Measuring and Predicting Sooting Tendencies of Oxygenates, Alkanes, Alkenes, Cycloalkanes, and Aromatics on a Unified Scale

Dhrubajyoti D. Das[a,1], Peter St. John[b,1], Charles S. McEnally[a,*], Seonah Kim[b], Lisa D. Pfefferle[a]

[a] *Yale University, Department of Chemical and Environmental Engineering, New Haven CT 06520*
[b] *National Renewable Energy Laboratory, Golden CO 80401*

Abstract

Soot from internal combustion engines negatively affects health and climate. Soot emissions might be reduced through the expanded usage of appropriate biomass-derived fuels. Databases of sooting indices, based on measuring some aspect of sooting behavior in a standardized combustion environment, are useful in providing information on the comparative sooting tendencies of different fuels or pure compounds. However, newer biofuels have varied chemical structures including both aromatic and oxygenated functional groups, making an accurate measurement or prediction of their sooting tendency difficult. In this work, we propose a unified sooting tendency database for pure compounds, including both regular and oxygenated hydrocarbons, which is based on combining two disparate databases of yield-based sooting tendency measurements in the literature. Unification of the different databases was made possible by leveraging the greater dynamic range of the color ratio pyrometry soot diagnostic. This unified database contains a substantial number of pure compounds ($\geq$ 400 total) from multiple categories of hydrocarbons important in modern fuels and establishes the sooting tendencies of aromatic and oxygenated hydrocarbons on the same numeric scale for the first time. Using this unified sooting tendency database, we have developed a predictive model for sooting behavior applicable to a broad range of hydrocarbons and oxygenated hydrocarbons. The model decomposes each compound into single-carbon fragments and assigns a sooting tendency contribution to each fragment based on regression against the unified database. The model's predictive accuracy (as demonstrated by leave-one-out cross-validation) is comparable to a previously developed, more detailed predictive model. The fitted model provides insight into the effects of chemical structure on soot formation, and cases where its predictions fail reveal the presence of more complicated kinetic sooting mechanisms. This work will therefore enable the rational design of low-sooting fuel blends from a wide range of feedstocks and chemical functionalities.

Keywords: soot, biofuels, color ratio pyrometry, Group contribution method

1. Introduction

A sooting tendency is a parameter that characterizes the chemical component of the propensity of a pure compound or fuel mixture to produce soot particles in a combustion environment. It

*Corresponding author: charles.mcenally@yale.edu
[1]These authors contributed equally to this work.

can be measured via several laboratory techniques: these include the height of the smoke point flame when the test fuel is burned in a wick-burner [1], the yield of soot in a flame whose fuel is doped with a small amount of the test fuel [2], and the amount of carbon deposited when the test fuel is pyrolyzed in a packed-bed reactor [3]. Numerous databases have been reported of sooting tendencies measured using these techniques [1–26]. Several semi-empirical predictive models have been developed from these experimental results that can be used to determine sooting tendencies of compounds that have not been studied experimentally [8, 13, 17, 25, 27–30].

The results from these different configurations agree reasonably well, which indicates that sooting tendency is a true fuel property. The underlying chemical mechanism is that different fuels produce different pools of primary reaction products, and these different pools grow to large aromatic hydrocarbons and soot at different rates [2, 6, 13]. For example, a single-ring aromatic fuel such as toluene will produce aromatic products that can grow directly to two-ring aromatics, whereas an alkane fuel such as *n*-heptane will produce aliphatic products that must grow to single-ring products before they can grow to two-ring aromatics [31]; thus single-ring aromatics have much greater sooting tendencies than alkanes [7].

The sooting tendency of the fuel is important because it strongly affects particulate formation and emissions from practical combustion devices. For example, the specifications for Jet A aviation fuel include a provision that its smoke point meet certain criteria; the purpose of this provision is to ensure that radiation heat transfer from soot particles does not overheat the combustor liner [32, 33]. Similarly, Yang and co-workers found that smoke point sooting tendencies were predictive of particulate concentrations in gas turbine exhausts [10]. In some systems other fuel properties may also affect soot formation, and sooting tendency has to be combined with these other properties to predict emissions. For example, in gasoline direct-injection engines, the fuel volatility affects the degree of mixing during the compression stroke and therefore the amount of particulates formed during the combustion phase; Aikawa and co-workers proposed and validated a quantity called Particulate Mass Index (PMI) to estimate emissions from the joint effects of fuel volatility and sooting tendency [34].

A major challenge for sooting tendency databases and models is the wide range of hydrocarbons found in modern fuels. Petroleum-derived fuels generally contain linear and branched alkanes, alkenes, cycloalkanes, and aromatics [35–37]. Furthermore, these fuels increasingly also contain oxygenated hydrocarbons derived from biomass. Currently most gasoline sold in the United States contains 10 volume % ethanol [38], and most gasoline sold in Brazil has contained 20 vol% or more ethanol for several decades [39]. The US Department of Energy's Co-optimization of Fuels and Engines program – one of the sponsors of this research – is examining the use of fuels with blendstocks other than ethanol. A wide range of promising blendstocks have been identified, including higher alcohols, esters, ethers, furans, and ketones [40].

Unfortunately, all of the existing sooting tendency databases contain either oxygenates or aromatics, but not significant numbers of both. Consequently predictive models of sooting tendency have been explicitly limited to one category or the other. The most significant exception is the work of Ladommatos and co-workers, which contains 3 aromatics (benzene, toluene, and 2-xylene) and 5 alcohols [9]. The underlying reason for the absence of this concurrent sooting data is that the sooting tendencies of aromatics and oxygenates differ greatly, which makes it difficult to obtain accurate measurements for both in the same experiment. For example, we have measured sooting tendencies of oxygenates and aromatics in previous work [2, 11, 12, 22], but the results are contained in two incompatible databases due to the limited dynamic range of the laser-induced incandescence

diagnostic used in that work.

The objective of the research reported here was to produce a unified sooting tendency database that contains substantial numbers of compounds from all of the categories important in modern fuels – including alkanes, cycloalkanes, aromatics, and oxygenates – and to develop a predictive model applicable to all of them. The database was created by stitching together the two earlier incompatible databases (see Table 1); we have implemented a color-ratio pyrometry diagnostic that has better dynamic range than the earlier diagnostics [23, 41], and used it to measure compounds from both databases in the same set of experiments. In an earlier work we generated a predictive model for sooting tendency based on molecular descriptors [29], however, the increased range of molecular sizes and sooting tendencies in the combined database in this work is better suited to an approach that naturally captures the increase in sooting tendency with an increase in the size of molecules. A new model was developed that is based on decomposing each compound into individual carbon-atom fragments, and assigning a sooting tendency contribution to each fragment based on regression against the unified database. The form of this group contribution model is similar to that used by Pepiot-Desjardins and co-workers to predict sooting tendency [13], and ultimately derives from the approach of Benson to estimating thermodynamic properties [42].

Sooting tendency in this study is defined by the amount of soot produced in a methane/air nonpremixed flame when a small quantity of the test compound of interest is added to the fuel. Since the absolute amount of soot is strongly dependent on the additive concentration and the burner configuration, we linearly rescale the soot concentration into a Yield Sooting Index (YSI) which is essentially independent of experimental details. To do this, we (1) choose two index compounds, A and B, (2) assign them YSI values, YSI_A and YSI_B, and (3) measure their maximum soot concentration in the flames, $f_{v,\mathrm{max},A}$ and $f_{v,\mathrm{max},B}$. Then the YSI for any other test compound i is defined by

$$\mathrm{YSI}_i = (\mathrm{YSI}_A - \mathrm{YSI}_B) \times \frac{f_{v,\mathrm{max},i} - f_{v,\mathrm{max},B}}{f_{v,\mathrm{max},A} - f_{v,\mathrm{max},B}} + \mathrm{YSI}_B \tag{1}$$

This procedure is analogous to octane rating, where the knock intensity measured for a fuel is translated onto a scale where $\mathrm{ON}_{\mathrm{heptane}} = 0$ and $\mathrm{ON}_{\mathrm{isooctane}} = 100$ [43]. In this study the index compounds are hexane and benzene, and the values assigned to them are $\mathrm{YSI}_{\mathrm{hexane}} = 30$ and $\mathrm{YSI}_{\mathrm{benzene}} = 100$; the rationale for these choices is discussed in Sections 3.3 and 3.4.

One subtlety of YSI is that the additive concentration can be defined by either the additive mole fraction or mass fraction in the fuel. We will henceforth refer to YSIs determined by these two definitions as YSI-molar and YSI-mass, respectively. YSI-molar is arguably more fundamental and we have used it in our studies of pure hydrocarbons, e.g., [2, 11, 12, 22]. However, adding a specified mole fraction of the test compound requires knowing its average molecular weight; for practical fuels that are complex mixtures of many hydrocarbons, the average molecular weight is typically not known accurately. Thus for these fuels YSI-mass is preferable, and we have used it in our studies of diesel and jet fuel surrogates [24]. Since the current study is focused on pure hydrocarbons, it uses YSI-molar.

The rest of the paper is organized as follows. The paper begins with an outline of the experimental approach adopted to unite the disparate YSI databases (Section 2.1) and lays out the framework of the linear regression model to predict YSI values (Section 2.2). The results of the unification efforts, verification of the unified YSI values, and a discussion of the rationale behind redefinition of the YSI scale are presented in the first part of Section 3. The performance of the predictive model and some insights into sooting chemistry are presented in the second part of Section 3. Finally, in

Table 1: Summary of literature sources of Yield Sooting Indices (YSI), the index (or reference) compounds used to define the YSI scale therein, and the YSI values for the respective index compounds.

Database	Literature sources	Index compounds	Assigned YSI
"High scale"	McEnally and Pfefferle (2007) [2], McEnally and Pfefferle (2009) [11]	benzene naphthalene	30 100
"Low scale"	McEnally and Pfefferle (2011) [12], Das *et al.* (2015) [22]	*n*-hexane benzene	0 100
"Unified scale"	This work	*n*-hexane benzene	30 100

Section 4, we discuss some limitations of the model in light of notable innaccurate predictions and degeneracies within the model regarding hydrocarbon isomers.

2. Methods

2.1. Experimental

The experimental approach and devices used here have been discussed in detail previously [23]. A brief summary of the approach along with some associated modifications are noted here.

2.1.1. Overall approach

Atmospheric pressure, axisymmetric, coflowing, nonpremixed, laminar methane/air flames were established on a Yale Coflow burner [44]. Sooting tendencies of pure compounds (listed in Tables 2 and 3) were determined by doping them separately into the fuel of a flame. In keeping with the approach adopted in our previous work [2, 11, 12, 22], the dopants were added to the fuel of the flame on a constant mole fraction basis. Spatially resolved two-dimensional soot volume fraction (f_v) maps in these flames were measured using color-ratio pyrometry. The peak-region soot volume fraction in these maps, $f_{v,\max}$ (defined in Sec 2.1.3 below), was used to determine a Yield Sooting Index (YSI) for each dopant as per Equation (1).

2.1.2. Burner geometry and flame details

The fuel mixture – methane (mole fraction $X_{CH_4} = 42\%$), nitrogen ($X_{N_2} = 57.9\%$), and a dopant ($X_{\text{dopant}} = 1000$ ppm) – flows out of a 4mm inner diameter tube and reacts with air that flows from the annular region between this tube and a 7.4 cm inner diameter concentric aluminum tube, which acts as a chimney. The nominal reactant flow rates were 238 cm^3/min of CH_4, 328 cm^3/min of N_2, and 50,000 cm^3/min of air. All dopants were liquids at room temperature, with liquid-phase flowrates ranging from 82 to 321 μL/hour. Dopant flowrates varied depending on the liquid-phase mass density of any individual dopant so as to establish a dopant mole fraction of 1000 ppm in the fuel. All reactants were obtained from Sigma Aldrich (liquid-phase dopants), Airgas (ultra-high purity grade CH_4 and N_2 cylinders), and laboratory dry air. All dopants were dispensed into the gas-phase fuel mixture through a syringe pump (pump = Cole Parmer EW-74900, syringe = Hamilton Gastight 1710). The needle of the syringe charged with the dopant was introduced into the fuel line through a septum in a stainless steel tee. The fuel line and fuel tube were heated to

145°C with temperature-controlled resistive tapes. At these thermal conditions, the vapor pressure [45] of 2,2'-dimethylbiphenyl ($\approx$ 3500 Pa), the least volatile of all dopants used, is approximately 35 times greater than its partial pressure in the fuel mixture ($\approx$ 100 Pa). Therefore, all liquid reactants vaporized upon injection and were swept as gases to the flame by other fuel components. Each dopant was allowed to flow for at least 30 min prior to data acquisition. Time-resolved flame luminosity measurements confirmed that all dopants achieved adsorption/desorption equilibrium with the walls of the fuel line and fuel tube within this 30 min interval.

2.1.3. Soot volume fraction measurements

Two-dimensional f_v maps of these flames were measured using color-ratio pyrometry. This technique of quantifying soot concentration in coflow flames has been used in our previous work [23, 24], and elsewhere in the literature [41, 46–48]; the approach employed in Das *et al.* (2017)[24] was adopted without modification to obtain f_v maps of all flames in this work. A "peak-region" soot volume fraction $f_{v,\max}$ was defined to provide a single quantity of interest for characterizing the soot load in these flames. In previous work $f_{v,\max}$ has typically been defined to be the maximum f_v as measured along the centerline of an axisymmetric flame, if f_v peaks along the centerline, or as the global maximum value of f_v, if the soot peaks off-axis or along the wings of a flame. In all flames considered in this work, f_v peaked along the centerline of the flame. Unfortunately, f_v profiles close to this symmetry axis are extremely susceptible to noise arising from the Abel deconvolution process [49], and a naïve determination of the maximum value of f_v along the centerline could lead to an erroneous characterization of the overall soot load. Therefore, to obtain a more robust measure of the overall soot load, we defined the "peak-region" $f_{v,\max}$ to be the soot concentration averaged over the sootiest parts of the flame, which we define to be regions with f_v greater than the 90^{th} percentile for each flame.

For illustrative purposes, Figure 1 shows the full 2d f_v maps for a series of benzene-doped flames at different dopant concentrations. The top row of f_v panels shows the soot distribution over the full region of the flame where soot was detectable. The panels in the bottom row isolate the region over which f_v is averaged to yield $f_{v,\max}$ in each flame. We argue that this $f_{v,\max}$ is representative of the actual maximum f_v along the flame centerline for the following three reasons -

1. The f_v in these regions is close to single-valued in comparison to the variation in f_v elsewhere in the flame. This independence of f_v with position is to be expected when f_v is near a local maxima.

2. The center-of-mass of the soot region isolated in the bottom panel of Figure 1 occurs at the same height above burner (41.0 $\pm$ 0.5 mm) independent of the dopant concentration. Though we have only shown these regions for a series of benzene-doped flames in Figure 1, this observation holds true for all flames that we studied in this work. This independence of the height of the peak f_v in doped flames has been previously observed in f_v measurements performed in our lab [2] and elsewhere [50] using other diagnostic techniques such as laser-induced incandescence (LII) that do not involve Abel deconvolution and therefore do not have noisy centerline profiles.

3. We find $f_{v,\max}$ derived from this region to be linearly correlated with the dopant concentration (X_{dopant}). This can be seen in Figure 2, where we plot $f_{v,\max}$ vs. X_{dopant} for the series of benzene-doped flames in Figure 1, as well as three other series of flames doped separately with 1-methylnaphthalene, *n*-hexane, and ethanol. These four dopants were chosen to span

the range of soot volume fractions that would be measured in this study. All of the $f_{v,\max}$ measured in this study are in the range of the data in Figure 2. The data closely fit the straight lines shown in the figure for all four dopants, indicating that the measured soot volume fractions, as quantified by $f_{v,\max}$ in these flames, depend linearly on X_{dopant} at these doping levels. This observation is consistent with measurements of variation in soot levels with dopant concentration in doped flames in previous work from our group [2, 22], thereby increasing our confidence in $f_{v,\max}$ as defined here as being an accurate measure of the overall soot level in these flames.

The measurements presented in Figure 2 also serve as a test of the dynamic range of this diagnostic procedure. From Figure 2 (B), we see that we are able to measure changes in $f_{v,\max}$ across more than two orders of magnitude with very high signal-noise ratio. This wide dynamic range of the color ratio pyrometry diagnostic enables us to characterize the sooting tendency of both very low sooting compounds (e.g. ethanol) and high sooting compounds (e.g. 1-methylnaphthalene) under identical conditions, and thereby unify the sooting tendency measurements of low and high sooting compounds onto a single scale.

2.2. Linear Regression Modeling

In this study, we propose a modified Benson group-increment model to predict the YSI of a compound from its molecular structure [51]. In this approach, the sooting tendency of a molecule is modeled as a linear sum of the YSI contributions from each of its component carbon-centered fragments.

2.2.1. Decomposition of molecules into fragments

The decomposition of molecules into fragments is handled similarly to that of previous group-contribution methods [42, 51, 52]. We specifically follow a slightly modified version of the method of Benson & Buss (BB) [51], in which molecule fragments are determined by carbon atoms together with their bonds and ligands. Unlike the BB method, divalent oxygen atoms were excluded; including these fragments lowered the overall identifiability of fragment parameters without substantially improving the fit quality. Carbon centers were further specified by whether or not they were aromatic or present in a ring structure. No explicit determinations were made between rings of different sizes. The 441 molecules present in the unified YSI database resulted in 66 unique fragment types.

2.2.2. Bayesian linear regression

One of the main goals of the group contribution regression is to compare the relative contributions of different carbon types towards soot formation. As such, we constructed the linear regression using Bayesian inference in order to determine our confidence in the relative contribution from different carbon types after observing unified YSI values from the complete database [53]. We assume the YSI measurements are independent, normally distributed, and have an error proportional to their estimated measurement error:

$$y|\,\beta,\sigma,X \sim \mathrm{N}\left(X\beta, \mathrm{diag}\left(\frac{1}{\sigma_{\text{exp}}^2}\right)\sigma^2\right) \tag{2}$$

Here, y are the experimental YSI measurements, X is the feature matrix of fragment counts and an intercept, β are the slope terms representing YSI contribution from each carbon type, σ_{exp} are the

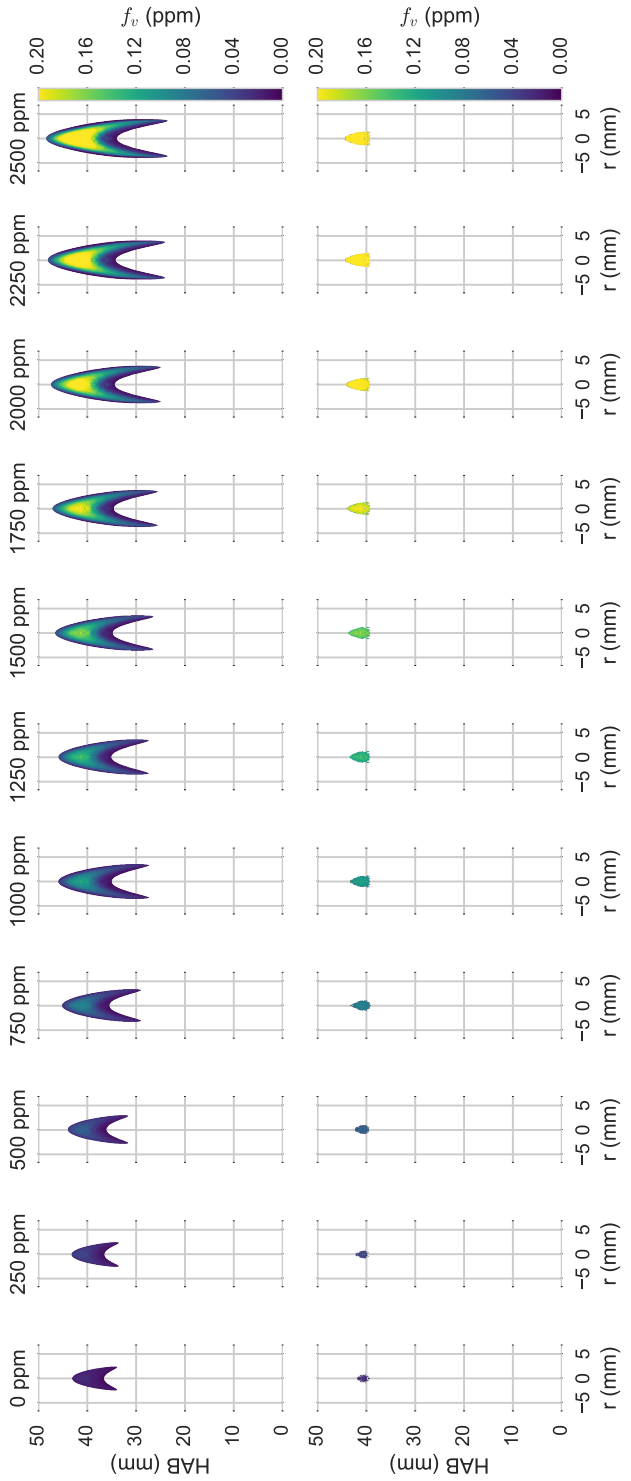

Figure 1: (Top row) 2D f_v maps of methane flames doped with different mole fractions of benzene. Doped benzene concentration in each flame is indicated by the title of each panel. (Bottom row) Same f_v maps as the top row, with regions with soot $\geq 90^{\text{th}}$ percentile isolated. The average f_v in regions isolated in the bottom row is defined as the peak-region soot volume fraction, $f_{v,\text{max}}$.

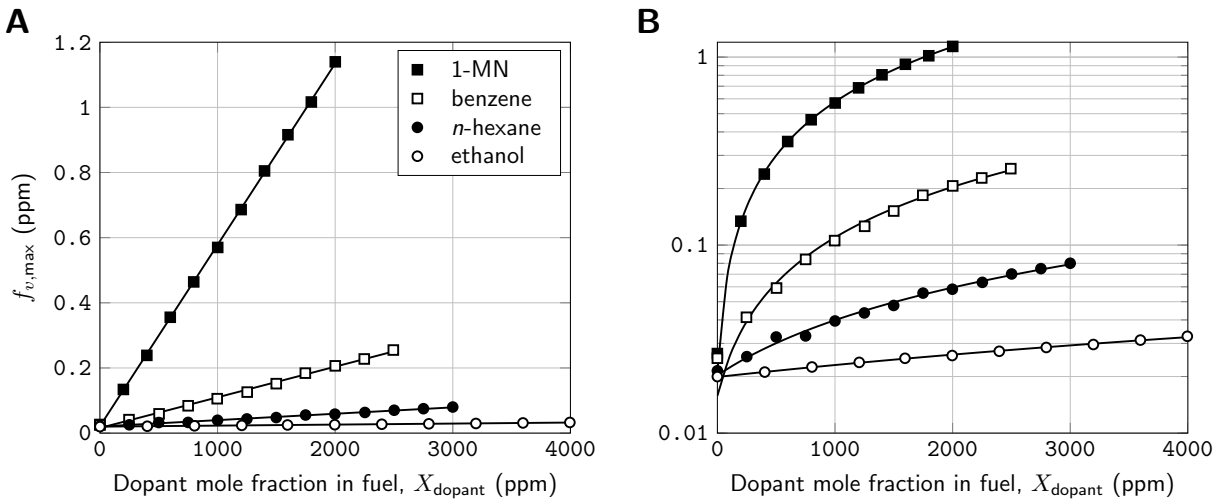

Figure 2: Maximum centerline soot volume fraction for flames doped with 1-methylnaphthalene (■), benzene (□), *n*-hexane (●), and ethanol (○) at different dopant concentrations in the fuel, graphed on a **(A)** linear and **(B)** semilog scale. The black lines are linear fits to the data points in both graphs. $R^2 > 0.995$ for all fits.

experimentally determined standard deviations, and σ is the unknown scale term. Additionally, we include informative priors on the slopes β, which function as a regularization penalty to improve model identifiability when groups of fragments are not linearly independent. These priors are formalized as artificial data points, with a 'measured' YSI of $y = 0$, $\sigma_{\text{exp}} = 25$, and a feature matrix $X = (0, I)$ [53]. The effect of these priors is to evenly distribute YSI contributions over groups of linearly dependent fragments. Model regressions were performed in Python; codes used are available from `https://github.com/pstjohn/ysi-fragment-prediction`.

2.2.3. Defining a domain of applicability

In any quantitative structure-activity relationship (QSAR) model, it is critical to establish the set of molecules for which predictions from the model can be deemed trustworthy [54]. In this work, we consider all molecules that are comprised of only carbon fragments that have previously been seen by the model to be inside the applicability domain. To make this definition more rigorous, we define the applicability domain to be any molecule whose fragment vector can be expressed as a linear combination of training set compound fragments. This definition prevents inconsistencies from arising from non-identifiable sets of fragments that always appear together.

2.2.4. Density functional theory

Quantum chemistry calculations were performed to help explain some of the discrepancies between QSAR-predicted and experimentally measured YSI values for isomers of methylcyclohexenes (see Section 4.4). Gaussian 09 [55] was used to fully optimize all stationary points using the composite G4 method for Density Functional Theory (DFT) calculations of methylcyclohexene isomers in Section 4.4. We performed optimizations with the hybrid GGA B3LYP density functions with the 6-31G(d) basis sets. Harmonic vibrational frequencies were calculated for all optimized structures including transition states to verify energy minima, possessing zero imaginary frequency

for optimized structures and one imaginary frequency for transition states. To verify corresponding reactants and products from transition state, the intrinsic reaction coordinate (IRC) calculations using B3LYP/6-31(G) were also performed.

3. Results

3.1. Creating the unified YSI scale

Table 2: List of compounds from different literature YSI databases whose sooting tendency was re-measured on a single basis to establish a unified YSI scale. Low scale and High scale refers to compounds featured in the YSI databases listed in Table 1.

Low scale [12, 22]	Original YSI[a]	Unified YSI[b]
n-hexane	0*	30*
benzene	100*	100*
ethanol	−31.1	14.6
1-pentanol	−7.7	25.3
1-octanol	16.7	38.8
trans-2-octene	40.8	54.7
1-decene	69.0	69.1
2,2,4,6,6-pentamethylheptane	106.9	100.3
High scale [2, 11]	**Original YSI[c]**	**Unified YSI[b]**
2-heptanone	17	30.0
benzene	30*	100*
ethylbenzene	53.6	233.8
1,2,3,4-tetrahydronaphthalene	75.1	347.5
1,2-dihydronaphthalene	100*	484.6
1-methylnaphthalene	135	643.9
1-ethylnaphthalene	151	709.8
2,2'-dimethylbiphenyl	199	928.0

a: experimental uncertainty ±2.0 YSI units

b: expt. uncertainty is the greater of ±4.0 YSI units or ±3.0%

c: expt. uncertainty ±3.0%

∗: by definition

In this work, the two older incompatible YSI databases were combined by re-measuring sets of compounds from each with the same diagnostic. In Figure 3, we have plotted the older literature YSI ("Low scale" and "High scale") against the newly-measured Unified YSI for the compounds listed in Table 2. We have also graphed the best fit lines for the "Low scale" vs "Unified scale" and "High scale" vs "Unified scale" datasets on the same figure. The best fit lines were determined through orthogonal distance regression, which is a form of least squares regression aimed at minimizing errors in both "x" (literature YSI) and "y" (Unified YSI) variables [56]. The equations of the best fit lines, Equations (3) and (4), are the stitching relationships used to map the YSI values for the compounds in the respective datasets on to the unified YSI scale. By applying these stitching

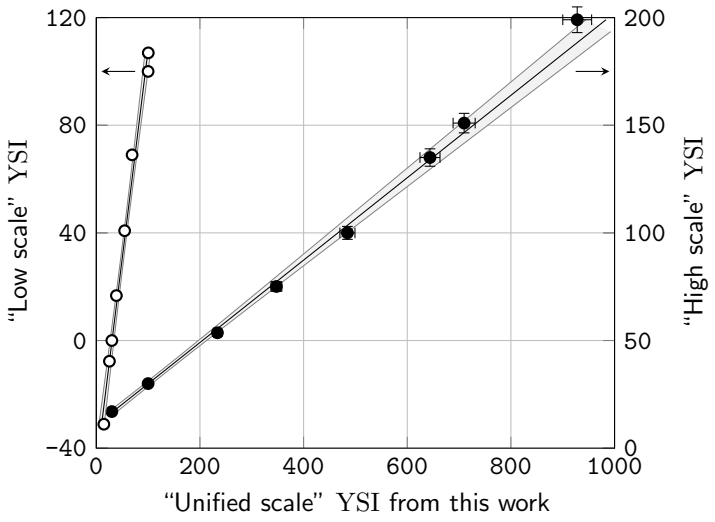

Figure 3: Comparison between literature values of the YSI and unified YSI values measured in this work for "Low scale" (○) and "High scale" (●) compounds listed in Table 2. For each dataset, the solid black line is the best fit line between the literature and unified YSI values as determined through orthogonal distance regression. The equation of the best fit lines is given in Equations (3) and (4). Shaded region represents the 95% confidence band for the regression line.

relationships to the set of YSI measurements in the Low scale and High scale databases, the YSI measurements of all compounds contained in the two databases can be put on the same Unified scale. The full set of values in the Unified scale are available online [57].

$$\mathrm{YSI_{Unified}} = 0.64^{[\pm 0.03]} \times \mathrm{YSI_{Low}} + 30.4^{[\pm 1.9]} \tag{3}$$

$$\mathrm{YSI_{Unified}} = 5.23^{[\pm 0.12]} \times \mathrm{YSI_{High}} - 56^{[\pm 5]} \tag{4}$$

In Figure 4, we schematically illustrate the application of the two stitching relationships to map the YSI measurements from the Low and High scales onto the Unified scale. We see that the low sooting compounds are compressed to the lower end of the Unified scale. The high sooting compounds, whose sooting tendency previously could not be directly compared to that of the low sooting compounds, now occupy most of the range on the Unified scale. The Unified scale itself spans more than three orders of magnitude.

3.2. Verification of the unification process

The validity of the process for unifying the YSI scales was verified in two ways. First, the unified YSIs of several hydrocarbons (see Table 3) from the Low and High scale databases were measured directly. These hydrocarbons are distinct from the set of hydrocarbons in Table 2, which were used to derive the stitching relationships, and therefore serve as an independent internal validation tool. The measured unified YSIs were compared to the unified YSIs expected for these compounds based on the stitching relationships, Equations (3) and (4). Figure 5 shows a comparison between the measured and expected unified YSIs for these hydrocarbons on a linear and logarithmic scale. It is

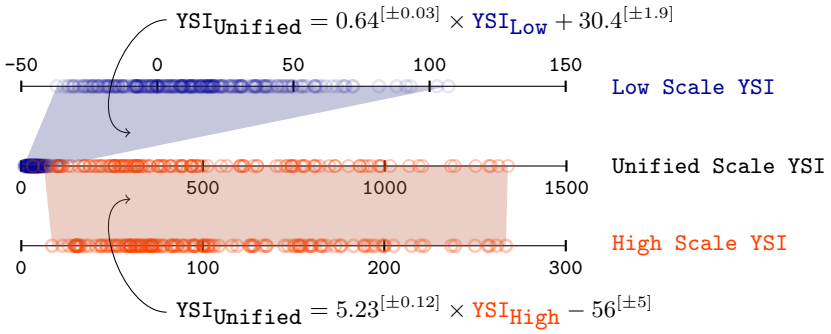

Figure 4: Schematic illustration of transformation of YSI values from the two extant incompatible YSI scales, "Low Scale" (○) and "High Scale" (○) in Table 1, onto a mutually compatible "unified" YSI scale with the stitching relationships, Equations (3) and (4). Colored circles represent YSIs of individual compounds in the respective databases. For references to color, the reader is referred to the web version of this article.

Table 3: List of dopants selected from YSI databases in Table 1 to validate the stitching relationships, Equations (3) and (4).

Database	Validation dopant	Original YSI	Expected $\mathrm{YSI_{unified}}$	Measured $\mathrm{YSI_{unified}}$
Low scale	1-propanol	−22.0	16.2 ± 2.7	18.9 ± 4.0
	2-methyl-1-butanol	3.0	32.3 ± 2.2	31.4 ± 4.0
	3-heptanol	12.6	38.5 ± 2.1	37.1 ± 4.0
	2,2-dimethylbutane	20.2	43.4 ± 2.0	39.8 ± 4.0
	2-ethyl-1-hexanol	31.1	50.4 ± 2.0	45.7 ± 4.0
	propylcyclohexane	60.3	69.2 ± 2.1	63.7 ± 4.0
High scale	decahydronaphthalene	31.0	105.5 ± 5.8	118.6 ± 4.0
	1,4-dimethylbenzene	51.2	211.1 ± 8.8	222.8 ± 6.7
	1,2,4-trimethylbenzene	69.8	308 ± 12	315.9 ± 9.5
	4-*tert*-butyltoluene	89.4	411 ± 16	396 ± 12

evident that the measured unified YSIs agree very well with the unified YSI values expected from transforming the literature YSI values with the stitching relationships.

In the second verification approach, three binary mixtures of *n*-dodecane and *iso*-butylbenzene were prepared. The YSI of *n*-dodecane, an aliphatic hydrocarbon, was reported on the Low scale in the literature [12] (original YSI = 64.2, expected Unified YSI = 71.7), while the YSI of *iso*-butylbenzene, an aromatic hydrocarbon, was reported on the High scale [2] (original YSI = 60.1, expected Unified YSI = 257.6). The unified YSI values of the three binary mixtures were measured experimentally and compared to the expected unified YSIs of the pure compounds. In Figure 6 the expected unified YSIs of the pure compounds and the measured unified YSIs of the binary mixtures are compared against the composition of the mixtures (represented by the mole fraction of the *n*-dodecane component). From the dashed line in Figure 6, it is evident that the expected unified YSI values of the pure compounds fall in line with the measured YSI values of the binary mixtures. This agreement indicates that (1) the literature YSI values of *n*-dodecane and *iso*-butylbenzene

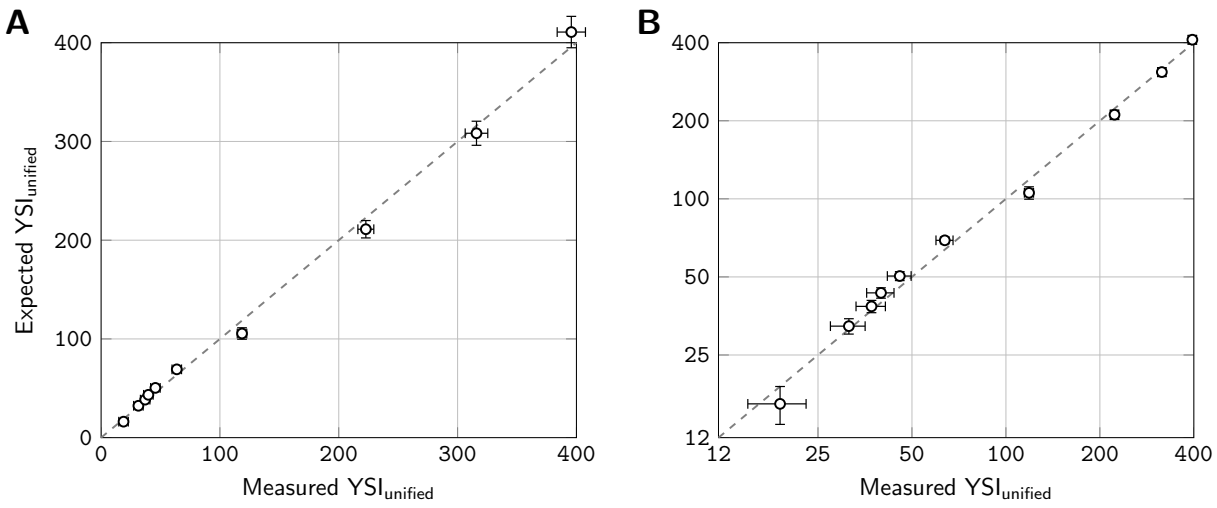

Figure 5: Comparison between the measured and expected unified YSIs for a selection of compounds from the Low and High scale databases listed in Table 3, graphed on **(A)** linear and **(B)** logarithmic axes. The dashed lines indicate $x = y$, or perfect agreement, in both graphs.

(and by extension, the other compounds in the respective databases, with the presumption that the reported YSI values are internally consistent) have been accurately transformed on to a unified YSI scale, and (2) there is no evidence of a non-linear or synergistic interaction between *n*-dodecane and *iso*-butylbenzene when the sooting tendency of mixtures of these compounds is measured in a doped methane flame at these concentrations.

3.3. Rationale for selecting YSI values of the index compounds in the Unified YSI scale

Defined YSI values for the index compounds (YSI_A and YSI_B in Equation (1)) are necessary to translate experimentally measured soot concentrations into numeric YSI values for various dopants. The YSI values of the index compounds defining the previous YSI scales were chosen to either nominally conform to existing alternative sooting tendency measures for similar compounds in the literature, or in the case of new compound families, for numerical convenience. The index compounds and their YSI values (Table 1) for the High scale were chosen to match older Threshold Sooting Index (TSI) values for benzene and naphthalene (30 and 100 respectively) from Olson *et al.* [7]. TSI is a measure of sooting tendency quantified on the basis of smoke point heights of flames fueled entirely by the pure compounds in a standard burner. Different compounds (*n*-hexane/benzene instead of benzene/naphthalene) were chosen as the indices for the Low scale YSI measurements in order to better match the significantly lower sooting tendency range of the primarily non-aromatic compounds featured in that database.

While any of the preexisting reference YSI values could have been used to define the Unified scale, we used this opportunity to redefine the YSI scale to satisfy the following two criteria, which were not satisfied with the older YSI scales - (1) Numerical YSI values for all compounds should be greater than 0, and (2) YSIs of homologous series of hydrocarbons should tend to a numerical value of 0 when the number of carbon atoms in the hydrocarbons is extrapolated to 0. This second condition implies that all YSI measured in the future will also be ≥ 0. By redefining the Unified YSI

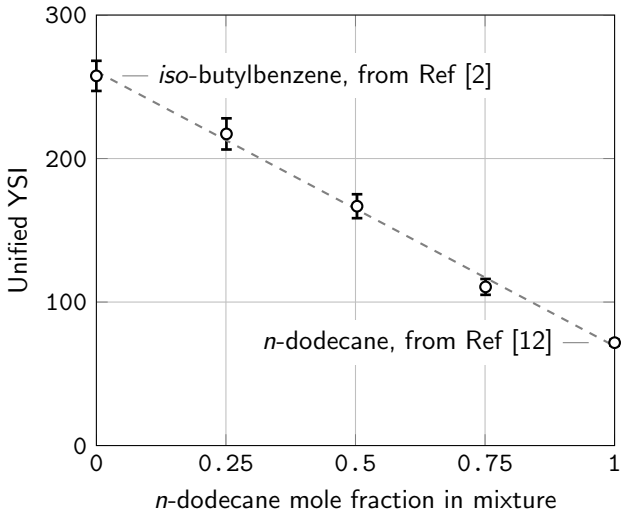

Figure 6: Variation in unified YSIs with composition of *n*-dodecane/*iso*-butylbenzene binary mixtures. The pure compound unified YSIs (*n*-dodecane mole fraction = 0 or 1) were transformed from the older YSI values using Equations (3) and (4), while the unified YSIs of the intermediate mixtures were measured directly. The dashed line is a linear fit to the data points. $R^2 > 0.995$.

scale such that the YSI of *n*-hexane ≡ 30 and the YSI of benzene ≡ 100, the YSI of the compound with the lowest sooting tendency measured to-date, methanol, is 6.6, thereby satisfying criteria #1 (the full YSI database can be accessed at [57]). Moreover, from Figure 7, we can see that if we graph the YSI of several homologous series of hydrocarbons against the number of carbon atoms in each compound, and extrapolate their YSIs to 0 carbon atoms, the Unified YSIs converge towards 0, thereby satisfying criteria #2.

3.4. Role of index compounds

As mentioned earlier in the Introduction, in order to define a relative scale such as the Unified YSI scale here, we need to specify at least two calibration points to set the "level" and "rate of change" for the scale. In this work, we have chosen *n*-hexane and benzene, with defined YSI values of 30 and 100, as calibration points. The choice of these two compounds, which are referred to as index compounds in this paper, was motivated primarily as a matter of practical convenience and is semi-arbitrary; we could have just as easily chosen any other pair of index compounds (for example, *n*-heptane ≡ 36 and toluene ≡ 171). We want to stress that our use of *n*-hexane and benzene as index compounds to unify the YSI scales should not be interpreted as an explicit or implicit recommendation that these two compounds be used as indices for all future YSI measurements. Any two compounds in the extant YSI database can be used as indices, and choosing indices other than *n*-hexane and benzene might be prudent when measuring the sooting tendency of new compounds or fuels in the future. As an example, in order to quantify the sooting tendency of large multi-ring aromatics, picking two index compounds with higher YSIs than *n*-hexane and benzene could be more appropriate from the perspective of minimizing experimental uncertainties. However, it is important to account for an additional source of uncertainty in measured YSI values when utilizing indices other than *n*-hexane/benzene. This uncertainty, which is sometimes called

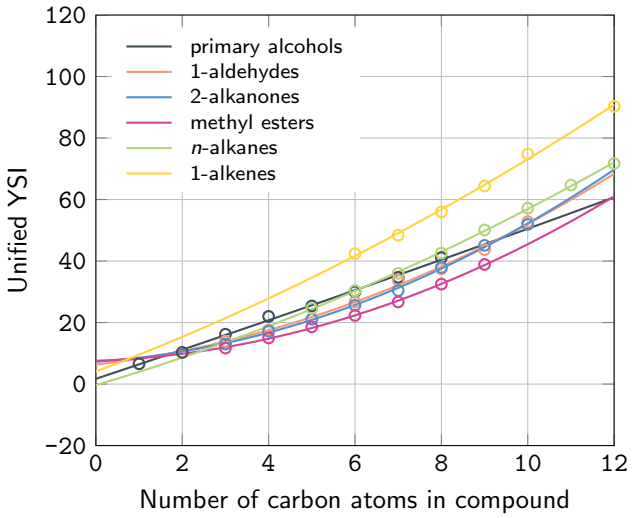

Figure 7: Variation in the Unified YSI of several (oxygenated) hydrocarbon families with size of compound, represented by the number of carbon atoms in the compound. Solid lines are quadratic fits to the existing YSI data, extrapolated down to zero carbon atoms in the compound. For references to color, the reader is referred to the web version of this article.

a "non-uniqueness" or inconsistency uncertainty, results from ambiguities in the scale definition when using new fixed-points (*i.e.*, new index compounds). See White and Saunders (2007) [58] for a good mathematical treatment on estimating and propagating this uncertainty, and Meyer and Tew (2006) [59] for an example application in the context of estimating the magnitude of this non-uniqueness uncertainty in the International Temperature Scale (ITS-90).

3.5. Development of a group increment model

In this work, a group-increment model to predict the YSI of a compound, using the number and types of different carbon-centered fragments in the molecule, was developed. Using these fragments (as defined in Section 2.2.1) and an intercept term, we fit a Bayesian multivariate linear regression to the observed Unified YSI data. Posterior distributions for the YSI contribution of the 66 carbon fragments present in the Unified YSI database are shown in Figure 8. This figure demonstrates that the regression matches our expectations, where the fragments that are observed most frequently (furthest to the right) have a correspondingly smaller 95% credible region, while those that appear only once are inferred with much lower confidence. Additionally, carbon fragments corresponding to aromatic molecules are among the sootiest carbon types, while those corresponding to oxygenates are typically lower than those corresponding to aliphatics.

A posterior predictive check, shown in Figure 9, was performed in order to ensure that the model appropriately weights errors across several orders of magnitude of YSI measurements. Using the inferred posterior distributions for β and σ, samples were drawn from the data distribution described in Equation (2). The 95% credible region for σ^2 was $[6.2, 8.1]$, which indicates that the variance in model predictions around the true YSI values is approximately 7 times the experimental variance. Additionally, 95% credible regions for predicted YSI values appropriately scale between compounds previously measured in the Low and High scale databases.

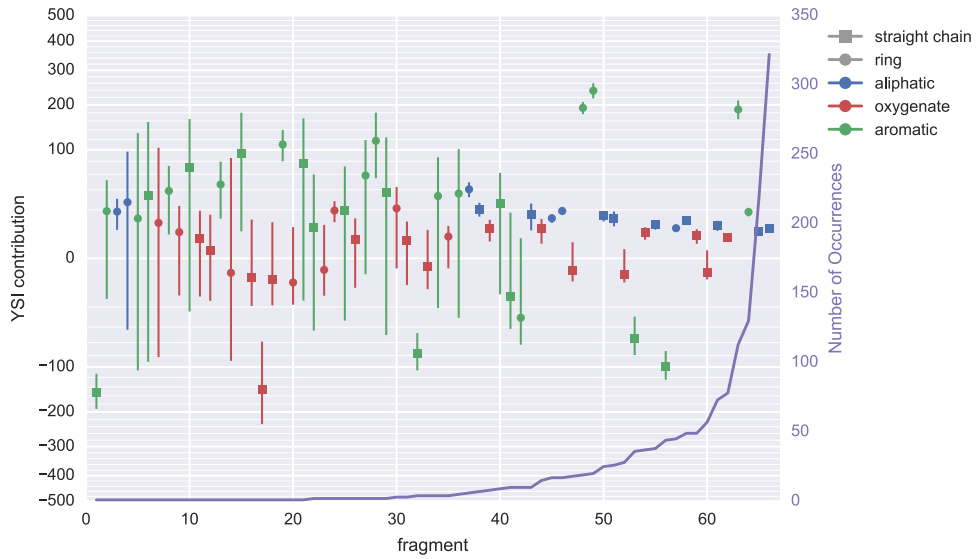

Figure 8: Median posterior estimates of the YSI contribution for each fragment type. Error bars extend to encompass the entire 95% credible region. Marker shape (square vs. circle) denotes whether the fragment center appears in a straight chain or ring, while marker color denotes whether the carbon is aromatic, aliphatic, or oxygenated. Fragments are sorted by the total number of times they appear in the dataset. Fragments which appear less frequently have higher associated uncertainty in their mean contribution.

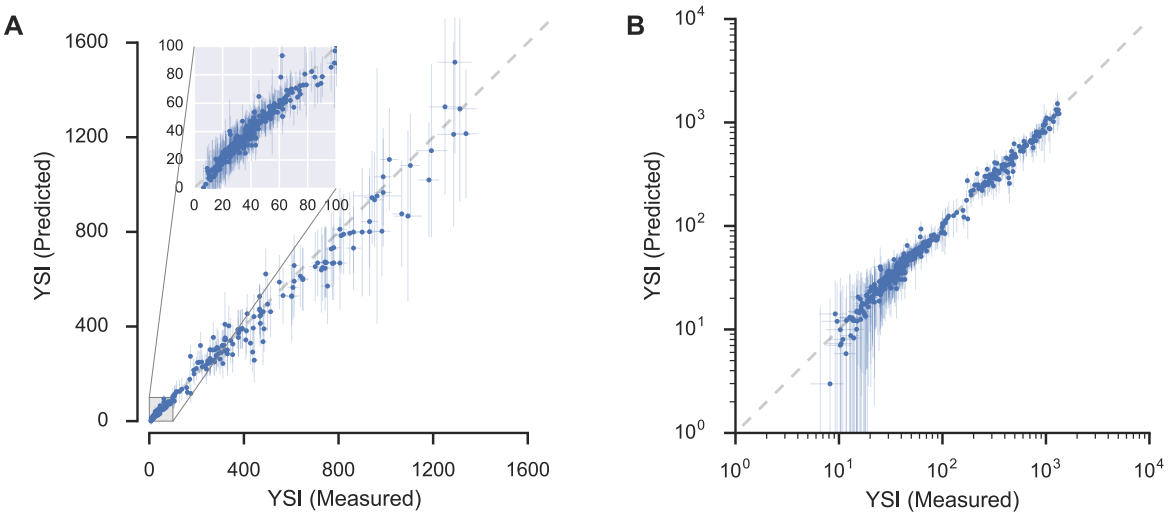

Figure 9: Posterior predictive interval of the YSI linear regression model. Predictions are shown on a **(A)** linear and **(B)** log scale. Error bars extend to 95% confidence intervals for measured YSI values, and to 95% highest posterior density intervals for predicted YSI values. Errors for the YSI predictions include the estimated model error.

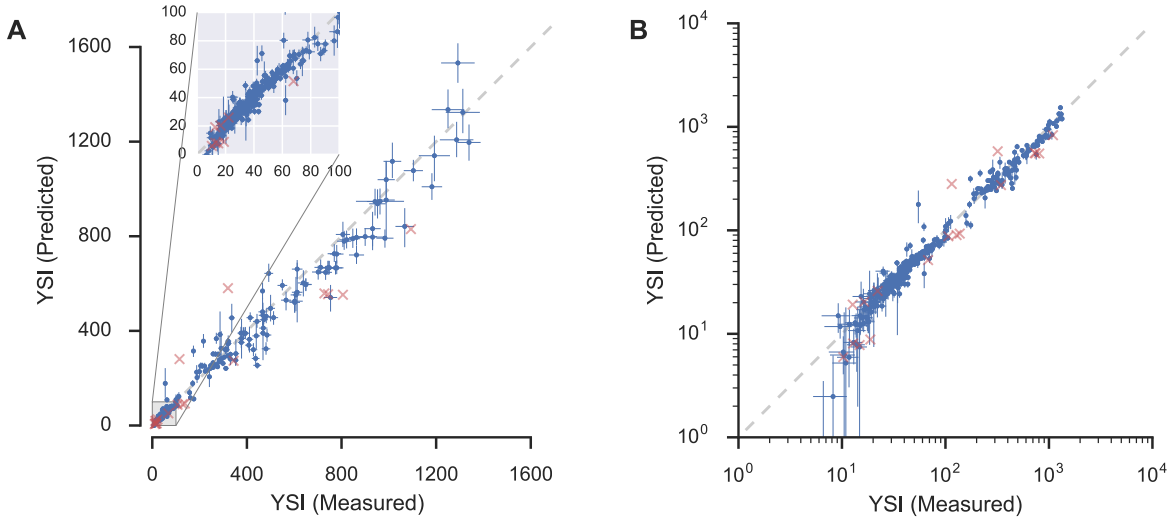

Figure 10: Performance of the group increment model under leave-one-out cross-validation. Predictions are shown on a **(A)** linear and **(B)** log scale. Error bars extend to 95% confidence intervals for measured YSI values, and to 95% highest posterior density intervals for the predicted means. Molecules with predictions that the fail the applicability domain test are marked with an ×.

3.6. Model verification

We verified that the model is capable of predicting new compounds outside of those found in the existing YSI database by performing a leave-one-out cross-validation. In this procedure, each molecule in sequence is withheld from the training data, the model is refit and used to predict the mean YSI of the withheld molecule. The results of this cross-validation are shown in Figure 10. Molecules that contain carbon fragments not found elsewhere in the database, and that therefore fail the applicability domain test, are marked with a red 'x'. The median absolute deviation for compounds previously in the Low scale database is 2.35, and 28.6 for compounds previously in the High scale database. The error observed for Low scale compounds is similar to the error observed in a previously developed QSAR for Low scale species only, indicating that broadening the scope of the regression has not deteriorated its predictive capability [29].

As an additional means of validating the model's predictive capability, we compare the YSIs for several aliphatic and aromatic hydrocarbons against their normalized smoke points (NSP) reported by Li and Sunderland [19]. These authors have published the most extensive dataset of smoke points of hydrocarbons in diffusion flames by collating and normalizing smoke point measurements from 12 past studies, yielding NSPs for 111 hydrocarbons. They also provided 95% confidence intervals for the 55 of these hydrocarbons that had been considered by more than one study. YSIs for 66 of these 111 hydrocarbons are in the unified database, while YSIs for the remaining 45 hydrocarbons were predicted using the group increment model in this work. In Figure 11 we plot the NSP against the measured or regressed YSI values from this work. Since lower smoke points denote sootier compounds, while the opposite is true for YSI, there is a rough agreement between sooting tendencies as expressed by either YSIs or smoke points for a broad set of hydrocarbons. YSIs for hydrocarbons predicted using the group increment model appear to fall in line with those measured experimentally. This successful prediction for the 45 hydrocarbons with missing experimental YSI

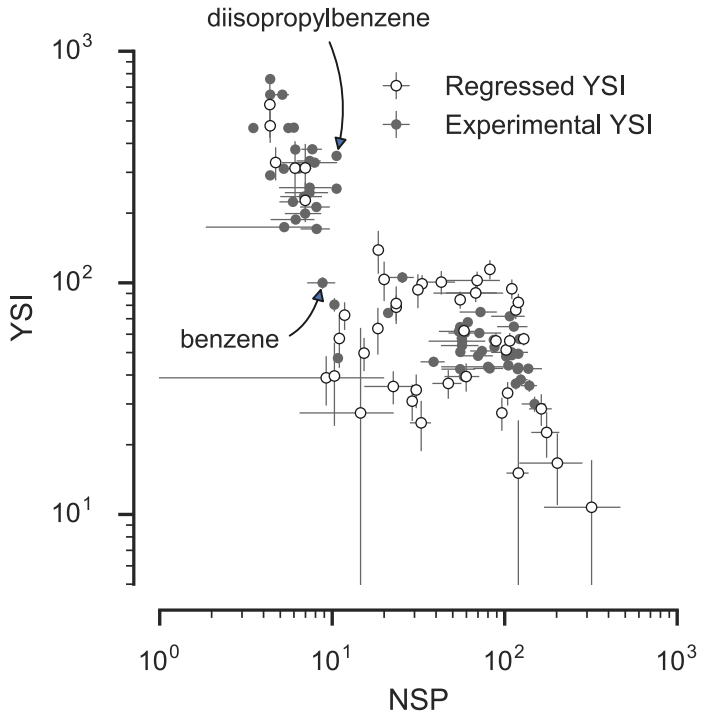

Figure 11: Comparison between unified YSI from this work and normalized smoke points (NSP, in mm) from Li and Sunderland [19], for 111 hydrocarbons, presented on logarithmic axes. YSIs for 45 of the 111 hydrocarbons determined using the group increment model (Regressed YSI) are marked with ∘, while those measured experimentally are marked with •. Error bars extend to 95% confidence intervals in both directions (where available).

data (and which therefore were not a part of the training set for the model) serves as a qualitative demonstration of the accuracy of the group increment model. We refrain from rigorously quantifying the performance of the model on this external validation set due to several inherent limitations of the smoke point database. The sources of the smoke point data range from several different authors, decades, and experimental equipment, and therefore results in a wide range of uncertainty. As an example, in Figure 11, we highlight diisopropylbenzene and benzene – while diisopropylbenzene has over three times the YSI of benzene as would be expected, it also has a higher smoke point (indicating a lower sooting tendency, which would not be expected). An additional limitation of the NSP database is that it does not contain any oxygenates.

4. Discussion

4.1. Insights into sooting chemistry

With the group increment model completed, we next look into the insights the model gives into how molecular structure influences soot formation. We look first at the estimated fragment contributions for the carbon types which appear most frequently in the database (Figure 12). Fragments containing oxygen functional groups (Figure 12A) tend to have a lower sooting contribution than those containing only carbon and hydrogen. Fragments with aromatic carbon centers (fragments 15-17) have a significantly higher sooting tendency than other fragment types.

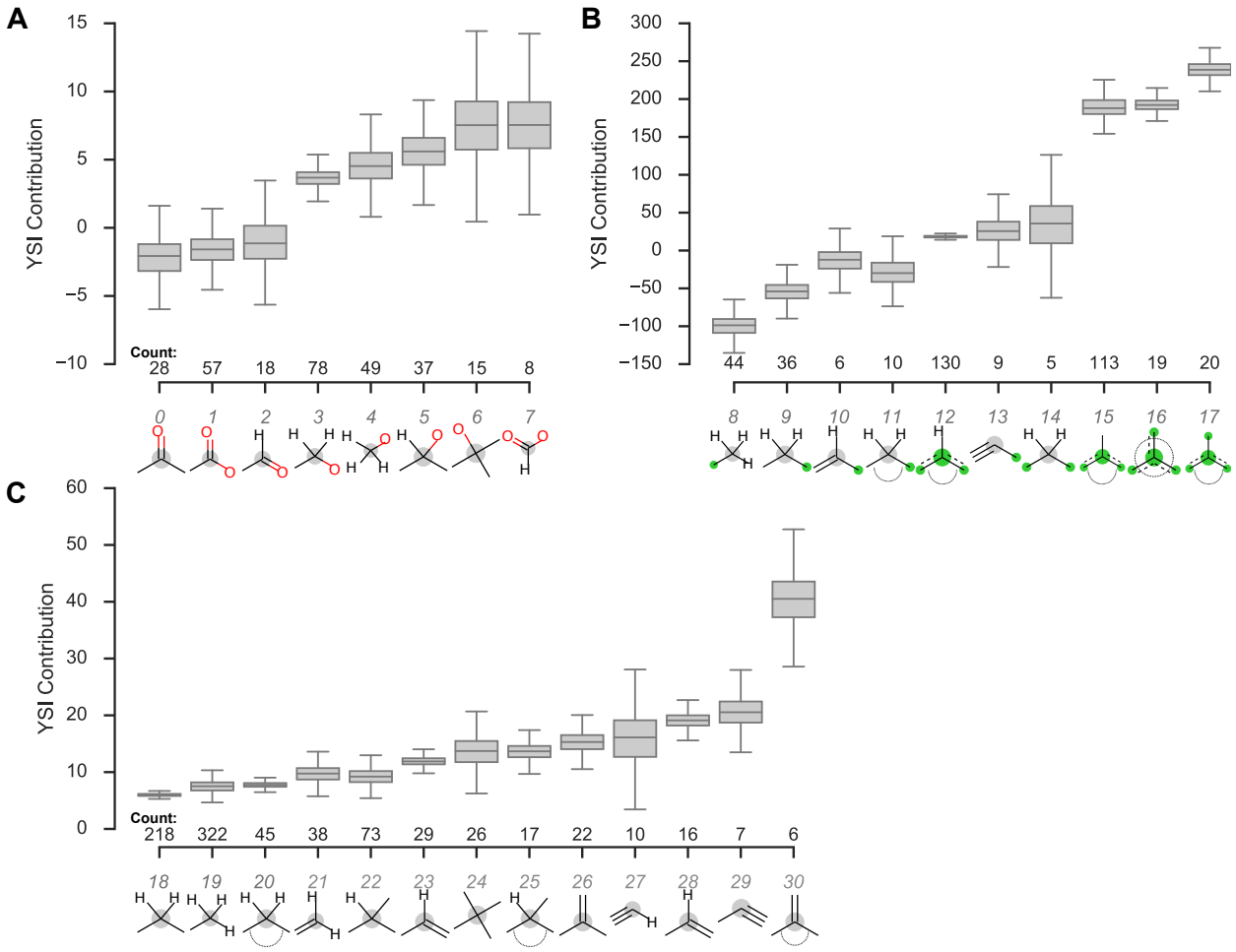

Figure 12: Mean YSI contributions for oxygenated **(A)**, aromatic **(B)**, or aliphatic **(C)** carbon types. Counts represent the number of molecules the particular fragment appears in. Carbons highlighted in green represent aromatic atoms, while grey circles indicate the central carbon is present in a ring. *Note:* Fragment numbering does not correspond to those in Figure 8.

Many of the fragments in Figure 12 appear together in molecules; for instance fragment 8 does not appear without being attached to fragment 15. To make interpretation of the results more straightforward, we have applied the group increment model to a set of example compounds, as shown in Figure 13. These compounds indicate that oxygen functional groups increase sooting tendency in the order – carboxylic acids, aldehydes, ketones, ethers, 1-alcohols, 2-alcohols, and 3-alcohols. All of the oxygenated functional groups are predicted to lower the YSI relative to the non-oxygenated counterpart. Even the tertiary alcohol (2-methylpentan-2-ol, species #8), while having a higher YSI than the straight chain compound n-hexane (#7), has a lower YSI than the branched non-oxygenated molecule 2-methylpentane (#9). The highest sooting tendencies among the six-carbon compounds come from highly unsaturated molecules, including alkynes, branched alkenes, and 5-carbon rings. For aromatic molecules a similar trend is observed, with unsaturated functional groups quickly increasing the YSI of the base benzene molecule. Interestingly, biphenyl

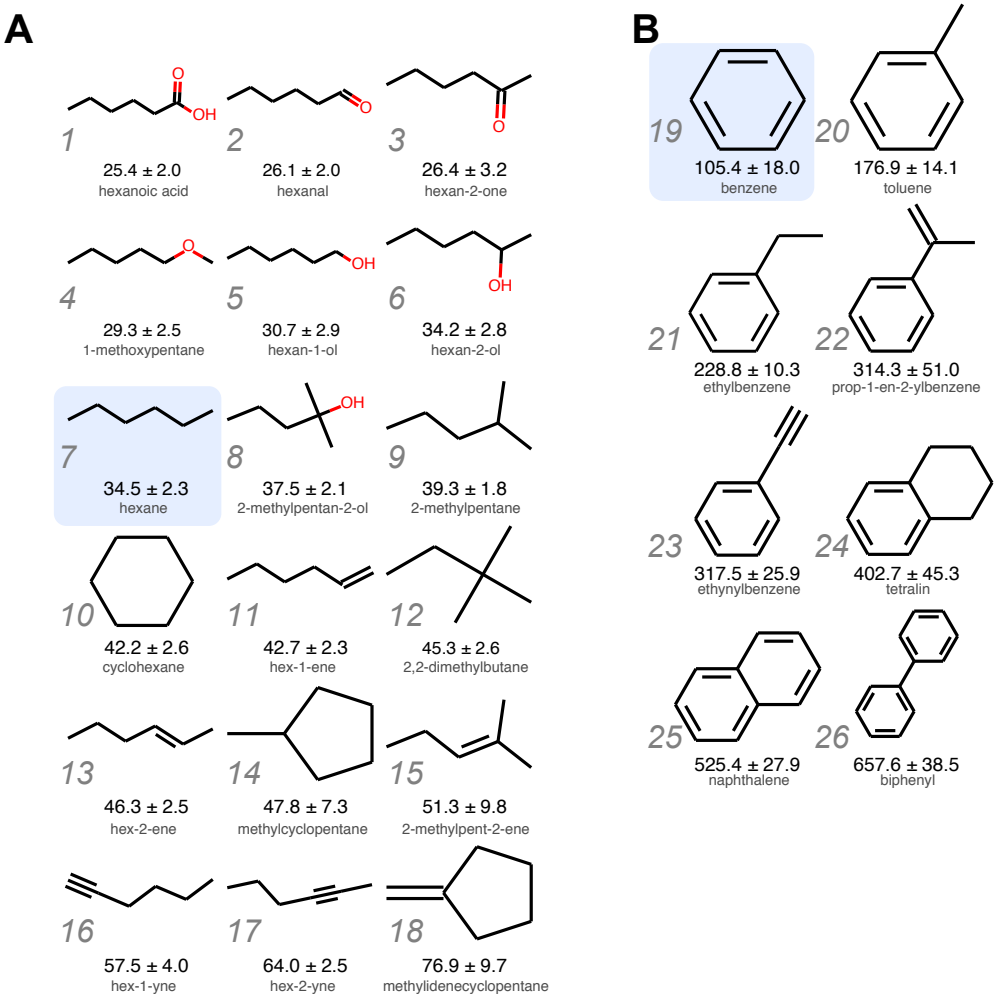

Figure 13: Demonstration of the model regression with whole-molecule examples consisting of different arrangements of **(A)** 6-carbon compounds, and **(B)** substituted benzene compounds. All YSIs reported are predicted with the group-increment model. Molecules are sorted in order of increasing YSI, and *n*-hexane and benzene (index compounds used in defining the Unified YSI scale) are highlighted.

(#26) has a higher YSI than naphthalene (#25), perhaps because of the ease with which the former can convert to the three-ring compound phenanthrene.

In order to evaluate the types of carbon chemistry that might be missed by the group increment model, we applied the model to the gdb-13 database [60], which enumerates 977 million organic molecules containing up to 13 atoms of C, N, O, S and Cl that follow simple chemical stability and synthetic feasibility rules. Of the $1,910,919$ compounds with formula $C_8H_xO_y$, $465,089$ pass the validity domain threshold defined in Section 2.2.3. Predicted YSIs range from 13.1 to 523, as shown in Figure 14A. While oxygenated molecules typically have a lower YSI than non-oxygenated ones, significant overlap exists between the oxygenated and aliphatic molecules. Fragment types were compared between the C-8 gdb-13 database and the unified YSI database (Figure 14B). Of the 130 fragment types found in the C-8 database, 53 are found in the unified YSI database. These are close

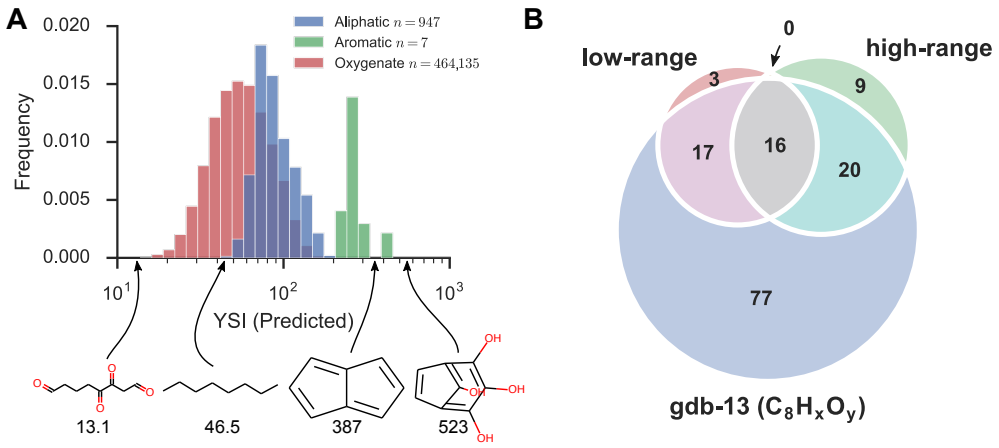

Figure 14: Prediction of the sooting tendency of $C_8H_xO_y$ compounds in the gdb-13 database. **(A)** Histograms of the resulting YSI predictions. Chemical structures shown on the horizontal axis represent the highest and lowest sooting molecules of the oxygenate and aliphatic categories. **(B)** Venn diagram of fragment types observed in the gdb-13 database and their coverage by the unified YSI database. "low-range" and "high-range" refers to fragments from molecules previously found in the respective Low and High scales (see Table 1).

to evenly distributed between fragments originating from the Low scale compounds and High scale compounds, which indicates the importance of the YSI unification in developing a generalizable soot formation model. Of the 77 fragments not covered by the unified YSI database, the most frequently encountered typically involve oxygen heterocycles. Future experimental work could therefore be targeted towards exploring species with rings that contain oxygen atoms.

4.2. Model prediction outliers

While the group increment model succeeds in accurately capturing the YSI of most compounds in the unified YSI database, it is useful to inspect where the model fails to predict sooting tendency correctly. Some of these cases may offer clues as to how to improve the model in future iterations, while others reveal interesting sooting mechanisms that are missed by a fragment-based description of a molecule's structure. In Table 4, we summarize the molecules with the highest relative YSI prediction error during cross-validation. Of the two non-aromatics in the list, 2-pentyne is likely poorly predicted due to the relative scarcity of the C≡C triple bond in the training data, with most alkynes appearing in the High scale YSI database. The other, 1-methyl-1-cyclohexene, belongs to the difficult-to-predict family of cyclohexenes that are discussed in more detail in Section 4.4. The most poorly predicted aromatic species, styrene and phenylacetylene, are unique because they are virtually the only aromatics that cannot directly react to resonantly stabilized radicals. Their sooting tendencies are therefore greatly over-predicted due to the influence of aromatics with similar structures that readily recombine to form polycyclic aromatic hydrocarbons, *i.e.*, 2-propynyl-benzene.

4.3. Analysis of degenerate compound groups

In general additivity schemes cannot account for non-nearest-neighbor interactions (NNIs) without ad hoc corrections [42], but can be helpful in identifying cases where such NNIs are important. One way of assessing the importance of NNIs in sooting tendency is to compare the

Table 4: Ten most poorly predicted molecules during leave-one-out cross-validation. Sorted by p-value of the difference in means (Z-test).

Species	CAS	Type	YSI (meas)	YSI (pred)	% Err
styrene	100-42-5	aromatic	174.0 ± 7.7	314.8 ± 6.9	-81%
phenylacetylene	536-74-3	aromatic	216.3 ± 9.3	356.6 ± 7.6	-65%
2-pentyne	627-21-4	alkynes	54.7 ± 2.0	178.0 ± 11.2	-225%
(2-propynyl)-benzene	10147-11-2	aromatic	443.7 ± 17.1	254.3 ± 5.0	+43%
naphthalene	91-20-3	aromatic	466.1 ± 8.4	568.5 ± 8.3	-22%
(2-methyl-1-propenyl)-benzene	768-49-0	aromatic	436.9 ± 16.6	283.4 ± 6.6	+35%
1-methyl-1-cyclohexene	591-49-1	cyclic alkenes	62.0 ± 3.0	108.4 ± 4.5	-75%
(1-butynyl)-benzene	622-76-4	aromatic	480.8 ± 18.4	324.2 ± 7.3	+33%
1,4-diethylbenzene	105-05-5	aromatic	270.7 ± 11.0	367.2 ± 6.8	-36%
azulene	275-51-4	aromatic	492.3 ± 19.0	643.1 ± 8.8	-31%

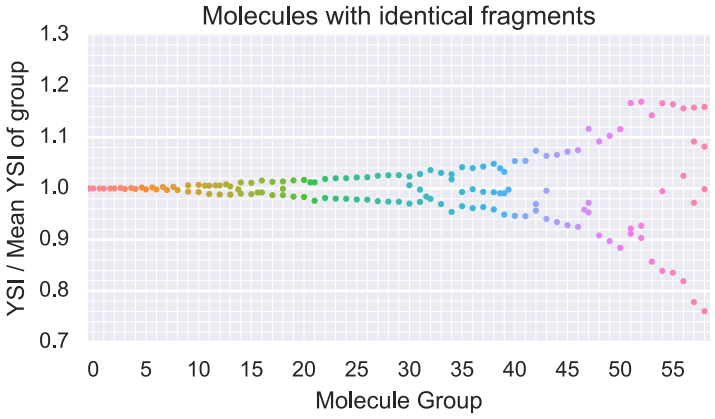

Figure 15: Distribution of measured YSIs for 59 groups of molecules with identical fragment decompositions. YSIs are scaled to the mean of the group, and groups are sorted by variance, defined as the difference between the maximum and minimum scaled YSI within each group, from lowest to highest.

measured YSIs for groups of hydrocarbons that are degenerate within the model; *i.e.*, sets of 2 or more hydrocarbons that are chemically distinct but contain the same collection of chemical fragments. This analysis also provides interesting insights into soot formation pathways.

Figure 15 shows the variation in the measured YSI for each of the 59 groups of hydrocarbons (numbered 0 through 58) in the experimental database that are degenerate within the predictive model. The groups are indexed by number and the molecules in each group are listed in Table S4 in Supplemental Information. The figure shows that the groups clearly fall into two categories: the first category – Groups 0 to roughly Group 30 – have variations of a few percent that can be attributed to experimental error, while the second category – roughly Group 31 to Group 55 – have much larger variations of ±10 to 25% that indicate true differences in the sooting tendency, presumably due to NNIs.

Figure 16 shows the 8 groups that have the smallest percentage variation between the maximum

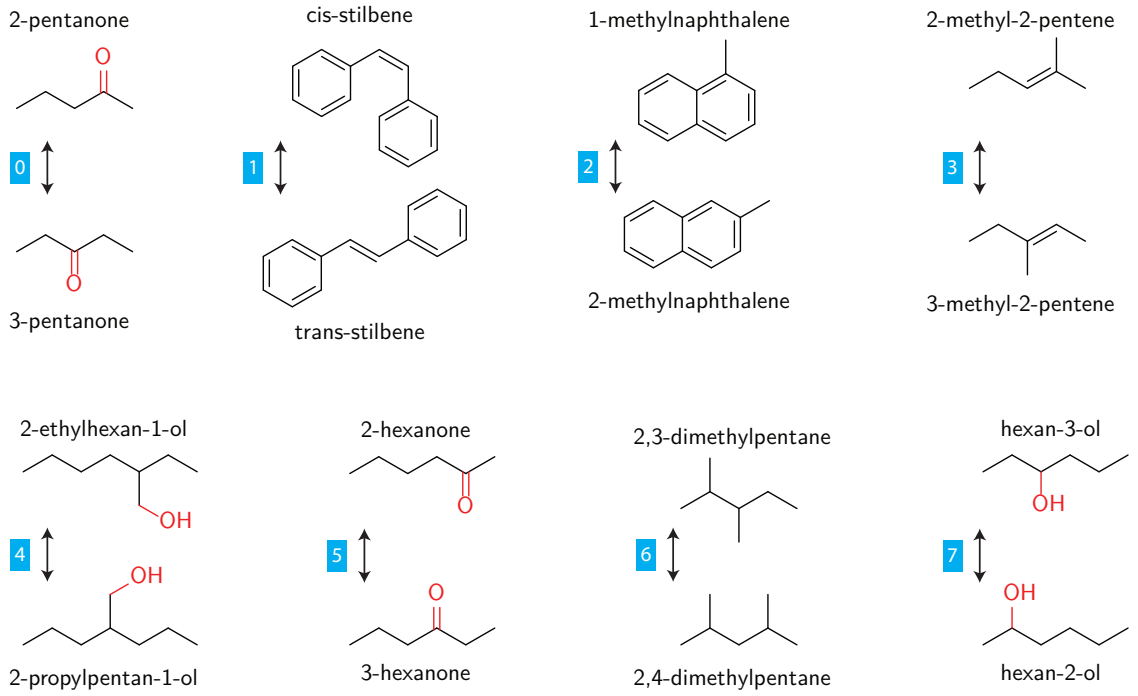

Figure 16: The 8 degenerate molecule groups that have the smallest percentage variation between the maximum and minimum measured YSIs.

and minimum measured YSI in the group. Group 1 is the stereoisomers *cis*- and *trans*-stilbene. Small variations were also observed for *cis*- and *trans*-2-hexene and for the chiral isomers (enantiomers) of 2-butanol; these results collectively suggest that stereoisomerism does not affect sooting tendency. Group 2 is the positional isomers 1- and 2-methylnaphthalene. Similar results are observed for positional isomers of aromatics with multiple methyl side-chains, including dimethylbenzenes, trimethylbenzenes, tetramethylbenzenes, and dimethylnaphthalenes. The observation that these isomers have similar sooting tendencies is a significant simplification since aromatics with one or more methyl side-chains are ubiquitous in practical fuels [61, 62]. Groups 0 and 3 to 7 are positional isomers of alkanes, alkenes, alcohols, and ketones where the functional group is at different positions on a carbon chain. Similar results are observed for many other groups, including heptanones, octanones, nonanones, and decanones; pentanols, heptanols, and octanols; and methylpentanes, methylhexanes, methylheptanes, and dimethylhexanes. Again, these observations are significant because branched alkanes such as those in Group 6 are important constituents of petroleum-derived fuels [61, 62], while ketones and alcohols are among the most promising biomass-derived blendstocks [40].

Figure 17 shows the 8 groups that have the largest percentage variation between the maximum and minimum measured YSI in the group. These are examples of positional isomers where NNIs do strongly affect sooting tendency. Groups 51, 52, and 54 are positional isomers of multiply-substituted aromatics where at least one of the groups is larger than methyl. In all of these cases the 1,2 isomer has a much greater sooting tendency than the other isomers. The likely explanation is that when the side-chains are attached to adjacent carbon atoms in the aromatic ring, they can link together to form five-membered rings and provide a short circuit to naphthalene. Consistent with

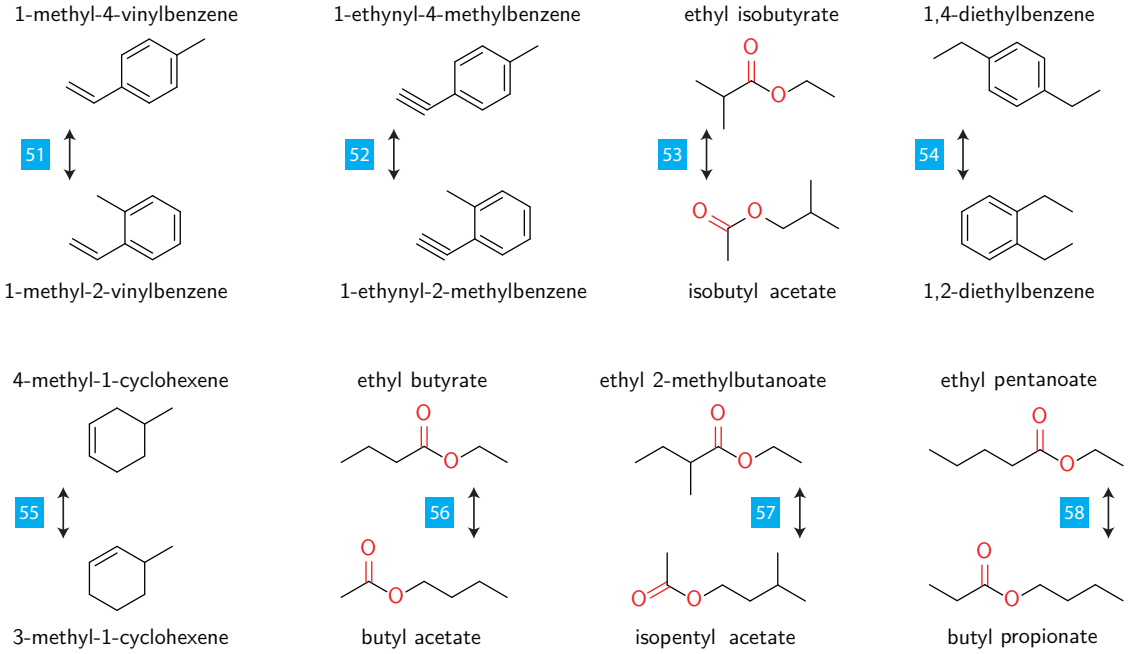

Figure 17: The 8 degenerate molecule groups that have the largest percentage variation between the maximum and minimum measured YSIs. The compound with maximum YSI is shown in the bottom in each case. For groups with more than two members, only the compounds with minimum and maximum YSI are shown.

this hypothesis, large differences are not observed in cases where both side-chains are methyl groups and do not collectively possess the 3 C-atoms necessary to form a five-membered ring. Furthermore, large differences are not observed for ethylmethylbenzenes, where the initial reaction is likely fission of the ethyl group [63], which only leaves 2 C-atoms in the side-chains. On the other hand, large differences are observed for diethylbenzenes, where fission of an ethyl group still leaves 3 C-atoms in the side-chains, and in the methylvinylbenzenes and methylethynylbenzenes, where the initial attack is likely on the methyl group instead of the C-2 side-chain. This type of NNI is likely to be of limited significance in practical cases: while diethylbenzenes are significant in some diesel fuels, most of the aromatics in practical fuels are dominated by methyl side-chains [61, 62].

Groups 53, and 56 to 58 are esters where different proportions of the overall C-atoms are located on either side of the ester functional group. The NNI in this case is likely to be 6-center elimination reactions that begin with a bond forming between the carbonyl O-atom and an H-atom on the C-atom β to the ether O-atom; these reactions are slow for the ethyl esters where the H-atom is attached to a primary C-atom, and fast for the propyl, isopropyl, butyl, isobutyl, pentyl, and isopentyl esters where it is attached to a secondary C-atom. These reactions accelerate soot formation because their products are large alkenes and carboxylic acids that use both O-atoms inefficiently to tie down a single C-atom [12, 64]. This mechanism has practical consequences: while traditional biodiesel fuels are mostly methyl esters, larger esters including isopropyl acetate, butyl acetate, isobutyl acetate, and isopentyl acetate have been identified as promising blendstocks due to their higher octane numbers [40].

Overall, the results for groups of degenerate compounds show that the model is applicable to most hydrocarbons contained in practical fuels without ad hoc corrections for NNIs. The most

significant exceptions, where NNIs are important, are esters with groups larger than ethyl on the ether side, and aromatics with multiple side-chains larger than a methyl group.

4.4. Positional isomers of methylcyclohexene

Group 55 of the degenerate molecule sets includes the three positional isomers of methylcyclohexene – 1-methyl-1-cyclohexene (1MCX), 3-methyl-1-cyclohexene (3MCX), and 4-methyl-1-cyclohexene (4MCX). These are a particularly interesting set of molecules since cyclohexene was also a prominent outlier in a previous Low scale-only QSAR model, primarily due to a unique retro-Diels-Alder mechanism [29]. In addition to high variance in measured YSI between molecules with identical fragments, the YSI of molecules in this group are also predicted particularly poorly during leave-one-out cross-validation (Table 5). Within the predictions, the relatively high predicted YSI for 1MCX compared to the other methylcyclohexenes is primarily due to the presence of the carbon fragment with both a double bond and a methyl group. This particular fragment is also found in other high-sooting compounds such as 2-methylindene (Unified YSI = 500.1) and 1-methyl-1,4-cyclohexadiene (Unified YSI = 175.6).

Table 5: Comparisons between experimental measured YSI and predicted YSI (via cross-validation) of cyclohexene and 3 methylcyclohexene isomers. YSI predictions are reported as medians and 95% HPD intervals (in brackets).

Compound	YSI (measured)	YSI (predicted)
Cyclohexene	45.6 ± 2.0	71.1 [65.6, 76.8]
1-methyl-1-cyclohexene	62.0 ± 3.0	108.4 [98.4, 118.8]
3-methyl-1-cyclohexene	85.0 ± 4.0	77.9 [72.9, 83.7]
4-methyl-1-cyclohexene	61.0 ± 3.0	80.4 [74.6, 85.7]

To investigate the differences among the 3 isomers, we performed DFT calculations to estimate the energy barriers to the retro-Diels-Alder pathway for each isomer at room temperature. As hypothesized in a previous study [29], cyclohexene's lower YSI (45.6) compared to cyclopentene and cycloheptene is due to a retro-Diels-Alder decomposition unique to cyclohexene which breaks the ring, thereby arresting the rate of growth of important polycyclic aromatic hydrocarbon (PAH) precursors such as cyclohexadiene and benzene. However, the availability of this kinetic mechanism does not sufficiently explain the differences between isomers of methylcyclohexene: energy barriers calculated using G4 for 1MCX, 3MCX, and 4MCX are 64.3, 64.2 and 65.4 kcal/mol, respectively. As these barriers are all very similar, we propose other possible mechanisms to produce dehydrogenated products more closely related to PAH precursors (shown in Figure 18).

For each methylcyclohexene isomer we computed C-H bond-dissociation energies (BDE) using the G4 method in order to make methylcyclohexadiene isomers through conjugated radical intermediates. In the first step, from methylcyclohexene isomers to conjugated radical intermediates, the C-H BDE of 3MCX was the lowest (80.0 kcal/mol) when compared with that of 1MCX and 4MCX. We also computed C-C BDE of the methyl group in each methylcyclohexene, as shown in top row of Figure 18, to compare their relative bond strengths. The C-C BDE of 3MCX is lower (71.9 kcal/mol) than that for 1MCX and 4MCX (98.6 and 86.8 kcal/mol, respectively). As with the C-H BDE, 3MCX can produce a resonance-stabilized radical intermediate that lowers the C-C BDE, while radical intermediates of the other two isomers do not show such resonance stabilization. These BDE differences, and therefore the relative ease of formation of resonance-stabilized radical

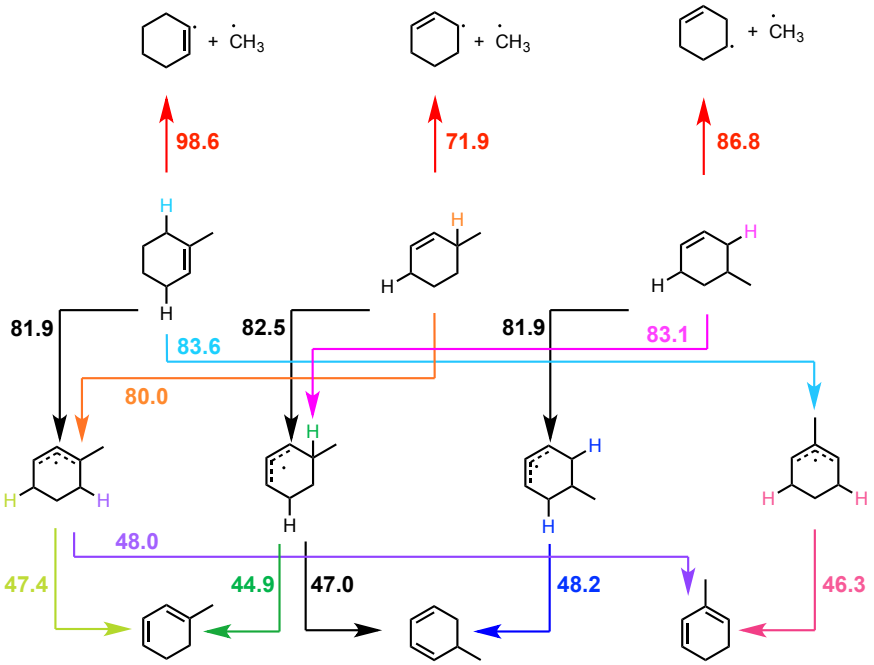

Figure 18: All C-H BDE energies (in kcal/mol using G4) from 3 methylcyclohexene to 3 methylcyclohexadiene isomers (from left to right, 1-methyl-1-cyclohexene, 3-methyl-1-cyclohexene, and 4-methyl-1-cyclohexene). All color codes of BDEs are matched with relevant H for C-H BDE (except red colors for C-C BDE).

intermediates for 3MCX, would lead to a greater accumulation of PAH precursors (cyclohexadiene or methylcyclohexadiene and then eventually benzene or methylbenzene) for that isomer, and manifest as a greater sooting tendency compared to 1MCX and 4MCX. While a full kinetic model of this reaction system is beyond the scope of this study, these results do highlight that methylcyclohexene combustion involves a competition between preservation of the cyclohexene ring (which leads to the formation of PAH precursors) and breaking of the cyclohexene ring through the retro-Diels-Alder decomposition. Higher sooting tendencies for 3-methyl-1-cyclohexene are likely due to a lower energy barrier for the initial hydrogen extraction and/or C−C bond breaking of methyl group, both of which would lead to faster formation of PAH precursors compared to the other two methylcyclohexene isomers. While the group-contribution model fails to pick up these differences between the isomers, it is nonetheless useful in highlighting the presence of the underlying NNIs in these types of compounds.

5. Conclusions

In this work we merged two different literature sooting tendency databases together into a single unified YSI database for pure compounds. The unified database represents the largest extant internally consistent database of sooting behavior for pure compounds: it includes ≥ 400 compounds; covers aliphatic, aromatic, cycloalkanes, and oxygenated hydrocarbons; and presents the sooting tendencies of oxygenates and aromatics on the same numeric scale. Unification of the databases was made possible by using the color-ratio pyrometry diagnostic, which allows accurate

measurements of f_v in flames doped with low-sooting oxygenated hydrocarbons and high-sooting aromatic hydrocarbons under identical experimental conditions.

A new scaling for the unified database was defined by setting the YSI of *n*-hexane and benzene to 30 and 100 respectively. With this definition of the YSI scale, all studied compounds had numeric YSI values greater than 0 (no negative YSI values), and YSI values for compounds tended towards 0 with the number of carbon atoms in the compound. YSI values spanned three orders of magnitude across the entire database, with non-aromatics YSI ranging from 6.6 (methanol) to 161 (1-*tert*-butyl-1-cyclohexene), and aromatics YSI ranging from 100 (benzene) to 1400 (1,2-diphenylbenzene).

We also developed a modified Benson group-increment model to predict the YSI of a compound from its molecular structure. The model predicts the YSI of a compound as a linear sum of contributions from each of its component carbon fragments. Contributions of component carbon fragments were determined through a Bayesian linear regression against the YSI values of compounds in the unified database developed in this work. The model was verified both internally through leave-one-out cross-validation and externally through comparisons to normalized smoke points of hydrocarbons in the literature.

Inspection of the different contributions of different carbon fragments provided insights into sooting chemistry. For example, fragments containing oxygen functional groups tended to have lower sooting contributions than fragments containing only carbon and hydrogen, while aromatic carbon centers had significantly greater sooting tendency contributions than other fragment types. To assess the extent of coverage of different functional groups in the model's applicability domain, we applied the model to a subset of the gdb-13 database. The most frequently encountered chemical fragment types in the gdb-13 database not currently covered by the model typically involve oxygen heterocycles, which could be a target for future experimental work addressing underrepresented fragment groups.

Prediction outliers (as characterized by relative YSI prediction error during cross-validation) are explained either by relative scarcity of training data (*i.e.*, alkynes), or by a presence of unique soot formation pathways (*i.e.*, styrene and phenylacetylene). The relative importance for sooting behavior of non-nearest-neighbor interactions (NNIs) between carbon atoms in a compound was characterized through model degeneracies (distinct sets of compounds with identical fragment decompositions) Multiply-substituted aromatics, esters, and methylcyclohexenes were found to have the most significant NNIs. Plausible chemical reaction pathways have been proposed to explain observed trends in sooting behavior for these compound families. Overall, the model is applicable to most hydrocarbons contained in practical fuels and is expected to enable the rational design of low-sooting fuel blends from a wide range of feedstocks and chemical functionalities.

6. Acknowledgements

We appreciate assistance from Thomas A. Kwan in conducting these experiments. This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Bioenergy Technologies Office (BETO) and Vehicle Technologies Office (VTO) Program Award Number DE-EE0007983. This material is based upon work supported by the National Science Foundation (NSF) under Grant No. CBET 1604983.

Disclaimer

References

[1] R. A. Hunt, Relation of Smoke Point to Molecular Structure, Industrial & Engineering Chemistry 45 (3) (1953) 602–606. doi:10.1021/ie50519a039.
URL http://pubs.acs.org/doi/abs/10.1021/ie50519a039

[2] C. S. McEnally, L. D. Pfefferle, Improved sooting tendency measurements for aromatic hydrocarbons and their implications for naphthalene formation pathways, Combustion and Flame 148 (4) (2007) 210–222. doi:10.1016/j.combustflame.2006.11.003.

[3] S. P. Crossley, W. E. Alvarez, D. E. Resasco, Novel Micropyrolyis Index (MPI) to Estimate the Sooting Tendency of Fuels, Energy & Fuels 22 (4) (2008) 2455–2464. doi:10.1021/ef800058y.
URL http://pubs.acs.org/doi/abs/10.1021/ef800058y

[4] R. L. Schalla, T. P. Clark, G. E. McDonald, Formation and combustion of smoke in laminar flames, Tech. rep., National Advisory Committee for Aeronautics, Cleveland, Ohio (1954).
URL http://hdl.handle.net/2060/19930092206

[5] H. Calcote, D. Manos, Effect of molecular structure on incipient soot formation, Combustion and Flame 49 (1-3) (1983) 289–304. doi:10.1016/0010-2180(83)90172-4.

[6] A. Gomez, G. Sidebotham, I. Glassman, Sooting behavior in temperature-controlled laminar diffusion flames, Combustion and Flame 58 (1) (1984) 45–57. doi:10.1016/0010-2180(84)90077-4.
URL http://linkinghub.elsevier.com/retrieve/pii/0010218084900774

[7] D. Olson, J. Pickens, R. Gill, The effects of molecular structure on soot formation II. Diffusion flames, Combustion and Flame 62 (1) (1985) 43–60. doi:10.1016/0010-2180(85)90092-6.

[8] Ö. L. Gülder, Influence of hydrocarbon fuel structural constitution and flame temperature on soot formation in laminar diffusion flames, Combustion and Flame 78 (2) (1989) 179–194. doi:10.1016/0010-2180(89)90124-7.

[9] N. Ladommatos, P. Rubenstein, P. Bennett, Some effects of molecular structure of single hydrocarbons on sooting tendency, Fuel 75 (2) (1996) 114–124. doi:10.1016/0016-2361(94)00251-7.
URL http://linkinghub.elsevier.com/retrieve/pii/0016236194002517

[10] Y. Yang, A. L. Boehman, R. J. Santoro, A study of jet fuel sooting tendency using the threshold sooting index (TSI) model, Combustion and Flame 149 (1-2) (2007) 191–205. doi:10.1016/j.combustflame.2006.11.007.

[11] C. S. McEnally, L. D. Pfefferle, Sooting tendencies of nonvolatile aromatic hydrocarbons, Proceedings of the Combustion Institute 32 (1) (2009) 673–679. doi:10.1016/j.proci.2008.06.197.

[12] C. S. McEnally, L. D. Pfefferle, Sooting tendencies of oxygenated hydrocarbons in laboratory-scale flames., Environmental Science and Technology 45 (6) (2011) 2498–503. doi:10.1021/es103733q.

[13] P. Pepiot-Desjardins, H. Pitsch, R. Malhotra, S. R. Kirby, A. Boehman, Structural group analysis for soot reduction tendency of oxygenated fuels, Combustion and Flame 154 (1-2) (2008) 191–205. doi:10.1016/j.combustflame.2008.03.017.

[14] K. M. Allan, J. R. Kaminski, J. C. Bertrand, J. Head, P. B. Sunderland, Laminar Smoke Points of Wax Candles, Combustion Science and Technology 181 (5) (2009) 800–811. doi:10.1080/00102200902935512.
URL http://www.tandfonline.com/doi/abs/10.1080/00102200902935512

[15] A. Mensch, R. J. Santoro, T. a. Litzinger, S.-Y. Lee, Sooting characteristics of surrogates for jet fuels, Combustion and Flame 157 (6) (2010) 1097–1105. doi:10.1016/j.combustflame.2010.02.008.

[16] E. J. Barrientos, A. L. Boehman, Examination of the Sooting Tendency of Three-Ring Aromatic Hydrocarbons and Their Saturated Counterparts, Energy & Fuels 24 (6) (2010) 3479–3487. doi:10.1021/ef100181s.

[17] E. J. Barrientos, M. Lapuerta, A. L. Boehman, Group additivity in soot formation for the example of C-5 oxygenated hydrocarbon fuels, Combustion and Flame 160 (8) (2013) 1484–1498. doi:10.1016/j.combustflame.2013.02.024.
URL http://linkinghub.elsevier.com/retrieve/pii/S0010218013000849

[18] E. J. Barrientos, J. E. Anderson, M. M. Maricq, A. L. Boehman, Particulate matter indices using fuel smoke point for vehicle emissions with gasoline, ethanol blends, and butanol blends, Combustion and Flame 167 (2016) 308–319. doi:10.1016/j.combustflame.2016.01.034.
URL http://www.sciencedirect.com/science/article/pii/S0010218016000493

[19] L. Li, P. B. Sunderland, An improved method of smoke point normalization, Combustion Science and Technology 184 (6) (2012) 829–841. doi:10.1080/00102202.2012.670333.

[20] M. Kashif, P. Guibert, J. Bonnety, G. Legros, Sooting tendencies of primary reference fuels in atmospheric laminar diffusion flames burning into vitiated air, Combustion and Flame 161 (6) (2014) 1575–1586. doi:10.1016/j.combustflame.2013.12.009.
URL http://linkinghub.elsevier.com/retrieve/pii/S0010218013004574

[21] M. Kashif, J. Bonnety, A. Matynia, P. Da Costa, G. Legros, Sooting propensities of some gasoline surrogate fuels: Combined effects of fuel blending and air vitiation, Combustion and Flame 162 (5) (2015) 1840–1847. doi:10.1016/j.combustflame.2014.12.005.
URL http://linkinghub.elsevier.com/retrieve/pii/S001021801400399X

[22] D. D. Das, C. S. McEnally, L. D. Pfefferle, Sooting tendencies of unsaturated esters in nonpremixed flames, Combustion and Flame 162 (4) (2015) 1489–1497. doi:10.1016/j.combustflame.2014.11.012.

[23] D. D. Das, W. J. Cannella, C. S. McEnally, C. J. Mueller, L. D. Pfefferle, Two-dimensional soot volume fraction measurements in flames doped with large hydrocarbons, Proceedings of the Combustion Institute 36 (1) (2017) 871–879. doi:10.1016/j.proci.2016.06.047.
URL http://linkinghub.elsevier.com/retrieve/pii/S1540748916301055

[24] D. D. Das, C. S. C. McEnally, T. T. A. Kwan, J. B. J. J. B. Zimmerman, W. W. J. Cannella, C. C. J. Mueller, L. D. L. L. D. Pfefferle, Sooting tendencies of diesel fuels, jet fuels, and their surrogates in diffusion flames, Fuel 197 (2017) 445–458. doi:10.1016/j.fuel.2017.01.099.
URL http://dx.doi.org/10.1016/j.fuel.2017.01.099

[25] R. Lemaire, D. Lapalme, P. Seers, Analysis of the sooting propensity of C-4 and C-5 oxygenates: Comparison of sooting indexes issued from laser-based experiments and group additivity approaches, Combustion and Flame 162 (9) (2015) 3140–3155. doi:10.1016/j.combustflame.2015.03.018.
URL http://linkinghub.elsevier.com/retrieve/pii/S0010218015001030

[26] G. D. J. Guerrero Peña, M. M. Alrefaai, S. Y. Yang, A. Raj, J. L. Brito, S. Stephen, T. Anjana, V. Pillai, A. Al Shoaibi, S. H. Chung, Effects of methyl group on aromatic hydrocarbons on the nanostructures and oxidative reactivity of combustion-generated soot, Combustion and Flame 172 (2016) 1–12. doi:10.1016/j.combustflame.2016.06.026.
URL http://dx.doi.org/10.1016/j.combustflame.2016.06.026

[27] M. P. Hanson, D. H. Rouvray, Novel applications of topological indices. 2. Prediction of the threshold soot index for hydrocarbon fuels, The Journal of Physical Chemistry 91 (11) (1987) 2981–2985. doi:10.1021/j100295a067.

[28] S. Yan, E. G. Eddings, A. B. Palotas, R. J. Pugmire, A. F. Sarofim, Prediction of Sooting Tendency for Hydrocarbon Liquids in Diffusion Flames, Energy & Fuels 19 (6) (2005) 2408–2415. doi:10.1021/ef050107d.

[29] P. C. St. John, P. Kairys, D. Das, C. Mcenally, L. Pfefferle, D. Robichaud, M. Nimlos, B. Zigler, R. McCormick, T. Foust, Y. Bomble, S. Kim, A quantitative model for the prediction of sooting tendency from molecular structure, Energy & Fuels (2017) (In Press)doi:10.1021/acs.energyfuels.7b00616.
URL http://dx.doi.org/10.1021/acs.energyfuels.7b00616

[30] L. S. Whitmore, R. W. Davis, R. L. McCormick, J. M. Gladden, B. A. Simmons, A. George, C. M. Hudson, Biocompoundml: A general biofuel property screening tool for biological molecules using random forest classifiers, Energy & Fuels 30 (10) (2016) 8410–8418. doi:10.1021/acs.energyfuels.6b01952.
URL http://dx.doi.org/10.1021/acs.energyfuels.6b01952

[31] C. S. McEnally, L. D. Pfefferle, B. Atakan, K. Kohse-Höinghaus, Studies of aromatic hydrocarbon formation mechanisms in flames: Progress towards closing the fuel gap, Progress in Energy and Combustion Science 32 (3) (2006) 247–294. doi:10.1016/j.pecs.2005.11.003.
URL http://linkinghub.elsevier.com/retrieve/pii/S0360128505000602

[32] ASTM, Standard Test Method for Smoke Point of Kerosine and Aviation Turbine Fuel, ASTM Standard D1322-15 97 (Reapproved) (2010) 1–8. doi:10.1520/D1322-15E01.

[33] ASTM, Standard Specification for Aviation Turbine Fuels, ASTM Standard D655-16cdoi:10.1520/D1655-16C.
URL http://www.astm.org/cgi-bin/resolver.cgi?D1655

[34] K. Aikawa, T. Sakurai, J. J. Jetter, Development of a Predictive Model for Gasoline Vehicle Particulate Matter Emissions, SAE International Journal of Fuels and Lubricants 3 (2) (2010) 610–622. doi:10.4271/2010-01-2115.

[35] C. J. Mueller, W. J. Cannella, T. J. Bruno, B. Bunting, H. D. Dettman, J. A. Franz, M. L. Huber, M. Natarajan, W. J. Pitz, M. A. Ratcliff, K. Wright, Methodology for formulating diesel surrogate fuels with accurate compositional, ignition-quality, and volatility characteristics, Energy & Fuels 26 (6) (2012) 3284–3303. doi:10.1021/ef300303e.

[36] G. M. Chupka, E. Christensen, L. Fouts, T. L. Alleman, M. A. Ratcliff, R. L. McCormick, Heat of Vaporization Measurements for Ethanol Blends Up To 50 Volume Percent in Several Hydrocarbon Blendstocks and Implications for Knock in SI Engines, SAE International Journal of Fuels and Lubricants 8 (2) (2015) 2015–01–0763. doi:10.4271/2015-01-0763.
URL http://papers.sae.org/2015-01-0763/

[37] F. L. Dryer, Chemical kinetic and combustion characteristics of transportation fuels, Proceedings of the Combustion Institute 35 (1) (2015) 117–144. doi:10.1016/j.proci.2014.09.008.

URL http://linkinghub.elsevier.com/retrieve/pii/S1540748914004258
[38] US-EIA, Almost all U.S. gasoline is blended with 10% ethanol.
URL https://www.eia.gov/todayinenergy/detail.php?id=26092
[39] J. Goldemberg, The Brazilian biofuels industry, Biotechnology for Biofuels 1 (1) (2008) 6. doi:10.1186/1754-6834-1-6.
URL http://biotechnologyforbiofuels.biomedcentral.com/articles/10.1186/1754-6834-1-6
[40] R. L. McCormick, G. Fioroni, L. Fouts, E. Christensen, J. Yanowitz, E. Polikarpov, K. Albrecht, D. J. Gaspar, J. Gladden, A. George, Selection Criteria and Screening of Potential Biomass-Derived Streams as Fuel Blendstocks for Advanced Spark-Ignition Engines, SAE International Journal of Fuels and Lubricants 10 (2) (2017) 2017–01–0868. doi:10.4271/2017-01-0868.
URL http://papers.sae.org/2017-01-0868/
[41] P. B. Kuhn, B. Ma, B. C. Connelly, M. D. Smooke, M. B. Long, Soot and thin-filament pyrometry using a color digital camera, Proceedings of the Combustion Institute 33 (1) (2011) 743–750. doi:10.1016/j.proci.2010.05.006.
[42] N. Cohen, S. W. Benson, Estimation of heats of formation of organic compounds by additivity methods, Chemical Reviews 93 (7) (1993) 2419–2438. doi:10.1021/cr00023a005.
URL http://pubs.acs.org/doi/abs/10.1021/cr00023a005
[43] ASTM, Standard Test Method for Motor Octane Number of Spark-Ignition Engine Fuel, ASTM Standard D2700-16adoi:10.1520/D2700-16A.
URL http://www.astm.org/cgi-bin/resolver.cgi?D2700
[44] J. Gau, D. Das, C. McEnally, D. Giassi, N. Kempema, M. Long, Yale coflow burner information and cad drawings, Figsharedoi:10.6084/m9.figshare.5005007.v1.
URL https://figshare.com/articles/Yale_Coflow_Burner_Information_and_CAD_Drawings/5005007
[45] C. L. Yaws, Yaws' Handbook of Thermodynamic and Physical Properties of Chemical Compounds, Knovel, 2003.
URL http://app.knovel.com/hotlink/toc/id:kpYHTPPCC4/yaws-handbook-thermodynamic/yaws-handbook-thermodynamic
[46] B. Ma, M. B. Long, Absolute light calibration using S-type thermocouples, Proceedings of the Combustion Institute 34 (2) (2013) 3531–3539. doi:10.1016/j.proci.2012.05.030.
[47] H. Guo, J. A. Castillo, P. B. Sunderland, Digital camera measurements of soot temperature and soot volume fraction in axisymmetric flames, Applied Optics 52 (33) (2013) 8040. doi:10.1364/AO.52.008040.
[48] D. R. Snelling, K. A. Thomson, G. J. Smallwood, O. L. Gulder, E. J. Weckman, R. A. Fraser, Spectrally Resolved Measurement of Flame Radiation to Determine Soot Temperature and Concentration, AIAA Journal 40 (9) (2002) 1789–1795. doi:10.2514/2.1855.
[49] P. S. Kolhe, A. K. Agrawal, Abel inversion of deflectometric data: comparison of accuracy and noise propagation of existing techniques, Applied Optics 48 (20) (2009) 3894. doi:10.1364/AO.48.003894.
URL https://www.osapublishing.org/abstract.cfm?URI=ao-48-20-3894
[50] D. Witkowski, K. Kondo, G. Vishwanathan, D. Rothamer, Evaluation of the sooting properties of real fuels and their commonly used surrogates in a laminar co-flow diffusion flame, Combustion and Flame 160 (6) (2013) 1129–1141. doi:10.1016/j.combustflame.2013.01.027.
[51] S. W. Benson, J. H. Buss, Additivity rules for the estimation of molecular properties. thermodynamic properties, The Journal of Chemical Physics 29 (3) (1958) 546âĂŞ572. doi:10.1063/1.1744539.
URL http://dx.doi.org/10.1063/1.1744539
[52] K. Joback, R. Reid, Estimation of pure-component properties from group-contributions, Chemical Engineering Communications 57 (1-6) (1987) 233âĂŞ243. doi:10.1080/00986448708960487.
URL http://dx.doi.org/10.1080/00986448708960487
[53] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, Bayesian Data Analysis, Third Edition (Chapman & Hall/CRC Texts in Statistical Science), Chapman and Hall/CRC, 2013.
[54] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR & Combinatorial Science 26 (5) (2007) 694–701. doi:10.1002/qsar.200610151.
URL http://dx.doi.org/10.1002/qsar.200610151
[55] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken,

C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, . Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09 Revision D.01, gaussian Inc. Wallingford CT 2009.

[56] P. T. Boggs, J. R. Donaldson, R. h. Byrd, R. B. Schnabel, Algorithm 676 ODRPACK: software for weighted orthogonal distance regression, ACM Transactions on Mathematical Software 15 (4) (1989) 348–364. doi:10.1145/76909.76913.
URL http://portal.acm.org/citation.cfm?doid=76909.76913

[57] C. S. McEnally, D. D. Das, L. D. Pfefferle, Yield sooting index database volume 2: Sooting tendencies of a wide range of fuel compounds on a unified scale (2017). doi:10.7910/DVN/7HGFT8.
URL http://dx.doi.org/10.7910/DVN/7HGFT8

[58] D. R. White, P. Saunders, The propagation of uncertainty with calibration equations, Measurement Science and Technology 18 (7) (2007) 2157–2169. doi:10.1088/0957-0233/18/7/047.
URL http://stacks.iop.org/0957-0233/18/i=7/a=047?key=crossref.38f2414d55d0b313b9cf1579fee1af7c

[59] C. W. Meyer, W. L. Tew, ITS-90 non-uniqueness from PRT subrange inconsistencies over the range 24.56K to 273.16K, Metrologia 43 (5) (2006) 341–352. doi:10.1088/0026-1394/43/5/002.
URL http://stacks.iop.org/0026-1394/43/i=5/a=002?key=crossref.3ab7593dceaac6f586b3d2d93740754f

[60] L. C. Blum, J.-L. Reymond, 970 million druglike small molecules for virtual screening in the chemical universe database gdb-13, Journal of the American Chemical Society 131 (25) (2009) 8732âĂŞ8733. doi:10.1021/ja902302h.
URL http://dx.doi.org/10.1021/ja902302h

[61] M. Alnajjar, B. Cannella, H. Dettman, C. Fairbridge, J. Franz, T. Gallant, R. Gieleciak, D. Hager, C. Lay, S. Lewis, M. Ratcliff, S. Sluder, J. Storey, H. Yin, B. Zigler, Chemical and Physical Properties of the Fuels for Advanced Combustion Engines (FACE) Research Diesel Fuels, Tech. Rep. July (2010).
URL https://crcao.org/reports/recentstudies2010/FACE-1/FACE1ChemandPhysPropsofFACEResearchDieselFuels.pdf

[62] W. Cannella, M. Foster, G. Gunter, W. Leppard, Face Gasolines and Blends With Ethanol: Detailed Characterization of Physical and Chemical Properties, CRC Report (July).

[63] A. Ergut, S. Granata, J. Jordan, J. Carlson, J. B. Howard, H. Richter, Y. A. Levendis, PAH formation in one-dimensional premixed fuel-rich atmospheric pressure ethylbenzene and ethyl alcohol flames, Combustion and Flame 144 (4) (2006) 757–772. doi:10.1016/j.combustflame.2005.07.019.
URL http://linkinghub.elsevier.com/retrieve/pii/S0010218005002695

[64] Y. Zhang, Y. Yang, A. L. Boehman, Premixed ignition behavior of C9 fatty acid esters: A motored engine study, Combustion and Flame 156 (6) (2009) 1202–1213. doi:10.1016/j.combustflame.2009.01.024.
URL http://linkinghub.elsevier.com/retrieve/pii/S001021800900042X