



Octane prediction from infrared spectroscopic data

| | |
|----------------|---|
| Item Type | Article |
| Authors | Al Ibrahim, Emad; Farooq, Aamir |
| Citation | Al Ibrahim, E., & Farooq, A. (2019). Octane Prediction from Infrared Spectroscopic Data. Energy & Fuels. doi:10.1021/acs.energyfuels.9b02816 |
| Eprint version | Post-print |
| DOI | 10.1021/acs.energyfuels.9b02816 |
| Publisher | American Chemical Society (ACS) |
| Journal | Energy & Fuels |
| Rights | This document is the Accepted Manuscript version of a Published Work that appeared in final form in Energy and Fuels, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see https://pubs.acs.org/doi/10.1021/acs.energyfuels.9b02816 . |
| Download date | 01/10/2022 21:18:49 |
| Link to Item | http://hdl.handle.net/10754/660233 |

Octane prediction from infrared spectroscopic data

Emad Al Ibrahim* and Aamir Farooq*

*Clean Combustion Research Center (CCRC), Physical Sciences and Engineering Division,
King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi
Arabia*

E-mail: emad.ibrahim@kaust.edu.sa; aamir.farooq@kaust.edu.sa

Abstract

A model for the prediction of research octane number (RON) and motor octane number (MON) of hydrocarbon mixtures and gasoline-ethanol blends has been developed based on infrared spectroscopy data of pure components. Infrared spectra for 61 neat hydrocarbon species were used to generate spectra of 148 hydrocarbon blends by averaging the spectra of their pure components on a molar basis. The spectra of 38 FACE (Fuels for Advanced Combustion Engines) gasoline blends were calculated using PIONA (Paraffin, Isoparaffin, Olefin, Naphthene, and Aromatic) class averages of the pure components. The study sheds light on the significance of dimensional reduction of spectra and shows how it can be used to extract scores with linear correlations to the following important features: molecular weight, paraffinic CH₃ groups, paraffinic CH₂ groups, paraffinic CH groups, olefinic -CH=CH₂ groups, naphthenic CH-CH₂ groups, aromatic C-CH groups, ethanolic OH groups, and branching index. Both scores and features can be used as input to predict octane numbers through nonlinear regression. Artificial Neural Network (ANN) was found to be the optimal method where the mean

absolute error on a randomly selected test set was within the experimental uncertainty of RON, MON, and octane sensitivity.

1. Introduction

Modern efforts in the transportation sector have been guided by the clash of increasing demand for mobility and climate change. The general consensus on the viability of meeting that demand while decreasing greenhouse emissions is through the development of more efficient vehicles. Co-optimization of fuels and engines has the potential to significantly enhance the performance of internal combustion engines. Fuel blends and future fuel formulations need to be tailored to specific engine design. For light duty passenger vehicles, the preferred fuel is gasoline which is made of hundreds of hydrocarbon molecules. Depending on the composition of gasoline, its combustion properties can be widely different. One of the most important properties of gasoline is its octane rating: a measure of the ignition quality of fuels. Fuels with high octane numbers can withstand higher compression ratios in spark-ignited (SI) internal combustion (IC) engines and thus are more efficient. Octane number is typically measured in a standardized single cylinder, four-stroke cycle, variable compression ratio, Cooperative Fuel Research (CFR) engine. The knocking tendency of test fuel is compared to a primary reference fuel (PRF) blend, a mix of 2,2,4-trimethylpentane (iso-octane) and n-heptane. If the CFR engine is run at standard conditions specified by ASTM D2699 [1], it gives the research octane number (RON), and if the engine is run with preheated fuel at higher engine speed as specified by ASTM D2700 [2], it gives the motor octane number (MON). These two standards are widely accepted in the combustion research community and are used heavily in literature. The difference of RON and MON is known as the octane sensitivity (S) of the fuel.

Using a CFR engine is not ideal for testing new fuel formulations due to its cost (\$200,000 price-tag, maintenance and calibration), size, time (20 minutes per test) and difficulty of op-

eration. Many efforts have been made to create models which can reduce the dependency on CFR engine for fuel testing. Efforts started with drawing simple correlations to predict octane number for pure components through quantitative structure property relations (QSPR) [3, 4]. Blending rules for mixtures of hydrocarbons, like PRFs, toluene PRF mixtures (TPRF), and ethanol blends, were developed based on simulated ignition delay times (IDT) [5], linear/nonlinear by-mole mixing [6], and response surface model mapping [7]. Similar efforts were done for naphtha [8] and gasoline [9] using multiple linear regression and nonlinear models, respectively, based on gas-chromatography composition. These studies explored the basics of octane prediction and blending behavior; however, these were limited to specific mixtures and thus could not be generalized.

Another frontier of research for fuel property prediction comprises of chemometric based studies. Analytical techniques such as FT-Raman spectroscopy [10], infrared spectroscopy [11–18], near-IR and laser-induced fluorescence [19], dispersive fiber-optic Raman spectroscopy [20], dielectric spectroscopy [21], flame spectroscopy [22], thermal wave interferometer [23], and distillation curves [24] have been used to predict octane number by multivariate methods. The standard in all of these studies is the partial least square regression (PLSR) algorithm or principal component regression (PCR). A review of advances in characterization of different fuels based on vibrational spectroscopy was reported by Kiefer [25]. He pointed out that most studies rely on established methods like PCR and PLSR but new approaches, including support vector machines (SVM) and artificial neural networks (ANN), are now being used to derive a multitude of fuel characteristics from spectroscopic data. A model using support vector regression (SVR) was shown by Kiefer and co-authors in a study where they related Raman difference spectroscopy to algal production of complex polysaccharides [26], concluding that SVR is capable of extracting important nonlinear features. Another study showed that SVR is far superior to PLS in predicting diesel flash point and cetane number from near-IR spectroscopy [27]. Another work compared ANN performance to PLS and found that ANN is superior even while using a smaller training set and spectral region

[28]. It also highlighted the importance of selecting a good spectral region since it was not viable to apply ANN on the entire spectra.

Recent efforts to predict physical and chemical properties of complex fuels have been based on surrogate formulation [29, 30]. A real fuel is made of hundreds of hydrocarbon molecules which makes it difficult to model all components of real fuels. A number of studies have proposed using a few representative molecules instead. It has been shown that simple blends can capture, to a good degree, gasoline thermo-physical and chemical kinetic properties. A more fundamental method of capturing the contribution of all constituents of a real fuel is the functional group approach. An ANN based functional group model was applied to pure components and gave reasonable performance [31]. A model was then developed by Jameel et al. [32] identifying 9 important features: paraffinic CH, paraffinic CH₂, paraffinic CH₃, olefinic -CH=CH₂, naphthenic CH-CH₂, aromatic C-CH, ethanolic OH, branching index (BI), and molecular weight (MW). These nine features can be calculated using knowledge of the pure components, and then averaged for blends by their composition. ¹H-NMR data were used to calculate the features for distillate fuels, such as FACE (Fuels for Advanced Combustion Engines) gasoline. Jameel et al. [32] used these features as inputs to an artificial neural network (ANN) and obtained results with errors almost within the experimental uncertainty, while the use of multiple linear regression (MLR) was shown to be inadequate at capturing nonlinear blending behavior. Depending on the components being blended, the effect on physical/chemical properties could be synergistic or antagonistic. Daly et al. [18] expressed the same concern in which they found that blending octane numbers are highly sensitive and nonlinear to the addition of certain species to their model. This hinders the development of predictive models and surrogate formulation. The nonlinear nature of the problem has led many researchers to employ ANNs in order to relate fuel properties to various variables [33–37]. However, majority of studies that make use of analytical techniques have employed linear methods (PLSR or PCR).

In this work, we investigate the application of nonlinear methods on gas-phase IR spec-

tra of hydrocarbons and ethanol to predict octane numbers of complex fuel mixtures. We first build a dataset of pure hydrocarbon components, comprising of their IR spectra, octane numbers and molecular properties. The spectra of fuel mixtures were calculated from their respective compositional breakdown on a molar basis. Class averaged spectra of pure components were used to represent PIONA (Paraffin, Isoparaffin, Olefin, Naphthene, and Aromatic) of distillate gasolines and to calculate their IR spectra. The model developed herein operates under the assumption that the composite spectra of a mixture can be calculated as the molar sum of the spectra of the pure components. This is a reasonable assumption for gas-phase spectra [38], in contrast to liquid spectra where excess absorbance has been reported previously, particularly for ethanol blends [39].

In our effort to bridge the gaps in literature, we first show that the IR spectra possess enough information to predict the 9 features specified by the functional group approach for hydrocarbon blends and FACE/ethanol blends. We demonstrate that these features can then be used as input to an artificial neural network (ANN), as proposed by Jameel et al. [32]. We then discuss the bigger picture of relating octane numbers to spectra, and discuss the different methods that can capture its nonlinear nature (see Fig. 1).

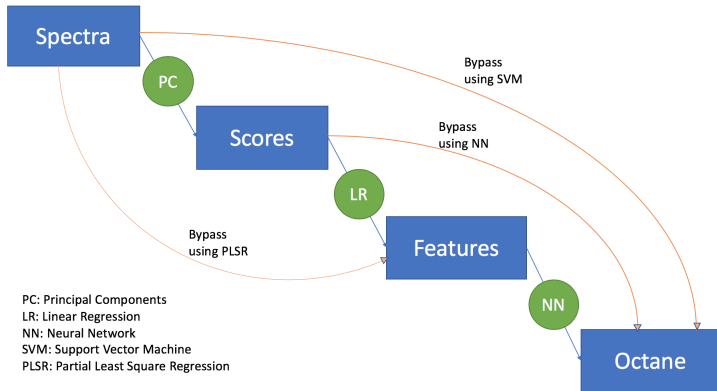


Figure 1: Model road-map

Since working with high-dimensional data like IR spectra makes the use of ANN unfeasible, we propose dimensional reduction to filter out unnecessary parts of the data. This eliminates the limitation encountered by Santos et al. [28] where only parts of the data were

considered. Dimensional reduction is done to extract relevant scores in a reduced subspace, which are then used as input to ANN to capture the nonlinear relation between scores and octane numbers. A similar approach was attempted by Brudzewski et al. [40] but the task was formulated as a classification problem. The fuels were split into six classes based on their octane numbers and, guided by the PC (principal component) of the spectra, the fuels were classified using SVM (support vector machine). Although they reported good classification accuracy, they concluded that further work is needed to assess the correlations between IR spectra and octane numbers. We believe that our predictive model addresses that need and provides a new tool in chemometrics.

2. Methodology

Data description:

Data for 61 pure components with values of RON, MON, molecular weight, PIONA class (Paraffin, Isoparaffin, Olefin, Naphthene, and Aromatic), and 5900 absorbance data points in the wavenumber range $600 - 6500\text{ cm}^{-1}$ were collected from literature (see Table 1 for details and Fig. 2 for the spectra). In Table 1, the RON and MON values are taken from the respective references listed for each pure component. The gas-phase spectra came from the Pacific Northwest National Laboratory (PNNL) IR spectroscopy database [41]. The spectra were measured by vaporizing the liquid hydrocarbons in a Bruker-66V FTIR at temperatures of 5 C, 25 C and 50 C. Here, we only make use of the spectra at 25 C. The pure components considered in our work were limited by the availability of their spectra in the PNNL database. It may be seen in Fig. 2 that the hydrocarbon spectral features primarily appear at wavenumbers below $\sim 3500\text{ cm}^{-1}$. These spectral features are related to fundamental vibrational motions of various chemical bonds, such as bond stretching and bending. Figure 2(b)-(c) shows averaged spectra of various molecular classes in two wavelength regions, where one can observe distinct spectral features of each molecular

group. Additionally, within one group (e.g., paraffins), the spectral intensity scales directly with the number of carbon atoms. We decided to utilize the entire spectral data and let our dimensional reduction methods extract key elements from the data.

Table 1: List of pure components considered in this work. The references indicate the source of RON and MON values.

| Name | RON | MON | Class | Ref | Name | RON | MON | Class | Ref |
|---------------------------|-------|------|-------------|------|----------------------------|-------|-------|------------|------|
| Ethane | 114.9 | 99 | nparaffin | [42] | 4-methyl-1-pentene | 95.7 | 80.9 | olefin | [31] |
| n-butane | 93.8 | 89.6 | nparaffin | [42] | 4-methyl-2-pentene | 99.7 | 84.5 | olefin | [31] |
| n-decane | -30 | -30 | nparaffin | [42] | cyclopentene | 93.3 | 69.7 | olefin | [42] |
| n-heptane | 0 | 0 | nparaffin | [42] | ethylene | 100.2 | 84.9 | olefin | [31] |
| n-hexane | 24.8 | 26 | nparaffin | [42] | isobutene | 106.3 | 90.3 | olefin | [31] |
| n-nonane | -20 | -20 | nparaffin | [42] | propylene | 100.2 | 84.9 | olefin | [31] |
| n-octane | -15 | -20 | nparaffin | [42] | cycloheptane | 38.9 | 40.8 | naphthenes | [31] |
| n-pentane | 61.7 | 62.6 | nparaffin | [31] | cyclohexane | 83 | 77.2 | naphthenes | [42] |
| propane | 112 | 97.1 | nparaffin | [31] | cyclohexene | 83.9 | 63 | naphthenes | [42] |
| 2-methylbutane | 92.3 | 90.3 | isoparaffin | [42] | cyclooctane | 71 | 51.2 | naphthenes | [31] |
| 2-methylpentane | 73.4 | 73.5 | isoparaffin | [42] | cyclopentane | 101.6 | 84.9 | naphthenes | [31] |
| 2,2-dimethylbutane | 91.8 | 93.4 | isoparaffin | [42] | 1,2,3,4-tetramethylbenzene | 100.5 | 100 | aromatic | [42] |
| 2,2,4-trimethylpentane | 100 | 100 | isoparaffin | [42] | 1,3,5-trimethylbenzene | 106 | 100.6 | aromatic | [42] |
| 2,3-dimethylbutane | 100 | 94.3 | isoparaffin | [42] | benzene | 105 | 102.5 | aromatic | [42] |
| 3-Methylhexane | 52 | 55.8 | isoparaffin | [42] | ethylbenzene | 100.8 | 97.7 | aromatic | [31] |
| 3-methylpentane | 74.5 | 74.3 | isoparaffin | [42] | isopropylbenzene | 102.1 | 99.3 | aromatic | [31] |
| isobutane | 102 | 97.6 | isoparaffin | [42] | m-ethyltoluene | 101.8 | 100 | aromatic | [42] |
| 1-butene | 98.8 | 79.9 | olefin | [31] | m-xylene | 104 | 102.8 | aromatic | [42] |
| 1-heptene | 54.5 | 50.7 | olefin | [42] | n-propylbenzene | 101.5 | 98.7 | aromatic | [31] |
| 1-hexene | 76.4 | 63.4 | olefin | [42] | o-ethyltoluene | 100.2 | 92.1 | aromatic | [42] |
| 1-octene | 28.7 | 34.7 | olefin | [42] | o-xylene | 105 | 100 | aromatic | [42] |
| 1-pentene | 90.9 | 77.1 | olefin | [42] | p-ethyltoluene | 102 | 97 | aromatic | [42] |
| 2-butene | 101.6 | 99.9 | olefin | [31] | p-xylene | 103.4 | 101.2 | aromatic | [31] |
| 2-methyl-1-butene | 100.2 | 81.9 | olefin | [42] | sec-butylbenzene | 100.7 | 95.7 | aromatic | [31] |
| 2-methyl-1-pentene | 94.2 | 81.5 | olefin | [42] | tert-butylbenzene | 103 | 100.8 | aromatic | [31] |
| 2-methyl-1,3-butadiene | 99.1 | 81 | olefin | [42] | tetralin | 96.5 | 81.9 | aromatic | [42] |
| 2-methyl-2-butene | 97.3 | 84.7 | olefin | [42] | toluene | 118 | 100.3 | aromatic | [42] |
| 2-methyl-2-pentene | 97.8 | 83 | olefin | [42] | ethanol | 108 | 90 | oxygenate | [42] |
| 2,4,4-trimethyl-1-pentene | 106 | 86.5 | olefin | [42] | tert-amyl methyl ether | 115 | 98 | oxygenate | [42] |
| 2,4,4-trimethyl-2-pentene | 100.3 | 86.2 | olefin | [42] | tert-Butyl methyl ether | 115.2 | 97.2 | oxygenate | [42] |
| 3-methyl-1-butene | 97.5 | 97.5 | olefin | [31] | | | | | |

The spectra S_b of hydrocarbon blends, of known compositions, are calculated by:

$$S_b = \sum_{n=1}^N w_n S_{pc}$$

where N is the number of components in the mixture, w_n is the molar fraction of a given component, and S_{pc} is the spectra of pure components. All available data in literature were used in efforts to create a robust model capable of capturing as many complex synergistic and antagonistic blending relations. The blends considered were 148 mixtures of 1-hexene,

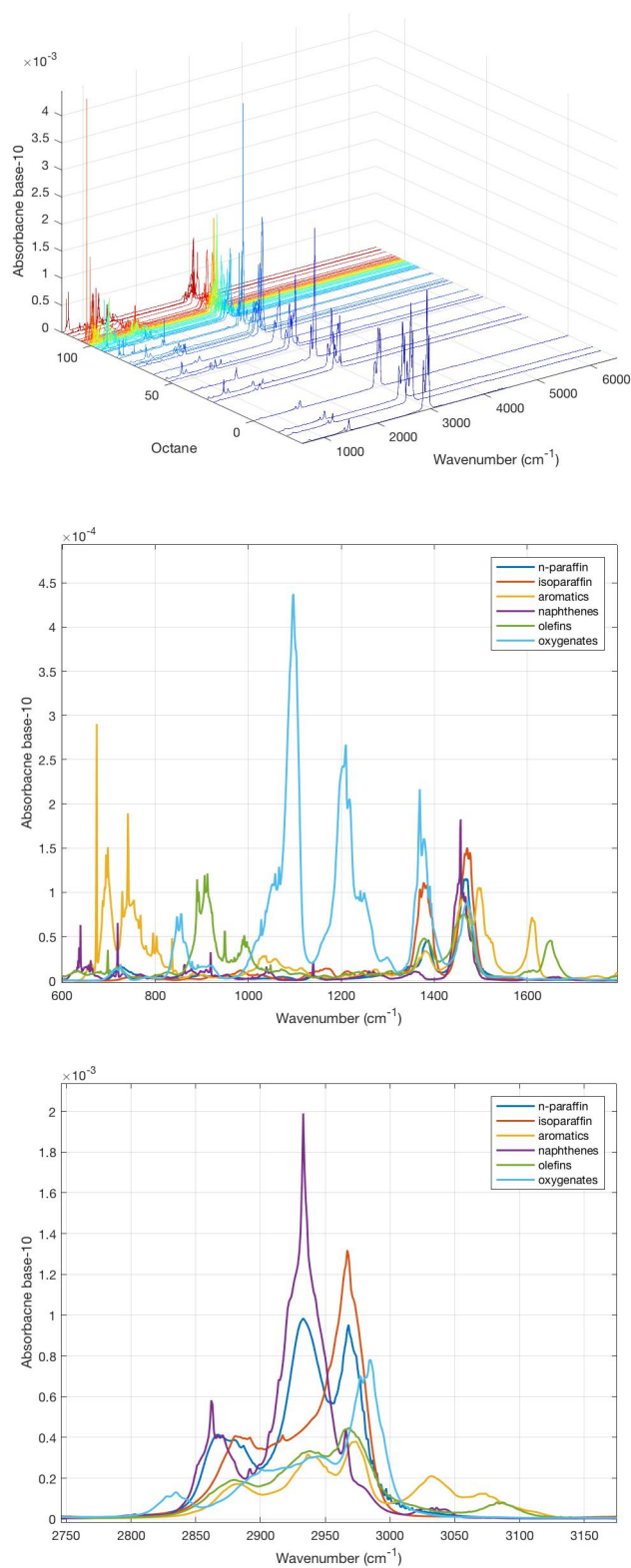


Figure 2: (a) Gas-phase IR spectra of 61 pure components at 298 K, along with their research octane number. Spectra source: PNNL database, (b) PIONA averaged spectra over 600-1700 cm^{-1} , (c) PIONA averaged spectra over 2750-3170 cm^{-1} .

n-heptane, 2,2,4 trimethylpentane (iso-octane), toluene, cyclopentane, and ethanol (see Supplemental Material for compositional details [6, 29, 43–46]).

For FACE gasoline and FACE/ethanol blends, we used a representative weighted sum of class averaged spectra where we have 9 n-paraffins, 8 iso-paraffins, 16 aromatics, 5 naphthenes, and 20 olefins. The weights come from detailed hydrocarbon analyses (DHA) summarized in Table 2. FACE/ethanol blends are also averaged by mole and summarized in Table 3. The linear-by-mole summation of IR spectra is a reasonable approach for gas-phase spectra.

$$S_{FACE} = \sum_{n=1}^N w_n \bar{S}_{class}$$

Table 2: FACE gasoline PIONA composition, obtained from the respective reference, by molar percent.

| FACE | n-Paraffins | iso-Paraffins | Aromatics | Naphthenes | Olefins | Ref |
|----------|-------------|---------------|-----------|------------|---------|------|
| A | 13.2 | 83.7 | 0.3 | 2.4 | 0.4 | [47] |
| B | 10.1 | 82.5 | 7.1 | 0.0 | 0.0 | [48] |
| C | 28.6 | 65.1 | 4.4 | 1.5 | 0.3 | [47] |
| F | 4.6 | 38.3 | 12.7 | 15.3 | 29.0 | [49] |
| G | 7.8 | 38.2 | 31.7 | 14.0 | 7.8 | [49] |
| I | 14.0 | 70.0 | 5.0 | 4.0 | 7.0 | [50] |
| J | 31.5 | 30.8 | 34.7 | 2.4 | 0.6 | [50] |

Brief description of algorithms:

In this section, we briefly discuss the different algorithms used here and metrics of error quantification. Details of these methods are available in literature.

- Dimensional Reduction:

Working on high dimensional data is difficult and makes the application of many algorithms challenging. Therefore, two methods to reduce dimensionality of spectral data were explored: Principal Component Analysis (PCA) and Singular Value Decomposition (SVD).

Table 3: FACE/ethanol blends by volume percentage. RON and MON values are from the respective references. Volume percentages were converted to molar percentages for calculating the composite spectra.

| no | Name | A | B | C | F | G | I | J | ethanol | RON | MON | Ref |
|----|-----------|-----|-----|-----|-----|-----|-----|-----|---------|-------|------|------|
| 1 | FACE A+0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 83.6 | 82.9 | [44] |
| 2 | FACE A+10 | 90 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 92 | 88 | [44] |
| 3 | FACE A+15 | 85 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 94.8 | 89.4 | [51] |
| 4 | FACE A+25 | 75 | 0 | 0 | 0 | 0 | 0 | 0 | 25 | 100.7 | 92.6 | [44] |
| 5 | FACE A+30 | 70 | 0 | 0 | 0 | 0 | 0 | 0 | 30 | 102.3 | 91.4 | [51] |
| 6 | FACE A+40 | 60 | 0 | 0 | 0 | 0 | 0 | 0 | 40 | 104.1 | 91.7 | [44] |
| 7 | FACE A+60 | 40 | 0 | 0 | 0 | 0 | 0 | 0 | 60 | 106 | 91.3 | [44] |
| 8 | FACE B+0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 95.8 | 92.4 | [51] |
| 9 | FACE B+10 | 0 | 90 | 0 | 0 | 0 | 0 | 0 | 10 | 101.1 | 93.5 | [51] |
| 10 | FACE B+15 | 0 | 85 | 0 | 0 | 0 | 0 | 0 | 15 | 103 | 93.5 | [51] |
| 11 | FACE B+30 | 0 | 70 | 0 | 0 | 0 | 0 | 0 | 30 | 106 | 93.3 | [51] |
| 12 | FACE C+0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 84.4 | 83 | [44] |
| 13 | FACE C+10 | 0 | 0 | 90 | 0 | 0 | 0 | 0 | 10 | 92.2 | 87.1 | [44] |
| 14 | FACE C+15 | 0 | 0 | 85 | 0 | 0 | 0 | 0 | 15 | 94.8 | 88.8 | [51] |
| 15 | FACE C+25 | 0 | 0 | 75 | 0 | 0 | 0 | 0 | 25 | 100.3 | 90.5 | [44] |
| 16 | FACE C+30 | 0 | 0 | 70 | 0 | 0 | 0 | 0 | 30 | 101.5 | 91.6 | [51] |
| 17 | FACE C+40 | 0 | 0 | 60 | 0 | 0 | 0 | 0 | 40 | 104.1 | 91 | [44] |
| 18 | FACE C+60 | 0 | 0 | 40 | 0 | 0 | 0 | 0 | 60 | 105.8 | 91.3 | [44] |
| 19 | FACE F+0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 94.2 | 87.4 | [44] |
| 20 | FACE F+10 | 0 | 0 | 0 | 90 | 0 | 0 | 0 | 10 | 98.9 | 88.5 | [44] |
| 21 | FACE F+25 | 0 | 0 | 0 | 75 | 0 | 0 | 0 | 25 | 103.2 | 89.5 | [44] |
| 22 | FACE F+40 | 0 | 0 | 0 | 60 | 0 | 0 | 0 | 40 | 104.7 | 90.3 | [44] |
| 23 | FACE F+60 | 0 | 0 | 0 | 40 | 0 | 0 | 0 | 60 | 105.7 | 90.5 | [44] |
| 24 | FACE G+0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 96.4 | 84.9 | [44] |
| 25 | FACE G+10 | 0 | 0 | 0 | 0 | 90 | 0 | 0 | 10 | 98.8 | 86.1 | [44] |
| 26 | FACE G+25 | 0 | 0 | 0 | 0 | 75 | 0 | 0 | 25 | 102.4 | 87.9 | [44] |
| 27 | FACE G+40 | 0 | 0 | 0 | 0 | 60 | 0 | 0 | 40 | 103 | 88.5 | [44] |
| 28 | FACE G+60 | 0 | 0 | 0 | 0 | 40 | 0 | 0 | 60 | 105 | 88.9 | [44] |
| 29 | FACE I+0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 69.5 | 69 | [44] |
| 30 | FACE I+10 | 0 | 0 | 0 | 0 | 0 | 90 | 0 | 10 | 79.9 | 78 | [44] |
| 31 | FACE I+25 | 0 | 0 | 0 | 0 | 0 | 75 | 0 | 25 | 89.8 | 85.3 | [44] |
| 32 | FACE I+40 | 0 | 0 | 0 | 0 | 0 | 60 | 0 | 40 | 98 | 88.3 | [44] |
| 33 | FACE I+60 | 0 | 0 | 0 | 0 | 0 | 40 | 0 | 60 | 103.6 | 89.7 | [44] |
| 34 | FACE J+0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 71.8 | 66.9 | [44] |
| 35 | FACE J+10 | 0 | 0 | 0 | 0 | 0 | 0 | 90 | 10 | 79 | 73.6 | [44] |
| 36 | FACE J+25 | 0 | 0 | 0 | 0 | 0 | 0 | 75 | 25 | 89.8 | 81.7 | [44] |
| 37 | FACE J+40 | 0 | 0 | 0 | 0 | 0 | 0 | 60 | 40 | 98 | 85.9 | [44] |
| 38 | FACE J+60 | 0 | 0 | 0 | 0 | 0 | 0 | 40 | 60 | 103.6 | 88.2 | [44] |

PCA reduces dimensions by choosing directions of maximum variance for the projected data, accomplished by solving for the eigenvectors and eigenvalues of the covariance matrix. SVD ranks the patterns/features by importance where this is done by maximizing the total projected squares yielding singular vectors. It can be shown that both methods essentially give the same results if the data are centered about the mean. The R functions *prcomp* and *svd* can be used for easy application.

- Partial Least Square Regression (PLSR)

The partial least square regression first appeared, as we know it, in the field of chemistry [52], which does not come as a surprise due to the nature of highly overlapping and difficult to interpret overtones inherent in spectroscopy. Regressing properties by spectra using the normal least-squares formulation ($B = [X^T X]^{-1} X^T Y$) is impossible since there are always more variables than observations in X (i.e., $X^T X$ is singular). To reduce dimensionality, PCA is applied on both X and Y, decomposing these into loadings and scores. The loadings and scores are then chosen to maximize $\text{cov}(X, Y)$ and correlated in a linear fit to yield the PLSR. The *pls* function can be accessed in R from the *pls library* [11].

- Support Vector Machines (SVM)

SVM was originally created for binary classification in 1995 [53] and then generalized for other purposes like regression. SVM constructs a set of hyper-planes in a high-dimensional space, and it aims to maximize the functional margins separating the data. This is accomplished by the use of kernel transformations, where computations are performed in the input space and nonlinearity is effectively resolved in the transformed higher dimension. The *svm* function can be accessed in R from the *e1071 library*.

- Artificial Neural Networks (ANN)

The popularity of multilayer perceptron (MLP) feedforward artificial neural networks (ANN) came after Werbos's work on backpropagation [54]. Backpropagation is the idea of updating the error back through layers of the model on the nodal level. Weights are updated based on a fixed learning rate and the partial derivatives. The choice of the learning rate

is essential to the performance of the algorithm since a large one will lead to oscillation and divergence while a small one will take many steps to converge. To counter this problem, adaptive algorithms tune the learning rate based on the behavior of the error function. Training in this study was done adaptively with resilient propagation (RPROP) [55], which tunes weights independent of the magnitude of the partial derivative to prevent "blurred adaptivity" caused by the unforeseeable behavior of the derivative itself. Instead, to accelerate convergence in shallow regions, update is increased if the partial derivative retains its sign or decreased if the sign has changed since a local minima has been skipped. The *neuralnet* function in R can be accessed from the *neuralnet library*.

- Error quantification

The performance of the models is quantified by commonly used error metrics and standardized by cross-validation. Error is described by the following equations:

Root Mean Squared Error:

$$RMS = \frac{1}{N} \sqrt{\sum_{i=1}^N (y_i - t_i)^2}$$

Mean Absolute Error:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - t_i|$$

Mean Percentage Error:

$$MPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - t_i}{t_i} \right|$$

Max Absolute Error:

$$MAX = \max(|y_i - t_i|)$$

Coefficient of determination:

$$R^2 = \text{cor}(y, t)^2$$

3. Results and discussion

Although neural nets are perfectly suited to deal with the non-linear nature of this problem, high dimensionality deems it impractical. Thus, dimensional reduction was first applied to extract important features from spectra. From past studies [31, 32], we know that the molecular weight, paraffinic CH₃ groups, paraffinic CH₂ groups, paraffinic CH groups, olefinic -CH=CH₂ groups, naphthenic CH-CH₂ groups, aromatic C-CH groups, ethanolic OH groups, and branching index are among the most important features. We noticed that PCA scores and SVD singular vectors produce the same relations between the features and octane number. For example, Fig. 3 shows the resemblance between molecular weight and the first principal component score for various PRF, TPRF, and 1-hexene/ethanol blends.

The PCA scores and SVD singular vectors show similar results which reaffirms our intuition that we are extracting physical features from the IR spectra. We tested this by splitting the dataset, comprised of 148 blends and 34 FACE/ethanol blends, to a random 70/30 training/test sets. A 10-fold cross-validation was done on the training set and the number of components was chosen to be 11. The predictions are very good with a maximum root mean square error of 0.094 (see Fig. 4).

We observed that all nine features have almost perfect linear relations with scores. Of course, this is to be expected for blends since both the features and spectra are linear sums of their pure components. Features for FACE gasoline, on the other hand, were obtained from ¹H-NMR data. The fact that we obtained good linear fits for FACE as well shows the feasibility of this methodology. The predicted features can then be used to predict octane numbers. We employed multiple linear regression (MLR) and artificial neural nets (ANN) for octane prediction. Much better prediction is obtained by ANN and these results (see Fig. 5) confirm that the octane blending behavior of hydrocarbons is in fact nonlinear. The mean absolute error for RON is almost cut by half from 4.69 to 2.53 for linear regression and neural nets, respectively, and similarly for MON the MAE reduces from 4.87 to 2.61, respectively.

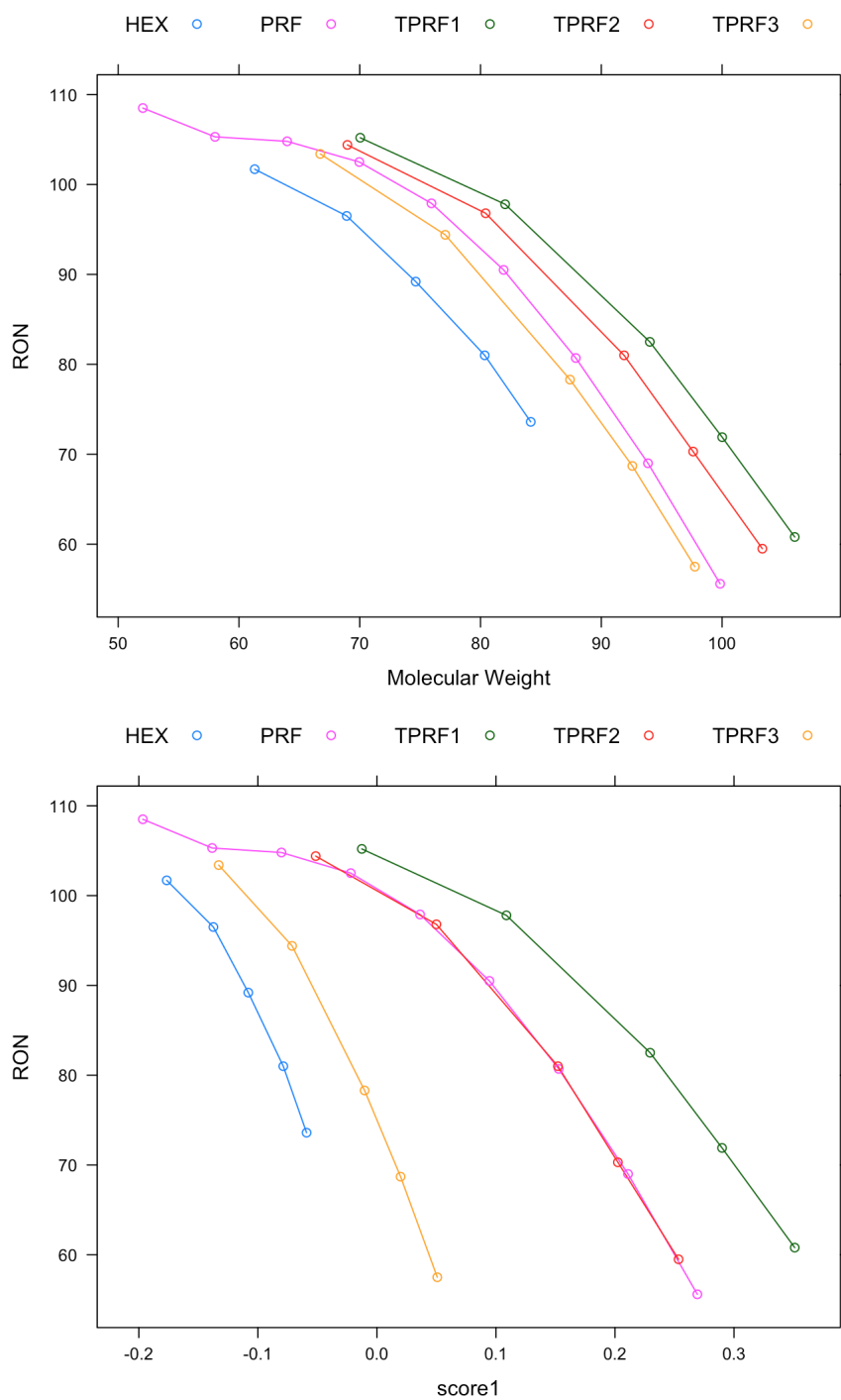


Figure 3: Resemblance in the correlation between the the molecular weight and octane number (top), and the first PCA score and octane number (bottom). See Supplementary Materials for the composition of these blends.

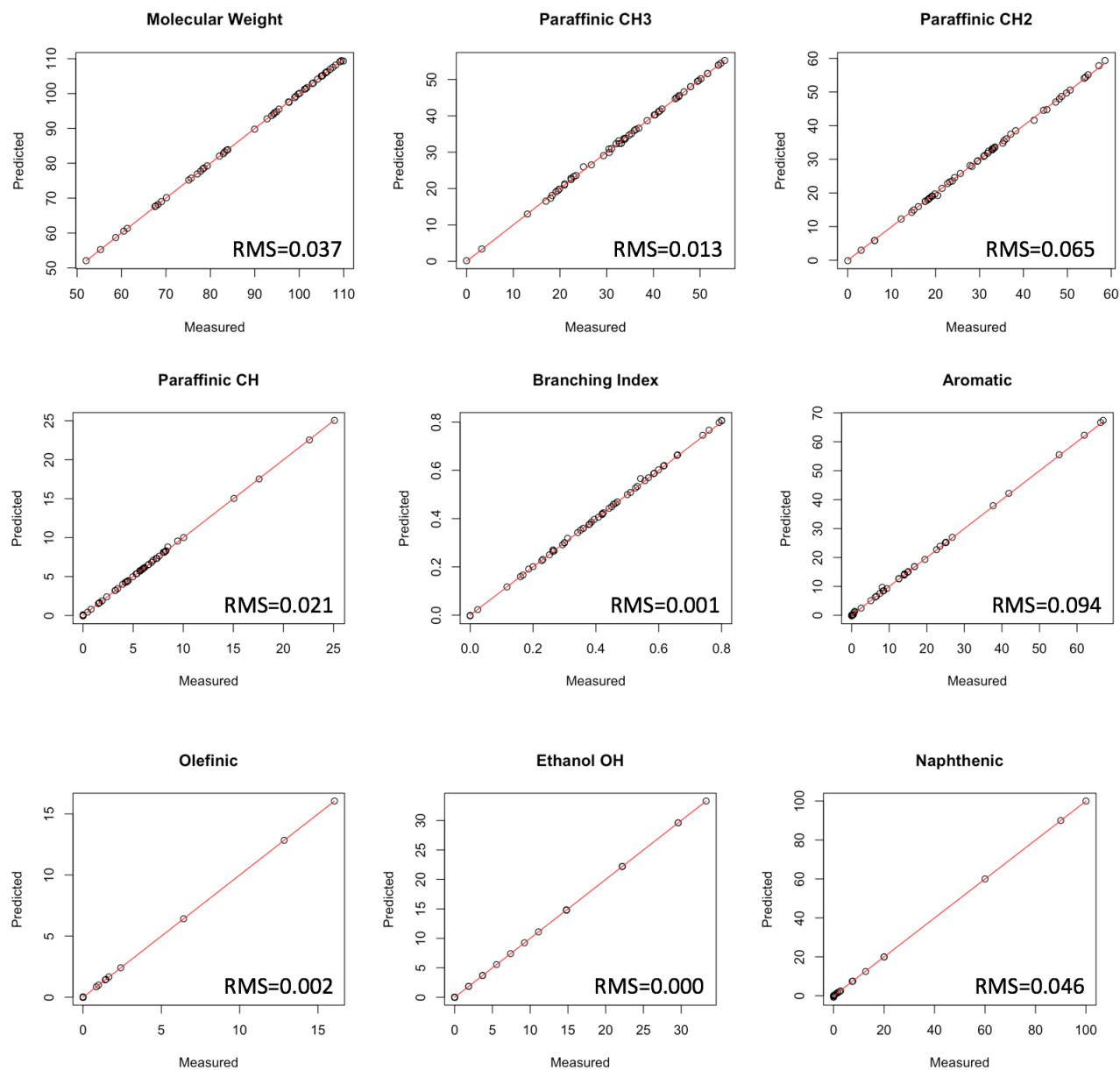


Figure 4: Comparison of predicted and measured features for 148 hydrocarbon blends and 34 FACE/ethanol blends. The prediction is based on PLSR of the IR spectra.

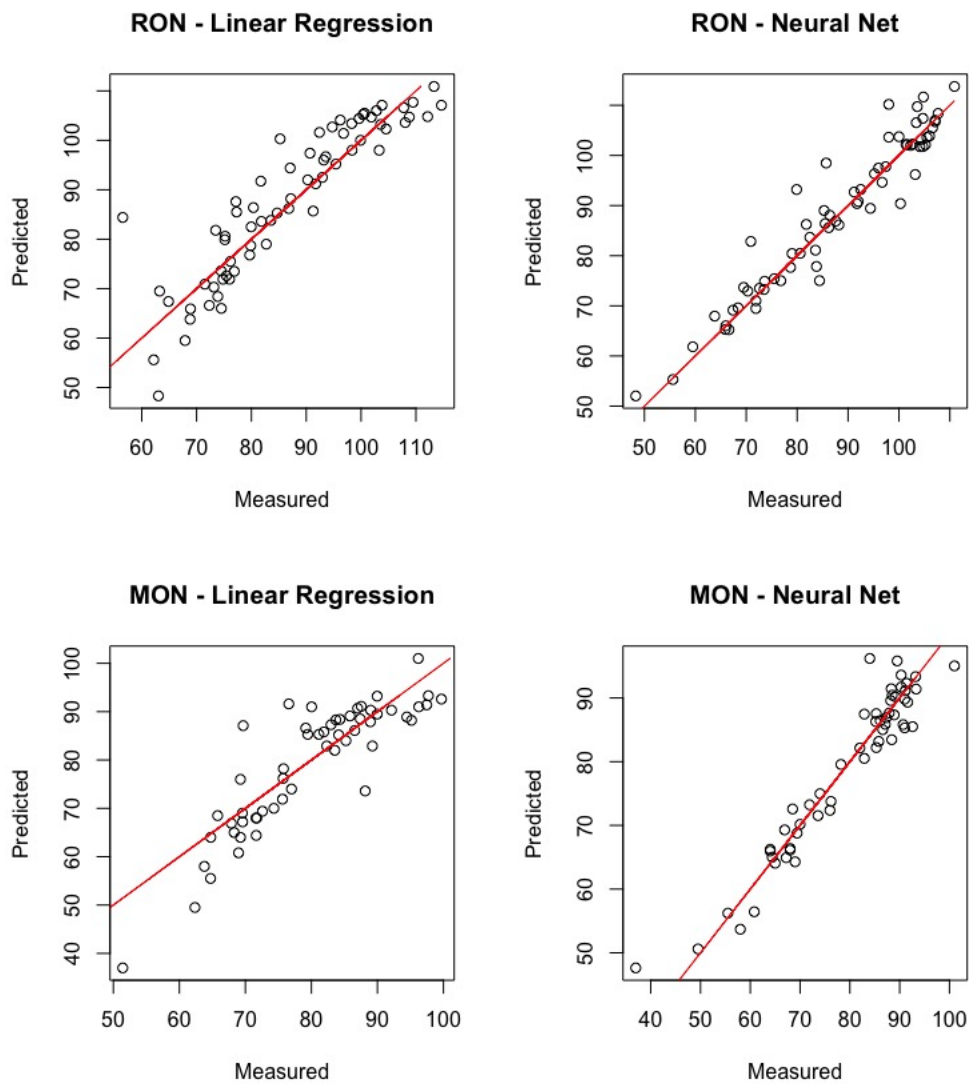


Figure 5: Comparison of predicted and measured octane numbers. Predicted octane numbers were obtained by applying MLR or ANN on the predicted features.

Octane predictions shown in Fig. 5 were obtained via a two-step model, i.e., in the first step, features were extracted from the IR spectra and these features were then used in the second step to predict octane numbers. Each step in the model comes with associated errors. It is preferred to implement one step models to avoid carrying forward the error from previous steps. Moreover, although features were extracted successfully, the scores might contain more information that is being discarded. Therefore, we now present the alternative model comprising of dimensional reduction and neural net implementation, while skipping the intermediate step of feature extraction. Since SVD gave slightly better results than PCA, so it was used in conjunction with ANN. For evaluation, we use a 70/30 split and compare the performance of ANN to PLSR and SVM. Testing is carried out for the prediction of RON, MON, and octane sensitivity. Sensitivity is defined as the difference between RON and MON. For modern engines, a fuel with higher sensitivity results in higher efficiency.

Model tuning was done through 10-fold cross-validation due to its good variance-bias balance. Model parameters to be considered are number of components for PLSR, gamma (free parameter of the Gaussian radial basis function) and cost (soft margin tolerance) for SVM, and the number of SVD vectors employed and the number of nodes in the hidden layer for ANN. Threshold on the partial derivatives in RPROP was chosen to be 0.1 with a learning rate range of [0.5-1.2]. Sum of squares error was minimized and a logistic function was chosen for activation. Summary of tuned parameters for RON, MON, and sensitivity is provided in Table 4. Results for both the 10-fold cross-validation on the training set and errors for the test set are given in Tables 5 - 7 for RON, MON, and sensitivity, respectively. Comparisons among PLSR, SVM, and ANN for RON and MON are plotted in Figs. 6-8.

Table 4: Summary of tuned parameters

| Model | PLSR | SVM | | ANN | |
|--------------------|----------------|--------------|-------------|------------------|--------------|
| Parameters | # Comp. | gamma | cost | # vectors | nodes |
| RON | 12 | 1.00E-04 | 100 | 9 | 105 |
| MON | 12 | 1.00E-04 | 100 | 9 | 190 |
| Sensitivity | 12 | 1.20E-04 | 101 | 8 | 135 |

Table 5: RON prediction results

| RON | PLSR | | SVM | | ANN | |
|----------------------|----------------|-------------|----------------|-------------|----------------|-------------|
| Error | 10-Fold | Test | 10-Fold | Test | 10-Fold | Test |
| RMS | 2.948 | 2.813 | 1.210 | 1.880 | 1.219 | 0.734 |
| MPE | 2.341 | 2.073 | 1.239 | 1.626 | 0.950 | 0.697 |
| MAE | 2.098 | 1.926 | 1.064 | 1.378 | 0.745 | 0.577 |
| MAX | 19.657 | 8.436 | 1.573 | 5.006 | 2.554 | 2.040 |
| R² | 0.963 | 0.965 | 0.995 | 0.983 | 0.994 | 0.999 |

Table 6: MON prediction results

| MON | PLSR | | SVM | | ANN | |
|----------------------|----------------|-------------|----------------|-------------|----------------|-------------|
| Error | 10-Fold | Test | 10-Fold | Test | 10-Fold | Test |
| RMS | 3.060 | 2.663 | 0.894 | 1.807 | 1.741 | 0.971 |
| MPE | 2.956 | 2.332 | 1.000 | 1.546 | 1.442 | 0.933 |
| MAE | 2.250 | 1.931 | 0.773 | 1.203 | 1.056 | 0.725 |
| MAX | 14.839 | 7.449 | 1.276 | 8.158 | 3.299 | 2.829 |
| R² | 0.936 | 0.945 | 0.994 | 0.973 | 0.978 | 0.993 |

Table 7: Octane sensitivity prediction results

| Sensitivity | PLSR | | SVM | | ANN | |
|----------------------|----------------|-------------|----------------|-------------|----------------|-------------|
| Error | 10-Fold | Test | 10-Fold | Test | 10-Fold | Test |
| RMS | 1.463 | 1.233 | 0.53 | 0.988 | 0.978 | 0.803 |
| MAE | 1.052 | 0.903 | 0.431 | 0.745 | 0.706 | 0.596 |
| MAX | 5.522 | 3.466 | 1.145 | 2.756 | 1.809 | 2.395 |
| R² | 0.913 | 0.94 | 0.987 | 0.963 | 0.963 | 0.976 |

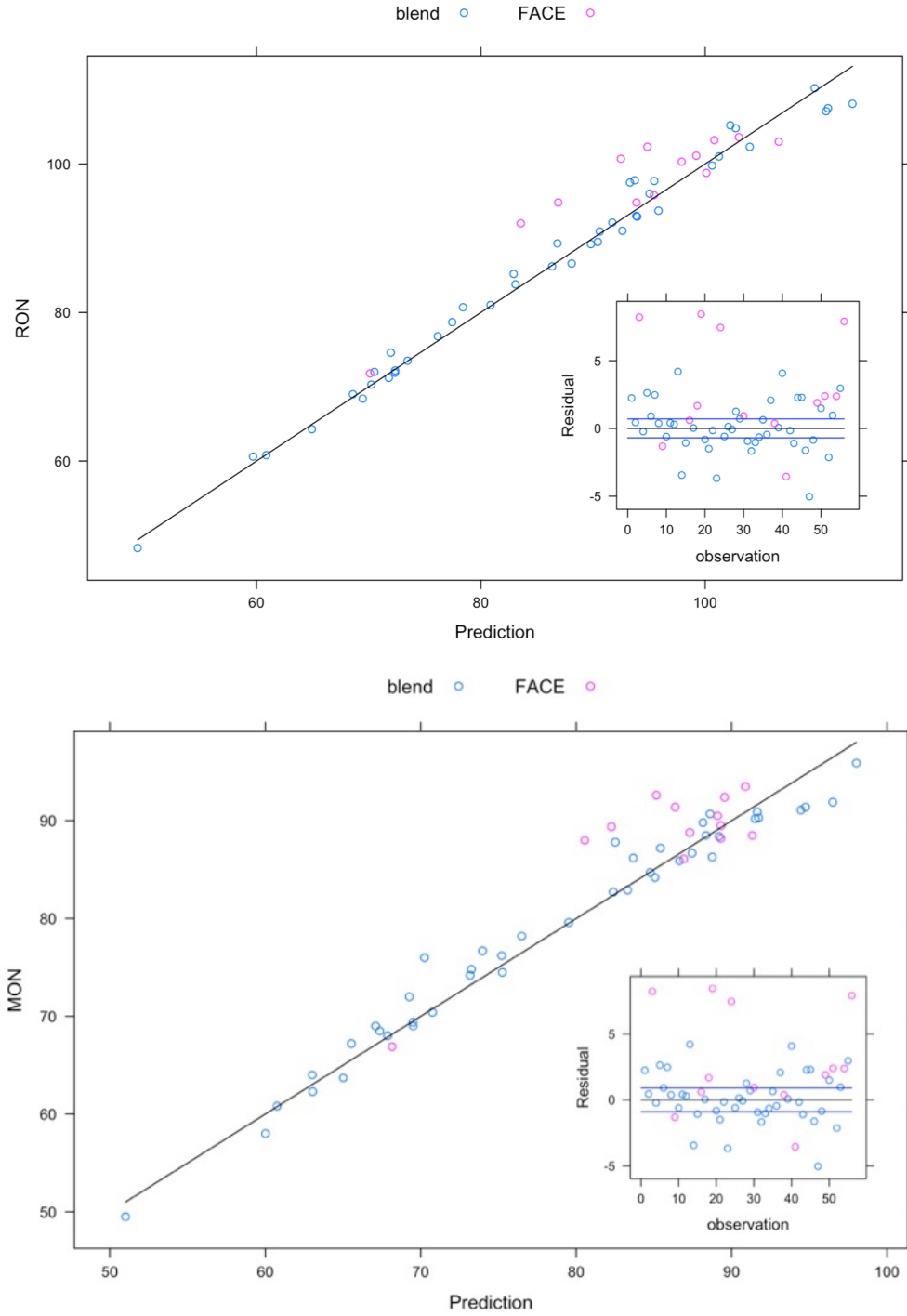


Figure 6: Prediction vs measured and residual using linear model, (a) RON, (b) MON. Horizontal lines in the residual plot indicate the uncertainty limits of experimental measurements.

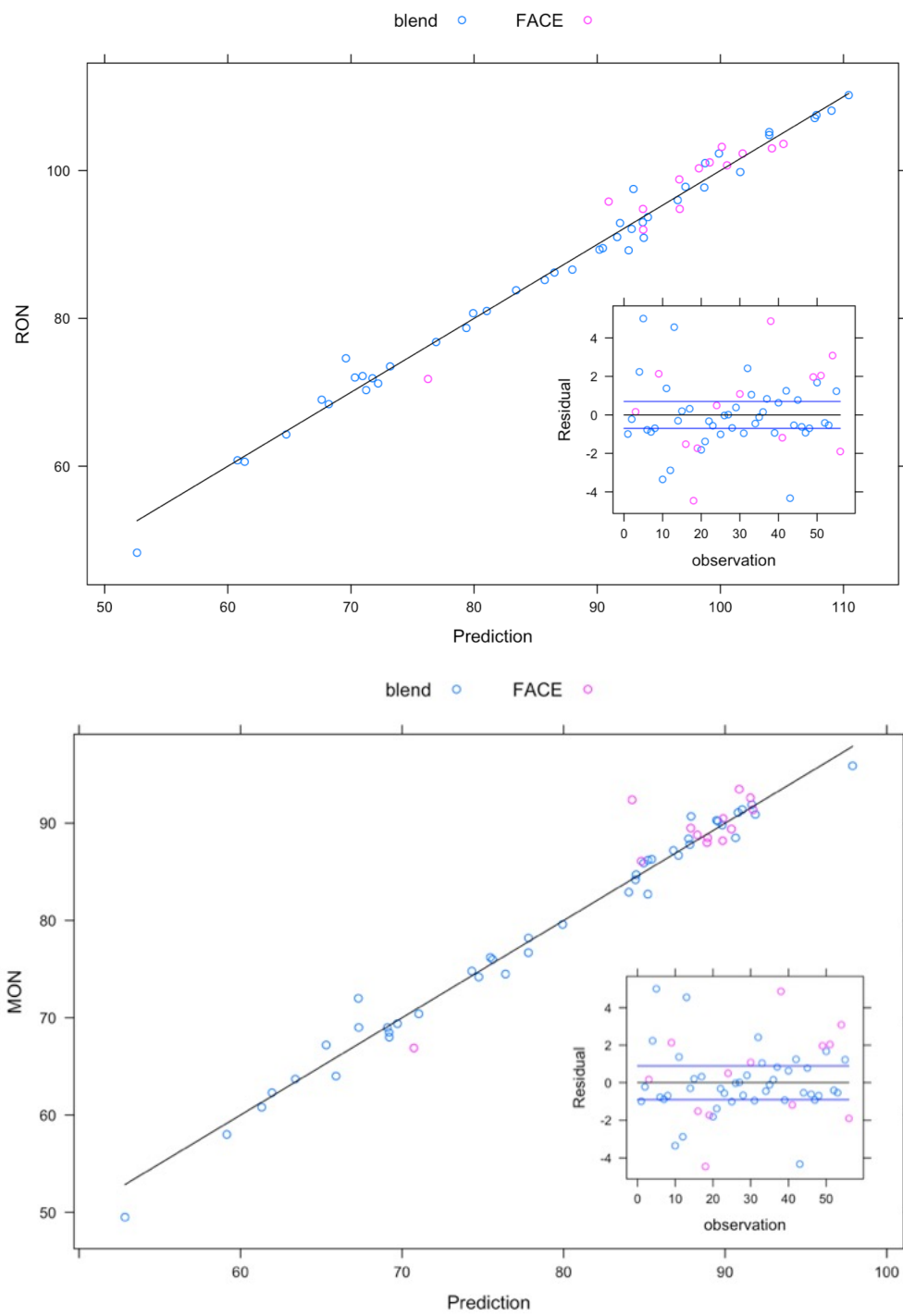


Figure 7: RON Prediction vs measured and residual using SVM, (a) RON, (b) MON.

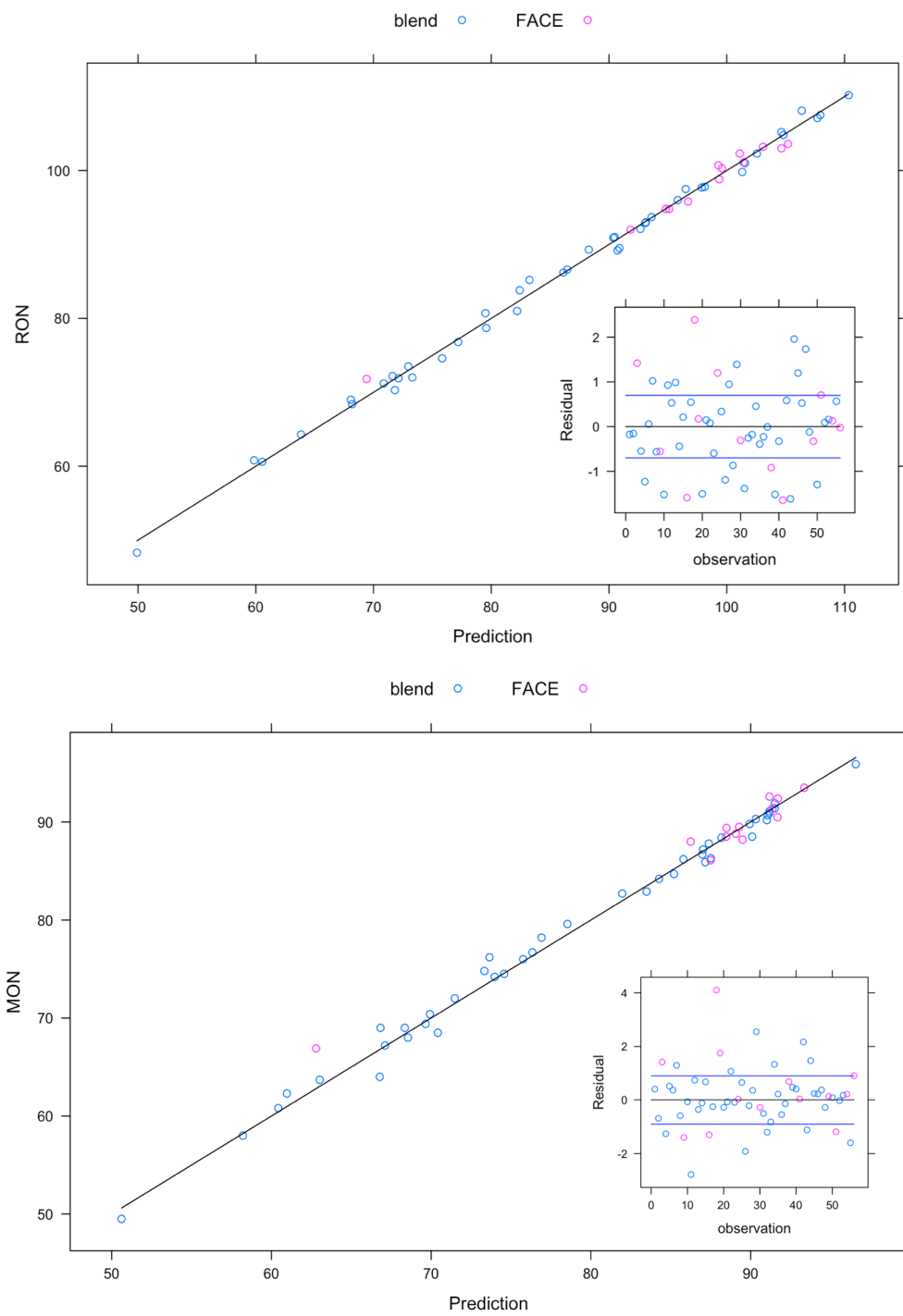


Figure 8: RON Prediction vs measured and residual using ANN, (a) RON, (b) MON.

Although linear model (PLSR) gives acceptable results, nonlinear methods give much better performance. It is seen that SVM performs consistently better than the linear model while ANN performs best among the three methods. It is also worth noting that SVM is very sensitive to tuning parameters which could be cumbersome. ANN also requires some tuning for optimum results but is worth the >60 percent enhancement in performance. Further tuning could have been explored by using more hidden layers, but it was not necessary since one hidden layer already gave results within experimental uncertainty. Experimental uncertainty limits are plotted along with the residuals in Figs. 6-8 (± 0.7 RON and ± 0.9 MON). These limits are stretched at higher octane ratings to as high as ± 3.5 [6]. Although comparing between models is always tricky since datasets are often dissimilar, our ANN model seems to give smaller error than other models containing blends and FACE gasolines [18, 32] and as good as studies on PRF/TPRF blends [5, 6]. Another advantage of using ANN is that it performed significantly better for FACE gasoline when compared to linear regression. This is to be expected since the composition of FACE fuels and their corresponding spectra are much more complex. Finally, compared to octane prediction methods utilizing NMR spectra, ignition quality tester (IQT) or experimental blending data, our methodology provides a simpler solution needing readily-acquired FTIR or laser-based spectra.

4. Conclusions

In this study, we have investigated several models that predict the octane number of blends of hydrocarbons and distillate FACE gasoline blends. The models use gas-phase IR spectroscopy data of neat hydrocarbons to simulate spectra for hydrocarbon blends and FACE fuels based on the ideal blending assumption. First, we implemented a two-step model that decomposes spectra to important scores, relates the scores to chemical features, and uses these features to predict octane numbers. The significance of the two-step model lies in

bridging the gap between the group contribution method and IR spectroscopy chemometric studies. We demonstrated that the nine features from group contribution method (molecular weight, paraffinic CH₃ groups, paraffinic CH₂ groups, paraffinic CH groups, olefinic -CH=CH₂ groups, naphthenic CH-CH₂ groups, aromatic C-CH groups, ethanolic OH groups, and branching index) can be extracted from IR spectra by applying PLSR. We also showed that nonlinear methods like ANN are more capable of capturing octane blending behavior. In the alternative approach, we used the scores directly as input to ANN and compared its performance to SVM and PLSR. Although the standard in most chemometric studies is the PLSR, our proposed ANN method produced errors smaller by $\sim 70\%$ and $\sim 60\%$ for RON and MON, respectively. Mean absolute error from ANN was found to be 0.56 and 0.73 for RON and MON, respectively, which is within the experimental uncertainty of octane testing.

Acknowledgement This work was funded by the Office of Sponsored Research at King Abdullah University of Science and Technology (KAUST). We are thankful to Prof. Mani Sarathy and Dr. Abdul Gani Abdul Jameel for helpful discussions.

References

- (1) *Standard Test Method for Research Octane Number of Spark-Ignition Engine Fuel*; tech. rep. ASTM D2699 - 15a; ASTM int., 2019, DOI: 10.1520/D2699-19.
- (2) *Standard Test Method for Motor Octane Number of Spark-Ignition Engine Fuel*; tech. rep. ASTM D2700-19; ASTM int., 2019, DOI: 10.1520/D2700-19.
- (3) Smolenskii, E. A.; G. V. Vlasova, L. L. A Study of the Structure-Octane Number Relationship for Hydrocarbons. *Dolkadi Physical Chemistry* **2004**, 397, 145–149, DOI: 10.1023/B:DOPC.0000035400.72438.a8.
- (4) Lapidus, A. L.; Smolenskii, E. A.; Bavykin, V. M.; Myshenkova, T. N. Models for the Calculation and Prediction of the Octane and Cetane Numbers of Individual Hydrocarbons. *Petroleum Chemistry* **2007**, 48, 277–286, DOI: 10.1134/S0965544108040051.

- (5) A.Badra, J.; Bokhumseen, N.; Mulla, N.; Sarathy, S. M.; Farooq, A.; Kalghatgi, G.; Gaillard, P. A methodology to relate octane numbers of binary and ternary n-heptane, iso-octane and toluene mixtures with simulated ignition delay times. *Fuel* **2015**, *160*, 458–469, DOI: 10.1016/j.fuel.2015.08.007.
- (6) AlRamadan, A. S.; Sarathy, S. M.; Khurshid, M.; Badra, J. A blending rule for octane numbers of PRFs and TPRFs with ethanol. *Fuel* **2016**, *180*, 175–186, DOI: 10.1016/j.fuel.2016.04.032.
- (7) Morgan, N.; Smalbonel, A.; Bhave, A.; Kraft, M.; Cracknell, R.; Kalghatgi, G. Mapping surrogate gasoline compositions into RON/MON space. *Combustion and Flame* **2010**, *157*, 1122–1131, DOI: 10.1016/j.combustflame.2010.02.003.
- (8) Lugo, H. J. Correlations between Octane Numbers and Catalytic Cracking Naphtha Composition. *Ind. Eng. Chem.* **1999**, *38*, , 2171–2176, DOI: 10.1021/ie980273r.
- (9) Nikolaou, N.; Papadopoulos, C.; Gaglias, I.; K.G.Pitarakis A new non-linear calculation method of isomerisation gasoline research octane number based on gas chromatographic data. *Fuel* **2004**, *83*, 517–523, DOI: 10.1016/j.fuel.2003.09.011.
- (10) Cooper, J. B.; Wise, K. L.; Groves, J.; Welch, W. T. Determination of Octane Numbers and Reid Vapor Pressure of Commercial Petroleum Fuels Using FT-Raman Spectroscopy and Partial Least-Squares Regression Analysis. *Analytical Chemistry* **1995**, *67*, 4096–4100.
- (11) Mevik, B.; Wehrens, R. The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software* **2007**, *18*, 517–523.
- (12) Kelly, J.; Barlow, C.; Jinguji, T.; Callis, J. Prediction of gasoline octane numbers from near-infrared spectral features in the range 660–1215 nm. *Analytical chemistry* **1989**, *4*, 313–20, DOI: 10.1021/ac00179a007.

- (13) Kardamakis, A. A.; Pasadakis, N. Autoregressive modeling of near-IR spectra and MLR to predict RON values of gasolines. *Fuel* **2010**, *89*, 158–161, DOI: 10.1016/j.fuel.2009.08.029.
- (14) Swarin, S.; Drumm, C. Prediction of gasoline properties with near-infrared spectroscopy and chemometrics. *SAE Technical Paper* **1991**, 313–20, DOI: 10.4271/912390.
- (15) Korolev, V. N.; Marugin, A. V.; Tsaregradskii, V. B. Estimation of the petroleum product knock rating by regression analysis of near-infrared absorption spectra. *SAE Technical Paper* **2000**, *45*, 1177–1181, DOI: 10.1134/1.1318105.
- (16) Fodor, G. E.; Kohl, K. B.; Mason, R. L. Analysis of Gasolines by FT-IR Spectroscopy. *Anal. Chem.* **1996**, *68*, 23–30, DOI: 10.1021/ac9507294.
- (17) Andrade, J. M.; Muniategui, S.; Prada, D. Prediction of clean octane numbers of catalytic reformed naphthas using FT-m.i.r. and PLS. *Fuel* **1997**, *76*, 1035–1042, DOI: 10.1016/S0016-2361(97)00095-1.
- (18) Daly, S. R.; Niemeyer, K. E.; Cannella, W. J.; Hagen, C. L. Predicting fuel research octane number using Fourier-transform infrared absorption spectra of neat hydrocarbons. *Fuel* **2016**, *183*, 359–365, DOI: 10.1016/j.fuel.2016.06.097.
- (19) IrisLitani-Barzilai; Sela, I.; Bulatov, V.; Zilberma, I.; Schechter, I. On-line remote prediction of gasoline properties by combined optical methods. *Analytica Chimica Acta* **1997**, *339*, 193–199, DOI: 10.1016/S0003-2670(96)00472-2.
- (20) E.Flecher, P.; Welch, W. T.; Albin, S.; Cooper, J. B. Determination of octane numbers and Reid vapor pressure in commercial gasoline using dispersive fiber-optic Raman spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **1997**, *53*, 199–206, DOI: 10.1016/S1386-1425(97)83026-0.

- (21) L.Guan; Feng, X.; Li, Z.; G.M.Lin Determination of octane numbers for clean gasoline using dielectric spectroscopy. *Fuel* **2009**, *88*, 1453–1459, DOI: 10.1016/j.fuel.2009.02.017.
- (22) Paulo, J. M.; E.M.Barro, J.; J.S.Barbeira, P. A PLS regression model using flame spectroscopy emission for determination of octane numbers in gasoline. *Fuel* **2016**, *176*, 216–221, DOI: 10.1016/j.fuel.2016.02.033.
- (23) Teixeira, L. S.; Dantas, M. S.; Guimaraes, P. R.; Teixeira, W.; Vargas, H.; A.P.Lima, J. Correlation of PVR, Octane Numbers and Distillation Curve of Gasoline with Data from a Thermal Wave Interferometer. *Computer Aided Chemical Engineering* **2009**, *27*, 759–764, DOI: 10.1016/S1570-7946(09)70347-5.
- (24) Helga, G. M.; Paulo, G. A.; Barbeira, J. Octane number prediction for gasoline blends. *Fuel* **2012**, *97*, 131–136, DOI: 10.1016/j.fuel.2012.01.058.
- (25) Kiefer, J. Recent Advances in the Characterization of Gaseous and Liquid Fuels by Vibrational Spectroscopy. *Energies* **2015**, *8*, 3165–3197, DOI: 10.3390/en8043165.
- (26) Noack, K.; Eskofier, B.; Kiefer, J.; Dilk, C.; Bilow, G.; Schirmer, M.; Buchholzbe, R.; Leipertz, A. Combined shifted-excitation Raman difference spectroscopy and support vector regression for monitoring the algal production of complex polysaccharides. *Analyst* **2013**, *138*, 5639–5646, DOI: 10.1039/C3AN01158E.
- (27) Cesar, J.; Alves, L.; Claudete; Henriques, B.; Poppi, R. J. Determination of diesel quality parameters using support vector regression and near infrared spectroscopy for an in-line blending optimizer system. *Fuel* **2012**, *97*, 710–717, DOI: 10.1016/j.fuel.2012.03.016.
- (28) O.Santos, V.; Oliveira, J. F. C.; Lima, D. G.; Petry, A. C.; Garcia, E.; Suarez, P. A.; C.Rubim, J. A comparative study of diesel analysis by FTIR, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis. *Analytica Chimica Acta* **2005**, *547*, 188–196, DOI: 10.1016/j.aca.2005.05.042.

- (29) Singh, E.; Badra, J.; Mehl, M.; Sarathy, M. Chemical Kinetic Insights into the Octane Number and Octane Sensitivity of Gasoline Surrogate Mixtures. *Energy and Fuels* **2017**, *31*, 1945–1960.
- (30) Sarathy, S.; Farooq, A.; Kalghatgi, G. T. Recent progress in gasoline surrogate fuels. *Progress in Energy and Combustion Science* **2018**, *65*, 67–108, DOI: 10.1016/j.peecs.2017.09.004.
- (31) Kubic, W. L.; Jenkins, J. R. W.; Moore, C. M.; Semelsberger, T. A.; Sutton, A. D. Artificial Neural Network Based Group Contribution Method for Estimating Cetane and Octane Numbers of Hydrocarbons and Oxygenated Organic Compounds. *Ind. Eng. Chem. Res.* **2017**, *56*, 12236–12245, DOI: 10.1021/acs.iecr.7b02753.
- (32) Jameel, A. G. A.; Oudenhoven, V. V.; Emwas, A.-H.; Sarathy, S. M. Predicting Octane Number Using Nuclear Magnetic Resonance Spectroscopy and Artificial Neural Networks. *Energy and Fuels* **2018**, *32*, 6309–6329, DOI: 10.1021/acs.energyfuels.8b00556.
- (33) Murty, B. N.; Rao, R. N. Global optimization for prediction of blend composition of gasolines of desired octane number and properties. *Fuel Processing Technology* **2004**, *85*, 1595–1602, DOI: 10.1016/j.fuproc.2003.08.004.
- (34) Pasadakis, N.; Gaganis, V.; Foteinopoulos, C. Octane number prediction for gasoline blends. *Fuel Processing Technology* **2006**, *87*, 505–509, DOI: 10.1016/j.fuproc.2005.11.006.
- (35) Elfghi, F. A hybrid statistical approach for modeling and optimization of RON: A comparative study and combined application of response surface methodology (RSM) and artificial neural network (ANN) based on design of experiment (DOE). *Chemical Engineering Research and Design* **2016**, *113*, 264–272, DOI: 10.1016/j.cherd.2016.05.023.

- (36) R.A., A.; J.Castillo, C.-G.; Hernandez, G. J. Direct neural network modeling for separation of linear and branched paraffins by adsorption process for gasoline octane number improvement. *Fuel* **2014**, *124*, 158–167, DOI: 10.1016/j.fuel.2014.01.080.
- (37) Albahri, T. A. Specific Gravity, RVP, Octane Number, and Saturates, Olefins, and Aromatics Fractional Composition of Gasoline and Petroleum Fractions by Neural Network Algorithms. *Petroleum Science and Technology* **2014**, *32*, 1219–1226, DOI: 10.1080/10916466.2011.647203.
- (38) Wang, Y.; Ding, Y.; Wei, W.; Cao, Y.; Davidson, D. F.; Hanson, R. K. On estimating physical and chemical properties of hydrocarbon fuels using mid-infrared FTIR spectra and regularized linear models. *Fuel* **2019**, *255*, 115715, DOI: 10.1016/j.fuel.2019.115715.
- (39) Corsetti, S.; Zehentbauer, F. M.; McGloin, D.; Kiefer, J. Characterization of gasoline/ethanol blends by infrared and excess infrared spectroscopy. *Fuel* **2015**, *151*, 136–142, DOI: 10.1016/j.fuel.2014.10.025.
- (40) Brudzewski, K.; Kesik, A.; Kolodziejczyk, K.; Zborowska, U.; Ulaczyk, J. Gasoline quality prediction using gas chromatography and FTIR spectroscopy: An artificial intelligence approach. *Fuel* **2006**, *85*, 553–558, DOI: 10.1016/j.fuel.2005.07.019.
- (41) Sharpe, S. W.; Johnson, T. J.; Sams, R. L. *Infrared Spectral Database*; tech. rep.; Pacific Northwest National Laboratory (PNNL), 2008.
- (42) Derfer, J. M.; Boord, C. E.; Burk, F. C.; Hess, R. E.; Lovell, W. G.; Randall, R. A.; Sabina, J. R. *Knocking characteristics of pure hydrocarbons*; tech. rep. 225; American Society for Testing Materials, 1958, DOI: 10.1520/STP225-EB.
- (43) Solaka, H.; Tuner, M.; Johansson, B.; Cannella, W. Gasoline Surrogate Fuels for Partially Premixed Combustion, of Toluene Ethanol Reference Fuels. *SAE* **2013**, DOI: 10.4271/2013-01-2540.

- (44) Badraa, J.; AlRamadan, A.; Sarathy, M. Optimization of the octane response of gasoline/ethanol blends. *Applied Energy* **2017**, *203*, 778–793, DOI: 10.1016/j.apenergy.2017.06.084.
- (45) Foong, T. M.; Morganti, K. J.; Brear, M. J.; da Silva, G.; Yang, Y.; Dryer, F. L. The octane numbers of ethanol blended with gasoline and its surrogates. *Fuel* **2014**, *115*, 727–739, DOI: 10.1016/j.fuel.2013.07.105.
- (46) Perez, P. L.; Boehman, A. L. Experimental Investigation of the Autoignition Behavior of Surrogate Gasoline Fuels in a Constant-Volume Combustion Bomb Apparatus and Its Relevance to HCCI Combustion. *Energy and Fuels* **2012**, *26*, 6106–6117, DOI: 10.1021/ef300503b.
- (47) Sarathy, S. M.; Kukkadapu, G.; Mehl, M.; Wang, W.; Javed, T.; Parka, S.; Oehlschlaeger, M. A.; Farooq, A.; Pitz, W. J.; Sung, C.-J. Ignition of alkane-rich FACE gasoline fuels and their surrogate mixtures. *Proceedings of the Combustion Institute* **2015**, *35*, 249–257, DOI: 10.1016/j.proci.2014.05.122.
- (48) Cameron, D. M. Autoignition Studies of Gasoline Surrogate Fuels in the Advanced Fuel Ignition Delay Analyzer., Ph.D. Thesis, University of Colorado at Boulder, 2017.
- (49) Sarathy, S. M. et al. Compositional effects on the ignition of FACE gasolines. *Combustion and Flame* **2016**, *169*, 171–193, DOI: 10.1016/j.combustflame.2016.04.010.
- (50) Javed, T.; Ahmed, A.; Lovisotto, L.; Issayeva, G.; Badra, J.; Sarathy, S. M.; Farooq, A. Ignition studies of two low-octane gasolines. *Combustion and Flame* **2017**, *185*, 152–159, DOI: 10.1016/j.combustflame.2017.07.006.
- (51) Cannella, W.; Foster, M.; Gunter, G.; Leppard, W. *FACE gasolines and blends with ethanol: detailed characterization of physical and chemical properties*; tech. rep. AVFL-24; Coordinating Research Council, 2014.
- (52) Wold, S. Personal memories of the early PLS development. *Chemometrics and Intelligent Laboratory Systems* **2001**, *58*, 83–84, DOI: 10.1016/S0169-7439(01)00152-6.

- (53) Cortes, C.; Vapnik, V. Support-vector networks. *Machine Learning* **1995**, *20*, 273–297, DOI: 10.1007/BF00994018.
- (54) Werbos, P. J., *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting, The Science of Microfabrication*; John Wiley: 1994.
- (55) Riedmiller, M.; Braun, H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. *IEEE* **1993**, 586–591, DOI: 10.1109/ICNN.1993.298623.