		0.04
Aromatics (wt%)		0.21		0.27		0.14

^{a)} Model label – see Table H22

Table H25 – FSN models - the engine dataset: Sandia heavy-duty optical engine. See section 7.5.6

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_lp</i>	<i>2a_lp</i>	<i>1a_hp</i>	<i>2a_hp</i>
<i>q2model</i>	0.412	0.857	0.467	0.929	-0.047	0.853
<i>rfit</i>	0.669	0.936	0.712	0.972	0.197	0.941
<i>RMSCV</i>	1.018	0.503	0.784	0.286	1.169	0.438
<i>RMS</i>	0.987	0.469	0.753	0.254	1.120	0.388
<i>Aopt</i>	1.0	3.0	1.0	3.0	1.0	3.0
Input	PLS regression coefficients					
inj rail press (MPa)	-0.720	-0.747	0.000	0.000	0.000	0.000
O2 mole fract (mol%)	-0.510	-0.518	-0.772	-0.789	-0.229	-0.240
Cetane Number		0.767		0.607		0.963
T90 (°C)		-0.225		-0.226		-0.255
Aromatics (wt%)		0.337		0.289		0.391
Intercept	-2.936	-2.932	-2.229	-2.221	-3.697	-3.688
Input	Variable importance in the projection (VIP)					
inj rail press (MPa)	1.15	1.23	0.00	0.00	0.00	0.00
O2 mole fract (mol%)	0.82	0.87	1.41	1.53	1.41	0.45
Cetane Number		1.24		1.17		1.72
T90 (°C)		0.50		0.45		0.73
Aromatics (wt%)		0.97		1.03		1.14

^{a)} Model label – see Table H22

Table H26 – ETAF models - the engine dataset: Sandia heavy-duty optical engine. See section 7.5.7

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_lp</i>	<i>2a_lp</i>	<i>1a_hp</i>	<i>2a_hp</i>
<i>q2model</i>	0.942	0.964	-0.087	0.229	-0.061	0.566
<i>rfit</i>	0.973	0.984	0.032	0.619	0.168	0.805
<i>RMSCV</i>	0.563	0.442	0.653	0.550	0.451	0.289
<i>RMS</i>	0.544	0.415	0.626	0.492	0.432	0.260
<i>Aopt</i>	1.0	3.0	1.0	2.0	1.0	3.0
Input	PLS regression coefficients					
inj rail press (MPa)	2.282	2.300	0.000	0.000	0.000	0.000
O2 mole fract (mol%)	0.028	-0.024	0.020	0.041	-0.075	-0.067
Cetane Number		-0.384		-0.432		-0.351
T90 (°C)		0.190		0.110		0.291
Aromatics (wt%)		0.043		0.061		0.037
Intercept	37.588	37.586	35.363	35.359	39.913	39.908
Input	Variable importance in the projection (VIP)					
inj rail press (MPa)	1.41	2.21	0.00	0.00	0.00	0.00
O2 mole fract (mol%)	0.02	0.06	1.41	0.14	1.41	0.61
Cetane Number		0.24		1.96		1.70
T90 (°C)		0.14		0.74		1.21
Aromatics (wt%)		0.14		0.76		0.53

^{a)} Model label – see Table H22

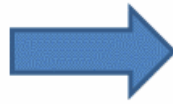
Table H27 – ETAC models - the engine dataset: Sandia heavy-duty optical engine.

	<i>1a_all</i> ^{a)}	<i>2a_all</i>	<i>1a_lp</i>	<i>2a_lp</i>	<i>1a_hp</i>	<i>2a_hp</i>
<i>q2model</i>	0.208	0.892	0.180	0.905	0.141	0.952
<i>rfit</i>	0.505	0.952	0.494	0.961	0.458	0.981
<i>RMSCV</i>	0.547	0.202	0.643	0.219	0.428	0.101
<i>RMS</i>	0.531	0.188	0.617	0.196	0.410	0.090
<i>Aopt</i>	1.0	3.0	1.0	3.0	1.0	3.0
Input	PLS regression coefficients					
inj rail press (MPa)	0.127	0.108	0.000	0.000	0.000	0.000
O2 mole fract (mol%)	0.282	0.271	0.354	0.333	0.214	0.205
Cetane Number		0.464		0.548		0.375
T90 (°C)		-0.039		-0.054		-0.013
Aromatics (wt%)		0.118		0.137		0.093
Intercept	98.839	98.840	98.707	98.711	98.968	98.971
Input	Variable importance in the projection (VIP)					
inj rail press (MPa)	0.58	0.44	0.00	0.00	0.00	0.00
O2 mole fract (mol%)	1.29	0.99	1.41	1.04	1.41	0.95
Cetane Number		1.65		1.67		1.69
T90 (°C)		0.66		0.59		0.80
Aromatics (wt%)		0.82		0.87		0.79

^{a)} Model label – see Table H22

**APPENDIX I: THE LIST OF INPUT/OUTPUT PARAMATERES COLLECTED DURING
SANDIA HEAVY-DUTY OPTICAL ENGINE EXPERIMENTS**

The original spreadsheet contains 238 columns. We split these columns into four groups.



	<i>Fuel descriptors</i>
	<i>Engine input variables</i>
	<i>Engine output variables</i>
	<i>Constant values</i>

No	Variable name [unit]	Variable details
1	Date	data acquisition date [YYMMDD]
2	Text Descr.	text descriptor [-]
3	Num. Descr.	numerical descriptor [-]
4	Process [-]	process data (0=no, ~0=yes) [-]
5	Plot Order [-]	order number for plotting [-]
6	run [-]	run number [-]
7	fnum [-]	fuel/calibration number [-]
8	cycnum	cycle number [-]
9	S [rpm]	engine speed [rpm]
10	fired [-]	0 = motored, 1 = fired [-]
11	SOLi [CAD]	indicated start of injection [CAD]
12	SOLi Delay [μs]	SOLi delay on SRS2 [us]
13	DOli [μs]	indicated duration of inj. [us]
14	fstop [-]	indicated SINL lens f-stop [-]
15	Gnum [-]	SINL detector gain: 1 = low, 2 = med., 3 = high [-]
16	fouling [-]	window fouling affects SINL? 0 = no, ~0 = yes [-]
17	inj. lag [s]	injector lag [s]
18	prail [MPa]	injector rail pressure [MPa]
19	ncharge [mol]	# of moles of charge in cylinder before SOLa [mol]
20	nfuel [mol]	# of moles of fuel injected [mol]
21	nprods [mol]	# of moles of mixture in cylinder after EOC [mol]
22	Nskip [cyc]	# of cycles per fired cycle [-]
23	tmotored [s]	time engine was motored before data acq. began [s]
24	seq period [s/seq]	time required to complete one <SH> sequence [s]
25	Nseq [seq]	# of <SH> sequences over which <PR> data acquired
26	ndotexh [mol/s]	molar flow rate of exh.-gas mixture [mol/s]
27	fW [-]	oxygen equivalence ratio [-]
28	(A/F)cell [-]	raw air/fuel ratio [-]
29	(A/F)corr [-]	air/fuel ratio accounting for oxygen in EGR [-]
30	EGR [%]	EGR rate (by mass) [%]
31	XAir setpt [%]	desired intake-air mole fraction [mol%]
32	XN2 setpt [%]	desired intake-N2 mole fraction [mol%]
33	XCO2 setpt [%]	desired intake-CO2 mole fraction [mol%]
34	XO2 setpt [%]	desired intake-O2 mole fraction [mol%]
35	CRgeo [-]	geometric compression ratio (from calcCR.m) [-]
36	CReff [-]	effective compression ratio (from calcCR.m) [-]
37	LII Filter ND [-]	LII optical filter ND value [-]
38	LII PMT Ctrl V [V]	LII PMT control voltage [V]
39	IMT [°C]	intake manifold temperature [°C]
40	IMAP [bar]	intake manifold abs. pressure [bar]
41	Tpintxd [°C]	intake pxd temperature [°C]
42	Vpintxd [V]	intake pxd voltage [V]

No	Variable name [unit]	Variable details
120	SOLa [CAD]	actual start of injection [CAD]
121	DOla [μs]	actual duration of injection [us]
122	DOla [CAD]	actual duration of injection [CAD]
123	EOla [CAD]	actual end of injection [CAD]
124	SOC [CAD]	start of combustion [CAD]
125	std SOC [CAD]	stdev of SOC [CAD]
126	SOCfit [CAD]	SOC from linear fit through low and high HR thresholds [CAD]
127	stdSOCfit [CAD]	stdev of SOCfit [CAD]
128	igndelay [CAD]	ignition delay = SOC - SOLa [CAD]
129	igndelay [us]	ignition delay [us]
130	EOPMB [CAD]	location of end of premixed burn [CAD]
131	std EOPMB [CAD]	stdev of EOPMB [CAD]
132	mfb [%]	mass fraction burned at EOCmfb [%]
133	EOCmfb [CAD]	end of combustion based on mfb [CAD]
134	std EOCmfb [CAD]	stdev of EOCmfb [CAD]
135	DOCmfb [CAD]	duration of combustion based on SOC and EOCmfb [CAD]
136	std DOCmfb [CAD]	stdev of DOCmfb [CAD]
137	Q1 [%]	fraction of heat rel. from SOC to EOPMB [% abs]
138	Q2 [%]	fraction of heat rel. from EOPMB to EOla [% abs]
139	Q3 [%]	fraction of heat rel. from EOla to EOCmfb [% abs]
140	std Q1 [% abs]	stdev of Q1 (SOC to EOPMB) [% abs]
141	std Q2 [% abs]	stdev of Q2 (EOPMB to EOla) [% abs]
142	std Q3 [% abs]	stdev of Q3 (EOla to EOCmfb) [% abs]
143	Q1dur [%]	duration of Q1 relative to DOC [% abs]
144	Q2dur [%]	duration of Q2 relative to DOC [% abs]
145	Q3dur [%]	duration of Q3 relative to DOC [% abs]
146	std Q1dur [% abs]	stdev of duration of Q1 relative to DOC [% abs]
147	std Q2dur [% abs]	stdev of duration of Q2 relative to DOC [% abs]
148	std Q3dur [% abs]	stdev of duration of Q3 relative to DOC [% abs]
149	gIMEP [bar]	gross IMEP fired cycles (during compr'n and exp'n only) [bar]
150	COV of gIMEP [%]	coeff. of variation of gIMEPf [%]
151	pIMEP [bar]	pumping IMEP (during gas exchange) [bar]
152	nIMEP [bar]	net IMEP = gIMEP + pIMEP [bar]
153	mf [mg]	mass of fuel injected [mg]
154	Wc,ig [J]	gross indicated work per cycle [J]
155	Wc,in [J]	net indicated work per cycle [J]
156	tauiq [N-m]	gross indicated torque [N-m]
157	Qtot [J]	integrated AHRR from SOC to EOC (see compSOC.m) [J]
158	etaf [%]	gross indicated fuel-conversion efficiency [%]
159	etac [%]	indicated combustion efficiency [%]
160	mean PkSINL [V]	mean of peak indicated SINL [V]
161	std PkSINL [V]	stdev of peak indicated SINL [V]

43	EMAP [bar]	exh. manifold abs. pressure [bar]
44	Tpexhxd [°C]	exh. pxd temperature [deg C]
45	Vpexhxd [V]	exh. pxd voltage [V]
46	Texh [°C]	exh.-gas temperature [°C]
47	etavol [%]	calculated volumetric efficiency [%]
48	mdot [g/s]	total mass flow rate of charge-gas [g/s]
49	mdotAir [g/s]	mass flow rate of air [g/s]
50	puAirSet [bar]	setpoint for air pressure upstream of orifice [bar]
51	puAir [bar]	actual air pressure upstream of orifice [bar]
52	TuAir [°C]	air temperature upstream of orifice [°C]
53	pdAir [bar]	air pressure downstream of orifice [bar]
54	mdotN2 [g/s]	mass flow rate of N2 [g/s]
55	puN2Set [bar]	setpoint for N2 pressure upstream of orifice [bar]
56	puN2 [bar]	actual N2 pressure upstream of orifice [bar]
57	TuN2 [°C]	N2 temperature upstream of orifice [°C]
58	pdN2 [bar]	N2 pressure downstream of orifice [bar]
59	mdotCO2 [g/s]	mass flow rate of CO2 [g/s]
60	puCO2Set [bar]	setpoint for CO2 pressure upstream of orifice [bar]
61	puCO2 [bar]	actual CO2 pressure upstream of orifice [bar]
62	TuCO2 [°C]	CO2 temperature upstream of orifice [°C]
63	pdCO2 [bar]	CO2 pressure downstream of orifice [bar]
64	pLubeOil [psig]	lube-oil pressure [psig]
65	TLubeOil [°C]	lube-oil temperature [°C]
66	pLPfuel [psig]	low-pressure fuel pressure [psig]
67	TFuelHX [°C]	fuel temperature in LP heat exchanger [°C]
68	TLinerIn [°C]	coolant temperature at liner inlet [°C]
69	TLinerOut [°C]	coolant temperature at liner outlet [°C]
70	DeltaLinerT [°C]	TLinerOut - TLinerIn [°C]
71	THead [°C]	coolant temperature at head outlet [°C]
72	TpHPfuelxd [°C]	high-pressure fuel pxd temperature [°C]
73	pHPfuel [MPa]	high-pressure fuel pressure [MPa]
74	THPFuelAccum [°C]	fuel temperature in HP accumulator [°C]
75	FLA [%]	dyno load as fraction of full-load current rating [%]
76	FLAAF [%]	mean FLA after firing [%]
77	maxFLAAF [%]	max FLA after firing [%]
78	VTDCmeas [cm^3]	engine clearance volume [cm^3]
79	piston # [-]	piston number [-]
80	cpxdrev [-]	cylinder pxd calibration rev. [-]
81	ipxdrev [-]	intake pxd calibration rev. [-]
82	epxdrev [-]	exhaust pxd calibration rev. [-]
83	Tpcylxd [°C]	cyl. pxd temperature [deg C]
84	f1 [Hz]	Fourier filter start frequency [Hz]

162	mean PkSINL [mW]	mean of peak actual SINL [mW]
163	std PkSINL [mW]	stdev of peak actual SINL [mW]
164	LPkSINL [CAD]	mean location of peak SINL [CAD]
165	std LPkSINL [CAD]	stdev of location of peak SINL [CAD]
166	Cycle 0 Peak SINL [V]	Cycle 0 peak indicated SINL [V]
167	Cycle 0 Peak SINL [mW]	Cycle 0 peak actual SINL [mW]
168	SINL @ EVO [mW]	average SINL curve value at EVO [mW]
169	Mean Pre-SOC SINL [mW]	average SINL signal before SOC [mW]
170	Int of SINL (SOC to EVO) [u]	integral of SINL reference curve [J]
171	mfb1 [%]	1st mfb threshold [%]
172	mfb2 [%]	2nd mfb threshold [%]
173	mfb3 [%]	3rd mfb threshold [%]
174	mfb4 [%]	4th mfb threshold [%]
175	mfb5 [%]	5th mfb threshold [%]
176	mfb6 [%]	6th mfb threshold [%]
177	mfb7 [%]	7th mfb threshold [%]
178	mfb8 [%]	8th mfb threshold [%]
179	mfb9 [%]	9th mfb threshold [%]
180	mfb10 [%]	10th mfb threshold [%]
181	mfb11 [%]	11th mfb threshold [%]
182	CAmfb1 [CAD]	1st mfb threshold location [CAD]
183	CAmfb2 [CAD]	2nd mfb threshold location [CAD]
184	CAmfb3 [CAD]	3rd mfb threshold location [CAD]
185	CAmfb4 [CAD]	4th mfb threshold location [CAD]
186	CAmfb5 [CAD]	5th mfb threshold location [CAD]
187	CAmfb6 [CAD]	6th mfb threshold location [CAD]
188	CAmfb7 [CAD]	7th mfb threshold location [CAD]
189	CAmfb8 [CAD]	8th mfb threshold location [CAD]
190	CAmfb9 [CAD]	9th mfb threshold location [CAD]
191	CAmfb10 [CAD]	10th mfb threshold location [CAD]
192	CAmfb11 [CAD]	11th mfb threshold location [CAD]
193	std CAmfb 1 [CAD]	stdev of 1st mfb thresh. loc. [CAD]
194	std CAmfb 2 [CAD]	stdev of 2nd mfb thresh. loc. [CAD]
195	std CAmfb 3 [CAD]	stdev of 3rd mfb thresh. loc. [CAD]
196	std CAmfb 4 [CAD]	stdev of 4th mfb thresh. loc. [CAD]
197	std CAmfb 5 [CAD]	stdev of 5th mfb thresh. loc. [CAD]
198	std CAmfb 6 [CAD]	stdev of 6th mfb thresh. loc. [CAD]
199	std CAmfb 7 [CAD]	stdev of 7th mfb thresh. loc. [CAD]
200	std CAmfb 8 [CAD]	stdev of 8th mfb thresh. loc. [CAD]
201	std CAmfb 9 [CAD]	stdev of 9th mfb thresh. loc. [CAD]
202	std CAmfb 10 [CAD]	stdev of 10th mfb thresh. loc. [CAD]
203	std CAmfb 11 [CAD]	stdev of 11th mfb thresh. loc. [CAD]

85	f2 [Hz]	Fourier filter end frequency [Hz]
86	filter xmission @ f2 [%]	Fourier filter transmission at end frequency [%]
87	pfuel [kg/m3]	fuel density [kg/m3]
88	LHVfuel [MJ/kg]	lower heating value of fuel [MJ/kg]
89	NC [cyc]	number of cycles in cycnum vector [-]
90	pf,BI [MPa]	mean fuel rail pressure before injection [MPa]
91	std pf,BI [MPa]	stdev of fuel rail pressure before injection [MPa]
92	pf,AI [MPa]	mean fuel rail pressure after injection [MPa]
93	std pf,AI [MPa]	stdev of fuel rail pressure after injection [MPa]
94	pegp [bar]	mean peg pressure [bar]
95	pexh [bar]	mean exhaust pressure [bar]
96	PCP [bar]	mean peak cylinder pressure (PCP) [bar]
97	std PCP [bar]	stdev of peak cylinder pressure [bar]
98	LPCP [CAD]	mean location of PCP [bar]
99	std LPCP [CAD]	stdev of location of PCP [bar]
100	Tmax [K]	mean peak bulk-gas-avg. in-cylinder temperature [K]
101	std Tmax [K]	stdev of peak bulk-gas-avg. in-cylinder temperature [K]
102	LTmax [CAD]	mean location of peak bulk-gas-avg. in-cylinder T [CAD]
103	std LTmax [CAD]	stdev of location of peak bulk-gas-avg. in-cylinder T [CAD]
104	dp/dCA max [MPa/ms]	mean peak pressure-rise rate (dpdCA) [MPa/ms]
105	dp/dCA max [bar/CAD]	mean peak dpdCA [bar/CAD]
106	std dp/dCA max [MPa/ms]	stdev of peak dpdCA [MPa/ms]
107	Ldp/dCA max [CAD]	mean location of peak dpdCA [CAD]
108	std Ldp/dCA max [CAD]	stdev of location of peak dpdCA [CAD]
109	HRmx [MW]	mean peak AHRR (dQdCA) [MW]
110	HRmx [J/CAD]	mean peak AHRR [J/CAD]
111	std HRmx [MW]	stdev of peak AHRR [MW]
112	LHRmx [CAD]	mean location of peak AHRR [CAD]
113	std LHRmx [CAD]	stdev of location of peak AHRR [CAD]
114	HRmx1 (avg. curve) [J/CAD]	1st HR peak value (from avg. HR curve) [J/CAD]
115	LHRmx1 (avg. curve) [CAD]	loc. of 1st HR peak (from avg. HR curve) [CAD]
116	HRmx2 (avg. curve) [J/CAD]	2nd HR peak value (from avg. HR curve) [J/CAD]
117	LHRmx2 (avg. curve) [CAD]	loc. of 2nd HR peak (from avg. HR curve) [CAD]
118	ncmpfit [-]	polytropic exponent during compression [-]
119	nexpfit [-]	polytropic exponent during expansion [-]

204	NOx [ppm]	exhaust NOx mole fraction [ppm]
205	HC [ppm]	exhaust HC mole fraction [ppm]
206	CO [ppm]	exhaust CO mole fraction [ppm]
207	CO2 [%]	exhaust CO2 mole fraction [%]
208	O2 [%]	exhaust O2 mole fraction [%]
209	Paper Blackening [-]	smoke meter paper blackening [-]
210	Smoke [FSN]	filter smoke number [FSN]
211	O2 Before Firing [%]	O2 mole fraction before firing [%]
212	CO Before Firing [ppm]	CO mole fraction before firing [ppm]
213	HC Before Firing [ppm]	HC mole fraction before firing [ppm]
214	CO2 Before Firing [%]	CO2 mole fraction before firing [%]
215	ISNOx [g/kW-hr]	indicated specific NOx emissions [g/kW-hr]
216	ISHC [g/kW-hr]	indicated specific HC emissions [g/kW-hr]
217	ISCO [g/kW-hr]	indicated specific CO emissions [g/kW-hr]
218	LII Before Firing [V]	raw LII voltage (i.e., raw PMT signal) before firing [V]
219	LII [a.u.]	exhaust LII signal adjusted for ND and PMT gain [a.u.]
220	gIMEPm [bar]	gross IMEP motored cycles [bar]
221	COV of gIMEPm [%]	coeff. of variation of gIMEPm [%]
222	Wc,ig,m [J]	gross indicated work per motored cycle [J]
223	pC[mg/m3]	carbon density from avl smoke meter[mg/m3]
224	mdotFuel [kg/s]	mass flow rate of fuel [kg/s]
225	EINOx[g/kg_fuel]	emission index NOx emissions [g/kg_fuel]
226	EIHC [g/kg_fuel]	emission index HC emissions [g/kg_fuel]
227	EICO [g/kg_fuel]	emission indexCO emissions [g/kg_fuel]
228	EISoot [g/kg_fuel]	indicated specific Soot emissions [g/kg_fuel]
229	mdotSoot[g/sec]	mass flow rate of soot [g/s]
230	ISSoot [g/kW-hr]	indicated specific Soot emissions [[g/kW-hr]
231	H at EOPMB [mm]	lift-off length at EOPMB [mm]
232	Mean H [mm]	mean lift-off length from EOPMB to EOla [mm]
233	H at EOla [mm]	lift-off length at EOla [mm]
234	phiH at EOPMB [-]	f at lift-off length at EOPMB [-]
235	Mean phiH [-]	mean f at lift-off length from EOPMB to EOla [-]
236	phiH at EOla [-]	f at lift-off length at EOla [-]
237	Mean LL [mm]	mean liquid length over injection event [mm]
238	Descr. for Baseline Cond. [-]	numerical descriptor for baseline condition [-]

