Sooting tendencies: Combustion science for designing sustainable fuels with improved properties

Lisa D. Pfefferle, [a] Seonah Kim, [b] Sabari Kumar, [b] Charles S. McEnally, [a,*] Raul Perez-Soto, [b] Zhanhong Xiang, [a] Yuan Xuan [c]

[a] *Department of Chemical and Environmental Engineering, Yale University, New Haven CT USA*

[b] *Department of Chemistry, Colorado State University, Fort Collins CO USA*

[c] *Department of Mechanical Engineering, Pennsylvania State University, University Park PA USA*

[*] Corresponding author: charles.mcenally@yale.edu

Abstract

The transition from fossil fuels to sustainable fuels offers a unique opportunity to select new fuel compositions that will not only reduce net carbon dioxide emissions, but also improve combustor performance and reduce emissions of other pollutants. A particularly valuable goal is finding fuels that reduce soot emissions. These emissions cause significant global warming, especially from aviation since soot particles are the nucleation site of contrails. Furthermore, soot contributes to ambient fine particulates, which are responsible for millions of deaths worldwide each year. Fortunately, soot formation rates depend sensitively on the molecular structure of the fuel, so fuel composition provides a strong lever for reducing emissions. Sooting tendencies measured in laboratory-scale flames provide a scientific basis for selecting fuels that will maximize this benefit. Recent work has developed new techniques that expand the range of compounds that can be tested by reducing the required sample volume and increasing the dynamic range. This has many benefits, but it is particularly essential for the development of structure-property relationships using machine learning algorithms: the accuracy and predictive ability of

these relationships depends strongly on the number of compounds in the training set and the coverage of structural features. This paper reviews: (1) these new techniques; (2) trends in sooting tendency versus molecular structure; (3) structure-property relationships for sooting tendency; and (4) interpretation of the observed trends based on first-principle chemical kinetic and molecular dynamic simulations.

Keywords: biofuels, synthetic fuels, smoke point, yield sooting index

Nomenclature

Variables

a, b	rescaling constants for TSI
a', b'	rescaling constants for OESI
C	number of carbon atoms in a molecule
d_a	particle aerodynamic diameter
H	number of hydrogen atoms in a molecule
h_{flame}	the height of a flame
$\dot{m}_{fuel}$	mass fuel flowrate
M_r	relative molecular mass
N	number of nitrogen atoms in a molecule
n_{oxy}	number of oxymethylene units in a POME
$\dot{n}_{fuel}$	molar fuel flowrate
O	number of oxygen atoms in a molecule
P_{443K}	vapor pressure at 443 K
Q_{fuel}	volumetric fuel flowrate
S	moles of oxidizer required to burn one mole of fuel
T_{ad}	adiabatic flame temperature
T_{flame}	characteristic temperature of a flame
T_o	initial temperature of the fuel
ρ	mass density
φ	equivalence ratio

Acronyms

AIC	aviation induced cloudiness/cirrus
ANN	artificial neural network
AQG	air quality guideline of the WHO
BC	black carbon
CRP	color ratio pyrometry
DBE	double bond equivalent
DSP	derived smoke point
FESI	fuel equivalent sooting index
FURTI	fuel uptake rate measurement with threshold imaging
GDI	gasoline direct injection
GCM	group contribution method
GPR	Gaussian process regression
GSAT	global surface air temperature
HC	hydrocarbon
IDSY	isolated droplet soot yield
LHV	lower heating value
LII	laser induced incandescence
MD	molecular dynamics
ML	machine learning
MLR	multivariate linear regression
MPI	micropyrolysis index
NAAQS	national ambient air quality standard of the US EPA
NSP	normalized smoke point
OESI	oxygen extended sooting index
$PM_{2.5}$	particulates with $d_a \leq 2.5$ µm
PM_{10}	particulates with $d_a \leq 10$ µm
PMI	particulate mass index
POME	poly(oxy)methylene ether
QSPR	quantitative structure property relationship
SAF	sustainable aviation fuels
ST	sooting tendency
SVM	support vector machines
TSI	threshold sooting index
UFP	ultrafine particles; $d_a \leq 0.1$ µm

VSP	virtual smoke point
YSI	yield sooting index

1. Introduction

This paper reviews recent research related to the sooting tendency (ST) of hydrocarbons. This topic has seen an explosion of interest in the past decade since it provides the necessary scientific basis for choosing sustainable fuel compositions that will reduce soot emissions. The remainder of this section describes this opportunity (Section 1.1), the motivation to pursue it due to the adverse effects of soot emissions on climate (Section 1.2) and human health (Section 1.3), the chemical kinetic basis that causes ST to depend strongly on fuel composition (Section 1.4), the relationship of ST to other fuel properties that affect soot emissions (Section 1.5), and strategies for designing fuels that simultaneously optimize soot and other fuel properties (Section 1.6). Section 2 reviews experimental techniques for measuring ST, including both older threshold-based approaches such as smoke point and newer yield-based approaches such as yield sooting index (YSI). Section 3 discusses some of the trends in measured STs, including the effects of unsaturation, heteroatoms, and isomerism. Section 4 describes the use of machine learning to convert these experimental results into systematic knowledge through structure-property relationships. Finally, Section 5 reviews numerical simulations of ST, which provides insight into the kinetic mechanisms of soot formation.

1.1 Sooting tendency research is a key enabler for sustainable fuels

The left side of Fig. *1* shows the main part of the world's current energy economy: fossil fuels are extracted from the ground, they are burned to produce energy, and the products are emitted into the air. This model is completely linear—whenever more energy is needed, more fossil fuels are harvested, and more CO_2 is released. It has enabled an extraordinary increase in human prosperity, life expectancy, and population. However, it has also led to geopolitical problems due to the uneven distribution of fossil fuels, and to global warming due to the accumulation of CO_2 in the atmosphere.

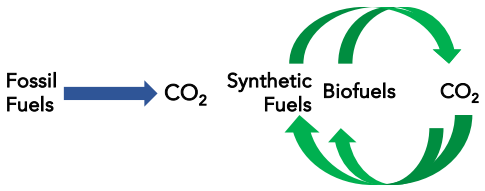


Figure 1. The current energy economy (left) and a circular energy economy based around sustainable fuels (right). The left-side emphasizes fossil fuels since they constituted 80% of the world's primary energy supply in 2021 [1].

These problems can be ameliorated by the circular energy economy shown on the right of Fig. *1*. In this model, hydrocarbon (HC) fuels would still be burned to produce energy in many contexts such as heavy-duty road transport, aviation, and shipping (upper half of the circle). However, the CO_2 produced from combusting these fuels—or an equivalent amount of atmospheric CO_2—would then be recycled into the next batch of fuels (lower half of the circle). CO_2 can be converted into these sustainable fuels with two broad strategies. The first is to grow plants, algae, or cyanobacteria and then convert their biomass into biofuels. These organisms use solar energy to take CO_2 directly from the atmosphere and fix it into HCs via photosynthesis. The second strategy is to use catalytic reactors to convert CO_2 from powerplant exhausts, fermentation, direct air capture, and other sources into synthetic fuels. These fuels are also called e-fuels or power-to-liquids since the thermochemical reactors are ideally powered by green electricity from solar and wind.

The circular energy economy raises a key question: what is the best chemical composition for these sustainable fuels [2, 3]? Since the starting material is CO_2, the fuel composition is not limited to the molecules in fossil fuels but can be anything in organic chemistry. Answering this question correctly is vital to enabling the circular energy economy. Sustainable fuels currently cost much more than fossil fuels, which is the primary factor limiting their deployment. The value proposition of sustainable fuels can be improved by choosing compositions that are less expensive to produce and that have better properties so users will pay more for them. Combustion science can contribute to the former by developing new technologies that allow a wider range of fuels to be used in practice, and to the latter by identifying desirable fuel properties and the molecules that possess them. Thus, sustainable fuel composition presents an opportunity for the combustion science community to move from reactively studying the behavior of sustainable fuels in case they are ever implemented, to proactively enabling their deployment in the first place.

Sooting tendency (ST) is among the most important fuel property targets for sustainable fuels. One reason is the mounting evidence that soot emissions are among the worst aspects of combustion due to their effects on climate change and human health—these issues are reviewed in Sections 1.2 and 1.3. A second reason is that ST depends strongly on molecular structure, so fuel composition presents a strong lever to reduce emissions. As an example, Fig. 2 shows a photograph of pool fires burning ethanol (left) and benzene (right). The yellow color is blackbody emission from soot particles in the flames. The ethanol flame is mostly blue, with only a small amount of soot production at its tip. In contrast, the benzene flame produces so much soot that the camera is saturated over the entire surface of the flame. Thus, this photograph shows that two fuels can produce very different amounts of soot even when burning under the same conditions.

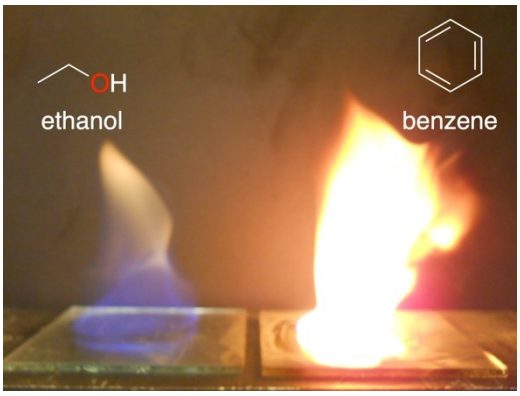


Figure 2. Photograph of pool fires burning ethanol (left) and benzene (right).

1.2. Effects of soot emissions on climate change

Soot emissions affect climate through three mechanisms [4, 5]:

1. Soot particles are "black", with emissivities close to one, so they absorb incident sunlight and heat the atmosphere. This warming mechanism is more direct than the greenhouse gas effect, where species like CO_2 do not interact with incident visible light, but absorb infrared light reradiated from the earth's surface. Atmospheric particles other than soot are mostly "white" and scatter or reflect incident sunlight. Overall, the net effect of aerosols is to cool the climate, but this is the sum of a warming effect due to soot and a cooling effect due to the other particles. Atmospheric scientists refer to soot as black carbon (BC) to emphasize the consequences of its unique optical properties.

2. Soot particles decrease the earth's reflectivity when they deposit on snow and ice. This change in albedo causes warming since it increases the proportion of incident light that is absorbed by the earth versus reflected to space.
3. Soot particles emitted from airplanes in the stratosphere create contrails by acting as nucleation sites for ice crystals. These contrails can then convert into aviation induced cloudiness/cirrus (AIC), which is relatively long-lived (i.e., longer than 10 minutes) [6, 7]. Water-based clouds at lower tropospheric altitudes tend to have a net cooling effect due to reflection of incident sunlight, but the ice-dominated stratospheric AIC tend to have a net warming effect due to the greenhouse mechanism [6].

The Intergovernmental Panel on Climate Change is a United Nations organization that comprehensively reviews and synthesizes the state-of-the-art in climate science via its periodic Assessment Reports. The Sixth Assessment Report includes estimates of the effect of specific climate forcing agents on the global surface air temperature (GSAT) in 2019 relative to 1750 [4, 5]. The values include +1.0 °C for CO_2, +0.1 °C for the BC direct effect, +0.04 °C for the BC albedo effect, and +0.02 °C for the BC contrail effect. Thus, soot emissions have a total effect of +0.16 °C. This amount is one-tenth of the 1.5 °C limit identified by the Paris Climate Accords, so reductions in ST will add significantly to the climate benefits of sustainable fuels.

The change in GSAT due to contrails is only a small part of the effect from soot, but it derives entirely from one sector of human activity—aviation—and it is the largest cause of climate change from that sector [4]. Voigt et al. demonstrated the role of soot particles and sustainable fuels on contrail formation through flight experiments where a lead jet used various fuels, and a chase plane sampled its exhaust [8]. When the lead plane burned sustainable aviation fuels (SAF) instead of a conventional Jet A-1 fossil fuel, its soot emissions decreased by about a factor of two, and the ice particle concentrations in its wake decreased by the same proportion. These results directly show that SAF can reduce the climate impact of contrails. Subsequent emissions modeling estimates that adoption of 100% SAF on all flights would reduce the climate impact of contrails by 44%, and that targeted adoption of 50% SAF to the most critical 2% of flights would still reduce it by 10% [9]. The ST research discussed in this review can be used to optimize the composition of SAF to maximize this benefit.

1.3. Effects of soot emissions on human health

Ambient particles are classified into categories based on their aerodynamic diameter d_a: coarse particulates ($d_a \leq 10$ μm, designated PM_{10}), fine particulates ($d_a \leq 2.5$ μm, $PM_{2.5}$), and ultrafine particulates ($d_a \leq 0.1$ μm, UFPs). Strictly speaking, these categories overlap, i.e. PM_{10} includes the two smaller categories; however, most particulate standards and guidelines are written in terms of mass—which scales with d_a^3—so the particles from the smaller categories contribute negligibly to the total. In terms of aerosol dynamics, the UFPs correspond to the nucleation mode of the initial particles formed by condensation processes, $PM_{2.5}$ corresponds to the accumulation mode of aggregated nucleation mode particles, and PM_{10} corresponds to the coarse mode of particles formed by frictional processes. Soot particles are formed by a chemical condensation process, so the primary particles are typically in the UFP range and the aggregates are in the $PM_{2.5}$ range [10, 11].

In 1993 the Six Cities Study showed that mortality in six U.S. cities correlated strongly with $PM_{2.5}$, but less so with PM_{10} and other types of air pollution [12]. Since then, overwhelming epidemiological evidence has verified that $PM_{2.5}$ contributes to poor human health through cardiopulmonary disease and lung cancer [13, 14]. While people tend to fixate on cancer, heart disease is the leading cause of death worldwide (one out of every three deaths) [15], so it is the contribution to heart disease that makes $PM_{2.5}$ exposure a particularly severe problem. Particles may cause heart disease by (1) triggering oxidative stress in the lungs and the release of inflammatory compounds into the bloodstream, (2) activating alveolar sensory receptors and disrupting the autonomic nervous system that regulates heart function, and (3) crossing directly over the lung epithelial barrier into the bloodstream [16].

Many governments and international agencies have implemented $PM_{2.5}$ air quality standards and guidelines, and these have tended to grow more stringent over time. For example, in 2021 the World Health Organization (WHO) reduced its $PM_{2.5}$ air quality guideline (AQG) from 10 to 5 μg/m³ [17]. Similarly, in February 2024 the U.S. Environmental Protection Agency (US EPA) reduced its annual $PM_{2.5}$ national ambient air quality standard (NAAQS) from 12.0 μg/m³ to 9.0 μg/m³ [18, 19].

Unfortunately, the exposures to fine particulates are enormous. Spatially resolved $PM_{2.5}$ concentrations for the entire earth can be derived from satellite aerosol optical depth measurements. The results show that in 2018 essentially the entire global population was exposed to concentrations that exceeded the WHO AQG, and more than 90% of people were exposed to levels above the US EPA NAAQS [20]. Moreover, for most of the population the levels were increasing over time, by up to 3 μg/m³ per year.

The most comprehensive effort to determine the health consequences of these exposures, and how they compare to other risk factors, is the 2019 Global Burden of Disease study (GBD 2019), which was a collaborative effort of over 5,000 researchers [21]. GDB 2019 concluded that $PM_{2.5}$ was responsible for 4.1 million deaths worldwide in 2019. The number of deaths had more than doubled from 2.0 million in 1990 due to higher exposures. $PM_{2.5}$ was the most significant environmental risk factor, and the sixth largest risk factor overall. Other studies have produced even larger totals; for example, Burnett et al. concluded that $PM_{2.5}$ was responsible for 8.9 million deaths in 2015 [22]. Ambient $PM_{2.5}$ includes many types of particles (soot, dust, sulfates, nitrates, etc.) from many sources (engines, furnaces, cookstoves, wildland fires, etc.). However, estimates of the annual deaths due to just fossil fuel combustion sources range from 1 to 10 million [23, 24]. Furthermore, estimates of the annual deaths due to just BC—the particle type definitely attributable to soot emissions—include 14,000 in the United States in 2010 and 265,000 in China in 2013 [25, 26]. These numbers are astonishing, and they show that adopting sustainable fuels with lower ST than existing fuels will have enormous health benefits for humanity.

1.4 The chemical kinetic basis for sooting tendencies

This review defines ST as the quantitative propensity of a fuel to produce soot relative to other fuels when burned. It is a property of the fuel and not the measurement procedure. Sections 2.1 and 2.2 show that ST can be measured by a wide range of methods, and section 2.3 shows that these methods give similar results for any given fuel when properly scaled. This section discusses the kinetic mechanism of ST with the goal of explaining why ST is a true fuel property, and why it depends so strongly on molecular structure.

As background, Fig. 3 presents a schematic overview of the soot formation process. As the fuel enters the flame, it encounters heat and small radicals generated at the main reaction front and it decomposes to a pool of primary hydrocarbon products (top). If the local C:O ratio is sufficiently low, then these products will be oxidized to CO_2 and H_2O (bottom left). However, if the local C:O ratio is too high, then these products will react with one another and produce larger hydrocarbons through a process of hydrocarbon growth (bottom right). These larger hydrocarbons will include aromatic molecules with more benzenoid rings than were present in the fuel, and they will eventually grow to soot particles.

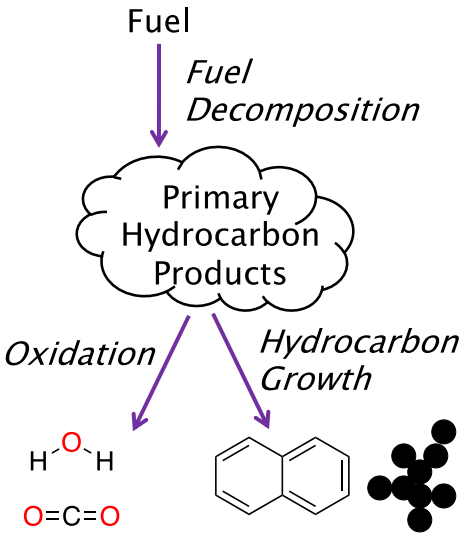


Figure 3. A schematic overview of the soot formation process.

ST is determined by the composition of the primary hydrocarbon products and how quickly they can grow to larger aromatics [27]. To illustrate this concept, Fig. *4* shows the soot formation process for ethanol and benzene, the two fuels that were compared in Fig. *2*. Ethanol decomposes to small HCs (methyl radical, ethylene, etc.) and small oxygenated HCs (formaldehyde, acetaldehyde, etc.) (left side) [28]. Benzene decomposes primarily to phenyl radical since benzenoid rings are stable under pyrolytic conditions (right side) [29]. Thus, two possible factors can explain why the ethanol pool fire produces so much less soot than the benzene pool fire in Fig. *2*. First, the primary products from ethanol must undergo the process of forming one-ring aromatic species, whereas the primary products from benzene already include one-ring species and skip this slow step. Second, the oxygenated products of ethanol may already be on an inescapable path to CO_2, especially since they include strong C=O double bonds, and so their carbon is not available to grow to aromatics. Crucially, both factors are very general and would apply to any fuel-rich combustion environment, which explains why ST is a property of the fuel and not the combustion system.

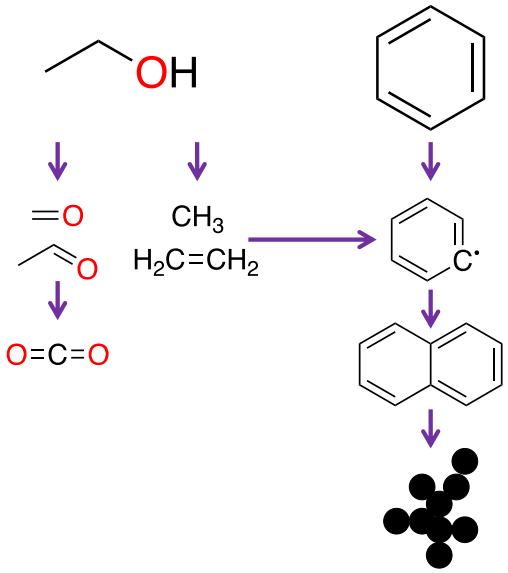


Figure 4. A schematic overview of the soot formation process for the two fuels that were compared in Fig. 2: ethanol (left) and benzene (right).

1.5. Predictive emission indices

Soot formation is a complex process whose rates depend on many factors including temperature, residence time, pressure, etc. [30, 31], but this review focuses on the effects of fuel composition. The overview of soot formation presented in Fig. 3 suggests that there are two primary mechanisms by which the fuel composition can affect the amount of soot formed in a particular combustion device. The first is the rate at which the fuel's decomposition products grow to large aromatic molecules, i.e., the effect quantified by ST. The second is the level of fuel-to-air mixing, which affects the C:O ratio and therefore the relative importance of hydrocarbon growth to soot (right side of the figure) versus oxidation to CO_2 and H_2O (left side). The level of mixing is primarily determined by the device technology, but it can also be affected by fuel properties.

Predictive emission indices are formulas that combine the direct effect of ST and the indirect effect of other fuel properties on mixing to predict the overall fuel-dependence of emissions from a specific combustion technology. The best-known example is particulate mass index (PMI), which was developed by Aikawa and co-workers in 2010 to predict soot emissions from gasoline direct-injection (GDI) engines [32]. PMI is defined by:

$$\mathrm{PMI} = \sum_j w_j * \frac{\mathrm{DBE}_j + 1}{P_{j,443\mathrm{K}}} \quad (1)$$

where the sum is over all the components j in the fuel, w_j is the mass fraction of component j, DBE_j is the double bond equivalent of component j, and $P_{j,443\mathrm{K}}$ is the vapor pressure of component j at 443 K. DBE is a measure of the number of rings and higher order bonds in a molecule; for the general hydrocarbon $\mathrm{C}_C\mathrm{H}_H\mathrm{O}_O\mathrm{N}_N$ it is given by:

$$\mathrm{DBE} = \frac{2C + 2 - H + N}{2} \quad (2)$$

The numerator of Eq. 1 accounts for the direct effect of fuel composition on soot formation. It uses DBE as a crude estimate of ST. This was necessary in 2010 when PMI was created: DBE can be immediately calculated for any molecule, while STs had only been measured for a limited range of molecules at that time. However, Section 3.1 shows that while ST roughly correlates with DBE, a wide range of ST exists for any given DBE. Fortunately, since 2010 ST has been measured for many more molecules (see Section 2), and this has enabled detailed quantitative structure-property relationships (QSPRs) that are much more precise than ST ~ DBE and just as easy to calculate (see Section 4).

The denominator of Eq. 1 accounts for the indirect effect of the fuel on soot formation via the impact of its properties on mixing. In a GDI engine, the fuel is injected directly into the cylinder during the intake stroke and forms a liquid layer on the cylinder walls and piston surface. During the compression stroke the fuel will evaporate and begin to mix with the air. If the fuel is more volatile (i.e., has a larger value of $P_{j,443\mathrm{K}}$), then it will evaporate more quickly and mix more thoroughly with the air by the end of the compression stroke. This mechanism is specific to GDI engines, and it does not occur in other devices such as diesel engines and gas turbines—thus, PMI is not applicable to these other devices.

In general, ST is a fundamental property that can be measured once for each fuel and then compiled into databases that can be applied to any combustion device. In contrast, predictive emission indices like PMI need to be developed separately for each combustion technology, but then they can use ST as input to produce more refined emissions predictions.

1.6 Rational design of sustainable fuels with improved properties

STs have been used for many purposes: (1) formulating surrogate mixtures that mimic the behavior of real fuels, (2) testing the ability of chemical kinetic mechanisms to predict soot formation, and (3) designing fuels with improved properties. Table *1* lists some noteworthy examples of these uses.

Table 1
Noteworthy publications that use ST. SP, TSI, and YSI are measures of ST defined in Sections 2.1 and 2.2.

Publication	Description
Aksit and Moss, 2005 [33]	SP was a property target for a jet fuel surrogate
Mensch et al., 2010 [34]	TSI was a property target for a jet fuel surrogate
Huo et al., 2019 [35]	YSI was a property target for selecting alkanes for improved diesel fuels
Huq et al., 2019 [36]	YSI was a property target for selecting ethers for improved diesel fuels
Kwon et al., 2020 [37]	20 YSIs were simulated to test kinetic mechanisms
Bartholet et al., 2021 [38]	YSI was a property target for selecting polyoxymethylene ethers for improved diesel fuels
Cai et al., 2022 [39]	YSI was a property target for selecting molecules for improved diesel fuels
Kuzhagaliyeva et al., 2022 [40]	YSI was a property target for selecting mixtures for improved spark-ignition fuels
Li et al., 2022 [41]	YSI was a property target for selecting molecules for improved internal combustion engine fuels
Fleitmann et al., 2023 [42]	YSI was a property target for selecting molecules for improved spark-ignition fuels

Given the purposes of this review, fuel design is of particular interest. The challenge is that the performance of a fuel is affected by many properties: ST, heat of combustion, ignition quality, melting point, boiling point, flash point, corrosivity, water solubility,

biodegradability, toxicity, and many others. No fuel will have ideal values of all these properties, so designing optimal fuels requires trade-offs.

Polyoxymethylene ethers (POMEs) provide a good example of these trade-offs. The structure of a POME consists of an initial end-group R, an ether O, n_{oxy} oxymethylene units (–CH_2–O–), and a final end-group R' (see Fig. 5A). POMEs with R = R' = methyl have been widely discussed as alternative diesel fuels [43, 44]. They have very low STs due to the absence of C-C bonds [45, 46, 47], and engine studies show that they have very low soot emissions [48, 49, 50]. However, they also have high water solubilities and low heats of combustion. Water-soluble fuels are a concern because they can dissolve into water layers that are often present in the fuel infrastructure, and they can contaminate groundwater when spilled. These properties can be improved by replacing the methyl end-groups with larger end-groups, but at the cost of increasing ST (see Fig 5B). Bartholet et al. optimized this trade-off by (1) defining a fuel design space containing 67 candidate POMEs with different combinations of R, R', and n_{oxy}, (2) setting required values for eight properties, (3) predicting these properties for every candidate with QSPRs, and (4) identifying the candidates that met all the property requirements [38]. The best candidate was dibutoxymethane (R = R' = *n*-butyl, n_{oxy} = 1; see Fig 5C). This compound has a ST that is higher than the methyl-terminated POMEs, but still much lower than conventional diesel fuel, and its water solubility is about 1,000 times lower than the methyl-terminated POMEs. The superior performance of this butyl-terminated POME was confirmed by subsequent testing [51].

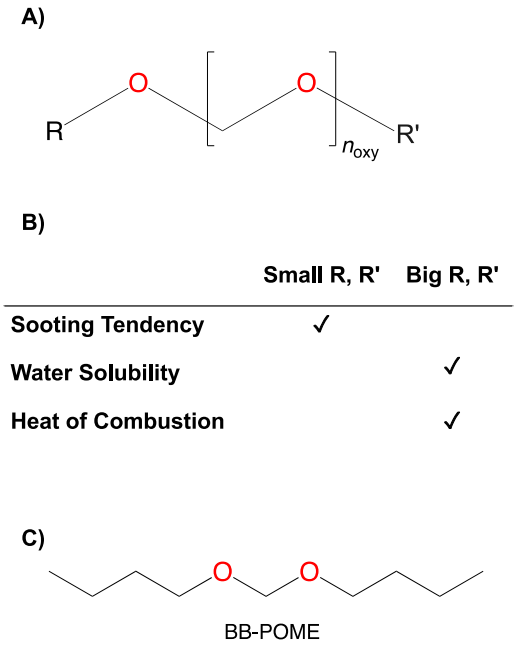


Figure 5. A) Generic POME structure. B) The trade-offs in fuel properties depending on the size of the end groups. C) The optimum POME structure identified by Bartholet et al. [38].

This example shows that fuel design studies require QSPRs that can accurately predict fuel properties for a wide range of candidate molecules—almost all the 67 POME candidates had never been synthesized, let alone tested for fuel properties. The accuracy of QSPRs depends critically on the size and comprehensiveness of the database used for training, so large databases with wide coverage of structural features are also essential. Section 2 reviews methods for measuring ST, with emphasis on newer yield-based methods that enable datasets that are large enough to accurately train QSPRs. Section 3 shows that measured STs cannot be predicted precisely with simple parameters such as C:H ratio or DBE. Section 4 discusses QSPRs that provide much better predictions of ST. Finally, Section 5 discusses computational simulations that can explain measured STs with chemical kinetic mechanisms and molecular dynamics.

2 Experimental measurements of sooting tendency

This section discusses the experimental methods that are used to measure ST. The emphasis is on methods that have been applied to large sets of compounds (> 10) and therefore provide detailed structure-property information. Sooting propensity has also been studied in standard kinetic systems such as flow reactors (e.g., [52]) and shock tubes

(e.g. [53]), but the focus in those studies is on the effects of temperature and other kinetic parameters, so they are outside the scope of this review.

The ST methods divide into two categories: threshold-based methods (Section 2.1) and yield-based methods (Section 2.2). The threshold methods do not require a soot measurement and can use simple, commercially available equipment. The yield methods require a custom setup but can be applied to a wider range of compounds and enable larger datasets.

2.1 Threshold-based sooting tendencies

The first ST technique to be widely adopted was smoke point (SP). It was developed around 1930 [54] and has been codified into the ASTM D1322 standard test method [55]. The ASTM specification for Jet A/A-1 aviation fuel, D7566 [56], requires SP > 25 mm (or SP > 18 mm if the volume percent of naphthalenes is less than 3%). Given this requirement, fuels are routinely tested for SP in the aviation industry and the apparatus for performing it is ubiquitous.

The SP procedure is to generate a coflow flame, increase the fuel flowrate until soot just breaks through the tip of the flame, and then measure the flame height at this threshold. "Smoke point" refers both to the flame being at this condition and to the measured height. The underlying idea is that if the fuel has a greater ST, then it will form more soot for a given fuel flowrate, and the flame will be shorter when the amount of soot formed exceeds the amount that can be oxidized. Thus, SP is inversely proportional to ST, which explains why ASTM D7566 requires a minimum (not maximum) value for SP. The ASTM D1322 apparatus generates the flame with a wick burner and varies the fuel flowrate by changing the amount of wick that projects from a metal tube.

Although SP is primarily used for certifying aviation fuels, it has also been an important tool for fundamental combustion research. Table 2 lists some noteworthy studies related to SP and to other methods that depend on identifying a threshold.

Table 2

Noteworthy publications related to threshold-based ST methods. Q_{fuel} and $\dot{m}_{fuel}$ *are the volumetric and mass fuel flowrates;* φ *is fuel-air equivalence ratio.*

Publication	Description
Hunt, 1953 [57]	SP was measured for HCs, sulfur-containing HCs, and nitrogenated HCs

Schalla and McDonald, 1953 [58]	Q_{fuel} at the SP was measured for HCs
Blazowski, 1980 [59]	φ at the onset of soot formation was measured for HCs in a jet-stirred reactor
Glassman and Yaccarino, 1981 [27]	Q_{fuel} at the SP was measured for HCs with the flame temperature controlled via N_2 dilution
Calcote and Manos, 1983 [60]	Threshold sooting index (TSI) was proposed and used to combine literature data from premixed and nonpremixed flames
Senkan et al., 1983 [61]	φ at the onset of soot formation was measured for chlorinated HCs in a flat flame burner
Takahashi and Glassman, 1984 [62]	φ at the onset of soot formation was measured for HCs in a Bunsen burner with the temperature controlled via inert dilution
Gill and Olson, 1984 [63]	TSI was measured for binary and ternary HC mixtures and shown to fit a linear mixing rule
Gomez et al., 1984 [64]	$\dot{m}_{fuel}$ at the SP was measured for HCs with the temperature controlled via inert dilution
Olson et al., 1985 [65]	SP was measured for HCs
Gülder et al., 1989 [66]	SP was measured for binary and ternary HC mixtures
Gülder et al., 1990 [67]	SP was measured for HCs, binary and ternary HC mixtures, and transportation fuels
Ladommatos et al., 1996 [68]	SP was measured for HCs
Yan et al., 2005 [69]	SP was measured for binary HC mixtures
Yang et al., 2007 [70]	SP was measured for jet fuels and their TSIs were shown to correlate with soot formation in a gas turbine
Berry Yelverton and Roberts, 2008 [71]	SP was measured for methane and ethylene at pressures up to 8 atm
Pepiot-Desjardins et al., 2008 [72]	SP was measured for oxygenated HCs mixed with a HC base fuel

Allan et al., 2009 [73]	SP was measured for HCs, oxygenated HCs, and waxes in candles
Barrientos and Boehman, 2010 [74]	SP was measured for multi-ring aromatics and their saturated analogues mixed with a HC base fuel
Mensch et al., 2010 [34]	SP was measured for jet fuel surrogate HCs
Dotson et al., 2011 [75]	SP was measured for HCs in microgravity
Li and Sunderland, 2012 [76]	Literature SPs were combined into a database of normalized smoke points (NSPs)
Barrientos et al., 2013 [77]	Oxygen extended sooting index (OESI) was proposed and applied to SPs measured for oxygenated HCs
Li and Sunderland, 2013 [78]	Linear mixing rules were developed for literature SPs of HC mixtures
Llamas et al., 2013 [79]	SP was measured for biokerosenes mixed with conventional kerosene
Watson et al., 2013 [80]	Fuel uptake rate measurement with threshold imaging (FURTI) was proposed and used to measure objective STs for HCs and HC mixtures
Wang and Chung, 2014 [81]	Oxygen and fuel mole fractions at the onset of soot formation were measured in counterflow nonpremixed flames
Graziano et al., 2018 [82]	SP was measured for HCs and HC mixtures using FURTI
Tan et al., 2018 [45]	SP was measured for POMEs mixed with diesel fuel
Rubio-Gomez et al., 2019 [83]	An automated SP procedure was proposed and applied to HCs
Cho et al., 2020 [84]	Virtual smoke point (VSP) was proposed and used to measure SPs that exceeded the upper limit of the apparatus
Corral-Gomez et al., 2020 [85]	An automated SP procedure was proposed and applied to HCs

Donoso et al., 2021 [86]	SP was measured for terpenes and hydrogenated terpene mixtures
Li et al., 2022 [87]	Oxygen mole fraction at the onset of soot formation was measured for oxygenated HCs mixed with ethylene in counterflow nonpremixed flames
Muelas et al., 2023 [88]	An automated SP procedure was proposed and applied to HCs

These publications can be grouped into several categories. First, several studies report SPs for large sets of compounds [34, 57, 58, 64, 65, 68]. Li and Sunderland have compiled these results into a database of normalized smoke points (NSPs) that includes 112 HCs [76]. As discussed in Section 2.2, larger databases are available for yield-based STs.

Second, several procedures have been proposed for normalizing SPs. The purposes are to (1) invert SP so that larger values correspond to sootier fuels, (2) factor out experimental differences such as varying wick diameters, and (3) clarify the intensive basis of SP. Regarding the latter, the most fundamental basis would be (soot/mole)—the amount of soot formed per mole of fuel burned—but SP is not directly related to this ratio since the proportionality constant between the flame height h_{flame} and the molar fuel flowrate $\dot{n}_{\text{fuel}}$ depends on the fuel. Roper has shown that they are related by [89]:

$$h_{\text{flame}} \sim \frac{\dot{n}_{\text{fuel}} * (T_0/T_{\text{flame}})^{0.67}}{D_0 * \ln(1 + 1/S)} \tag{3}$$

where T_0 is the initial temperature of the fuel, T_{flame} is a characteristic temperature for the flame, D_0 is a characteristic diffusivity for the oxidizer into the fuel at T_0, and S is the moles of oxidizer required to stoichiometrically burn one mole of fuel. For the general hydrocarbon $C_CH_HO_ON_N$ burning in air (O_2 + 3.77N_2):

$$S = 4.77 * (C + \frac{H}{4} - \frac{O}{2}) \tag{4}$$

For a pure hydrocarbon if $C \geq 3$, then $S \geq 14.3$ and $\ln(1+1/S) = 1/S$ to within 3.5%. Thus, if T_0, T_{flame}, and D_0 are assumed to be constant, then:

$$h_{\text{flame}} \sim S * \dot{n}_{\text{fuel}} \tag{5}$$

If we assume that the SP occurs when the amount of soot formed in the flame reaches a fuel-independent critical value, then (soot/mole) ~ $1/\dot{n}_{\mathrm{fuel,SP}}$ where $\dot{n}_{\mathrm{fuel,SP}}$ is the molar fuel flowrate at the SP. Combining this with Eq. (5) yields:

$$\left(\frac{\mathrm{soot}}{\mathrm{mole}}\right) \sim \frac{S}{\mathrm{SP}} \tag{6}$$

Based on this type of argument, Calcote and Manos proposed normalizing SP into a threshold sooting index (TSI) defined by [60]:

$$\mathrm{TSI} = a\left(\frac{M_{\mathrm{r}}}{\mathrm{SP}}\right) + b \tag{7}$$

where M_{r} is the relative molecular mass (i.e., molecular weight) of the fuel, and a and b are constants chosen so that the TSI(1-methylnaphthalene) = 100 and TSI(n-hexane) = 2. They recognized that S would be the more appropriate normalizing factor but argued that $M_{\mathrm{r}} \sim S$ with sufficient accuracy given the uncertainties in SP. They included the rescaling with the constants a and b so that data from a wide range of experimental configurations could be directly compared. This data included both SPs measured in nonpremixed flames and fuel-air equivalence ratios φ for the onset of soot formation in premixed flames.

Subsequently, Barrientos et al. noted that $M_{\mathrm{r}} \sim S$ fails for oxygenated fuels due to the $-O/2$ term in Eq. (4), and they proposed an oxygen extended sooting index (OESI) defined by [77]:

$$\mathrm{OESI} = a'\left(\frac{S}{\mathrm{SP}}\right) + b' \tag{8}$$

A third category in Table 2 are studies that extend SP to new configurations. Glassman and Yaccarino tested gaseous fuels diluted by N_2 and showed that SPs can be strongly affected by a fuel's adiabatic flame temperature T_{ad} [27]. For example, pure acetylene had a smaller SP than pure C3 and C4 alkenes, but the order reversed when the acetylene was diluted enough to equalize the temperature in all the flames. Allan et al. measured SPs of solid hydrocarbons by (1) forming them into candles, (2) drilling holes through the centerline, and (3) inserting wicks whose position relative to the candle could be varied [73]. Wang and Chung have defined a series of threshold STs for counterflow

flames [81]. Perhaps most exotically, Dotson et al. used the International Space Station to measure SPs in microgravity [75].

A fourth category are studies that seek to improve the SP method. The standard method requires the experimenter to subjectively determine when the flame is at the SP. Watson et al. proposed an alternative procedure, fuel uptake rate measurement with threshold imaging (FURTI), where h_{flame} is measured as a function of the fuel mass flowrate $\dot{m}_{fuel}$ and then SP is objectively determined as the h_{flame} where $\partial(h_{flame})/\partial(\dot{m}_{fuel})$ is a maximum [80]. Their results showed that FURTI significantly reduced the uncertainty of the measured SPs, and Graziano et al. independently verified this conclusion [82].

Another issue with SP is that it has a narrow dynamic range: if the fuel has a high ST, then the SP will be too short to determine accurately (e.g., only 3.5 mm for naphthalene [76]), and if the fuel has a low ST, then the SP will exceed the limit of the instrument. Haas and co-workers have addressed the latter problem with a virtual smoke point (VSP) procedure where (1) the low ST target compound is mixed at varying ratios with a high ST compound, (2) SPs are measured for the mixtures, and (3) the results are extrapolated to the case of the pure target compound [84].

2.2 Yield-based sooting tendencies

Direct soot measurements were impossible when ST first became important at the beginning of jet aviation in the 1930s. However, many techniques have been developed since then that enable in-situ quantification of soot concentrations. These techniques cover a wide range of complexity:

1. In laser-induced incandescence (LII) a laser pulse heats soot particles in the flame to a much higher temperature than the surrounding gases, and then a detector records the intense blackbody emission from this heating [90, 91, 92]. Even at modest laser fluences, the particles can be heated to the various carbon sublimation temperatures in the 4000 to 4500 K range, so the signals are strong and easily discriminated from the background flame emission. LII can determine both the mass concentration of soot from the magnitude of the signal and the particle size from the cooling rate after the laser pulse ends. It offers the usual benefits of laser diagnostics including the ability to perform spatially resolved point or 2D imaging measurements.
2. In laser extinction (LE) a laser beam propagates through the flame, the transmitted signal is measured, and then the soot concentration is determined from the amount of

absorption [93, 94]. LE requires much less expensive lasers than LII, and it is often used to calibrate LII data since its data is much simpler to interpret.

3. In color ratio pyrometry (CRP) a color camera photographs the soot emission in several wavelength regions, the soot temperature is calculated from the ratios of the color channels with Planck's Law, and the soot concentration is determined from an absolute light calibration with a thermocouple [95, 96]. The only equipment required is a consumer grade digital camera. The line-of-sight data must be converted to spatially resolved data with an Abel inversion, but techniques have been developed for this transformation that suit CRP [97, 98].

As early as the 1980s researchers noted that the TSI of a fuel correlated with the maximum soot concentration in its SP flame [e.g., 65]. Since 2000, several methods have been formulated that determine the ST for a fuel from the amount of soot it forms in a particular system. Table 3 lists noteworthy publications related to these methods.

Table 3

Noteworthy publications related to yield-based ST methods.

Publication	Description
McEnally and Pfefferle, 2007 [99]	Yield sooting index (YSI) was proposed and measured for HCs
Crossley et al., 2008 [100]	Micropyrolysis index (MPI) was proposed and measured for HCs
McEnally and Pfefferle, 2009 [101]	YSI was measured for large aromatics
McEnally and Pfefferle, 2011 [102]	YSI was measured for HCs and oxygenated HCs
Kashif et al., 2014 [103]	YSI was measured for *n*-heptane/toluene and isooctane/toluene mixtures
Kashif et al., 2015 [104]	YSI was measured for *n*-heptane/toluene and isooctane/toluene mixtures
Das et al., 2015 [105]	YSI was measured for unsaturated esters
Lemaire et al., 2015 [106]	Fuel equivalent sooting index (FESI) was measured for oxygenated HCs sprayed into a turbulent diesel surrogate flame

Abboud et al., 2017 [107]	YSI was measured for methyl decanoate mixed with a diesel surrogate
Das et al., 2017 [108]	YSI was measured for jet and diesel fuels and their surrogates
Abboud et al, 2018 [109]	YSI was measured for esters mixed with a diesel surrogate
Das et al., 2018 [110]	A unified YSI scale was created to replace the incompatible scales in [99], [101], and [102]
Staples et al., 2018 [111]	YSI was measured for dioxolanes
Staples et al., 2019 [112]	YSI was measured for camphorane
Ford Ryan et al., 2020 [113]	YSI was measured for cycloalkanes
McEnally et al., 2019 [114]	YSI was measured for gasolines and their surrogates
Montgomery et al., 2019 [115]	YSI was shown to be insensitive to partial premixing and temperature variation in the base flame
Carlson et al., 2020 [116]	YSI was measured for alkyl ethers
Arellano-Treviño et al., 2021 [51]	YSI was measured for butyl-exchanged POME mixtures
Cosimbescu et al., 2021 [117]	YSI was measured for bicyclic and multicyclic HCs
Huq et al., 2021 [118]	YSI was measured for a wet waste-derived SAF
Gu et al., 2021 [119]	Soot yield was measured for toluene and *n*-heptane doped into CH_4 flame at pressures up to 8 atm
Montgomery et al., 2021 [120]	YSI was measured for amines
Yang and Gülder, 2021 [121]	Soot yield was measured for benzene, cyclohexane, and *n*-hexane doped into a CH_4 flame at pressures up to 10 bar
Yang and Gülder, 2021 [122]	Soot yield was measured for ethylbenzene, *p*-xylene, *o*-xylene, and *n*-octane doped into a CH_4 flame at pressures up to 10 bar
Dagle et al., 2022 [123]	YSI was measured for an iso-olefin gasoline

Lucas et al., 2022 [46]	YSI was measured for POMEs
Zhu et al., 2022 [124]	YSI was measured for HCs, surrogate mixtures, and transportation fuels
Gleason and Gomez, 2023 [125]	Absolute soot production rates were measured for HCs in counterflow flames
Jalain et al., 2023 [126]	YSI was measured for alcohols in an ethylene base flame
Monroe et al., 2023 [127]	YSI was measured for alkoxyalkanoates
Singh and Tsolas, 2023 [128]	YSI was measured for isopropanol-butanol-ethanol mixed with a diesel surrogate
Zhu et al., 2023 [129]	YSI was measured for terpenes and hydrogenated terpenes
Muelas et al., 2023 [88]	IDSY (isolated droplet soot yield) was proposed and measured for HCs
Arellano-Treviño et al., 2024 [130]	YSI was measured for POMEs with branched end-groups
Jung et al., 2024 [131]	YSI was measured for benzene substituted with hydroxy, formyl and methoxy groups
Xiang et al., 2024 [132]	YSI was measured for lactones

The most widely adopted yield-based ST is yield sooting index (YSI). We proposed this method in 2007 [99] and have employed it extensively since. Several other research groups have published YSI measurements [e.g., 124, 103, 128]. Several studies have formulated sustainable fuels with YSI as one of the fuel property targets (see Section 1.6) and many structure-property relationships have been developed to predict YSI (see Section 4).

The YSI procedure is to dope a small concentration of the test compound into the fuel of a methane/air coflow flame and measure the maximum soot concentration f in the doped flame. The underlying idea is that if the test fuel has a larger ST, it will produce more soot, and a larger f will be measured. f has been measured with LII [99, 124], LE [103], CRP [108], and flame emission [114]. The base flame is a CH_4/air flame since it produces a low background f, but also has a more representative chemical environment than an H_2 flame.

YSIs measured in CH_4/air flames agree well with SP and other STs measured in pure-fueled flames (see Section 2.3), which shows that the YSI results are not sensitive to the background fuel.

In analogy to TSI—and to other fuel properties such as octane and cetane rating—YSI is rescaled relative to two endpoint species. The rescaling equation is:

$$\mathrm{YSI_{TF}} = (\mathrm{YSI_U} - \mathrm{YSI_L}) * \frac{f_{\mathrm{TF}} - f_{\mathrm{L}}}{f_{\mathrm{U}} - f_{\mathrm{L}}} + \mathrm{YSI_L} \quad (9)$$

Where the subscripts TF, U, and L denote the test fuel, the upper endpoint species, and the lower endpoint species. $\mathrm{YSI_U}$ and $\mathrm{YSI_L}$ are constants that define the scale; normally U is toluene, $\mathrm{YSI_U} \equiv 170.9$, L is *n*-heptane, and $\mathrm{YSI_L} \equiv 36.0$ [110]. Equation 9 shows that YSI depends on the ratio of *f* between flames, not its absolute value in a single flame; therefore, many sources of uncertainty cancel out, including the absolute calibration and the effect of soot optical properties on the diagnostic.

YSI offers several benefits:

1. It eliminates the subjectivity associated with SP: the experimentalist calculates the flowrates necessary to achieve the specified dopant concentration, generates the flame, and then triggers the soot diagnostic. The results for the objective versions of SP show that the uncertainties are significantly reduced when the subjectivity is eliminated [80, 82] (see Section 2.1).
2. Since the test compound is added to the flame at low concentrations—typically 1000 μmol/mol (1000 ppm)—the required sample volume is very small. Each measurement takes less than 50 μL of sample, which is 200 times less than the 10 mL required by ASTM D1322 for SP [55] This difference has enabled YSI to be measured for many hydrocarbons that are not commercially available and had to be custom-synthesized (e.g., [111] and many others). The SP method requires 10 mL because the wick must be fully saturated with the test fuel to obtain an accurate measurement; we are not aware of any studies that have attempted to reduce this requirement.
3. It has a large dynamic range, both because it is an objective measurement that uses precise diagnostics and because the dopant concentration can be increased or decreased to suit test compounds with small or large ST. Measured values of YSI range from 0.5 for 2,4,6,8-tetraoxanonane [46], a POME, to 1250 for pyrene [110], a four-ring polycyclic aromatic hydrocarbon.

4. Since the test compound is doped into the base flame at a small concentration, the ST of every compound is measured at the same flame temperature and fuel concentration. As noted in Section 2.1, Glassman and Yaccarino studied fuels diluted with N_2 and concluded that the fuel's T_{ad} could significantly impact its SP [27]. Later Axelbaum et al. showed that while N_2 dilution could normalize the flame temperature, the differences in fuel concentration for fuels with different T_{ad} also affected their SP [133]. The YSI approach removes the effect of T_{ad} without introducing changes in fuel concentration.
5. In principle the measurements can be performed at high pressure. Although no high-pressure YSIs have been reported, Gülder and co-workers have shown that soot yields can be measured in fuel-doped CH_4/air flames at pressures up to 10 atm [119, 121, 122].
6. The results can be compared to simulations with detailed mechanisms (see Section 5). Simulating a YSI requires computing a single flame, not a series of flames as would be required with SP to locate the threshold. The pre-vaporized coflow YSI flames have much simpler boundary conditions than wick burner flames and can be simulated with high accuracy (e.g., [134]). Most importantly, the base flame can be solved with methane kinetic mechanisms, which are relatively small, then the YSI flames can be solved from the base flame using perturbation methods [135].
7. Simulations can also be used to determine the effects of flame parameters and other fuel properties that are difficult to vary experimentally. For example, the YSIs of alkanes and aromatics have been simulated at different pressures [136]. The results indicate that while the absolute soot concentration f depends strongly on pressure, this dependence largely factors out when f is converted to YSI with Eq. 9.
8. It has a straightforward intensive basis: since the test fuel is added to the flame at a fixed mole fraction, YSI scales directly with (soot/mole). Other intensive ratios can be obtained from simple manipulations:

$$\left(\frac{\text{soot}}{\text{mole}}\right) \sim \text{YSI} \tag{10}$$

$$\left(\frac{\text{soot}}{\text{mass}}\right) \sim \frac{\text{YSI}}{M_\text{r}} \tag{11}$$

$$\left(\frac{\text{soot}}{\text{liquid volume}}\right) \sim \frac{\text{YSI} * \rho}{M_\text{r}} \tag{12}$$

$$\left(\frac{\text{soot}}{\text{energy}}\right) \sim \frac{\text{YSI}}{\text{LHV}} \tag{13}$$

where ρ is mass density and LHV is the lower heating value (in units of J/mol). (In some previous studies we have measured (soot/mass) directly and called it "YSI" [e.g., 108], but we recognize that practice is confusing, and we are deprecating it.)

The net result of these benefits is that YSIs have been measured for a large total number of hydrocarbons from many chemical families. For example, we have posted a canonical database that contains fully vetted and published YSIs for 447 compounds to the Harvard Dataverse online depository [137]. We have unpublished YSIs for an additional 216 compounds; the combined database of 663 compounds is included in the Supplemental Information (SI) to this review. This database includes values for 257 regular HCs, which is more than twice the number of values in the NSP database [76].

As discussed in Section 1.6, a major purpose of STs is to train QSPRs that can guide the development of sustainable fuels. The accuracy of the QSPRs depends strongly on the total number of data points in the training set and how widely it covers structural features. Thus, YSI databases are ideal for training QSPRs, and they have been widely used for this purpose (see Section 4).

2.3 Comparison of sooting tendencies

Sections 2.1 and 2.2 have shown that ST has been measured in a wide range of experimental systems. In general, the results agree across these systems, which supports the hypothesis that ST is a fundamental property of a fuel, and not of the measurement system (see Section 1.4). For example, Calcote and Manos stated that with normalization to TSI "all of the data in the literature on premixed and diffusion flames, taken by many techniques, can be successfully correlated with respect to molecular structure" [60]. Mensch et al. observed strong linear correlations between their measured TSIs and literature YSIs—in fact, these correlations were stronger than most of the correlations with literature TSIs [34]. Crossley et al. found that their MPIs—which were measured for pre-vaporized fuels in an 1120 K non-flame pyrolytic reactor—correlated with literature TSIs measured in flames [100]. Similarly, Muelas et al. observed a good correlation between their TSIs measured in coflow flames and their IDSYs measured for droplets in a 1700 K inert environment [88].

Recently, Zhu et al. quantitatively compared SP and YSI using many more data points than in previous comparisons [129]. From Eqs. 6 and 10, YSI/S ~ 1/SP. Figure *6*

shows YSI/S plotted against 1/SP for 80 HCs and 7 HC mixtures from many chemical families. The plot uses a log-log scale since the STs vary over a range of about 100. The data correlates reasonably well with the linear least-squares fit, which is shown by the dashed line. The largest scatter occurs for the aromatics, but in this case the SPs are all less than 10 mm and the challenges of measuring SP in such short flames are well-known [80].

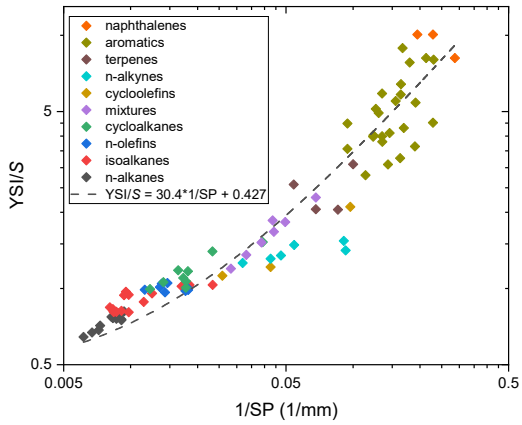


Figure 6. Comparison of YSIs and SPs. YSIs are from the database in the SI. SPs are from [55], [76], and [86].

The linear fit allows YSIs to be converted to derived smoke points (DSPs):

$$\frac{30.4}{\mathrm{DSP}} = \frac{\mathrm{YSI}}{S} - 0.427 \tag{14}$$

Melder et al. used Eq. 14 to demonstrate that the ST of complex fuels can be accurately predicted from the fuel's detailed composition [138]. They (1) characterized the molecular composition of 20 gasolines with a novel 2D gas chromatography technique; (2) calculated the YSI of these compositions using measured YSIs which were available for 95% of the components, extrapolated YSIs for the other 5% of components, and a linear mixing rule; (3) converted the YSIs to DSPs; and (4) showed that the DSPs agreed well with SPs measured directly for the fuels.

3 Effects of molecular parameters on sooting tendency

The goal of this section is to interrogate the large YSI database that is included in the SI to determine how ST depends on molecular details such as unsaturation, heteroatoms, and isomerism.

3.1 Effects of unsaturation

Combustion research has long since established that unsaturated hydrocarbons tend to soot more than their saturated analogues, and unsaturation is often used as a crude indicator of ST (e.g., in PMI, which is discussed in Section 1.5). Unsaturation can occur because of rings or because of double and triple C-C bonds. Figure 7 plots the measured YSIs as a function of C:H ratio, with separate data series for pure HCs (C and H only) and for HCs with heteroatoms (either O or N). The YSIs are plotted on a log scale since they vary over almost 10^4. In general, YSI increases as C:H increases, especially for C:H below 0.8. However, for any given C:H the YSIs vary by around a factor of four. For example, the YSIs for C:H = 0.5 range from 39 (cyclopentane) to 158 (1,2-dimethyl-4-(4-methylpentyl)cyclohexane), for C:H = 0.6 from 46 (cyclohexene) to 304 (*n*-dodecylbenzene), and for C:H = 1 from 100 (benzene) to 984 (2,2'-dimethylbiphenyl). A similar conclusion holds for YSI/*C*; its values range from 6.9 to 11.3 for C:H = 0.5, from 7.6 to 16.9 for C:H = 0.6, and 16.7 to 70.3 for C:H = 1.0.

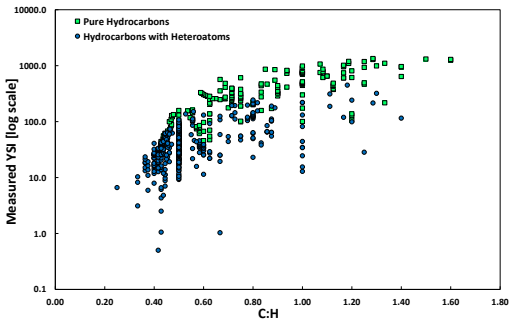


Figure 7. Measured YSIs as a function of the C:H ratio. YSIs are from the database in the SI.

A better indicator of unsaturation is DBE (defined in Eq. 2): alkanes and other fully saturated HCs have varying C:H but they all have DBE = 0. Figure 8 plots the measured YSIs as a function of DBE. As with C:H, YSI tends to increase as DBE increases, but for any given

DBE the YSIs vary by around a factor of four. For example, the YSIs for DBE = 0 range from 25 (*n*-pentane) to 133 (7-butyl-6-pentyltridecane), for DBE = 1 from 39 (cyclopentane) to 158 (1,2-dimethyl-4-(4-methylpentyl)cyclohexane), and for DBE = 2 from 46 (cyclohexene) to 161 (1-*tert*-butyl-cyclohexene). Again, the same conclusion holds for YSI/*C*; its values range from 4.9 to 8.3 for DBE = 0, from 6.9 to 11.3 for DBE = 1, and from 7.6 to 16.1 for DBE = 2. These results illustrate the need for QSPRs to predict sooting tendency.

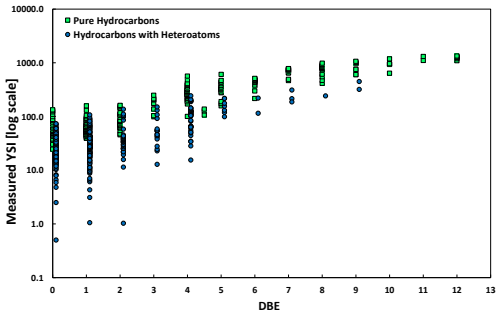


Figure 8. Measured YSIs as a function of DBE. DBE is defined in Eq. 2. YSIs are from the database in the SI. The data points for Hydrocarbons with Heteroatoms have been offset by +0.1 DBE for clarity.

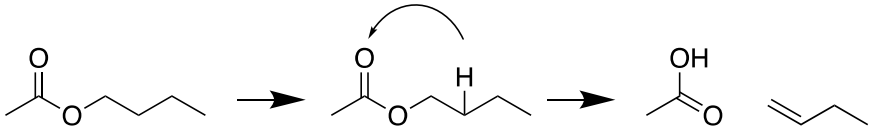
3.2 Effects of heteroatoms

Conventional fossil fuels are comprised almost exclusively of regular HCs (C and H only). However, sustainable fuels may include large amounts of oxygenated HCs since the raw material—biomass or CO_2—contains O. Fuel specifications often limit the O content, but relaxing these requirements may be a valuable way to improve the economics of sustainable fuels by reducing the amount of fuel processing.

Oxygenated HCs are also of interest because they frequently have lower ST than pure HCs. For example, Pepiot-Desjardins et al. observed that ST was reduced when oxygenated HCs were added to *n*-heptane/toluene or diesel mixtures [72]. By fitting the measured results to a QSPR, they were able to show that while some of the reduction came from diluting the high ST molecules in the base fuel, a large part of it came from a chemical effect due to the oxygenated moieties. The methyl-terminated POMEs discussed in Section 1.6 are an example of oxygenated HCs that have ST ≈ 0 [45, 46]. Another interesting case is the addition of hydroxy (–OH) groups to benzene [131, 139]: one group reduces the YSI from 100

(benzene) to 81 (phenol), two groups can reduce it to 34 (1,3-dihydroxybenzene), and three groups can reduce it to 10 (1,3,5-trihydroxybenzene)—a tenfold decrease.

However, the chemical effect of O is complex, and it can increase ST in some cases. A good example is esters [110]. Methyl pentanoate ($C_6H_{12}O_2$) has a lower YSI than *n*-hexane (C_6H_{14}), 22 vs. 30, but its isomer butyl acetate has a higher YSI than n-hexane, 36 vs. 30. A likely reason is that butyl acetate has four Cs next to the ether O, so it can undergo a six-centered H transfer [102]:



This reaction promotes soot formation since the first product (acetic acid) uses the O atoms inefficiently, and the second product is a large unsaturated species (1-butene). Methyl pentanoate cannot undergo this reaction since it only has one C next to the ether O.

To provide a global view of the effects of O, Figs. *7* and *8* above use different data series for pure HCs and HCs containing O or, in a few cases, N. On average, the YSIs for HCs with heteroatoms are smaller than those for pure HCs, but the two categories overlap significantly. For example, for DBE = 0, the YSIs of the pure HCs range from 25 to 133, while the YSIs of the heteroatom HCs range from 0.5 (2,4,6,8-tetraoxononane) to 73 (4-hexoxyheptane); for DBE = 1, the pure HCs range from 39 to 158, while the heteroatom HCs range from 1 (isopropyl nitrate) to 107 (isopentyl 2,2-bis(isopentyloxy)acetate); and for DBE = 2, the pure HCs range from 46 to 161, while the heteroatom HCs range from 1 (acetonitrile) to 137 (methyl oleate).

3.3 Effects of isomers

Isomers are a rich area for tuning fuel properties and for learning about combustion chemistry [28, 140, 141]. They contain the same set of atoms but connected in different ways, which often causes very specific changes in their properties and reactions. In this section we discuss various levels of isomerism and their observed effect on ST.

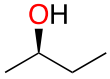
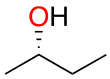
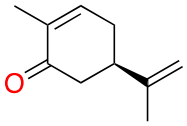
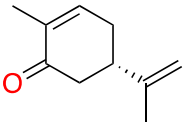
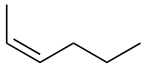
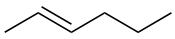
Chiral isomers have a "handedness" that cannot be interconverted by rotation or translation. Also called enantiomers, they are among the subtlest form of isomers that can be stably separated at room temperature. HCs will display chirality if one of their carbon atoms is attached to four distinct ligands. A classic example is 2-butanol, where the #2 C is

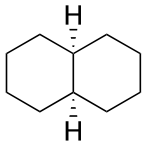
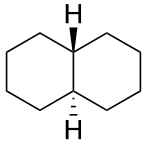
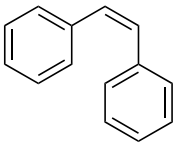
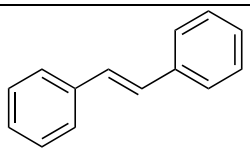
attached to –H, –OH, $–CH_3$, and $–C_2H_5$; Table *4* shows the two enantiomers. Chirality can strongly affect biological processes since enzymes often contain a three-dimensional "lock" that can only accept a specific enantiomer as the "key". However, it is unlikely to affect the high-temperature abstraction and dissociation reactions involved in soot formation. The YSI method allows this hypothesis to be directly tested since it needs only a small sample volume. Table *4* shows two cases of YSIs measured for purified enantiomers. The results show that, as expected, chirality has a negligible impact on ST.

Geometric isomerism occurs when the orientation of ligands is restricted by a structural feature that does not allow rotation such as a double C-C bond or a fused bicycle. For example, in any 1,2-substituted ethylene there is a *cis* isomer with the two substituents on the same side of the plane containing the π-bond, and a *trans* isomer with them on opposite sides of the plane. Table *4* shows one case of SPs and two cases of YSIs measured for purified *cis* and *trans* isomers. The results show that geometric isomerism, like chirality, does not affect ST. These conclusions greatly simplify the formulation of structure-property relationships for ST.

Table 4

Measured STs of optical isomer pairs.

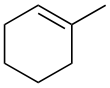
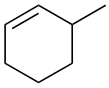
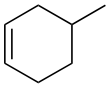
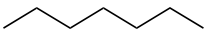
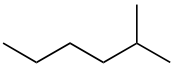
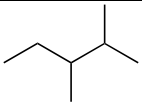
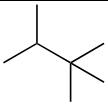
$C_4H_{10}O$ Enantiomers [110]	
	
I-2-butanol	(*S*)-2-butanol
YSI = 25.2	YSI = 25.3
$C_{10}H_{14}O$ Enantiomers [129]	
	
I-carvone	(*S*)-carvone
YSI = 127.7	YSI = 126.4
C_6H_{12} Geometric Isomers [110]	
	
cis-2-hexene	*trans*-2-hexene
YSI = 44.7	YSI = 45.8
$C_{10}H_{18}$ Geometric Isomers [74]	

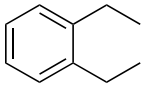
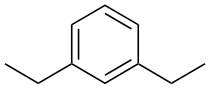
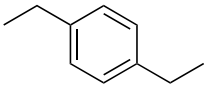
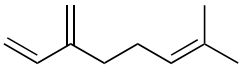
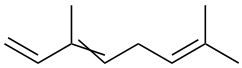
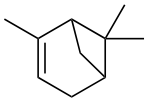
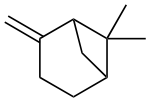
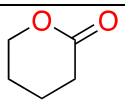
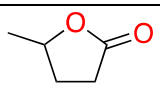
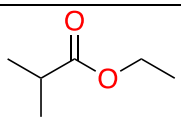
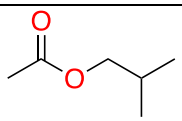
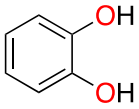
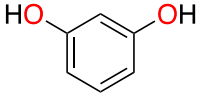
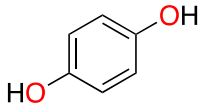
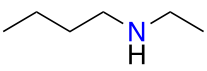
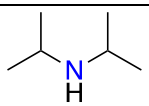
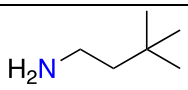
SPs measured for 5 wt% added to a 65 volume% *n*-heptane/35 volume% toluene mixture	
	
cis-decalin	*trans*-decalin
SP = 21.3 mm	SP = 21.3 mm
$C_{14}H_{12}$ Geometric Isomers [110]	
	
cis-1,2-diphenylethylene	*trans*-1,2-diphenylethylene
YSI = 602.0	YSI = 602.0

Structural isomers are sets of molecules that have the same atomic formula, but the atoms are connected in different patterns. The earliest SP measurements showed that these isomers can have large differences in ST, especially for branching in alkanes [57, 58]. More recent measurements have identified many additional cases. Table 5 lists several examples.

Table 5

Structural isomers with large differences in their measured ST.

C_7H_{12} Isomers [110]		
		
2-methyl-1-cyclohexene	3-methyl-1-cyclohexene	4-methyl-1-cyclohexene
YSI = 62	YSI = 85	YSI = 61
C_7H_{16} Isomers [76, 110]		
		
n-heptane	2-methyl-hexane	2,3-dimethyl-pentane
NSP = 139	NSP = 119	NSP = 107
YSI = 36	YSI = 42	YSI = 49
		
		2,2,3-trimethylbutane
		YSI = 55
$C_{10}H_{14}$ Isomers [110]		

		
1,2-diethylbenene	1,3-diethylbenzene	1,4-diethylbenzene
YSI = 376	YSI = 321	YSI = 271
$C_{10}H_{16}$ Isomers [129]		
			
myrcene	β-ocimene	α-pinene	β-pinene
YSI = 104	YSI = 214	YSI = 207	YSI = 137
$C_5H_8O_2$ Isomers [132]			
			
delta-valerolactone		gamma-valerolactone	
YSI = 21		YSI = 35	
$C_6H_{12}O_2$ Isomers [110]			
			
ethyl isobutyrate		isobutyl acetate	
YSI = 26		YSI = 35	
$C_6H_6O_2$ Isomers [131]			
		
catechol	resorcinol	hydroquinone
YSI = 57	YSI = 34	YSI = 38
$C_6H_{15}N$ Isomers [120]		
		
N-ethyl-*n*-butylamine	diisopropylamine	3,3-dimethylbutylamine
YSI = 22	YSI = 32	YSI = 41

4 Structure-property relationships for sooting tendency

QSPRs and machine learning (ML) techniques have been used extensively in chemistry [142], and they are increasing being used for combustion science [143]. This section reviews their application to predicting ST. As noted above in Section 1.6, QSPRs that can accurately predict soot for a wide range of compounds is essential for fuel design studies.

Historically, one of the most effective ways to tie the effects of different chemical moieties to chemical behavior has been utilizing group contribution methods (GCMs). Here, the effects of distinct functional groups and chemical structures on the properties of interest are parametrized independently; then, the value of the property for the compound of interest can be calculated by summing contributions from the compound backbone and the individual functional group constituents [144]. By applying ensemble statistical techniques, substituent effects can be scaled across different chemical families [145], offering a computationally cheap method to predict properties of interest rapidly. However, the generalizability of such linearly additive methods is often lacking, requiring increasingly unwieldy parametric forms to describe diverse groups of compounds across chemical space, if possible.

Modern machine learning technologies present an attractive alternative: methods like multivariate linear regression (MLR), support vector machines (SVM), Gaussian process regression (GPR), and decision tree methods can achieve high accuracies when predicting chemical properties. Further, recent advances in neural networks have leveraged the abilities of these models to describe highly nonlinear phenomena, achieving accuracies rivaling experimental uncertainty when modeling chemical systems. However, this comes at the tradeoff of dataset size; many of these models require at least hundreds of high-fidelity training data points to produce accurate and generalizable results.

The first step in building a machine learning model to predict chemical properties is to assemble a dataset and determine how to describe the chemical information best numerically to feed it as inputs to the model. Due to the complex nature of the soot formation process, computational techniques to quantitatively simulate soot formation with detailed kinetic mechanisms are either too demanding, inaccurate, or unfeasible. Consequently, experimental datasets remain the gold standard for sooting tendency predictions. However, the unique difficulties of standardizing experimental procedures across different experimental setups used in measuring sooting behavior often limit the size of available experimental datasets. Figure 9 shows that the experimental datasets used

for ST predictions have increased in size over the last 36 years, but remain the order of the hundreds

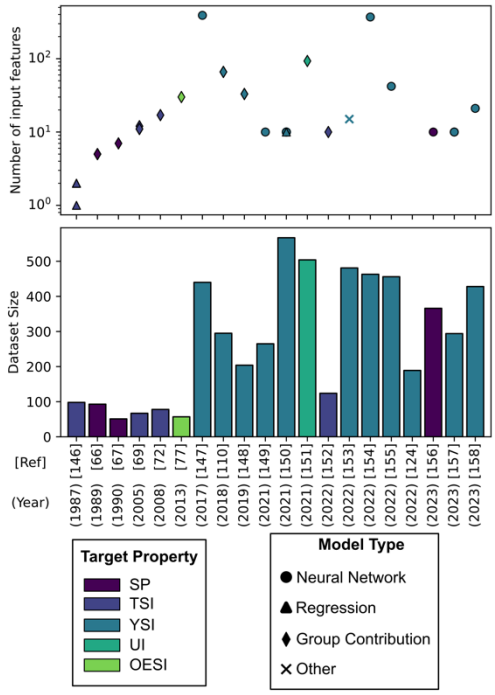


Figure 9. Reported dataset sizes and input features of works between 1987 and 2023 aiming to predict the sooting tendency of pure compounds and/or mixtures. For representation purposes, models shown as Group Contribution correspond to non-Neural Network models with most descriptors corresponding to functional group/atom/bond counts.

Once a suitable database of prediction target measurements has been established, the next step is to determine a suitable molecular representation. A commonly used technique relies on generating molecular descriptors, a series of calculated properties that describe molecular structure. These descriptors are widely classified into 0D, 1D, 2D, 3D, and 4D descriptors based on the geometric properties they exploit; common examples of such descriptors can be found in Table 6. The higher the descriptor dimension, the more computationally costly it is to calculate.

Table 6

Descriptor classification based on dimensionality.

Descriptor Type	Description	Example
0D	Aggregate properties; no structural information	Relative molecular mass; H bond count
1D	Based on the presence/absence/count of substructures, such as functional groups	Molecular fingerprints (Extended connectivity fingerprint, Morgan fingerprint, etc.)
2D	Molecular connectivity based	adjacency matrix, Coulomb matrix, distance matrix
3D	Based on 3D molecular geometry	3D MoRSE descriptors, sterimol
4D	based on dividing 3D space around molecules into discrete grids	Fukui function, CoMFA, GRID

Another means of representing molecules utilizes the inherent atomic connectivity described by bond information as a scaffold on which information is propagated; such models use the molecular graph itself as inputs, with atoms represented as graph nodes and bonds represented as graph edges. This method allows the unique neighborhood environments of the constituent atoms to be well resolved and is well suited to be used as the input for a type of neural network known as a Graph Neural Network. Such networks have been used with great success in chemical property prediction applications [159, 160]. With a representation, it is then possible to construct a predictive model. With recent rapid advances in computing hardware technologies, various machine learning models have become easily accessible to computational researchers.

Historically, linear regression models have been used to build QSPR models to estimate chemical properties. These models are easily interpretable, as the significance of each input parameter in determining the model output is reflected in the scalar weight of the term in the regression formula: the larger the weight, the more impactful the input parameter. However, since the model consists of only linear combinations of input terms, linear regression models tend to perform poorly when used to describe nonlinear phenomena. Careful feature engineering is necessary to overcome this issue.

Thus, machine learning models that can inherently describe nonlinear phenomena have become popular in recent years. One popular class of nonlinear models utilizes decision trees to make predictions, in which a flowchart-like process is used to make decisions on learned threshold values of input descriptors. Some such models, like random forests [161] or XGBoost [162], use ensembles of such "trees" to reduce overfitting. Such decision tree models perform well with smaller dataset sizes (100 data points). A recent study used decision tree models to design a general framework for biofuel compound screening, predicting the probability that a given input compound has a property above/below a threshold value [163]. The research octane number (RON), TSI, and melting point (MP) are selected as target properties; the model achieves test set predictive accuracies of 88%, 87%, and 94%, respectively, using a selection of 1D and 2D descriptors generated by the PaDEL-Descriptor software [164]. Since the parameter threshold values for generating predictions can be obtained from the trained model, an importance score can be assigned to each input descriptor, ensuring model interpretability.

Comesana et al. [153] deliberately uses this importance score to drive experimental design, studying the influence of chemical input descriptors on predictions of MP, boiling point (BP), flash point (FP), YSI, and net heat of combustion. The Tree-based Pipeline Optimization Tool (TPOT) [165] generates the decision tree architecture to avoid human bias in model selection. Underneath the TPOT software, a genetic algorithm-based selection is carried out to choose the optimal tree-based model for a given task. The authors found that the most important descriptors selected by the ML algorithms recover chemical intuition: for example, the most important descriptor to predict the YSI of a compound was the number of aromatic bonds. The atom bond connectivity (ABC) index, the second most important descriptor, is a measure of branching that correlates with stability for linear compounds and strain energy for cyclic compounds and is known to indicate thermodynamic stability [166].

Deep learning models using neural networks have also become popular for their ability to model complex nonlinear phenomena. In the chemical context, calculated descriptors for a given molecular structure are fed as inputs into the neural network and are propagated forward through the network, resulting in a target property prediction that is compared to a known value for the given structure. The error between the prediction and the known value is minimized by adjusting the network weights using a gradient descent procedure. The process is repeated for all structures in the training dataset, resulting in a trained neural network to make property predictions for an unseen molecular structure input. These models typically require large scale datasets of thousands or even millions of

data points to make successful predictions due to the large number of parameters that must be fit during the training process.

Recent work has used artificial neural networks (ANNs) to accurately predict ST from molecular structure descriptors. Several chemical moieties were selected, and their count in each molecule was used as a descriptor: Ahmed Qasem et al. [156] use the numbers of paraffinic CH_3, CH_2, CH, carbons, the number of olefinic CH–CH_2, naphthenic CH–CH_2, and aromatic C–CH bonds, and the number of alcohol and ether groups as inputs. The compound molecular weight and branching index are also used, resulting in 10 distinct model inputs. The authors train an ANN model on a curated dataset of 366 experimental smoke point measurements, achieving a model mean absolute error (MAE) of 4.5 units with an R^2 of 0.98. Alboqami et al. [157] use the same input descriptors to predict the YSI of 294 hydrocarbons and oxygenates curated from the work of Das et al. [110] with both an ANN and an adaptive network based fuzzy inference system (ANFIS) model. The ANFIS model incorporates fuzzy logic operations into the input structure of an ANN model, allowing a degree of ambiguity to be incorporated into the model inputs. It was found that the ANN model achieves a test set root mean square error (RMSE) of 15.83, with an R^2 of 0.99, while the ANFIS model achieves a test set RMSE of 58.78, with an R^2 of 0.92. Notably, the ANFIS model achieves a training set RMSE of 3.98, indicating that the ANFIS model cannot generalize training results to the unseen test set data.

Chen et al. [158] expand the input descriptor space using an ANN model trained on 428 compounds in the dataset published by Das et al. [110] to predict the YSI of species relevant to gasoline spark ignition engines. In addition to the descriptors used by Ahmed Qasem et al., [156] inputs representing frequencies of halogens, ketones, cyanides, esters, carboxyls, and aldehydes were added, resulting in a total of 21 input descriptors. A model MAE of 19.20 with an R^2 of 0.98 was obtained. By studying the sensitivity of the model predictions to changes in the descriptors, it was found that the length of the carbon chain connecting benzene rings significantly impacts the YSI. In keeping with chemical intuition, the YSI was found to show a positive sensitivity on carbon-centered groups, while groups with oxygen atoms showed a negative impact on YSI.

In contrast, Kessler et al. [167] train an ANN model to predict the CN, YSI, and lower heating value (LHV) of 24 test terpene molecules using QSPR descriptors. A starting dataset comprising 460, 463, and 388 data points for each property was compiled from published experimental data, and 5305 QSPR descriptors were calculated for each molecular structure using the alvaDesc software package. [168]. Separate principal component analysis (PCA) transformations were utilized as a dimensionality reduction

technique to reduce the number of descriptors used as model inputs, resulting in 368, 370, and 310 input descriptors for each property model. Model accuracies of 3.25 CN units, 1.68 YSI units, and 0.26 MJ/kg were achieved, and the CN, YSI, and LHV of the 24 test terpenes predicted; two of the test terpenes, geranial and citronellal, showed promising predicted CN. However, the YSIs of all predicted terpenes were predicted to be higher than traditional terpenes. The correlation between the CN, YSI, and LHV was also studied; it was found that YSI and LHV may be correlated. Further work investigating this correlation is necessary to establish the degree of interdependence between these two quantities.

St. John et al. [147] use a similar descriptor selection technique to build a YSI predictor model from 297 experimentally measured values. 5270 molecular descriptors were initially generated using the Dragon7 software package [169]. A variance threshold for each descriptor was used to throw out descriptors that showed little change across dataset constituent molecules, leading to 2414 descriptors being removed. The remaining 2856 descriptors were then subject to a recursive feature selection procedure. Finally, the remaining 390 descriptors were inputted into an ANN model. A detailed functional group-based cross-validation strategy was performed to analyze model performance and generalizability, and model outliers were investigated via manual quantum mechanical calculations. The model achieved an MAE of 5.47 YSI units and was then applied to predict the YSI for compounds with a known RON; ethanol, 2-propanone, and ethyl acetate were identified as the best-performing oxygenate molecules by RON and YSI.

Notably, the ANN methods presented so far have been descriptor based. While these models can achieve respectable predictive accuracies, the descriptor inputs neglect to consider molecular connectivities. On the other hand, recently developed graph neural network (GNN) models can explicitly use this information in addition to atom-wise, bond-wise, and global molecular descriptors.

Kessler et al. [150] compare the performance of three different computational prediction models for the YSI: the predictive accuracies of multivariate linear models, ANNs, and GNNs were compared using a dataset of 567 fuel compounds. To train the ANN models, 5305 descriptors were obtained from the alvaDesc software, [168] and the feature importance score of a random forest model trained on the dataset YSI was used to perform feature selection, leading to 1800 final input descriptors. GNN models were trained using the atomic symbol, degree of bonding, and if the atom exists within a ring as atom parameterizations, and discrete vectors for bond order (single, double, triple, or aromatic) as bond parameterizations. Following Jørgensen et al. [170], a message-passing scheme was implemented to allow information about distinct atomic environments to be exchanged

across the molecular structure. A multivariate linear model using the Levenberg-Marquardt method of least-squares regression on the piPC05, piPC04, and piPC03 descriptors was created as a baseline for comparison. The ANN and GNN methods achieve similar MAEs of 4.34 and 4.82 YSI units on the test set; notably, both methods drastically outperform the multivariate linear model, which reaches a test set accuracy of 29.12 YSI units.

Kim et al. [171] utilize a GNN model to predict the CN of biofuel compounds. A database comprising experimental CNs measured using several different techniques was collected. To account for the differing experimental accuracies of the various testing methods, a data weighting strategy was implemented that assigns a lower sample weight to data points collected using methods with higher errors, thus reducing their contributions during model training. In addition to the atom and bond state descriptors used by Kessler et al., [150] hydrogen bond donor/acceptor pairs were used as global features. The resultant model achieves a predictive accuracy (MAE) of 2.44 CN units. Notably, a parallel model omitting the global features was trained on the same dataset; it achieves an accuracy of 4.29 CN units, highlighting the importance of the chemically informed hydrogen bond features in ensuring predictive accuracy.

The models discussed so far have excelled at predicting the sooting tendencies of individual compounds; to state the obvious, however, most commercially utilized fuels are not single compounds and are tailored mixtures in which each component molecule influences the sooting behavior. Thus, novel methodologies must be used to model such blending behavior properly – one such methodology is a mixing operator.

Kuzhagaliyeva et al. [40] introduce a hybrid ANN/recurrent neural network (RNN) model to predict the blending behavior of a mixture of RON, MON, and YSI using a similar mixing operator approach. A framework is introduced to generate novel fuel compounds using the predicted values in an inverse-design problem. A database of 813, 690, and 491 data points was curated for each property; due to the scarcity of blend YSI measurements, a conversion formula was used to scale reported YSIs to an apparatus-independent value. A fingerprint representation of the target molecule is combined with calculated descriptors obtained from the MORDRED software package [172] to yield a total molecular representation; this per-molecule representation is then combined based on input concentrations using a mixing operator approach, and the result is scaled to account for differing YSI scales in the database. By exploiting the continuity of the neural network model's latent space, greedy and full-scope exploration strategies are employed to generate

new fuel blends with target YSIs; the authors present several fuel blends predicted to serve as viable fuel alternatives.

5 Numerical simulations of sooting tendencies

Numerical predictions of sooting tendencies have been carried out in numerous previous studies mainly for three purposes: (1) validation of chemical models and soot models, (2) design and validation of surrogate formulations to represent real fuels, and (3) screening of fuel additives to meet certain emission criteria. In this section, we review previous numerical predictions of sooting tendencies based on smoke point and soot yield. We also summarize recent efforts in predicting fuel sooting behaviors using reactive MD simulations.

5.1 Predictions of threshold based sooting tendencies

Previous numerical studies on smoke point based sooting tendencies of various fuels have focused mainly on two flame configurations, coflow diffusion flames and counterflow diffusion flames [30]. In coflow diffusion flames, the residence time is typically long [173], and once soot is formed, it will be convected and subsequently oxidized in downstream oxidation zones. Therefore, the smoke point in coflow diffusion flames is essentially where soot formation and oxidation are balanced. In contrast, in counterflow diffusion flames, smoke point is largely determined by soot formation without much interference from soot oxidation.

In counterflow nonpremixed flames, sooting tendencies of different fuels have typically been characterized in forms of sooting limit curves, where the flame transitions from non-sooting to sooting [174], as reviewed in [30]. In order to capture such sooting threshold, simulations typically require solving the full set of one-dimensional (1D) conservation equations with detailed chemical kinetics and soot formation models. Detailed simulations have been used to predict sooting limits for different fuels [174, 175, 176], oxygen levels in the oxidizer [135], exhaust gas recirculation levels [177], and strain rates [178]. These simulations can qualitatively predict how sooting limits change with varied operating conditions, but accurate quantitative sooting limit prediction is still a challenge [174, 175, 176, 177, 178, 179]. In addition, these predictions have not been used to directly calculate smoke point based sooting tendencies but have been used to infer mixing rules for TSIs of fuel mixtures [174].

Predicting smoke point or TSI data in coflow diffusion flames is computationally more expensive than in counterflow flames since a series of two-dimensional (2D) detailed flame simulations with varied flow rates are required. These simulations need to incorporate detailed species transport and finite-rate chemistry, with soot transport, growth, and oxidation models. For these reasons, to the authors' best knowledge, there is no TSI calculation reported in the literature that is directly from detailed numerical simulations. Instead, previous numerical studies on laminar coflow smoking flames have mainly focused on the predictions of smoke point height, for instance in flames fueled with ethylene, acetylene, propane, butadiene, butene, and diesel and bio-diesel surrogates [180, 181, 182, 183, 184, 185]. These studies have focused on investigating the effects of pressure [180], dilution [181], and fuel [180, 181, 182, 183, 184, 185] on the smoke point height and have primarily used these flames as modeling targets to validate empirical or phenomenological soot models specifically designed for smoke point prediction [180, 181, 182, 183, 184, 185].

5.2 Predictions of yield based sooting tendencies

The most direct way to predict YSI is to use two-dimensional detailed numerical simulations of the YSI flame doped with test fuels. These simulations have been demonstrated to be a reliable tool in reproducing axisymmetric laminar co-flow diffusion flames with various burner configurations and fuel compositions [134, 186, 187, 188, 189, 190, 191]. However, these simulations are computationally expensive, since all species transport equations need to be solved with detailed finite-rate chemistry, simultaneously with the Navier–Stokes equations in 2D. In addition, a different simulation needs to be carried out for the flame doped with each test hydrocarbon to predict the YSI of the test fuel [190, 191]. The factors make YSI predictions based on 2D detailed simulations expensive and inefficient for fast screening of fuel sooting properties.

Xuan and Blanquart [135] proposed a computationally-efficient, 1D flamelet-based YSI simulation framework by acknowledging that the YSI concept is a perturbation-based approach with nearly identical temperature and velocity fields in the YSI flames compared to those in the undoped methane/air flame. Therefore, only one detailed 2D simulation would be required for a single well-defined undoped flame, and the computations for the doped flames might be simplified to allow large kinetic mechanisms to be used without mechanism reduction. In contrast, modeling smoke point height or TSI data would require simulating a series of pure-fueled flames with different fuel flowrates and different computational grids to determine the threshold where soot is emitted from the flame.

In the flamelet-based framework [135], modified 1D flamelet equations were specially derived on the centerline of the YSI flames to account for the effects of multi-dimensional flow and differential diffusion on soot precursors. It took as input the temperature, convective velocity, and scalar dissipation rate profiles extracted from the direct simulation of the undoped flame. Soot transport model was excluded in these flamelet-based calculations for both simplicity and to be unbiased from the choice of soot model used. Sooting tendencies were estimated exclusively from the increment of polycyclic aromatic hydrocarbon (PAH) dimer production rate along the centerline where the flame was doped.

Xuan and Blanquart [135] performed the first set of chemical kinetic-based YSI calculations using this framework for a series of non-aromatic and aromatic test fuels and achieved good agreement with measurements. Thereafter, this framework has been used to predict the YSI of various gasoline surrogates [114], validate chemical kinetic models for prediction soot formation from bio-derived fuel additives for spark-ignition engines [37], and investigate the effects of elevated pressures on YSI for *n*-alkanes and aromatic fuels [136]. This framework was originally implemented in the FlameMaster software [192], and later combined with a more computationally efficient solver in Zero-RK [193], an open-source software package that simulates chemically reacting systems, hosted by the Lawrence Livermore National Laboratory [194, 195]. Overall, the flamelet-based YSI solver is efficient, which enables fast sooting tendency predictions using large, more accurate chemical kinetic mechanisms [193]. It also enables sensitivity analysis and uncertainty quantification which can help identify reactions particularly important for sooting tendency predictions [37, 193]. The main limitation of this approach is that the YSI prediction can only be performed for fuels with known chemical kinetics, which renders this approach inapplicable to many bio-derived fuels with unknown decomposition chemistry.

5.3 Application of MD simulations in yield based sooting tendency predictions

Since the combustion chemistry of bio-synthesized sustainable fuels for ground transportation and aviation is typically unknown, ReaxFF-based reactive Molecular Dynamics (MD) simulations have been recently employed for YSI predictions, and more generally to analyze fuel effects on soot formation. ReaxFF-based MD is commonly used to discover reaction networks in complex and large chemical systems over time scales longer than quantum mechanics (QM) [196, 197], since they use empirical force fields trained against QM-based data and therefore require much lower computational cost. Recent applications of the ReaxFF-based MD method in the broader combustion- and energy-

related fields are reviewed in [198]. ReaxFF-based MD simulations have been applied successfully to describe pyrolysis and oxidation processes for hydrocarbons, oxygenated, and nitrogenated hydrocarbons [199, 200, 201, 202, 203, 204], and soot formation [204, 205, 206]. A summary of recent studies using ReaxFF MD for YSI predictions and for quantifying fuel effects on soot formation is provided in Table 7. These studies investigated a wide spectrum of fuels ranging from hydrogen, methane, to gasoline/diesel fuel surrogates and bio-derived oxygenated fuels. Due to the limitations of simulation duration, they typically relied on the yield of premature soot nanostructures, PAH, or highly unsaturated non-aromatic soot precursors to compute sooting tendencies or to compare sooting behaviors of fuels and fuel mixtures. Although the ReaxFF MD simulations are typically carried out at very high density and pressure to accelerate the reaction dynamics, the simulation results were generally shown to qualitatively agree with available measurements, and the reaction pathways and chemical kinetic information extracted from these ReaxFF simulations can be used as starting point for further chemical kinetic investigations using higher fidelity methods, such as QM calculations based on density function theory.

Table 7

Summary of recent ReaxFF MD applications for YSI predictions and for fuels effect quantification on soot formation.

Publication	Fuels/Fuel Mixtures	Soot Indicator	Major Outcome
Zhang et al. [207]	Ethylene/ammonia mixtures	Large PAH	Quantified the effects of fuel ammonia content on soot suppression
Kwon et al. [208]	Bio-derived polycyclic alkanes	Non-aromatic soot precursors	Examined the effects of ring fusion on fuel soot propensity
Kwon et al. [209]	Bio-derived dioxolanes	Non-aromatic soot precursors	Examined side chain effects on soot propensity
Zhang et al. [210]	Butanol and butane isomers	PAH	Examined fuel molecular structure effects on soot propensity

Wang et al. [211]	Methane and ethylene with hydrogen	PAH	Examined the effects of hydrogen addition on PAH and soot formation
Kwon et al. [212]	Hexylamine isomers	Non-aromatic soot precursors	Examined fuel molecular structure effects on soot propensity
Chen et al. [213]	Diesel surrogates and oxygenated additives	PAH and nascent soot particles	Analyzed pyrolysis pathways and soot propensity of diesel fuels
Kwon et al. [139]	Toluene and phenol	PAH	Performed the first set of YSI calculations using ReaxFF

6 Conclusions and future directions

The tendency to produce soot during combustion of a fuel depends strongly on its molecular structure. Therefore, as we transition towards more sustainable fuels we have the opportunity to tailor their compositions to improve combustion performance and reduce emissions of soot. Sooting tendencies measured in laboratory-scale flames provide a scientific database or selecting low soot fuels. However, large databases require a systematic approach to arrive at quantitative structure property relationships (QSPR) that allow data to be extrapolated to a wide variety of fuel structures not measured. This is essential for tuning the synthesis strategies for these new fuels considering the large range of possibilities. Recently, a wide range of machine learning algorithms are being applied to relate fuel structures to desired combustion properties. These can apply of host of different statistical methods but the accuracy and predictive ability of these techniques depends strongly on the coverage of compounds and structural features in the database used both to develop and test the model. These factors suggest that future research should be directed at enlarging the databases of sooting tendencies. Compounds containing oxygen and nitrogen are underrepresented in the databased. Most oxygenated compounds reduce soot but not all. Nitrogen containing compounds can also reduce soot without harming other performance features but are not widely covered. In addition to statistically-based methods for developing QSPRs, STs can be used to test chemical mechanisms. Molecular dynamics simulations can provide useful insights into ST at a

lower computational cost than complex chemical models without requiring the same chemical reaction input.

Finally, we offer the following suggestions for future directions:

- Measure sooting tendencies for additional hydrocarbons.
- Use simulations to examine possible confounding effects on ST measurement due to other fuel properties such as mass diffusivity.
- Develop predictive emissions indices analogous to PMI for other systems such as diesel engines and gas turbines.
- Improve QSPRs, especially for categories like aromatic hydrocarbons and larger oxygenated hydrocarbons.
- Use kinetic and MD simulations to explain measured YSIs; for example, the isomer effects in Table 5.
- Use simulations to explain why QSPR relationships work for ST.

Acknowledgements

The authors' work on sooting tendencies has been supported by the U.S. National Science Foundation, U.S. Department of Energy (DOE) Vehicle Technologies Office, and the U.S. DOE Bioenergy Technologies Office. We would also like to thank the many researchers who have worked with us to provide fuel samples and make the YSI methodology useful and accessible, including those at the China Lake Naval Air Weapons Station, Los Alamos National Laboratory, the National Renewable Energy Laboratory, Oak Ridge National Laboratory, Pacific Northwest National Laboratory, and Sandia National Laboratory.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplemental Information

The YSI database discussed in Section 2.2 is available as a comma-separated-values file ("ysi_comma_separated.csv") and a semicolon-separated values file

("ysi_semicolon_separated.txt"). Both files identify the entries by name and by canonical SMILES string—the latter provides a definitive means for searching for specific compounds that is not subject to the ambiguity of chemical names.

References

[1] International Energy Agency, https://www.iea.org/data-and-statistics/data-tools/energy-statistics-data-browser?country=WORLD&fuel=Energy%20supply&indicator=TESbySource, accessed May 31, 2024.

[2] W. Leitner, J. Klankenmayer, S. Pischinger, H. Pitsch, K. Kohse-Höinghaus, Advanced biofuels and beyond: chemistry solutions for propulsion and production, Angew. Chem., Int. Ed. 56 (2017) 5412–5452.

[3] J.P. Szybist, S. Busch, R.L. McCormick, J.A. Pihl, D.A. Splitter, M.A. Ratcliff, C.P. Kolodziej, J.M.E. Storey, M. Moses-DeBusk, D. Vuilleumier, M. Sjöberg, C.S. Sluder, T. Rockstroh, P. Miles, What fuel properties enable higher thermal efficiency in spark-ignited engines?, Prog. Energy Combust. Sci. 82 (2020) 100876.

[4] S. Szopa, V. Naik, B. Adhikary, P. Artaxo, T. Berntsen, W.D. Collins, S. Fuzzi, L. Gallardo, A. Kiendler-Scharr, Z. Klimont, H. Liao, N. Unger, and P. Zanis, 2021: Short-Lived Climate Forcers. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 817–922.

[5] P. Forster, T. Storelvmo, K. Armour, W. Collins, J.-L. Dufresne, D. Frame, D.J. Lunt, T. Mauritsen, M.D. Palmer, M. Watanabe, M. Wild, and H. Zhang, 2021: The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 923–1054.

[6] B. Kärcher, Formation and radiative forcing of contrail cirrus, Nat. Commun. 9 (2018) 1824.

[7] R. Teoh, U. Schumann, E. Gryspeerdt, M. Shapiro, J. Molloy, G. Koudis, C. Voigt, M.E.J. Stettler, Aviation contrail climate effects in the North Atlantic from 2016 to 2021, Atmos. Chem. Phys. 22 (2022) 10919–10935.

[8] C. Voigt, J. Kleine, D. Sauer, R.H. Moore, T. Bräuer, P. Le Clercq, S. Kaufmann, M. Scheibe, T. Jurkat-Witschas, M. Aigner, U. Bauder, Y. Boose, S. Borrmann, E. Crosbie, G.S. Diskin, J. DiGangi, V. Hahn, C. Heckl, F. Huber, J.B. Nowak, M. Rapp, B. Rauch, C. Robinson, T. Schripp, M. Shook, E. Winstead, L. Ziemba, H. Schlager, B.E. Anderson, Cleaner burning aviation fuels can reduce contrail cloudiness, Communications Earth & Environment 2 (2021) 114.

[9] R. Teoh, U. Schumann, C. Voigt, T. Schripp, M. Shapiro, Z. Engberg, J. Molloy, G. Koudis, M.E.J. Stettler, Targeted use of sustainable aviation fuel to maximize climate benefits, Environ. Sci. Technol. 56 (2022) 17246–17255.

[10] R.A. Dobbins, C.M. Megaridis, Morphology of flame-generated soot as determined by thermophoretic sampling, Langmuir 3 (1987) 254–259.

[11] Ü.Ö. Köylü, G.M. Faeth, T.L. Farias, M.G. Carvalho, Fractal and projected structure properties of soot aggregates, Combust. Flame 100 (1995) 621–633.

[12] D.W. Dockery, C.A. Pope III, X. Xu, J.D. Spengler, J.H. Ware, M.E. Fay, B.G Ferris, Jr., F.E. Speizer, An association between air pollution and mortality in six U.S. cities, N. Engl. J. Med. 329 (1993) 1753–1759.

[13] A. Vodonos, Y. Abu Awad, J. Schwartz, The concentration-response between long-term PM2.5 exposure and mortality; a meta-regression approach, Environ. Res. 166 (2018) 677–689.

[14] C.A. Pope III, J.S. Lefler, M. Ezzati, J.D. Higbee, J.D. Marshall, S.-Y. Kim, M. Bechle, K.S. Gilliat, S.E. Vernon, A.L. Robinson, R.T. Burnett, Mortality risk and fine particulate air pollution in a large, representative cohort of U.S. adults, Environ. Health Perspect. 127 (2019) 077007.

[15] G.A. Roth, G.A. Mensah, V. Fuster, The global burden of cardiovascular diseases and risks: A compass for global action, J. Am. Coll. Cardiol. 76 (2020) 2980–2981.

[16] M.R. Miller, D.E. Newby, Air pollution and cardiovascular disease: car sick, Cardiovasc. Res. 116 (2020) 279–294.

[17] WHO global air quality guidelines. Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva: World Health Organization; 2021.

[18] United States, Environmental Protection Agency. “Reconsideration of the National Ambient Air Quality Standards for Particulate Matter.” 88 Fed. Reg. 5558–5719 (Jan. 27, 2023).

[19] United States Environmental Protection Agency. “Reconsideration of the National Ambient Air Quality Standards for Particulate Matter.” 89 Fed. Reg. 16202–16406 (March 6, 2024).

[20] M.S. Hammer, A. van Donkelaar, C. Li, A. Lyapustin, A.M. Sayer, N.C. Hsu, R.C. Levy, M.J. Garay, O.V. Kalashnikova, R.A. Kahn, M. Brauer, J.S. Apte, D.K. Henze, L. Zhang, Q. Zhang, B. Ford, J.R. Pierce, R.V. Martin, Global estimates and long-term trends of fine particulate matter concentrations (1998–2018), Environ. Sci. Technol. 54 (2020) 7879–7890.

[21] GBD 2019 Risk Factor Collaborators, Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019, Lancet 396 (2020) 1223–1249.

[22] R. Burnett, H. Chen, M. Szyszkowicz, N. Fann, B. Hubbell, C.A. Pope III, J.S. Apte, M. Brauer, A. Cohen, S. Weichenthal, J. Coggins, Q. Di, B. Brunekreef, J. Frostad, S.S. Lim, H. Kan, K.D. Walker, G.D. Thurston, R.B. Hayes, C.C. Lim, M.C. Turner, M. Jerrett, D. Krewski, S.M. Gapstur, W.R. Diver, B. Ostro, D. Goldberg, D.L. Crouse, R.V. Martin, P. Peters, L. Pinault, M. Tjepkema, A. van Donkelaar, P.J. Villeneuve, A.B. Miller, P. Yin, M. Zhou, L. Wang, N.A.H. Janssen, M. Marra, R.W. Atkinson, H. Tsang, T.Q. Thach, J.B. Cannon, R.T. Allen, J.E. Hart, F. Laden, G. Cesaroni, F. Forastiere, G. Weinmayr, A. Jaensch, G. Nagel, H. Concin, J.V. Spadaro, Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) 9592–9597.

[23] E.E. McDuffie, R.V. Martin, J.V. Spadaro, R. Burnett, S.J. Smith, P. O’Rourke, M.S. Hammer, A. van Donkelaar, L. Bindle, V. Shah, L. Jaeglé, G. Luo, F. Yu, J.A. Adeniran, J. Lin, M. Brauer, Source sector and fuel contributions to ambient PM2.5 and attributable mortality across multiple spatial scales, Nat. Commun. 12 (2021) 1–12.

[24] K. Vohra, A. Vodonos, J. Schwartz, E.A. Marais, M.P. Sulprizio, L.J. Mickley, Global mortality from outdoor fine particle pollution generated by fossil fuel combustion: results from GEOS-Chem, Environ. Res. 195 (2021) 110754.

[25] Y. Li, D.K. Henze, D. Jack, B.H. Henderson, P.L. Kinney, Assessing public health buden associated with exposure to ambient black carbon in the United States, Sci. Total Environ. 539 (2016) 515–525.

[26] Y. Wang, X. Li, Z. Shi, L. Huang, J. Li, H. Zhang, Q. Ying, M. Wang, D. Ding, X. Zhang, J. Hu, Premature mortality associated with exposure to outdoor black carbon and its sourc contributions in China, Resour., Conserv. Recycl. 170 (2021) 105620.

[27] I. Glassman, P. Yaccarino, The temperature effect in sooting diffusion flames, Proc. Combust. Inst., 18 (1981) 1175–1183.

[28] S.M. Sarathy, P. Oßwald, N. Hansen, K. Kohse-Höinghaus, Alcohol combustion chemistry, Prog. Energy Combust. Sci. 44 (2014) 40–102.

[29] C. Saggese, A. Frassoldati, A. Cuoci, T. Faravelli, E. Ranzi, A wide range kinetic modeling study of pyrolysis and oxidation of benzene, Combust. Flame 160 (2013) 1168–1190.

[30] Y. Wang, S.H. Chung, Soot formation in laminar counterflow flames, Prog. Energy Combust. Sci. 74 (2019) 152–238.

[31] J.W. Martin, M. Salamanca, M. Kraft, Soot inception: carbonaceous nanoparticle formation in flames, Prog. Energy Combust. Sci. 88 (2022) 100956.

[32] K. Aikawa, T. Sakurai, J.J. Jetter, Development of a predictive model for gasoline vehicle particulate matter emissions, SAE Int. J. Fuels Lubr. 3 (2010) 610–622.

[33] I.M. Aksit, J.B. Moss, Model fuels to reproduce the sooting behavior of aviation kerosene, Fuel 84 (2005) 239–245.

[34] A. Mensch, R.J. Santoro, T.A. Litzinger, S.-Y. Lee, Sooting characteristics of surrogates for jet fuels, Combust. Flame 157 (2010) 1097–1105.

[35] X. Huo, N.A. Huq, J. Stunkel, N.S. Cleveland, A.K. Starace, A.E. Settle, A.M. York, R.S. Nelson, D.G. Brandner, L. Fouts, P.C. St. John, E.D. Christensen, J. Luecke, J.H. Mack, C.S. McEnally, P.A. Cherry, L.D. Pfefferle, T.J. Strathmann, D. Salvachúa, S. Kim, R.L. McCormick, G.T. Beckham, D.R Vardon, Tailoring diesel bioblendstock from integrated catalytic upgrading of carboxylic acids: a “fuel property first” approach, Green Chem. 21 (2019) 5813–5827.

[36] N.A. Huq, X. Huo, G.R. Hafenstine, S.M. Tifft, J. Stunkel, E.D. Christensen, G.M. Fioroni, L. Fouts, R.L. McCormick, P.A. Cherry, C.S. McEnally, L.D. Pfefferle, M.R. Wiatrowski, P.T. Benavides, M.J. Biddy, R.M. Connatser, M.D. Kass, T.L. Alleman, P.C. St. John, S. Kim, D.R. Vardon, Performance-advantaged ether diesel bioblendstock production by a priori design, Proc. Natl. Acad. Sci. U. S. A. 116 (2019) 26421–26430.

[37] H. Kwon, S. Lapointe, K. Zhang, S.W. Wagnon, W.J. Pitz, J. Zhu, C.S. McEnally, L.D. Pfefferle, Y. Xuan, Sooting tendencies of 20 bio-derived fuels for advanced spark-ignition engines, Fuel, 276 (2020) 118059.

[38] D.L. Bartholet, M.A. Arellano-Treviño, F. Liang Chan, S. Lucas, J. Zhu, P.C. St. John, T.L. Alleman, C.S. McEnally, L.D. Pfefferle, D.A. Ruddy, B. Windom, T.D. Foust, K.F. Reardon,

Property predictions demonstrate that structural diversity can improve the performance of polyoxymethylene ethers as potential bio-based diesel fuels, Fuel 295 (2021) 120509.
[39] G.-Q. Cai, L.-Z. Zhang, Systematic diesel molecular performance evaluation based on quantitative structure-property relationship model, Pet. Sci. 19 (2022) 809–818.
[40] N. Kuzhagaliyeva, S. Horváth, J. Williams, A. Nicolle, S.M. Sarathy, Artificial intelligence-driven design of fuel mixtures, Commun. Chem. 5 (2022) 1–10.
[41] R. Li, J.M. Herreros, A. Tsolakis, W. Yang, Integrated machine learning-quantitative structure property relationship (ML-QSPR) and chemical kinetics for high throughput fuel screening toward internal combustion engine, Fuel 307 (2022) 121908.
[42] L. Fleitmann, P. Ackermann, J. Schilling, J. Kleinekorte, J.G. Rittig, F. vom Lehn, A.M. Schweidtmann, H. Pitsch, K. Leonhard, A. Mitsos, A. Barlow, M. Dahmen, Molecular design of fuels for maximum spark-ignition engine efficiency by combining predictive thermodynamics and machine learning, Energy Fuels 37 (2023) 2213–2229.
[43] K. Kohse-Höinghaus, Combustion, Chemistry, and Carbon Neutrality, Chem. Rev. 123 (2023) 5139–5219.
[44] H. Pitsch, D. Goeb, L. Cai, W. Willems, Potential of oxymethylene ethers as renewable diesel substitute, Prog. Energy Combust. Sci. 104 (2024) 101173.
[45] Y.R. Tan, M.L. Botero, Y. Sheng, J.A.H. Dreyer, R. Xu, W. Yang, M. Kraft, Sooting characteristics of polyoxymethylene dimethyl ether blends with diesel in a diffusion flame, Fuel 224 (2018) 499–506.
[46] S.P. Lucas, F. Liang Chan, G.M. Fioroni, T.D. Foust, A. Gilbert, J. Luecke, C.S. McEnally, J.J.A. Serdoncillo, A.J. Zdanowicz, J. Zhu, B. Windom, Fuel properties of oxymethylene ethers with terminating groups from methyl to butyl, Energy Fuels 36 (2022) 10213–10225.
[47] J. Liu, L. Wang, P. Wang, P. Sun, H. Liu, Z. Meng, L. Zhang, H. Ma, An overview of polyoxymethylene dimethyl ethers as alternative fuel for compression ignition engines, Fuel 318 (2022) 123582.
[48] L. Pellegrini, M. Marchionna, R. Patrini, C. Beatrice, N. Del Giacomo, C. Guido, Combustion behavior and emissions performance of neat and blended polyoxymethylene dimethyl ethers in a light-duty diesel engine, SAE Technical Paper 2012-01-1053, 2012, https://doi.org/10.4271/2012-01-1053.
[49] M. Härtl, P. Seidenspinner, E. Jacob, G. Wachtmeister, Oxygenate screening on a heavy-duty diesel engine and emission characteristics of highly oxygenated oxymethylene ether fuel OME_1, Fuel 153 (2015) 328–335.
[50] J.L. Burton, J.A. Martin, G.M. Fioroni, T.L. Alleman, C.K. Hays, M.A. Ratcliff, M.R. Thorson, A.J. Schmidt, R.T. Hallen, T.R. Hart, J.M. Billing, S. Fox, D.J. Gaspar, J. Zhu, C.

Kima, L.D. Pfefferle, C.S. McEnally, R.L. McCormick, Fuel property effects of a broad range of potential biofuels on mixing control compression ignition engine performance and emissions, SAE Technical Paper 2021-01-0505, 2021, https://doi.org/10.4271/2021-01-0505.

[51] M.A. Arellano-Treviño, D. Bartholet, A. The To, A.W. Bartling, F.G. Baddour, T.L. Alleman, E.D. Christensen, G.M. Fioroni, C. Hays, J. Luecke, J. Zhu, C.S. McEnally, L.D. Pfefferle, K.F. Reardon, T.D. Foust, D.A. Ruddy, Synthesis of butyl-exchanged polyoxymethylene ethers as renewable diesel blendstocks with improved fuel properties, ACS Sustainable Chem. Eng. 9 (2021) 6266–6273.

[52] K. Alexandrino, J. Salinas, Á. Millera, R. Bilbao, M.U. Alzueta, Sooting propensity of dimethyl carbonate, soot reactivity and characterization, Fuel 183 (2016) 64–72.

[53] S. Barak, R.K. Rahman, S. Neupane, E. Ninnemann, F. Arahin, A. Laich, A.C. Terraciano, S.S. Vasu, Measuring the effectiveness of high-performance Co-Optima biofuels on suppressing soot formation at high temperature, Proc. Natl. Acad. Sci. U. S. A. 117 (2020) 3451–3460.

[54] 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, Combust. Flame 167 (2016) 308–319.

[55] ASTM International, ASTM D1322-23: Standard test method for smoke point of kerosene and aviation turbine fuel, ASTM International, West Conshohocken, PA, USA 2023.

[56] ASTM International, ASTM D7566-23b: Standard specification for aviation turbine fuel containing synthesized hydrocarbons, ASTM International, West Conshohocken, PA USA, 2023.

[57] R.A. Hunt, Relation of smoke point to molecular structure, Ind. Eng. Chem. 45 (1953) 602–606.

[58] R.L. Schalla, G.E. McDonald, Variation in smoking tendency among hydrocarbons of low molecular weight, Ind. Eng. Chem. 45 (1953) 1497–1500.

[59] W.S. Blazowski, Dependence of soot production on fuel structure in backmixed combustion, Combust. Sci. Technol. 21 (1980) 87–96.

[60] H.F. Calcote, D.M. Manos, Effect of molecular structure on incipient soot formation, Combust. Flame 49 (1983) 289–304.

[61] S.M. Senkan, J.M. Robinson, A.K. Gupta, Sooting limits of chlorinated hydrocarbon-methane-air premixed flames, Combust. Flame 49 (1983) 305–314.

[62] F. Takahashi, I. Glassman, Sooting correlations for premixed flames, Combust. Sci. Technol. 37 (1984) 1–19.

[63] R.J. Gill, D.B. Olson, Estimation of soot thresholds for fuel mixtures, Combust. Sci. Technol. 40 (1984) 307–315.
[64] A. Gomez, G. Sidebotham, I. Glassman, Sooting behavior in temperature-controlled laminar diffusion flames, Combust. Flame 58 (1984) 45–57.
[65] D.B. Olson, J.C. Pickens, R.J. Gill, The effects of molecular structure on soot formation II. Diffusion flames, Combust. Flame 62 (1985) 43–60.
[66] Ö.L. Gülder, B. Glavinčevski, S. Das, Effect of molecular structure on soot formation characteristics of aviation turbine fuels, J. Eng. Gas Turbine Power 111 (1989) 77–83.
[67] Ö.L. Gülder, B. Glavinčevski, M.F. Baksh, Fuel molecular structure and flame temperature effects on soot formation in gas turbine combustors, J. Eng. Gas Turbine Power 112 (1990) 52–59.
[68] N. Ladommatos, P. Rubenstein, P. Bennett, Some effects of molecular structure of single hydrocarbons on sooting tendency, Fuel 75 (1996) 114–124.
[69] 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 (2005) 2408–2415.
[70] Y. Yang, A.L. Boehman, R.J. Santoro, A study of jet fuel sooting tendency using the threshold sooting index (TSI) model, Combust. Flame 149 (2007) 191–205.
[71] T.L. Berry Yelverton, W.L. Roberts, Effect of dilution, pressure, and velocity on smoke point in laminar jet flames, Combust. Sci. Technol. 180 (2008) 1334–1346.
[72] P. Pepiot-Desjardins, H. Pitsch, R. Malhotra, S.R. Kirby, A.L Boehman, Structural group analysis for soot reduction tendency of oxygenated fuels, Combust. Flame 154 (2008) 191–205.
[73] K.M. Allan, J.R. Kaminski, J.C. Bertrand, J. Head, P.B. Sunderland, Laminar smoke points of wax candles, Combust. Sci. Technol. 181 (2009) 800–811.
[74] E.J. Barrientos, A.L. Boehman, Examination of the sooting tendency of three-ring aromatic hydrocarbons and their saturated counterparts, Energy Fuels 24 (2010) 3479–3487.
[75] K.T. Dotson, P.B. Sunderland, Z.-G. Yuan, D.L. Urban, Laminar smoke points of coflowing flames in microgravity, Fire Saf. J. 46 (2011) 550–555.
[76] L. Li, P.B. Sunderland, An improved method of smoke point normalization, Combust. Sci. Technol. 184 (2012) 829–841.
[77] E.J. Barrientos, M. Lapuerta, A.L. Boehman, Group additivity in soot formation for the example of C-5 oxygenated hydrocarbon fuels, Combust. Flame 160 (2013) 1484–1498.
[78] L. Li, P.B. Sunderland, Smoke points of fuel-fuel and fuel-inert mixtures, F. Saf. J. 61 (2013) 226–231.

[79] A. Llamas, M. Lapuerta, A.-M. Al-Lal, L. Canoira, Oyxgen extended sooting index of FAME blends with aviation kerosene, Energy Fuels 27 (2013) 6815–6822.

[80] R.J. Watson, M.L. Botero, C.J. Ness, N.M. Morgan, M. Kraft, An improved methodology for determining threshold sooting indices from smoke point lamps, Fuel 111 (2013) 120–130.

[81] Y. Wang, S.H. Chung, Effect of strain rate on sooting limits in counterflow diffusion flames of gaseous hydrocarbon fuels: sooting temperature index and sooting sensitivity index, Combust. Flame 161 (2014) 1224–1234.

[82] B. Graziano, T. Ottenwälder, D. Manderfeld, S. Pischinger, G. Grünefeld, Advanced methodology for the detection of smoke point heights in hydrocarbon flames, Energy Fuels 32 (2018) 3908–3919.

[83] G. Rubio-Gomez, L. Corral-Gomez, J.A. Soriano, A. Gomez, F.J. Castillo-Garcia, Vision based algorithm for automated determination of smoke point of diesel blends, Fuel 235 (2019) 595–602.

[84] C.H. Cho, K.R. Han, C.H. Sohn, F.M. Haas, Sooting propensity estimation of jet aviation fuel surrogates and their *n*-alkane components by the virtual smoke point method, Energy Fuels 34 (2020) 15072–15076.

[85] L. Corral-Gomez, D. Rodriguez-Rosa, S. Jaurez-Perez, A. Martin-Parra, G.R. Gomez, F. Moya-Fernandez, A novel device for automated determination of the smoke point with non-invasive adaptation of ASTM D1322 normalized lamps, Meas. Sci. Technol. 31 (2020) 115004.

[86] D. Donoso, R. Ballesteros, D. Bolonio, M.-J. García-Martínez, M. Lapuerta, L. Canoira, Hydrogenated turpentine: A biobased component for jet fuel, Energy Fuels 35 (2021) 1465–1475.

[87] Z. Li, P. Liu, C. Chu, S.H. Chung, W.L. Roberts, Incipient sooting tendency of oxygenated fuels doped in ethylene counterflow diffusion flames, Combust. Flame 244 (2022) 112284.

[88] Á. Muelas, S. Sanz, J. Ballester, Study of sooting propensity and related indices based on tests with isolated droplets, J. Energy Resour. Technol. 145 (2023) 071701.

[89] F.G. Roper, The prediction of laminar jet diffusion flame sizes: Part I. Theoretical model, Combust. Flame 29 (1977) 219–226.

[90] C.R. Shaddix, K.C. Smyth, Laser-induced incandescence measurements of soot production in steady and flickering methane, propane, and ethylene diffusion flames, Combust. Flame 107 (1996) 418–452.

[91] P. Desgroux, X. Mercier, K.A. Thomson, Study of the formation of soot and its precursors in flames using optical diagnostics, Proc. Combust. Inst. 34 (2013) 1713–1738.

[92] F. Goulay, P.E. Schrader, X. López-Yglesias, H.A. Michelsen, A data set for validation of models of laser-induced incandescence from soot: temporal profiles of LII signal and particle temperature, Appl. Phys. B 112 (2013) 287–306.

[93] R.J. Santoro, H.G. Semerjian, R.A. Dobbins, Soot particle measurements in diffusion flames, Combust. Flame 51 (1983) 203–218.

[94] G. Legros, Q. Wang, J. Bonnety, M. Kashif, C. Morin, J.-L. Consalvi, F. Liu, Simultaneous soot temperature and volume fraction measurements in axis-symmetric flames by a two-dimensional modulated absorption/emission technique, Combust. Flame 162 (2015) 2705–2719.

[95] P.B. Kuhn, B. Ma, B.C. Connelly, M.D. Smooke, M.B. Long, Soot and thin-filament pyrometry using a color digital camera, Proc. Combust. Inst. 33 (2011) 743–750.

[96] H. Guo, J.A. Castillo, P.B. Sunderland, Digital camera measurements of soot temperature and soot volume fraction in axisymmetric flames, Appl. Opt. 52 (2013) 8040–8047.

[97] J.A.H. Dreyer, R.I. Slavchov, E.J. Rees, J. Akroyd, M. Salamanca, S. Mosbach, M. Kraft, Improved methodology for performing the inverse Abel transform of flame images for color ratio pyrometry Appl. Opt. 58 (2019) 2662–2670.

[98] D.D. Hickstein, S.T. Gibson, R. Yurchak, D.D. Das, N. Ryazanov, A direct comparison of high-speed methods for the numerical Abel transform, Rev. Sci. Instrum. 90 (2019) 065115.

[99] C.S. McEnally, L.D. Pfefferle, Improved sooting tendency measurements for aromatic hydrocarbons and their implications for naphthalene formation pathways, Combust. Flame 148 (2007) 210–222.

[100] S.P. Crossley, W.E. Alvarez, D.E. Resasco, Novel micropyrolysis index (MPI) to estimate the sooting tendency of fuels, Energy Fuels 22 (2008) 2455–2464.

[101] C.S. McEnally, L.D. Pfefferle, Sooting tendencies of nonvolatile aromatic hydrocarbons, Proc. Combust. Inst. 32 (2009) 673–679.

[102] C.S. McEnally, L.D. Pfefferle, Sooting tendencies of oxygenated hydrocarbons in laboratory-scale flames, Env. Sci. Technol. 45 (2011) 2498–2503.

[103] M. Kashif, P. Guibert, J. Bonnety, G. Legros, Sooting tendencies of primary reference fuels in atmospheric laminar diffusion flames burning into vitiated air, Combust. Flame 161 (2014) 1575-1586.

[104] 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, Combust. Flame 162 (2015) 1840–1847.

[105] D.D. Das, C.S. McEnally, L.D. Pfefferle, Sooting tendencies of unsaturated esters in nonpremixed flames, Combust. Flame 162 (2015) 14891497.

[106] 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, Combust. Flame 162 (2015) 3140–3155.

[107] J. Abboud, J. Schobing, G. Legros, J. Bonnety, V. Tschamber, A. Brillard, G. Leyssens, V. Lauga, E.E. Iojoiu, P. Da Costa, Impacts of oxygenated compounds concentration on sooting propensities and soot oxidative reactivity: application to diesel and biodiesel surrogates, Fuel 193 (2017) 241–253.

[108] D.D. Das, C.S. McEnally, T.A. Kwan, J.B. Zimmerman, W.J. Cannella, C.J. Mueller, L.D. Pfefferle, Sooting tendencies of diesel fuels, jet fuels, and their surrogates in diffusion flames, Fuel 197 (2017) 445–458.

[109] J. Abboud, J. Schobing, G. Legros, A. Matynia, J. Bonnety, V. Tschamber, A. Brillard, G. Leyssens, P. Da Costa, Impacts of ester's carbon chain length and concentration on sooting propensities and soot oxidative reactivity: Applications to diesel and biodiesel surrogates, Fuel 222 (2018) 586–598.

[110] D.D. Das, P.C. St. John, C.S. McEnally, S. Kim, L.D. Pfefferle, Measuring and predicting sooting tendencies of oxygenates, alkanes, alkenes, cycloalkanes, and aromatics on a unified scale, Combust. Flame 190 (2018) 349–364.

[111] O. Staples, C.M. Moore, J.H. Leal, T.A. Semelsberger, C.S. McEnally, L.D. Pfefferle, A.D. Sutton, A simple, solvent free method for transforming bio-derived aldehydes into cyclic acetals for renewable diesel fuels, Sustainable Energy Fuels 2 (2018) 2742–2746.

[112] O. Staples, J.H. Leal, P.A. Cherry, C.S. McEnally, L.D. Pfefferle, T.A. Semelsberger, A.D Sutton, C.M. Moore, Camphorane as a renewable diesel blendstock produced by cyclodimerization of myrcene, Energy Fuels 33 (2019) 9949–9955.

[113] C. Ford Ryan, C.M. Moore, J.H. Leal, T.A. Semelsberger, J.K. Banh, J. Zhu, C.S. McEnally, L.D. Pfefferle, A.D. Sutton, Synthesis of aviation fuel from bio-derived isophorone, Sustainable Energy Fuels 4 (2020) 1088–1092.

[114] C.S. McEnally, Y. Xuan, P.C. St. John, D.D. Das, A. Jain, S. Kim, T.A. Kwan, L.K. Tan, J. Zhu, L.D. Pfefferle, Sooting tendencies of co-optima test gasolines and their surrogates. Proc. Combust. Inst. 37 (2019) 961–968.

[115] M.J. Montgomery, D.D. Das, C.S. McEnally, L.D. Pfefferle, Analyzing the robustness of the yield sooting index as a measure of sooting tendency, Proc. Combust. Inst. 37 (2019) 911–918.

[116] J.S. Carlson, E.A. Monroe, R. Dhaoui, J. Zhu, C.S. McEnally, S. Shinde, L.D. Pfefferle, A. George, R.W. Davis, Biodiesel ethers: fatty acid-derived alkyl ether fuels as improved bioblendstocks for mixing-controlled compression ignition engines, Energy Fuels 34 (2020) 12646–12653.

[117] L. Cosimbescu, K.B. Campbell, S. Subramanian, M.S. Swita, N. Hao, C.M. Moore, K.K. Ramaswamy, A.D. Sutton, C.S. McEnally, L.D. Pfefferle, J. Zhu, The properties of bicyclic and multicyclic hydrocarbons as bio-derived compression ignition fuels that can be prepared via efficient and scalable routes from biomass, Sustainable Energy Fuels 5 (2021) 3143–3159.

[118] N.A. Huq, G.R. Hafenstine, X. Huo, H. Nguyen, S.M. Tifft, D.R. Conklin, D. Stück, J. Stunkel, Z. Yang, J.S. Heyne, M.R. Wiatrowski, Y. Zhang, L. Tao, J. Zhu, C.S. McEnally, E.D. Christensen, C. Hays, K.M. Van Allsburg, K.A. Unocic, H.M. Meyer III, Z. Abdullah, D.R. Vardon, Toward net-zero sustainable aviation fuel with wet waste-derived volatile fatty acids, Proc. Natl. Acad. Sci. U. S. A. 118 (2021) e2023008118.

[119] M. Gu, F. Liu, J.-L. Consalvi, Ö.L. Gülder, Effects of pressure on soot formation in laminar coflow methane/air diffusion flames doped with *n*-heptane and toluene between 2 and 8 atm, Proc. Combust. Inst. 38 (2021) 1403–1412.

[120] M.J. Montgomery, J. Zhu, L.D. Pfefferle, C.S. McEnally, Amines have lower sooting tendencies than analogous alkanes, alcohols, and ethers, Combust. Flame 227 (2021) 335–345.

[121] S.S. Yang, Ö.L. Gülder, Effects of benzene, *cyclo*-hexane and *n*-hexane addition to methane on soot yields in high-pressure laminar diffusion flames, Proc. Combust. Inst. 38 (2021) 1107–1114.

[122] S.S. Yang, Ö.L. Gülder, Sooting propensity dependence on pressure of ethylbenzene, *p*-xylene, *o*-xylene and *n*-octane in laminar diffusion flames, Combust. Flame 227 (2021) 202–213.

[123] V.L. Dagle, M. Affandy, J.S. Lopez, L. Cosimbescu, D.J. Gaspar, S.S. Goldsborough, T. Rockstroh, S. Cheng, T. Han, C.P. Kolodziej, A. Hoth, S.S. Majumdar, J.A. Pihl, T.L. Alleman, C. Hays, C.S. McEnally, J. Zhu, L.D. Pfefferle, Production, fuel properties and combustion testing of an iso-olefins blendstock for modern vehicles, Fuel (2022) 122314.

[124] L. Zhu, Z. Gao, X. Cheng, F. Ren, Z. Huang, An assessment of surrogate fuel using Bayesian multiple kernel learning model in sight of sooting tendency, Front. Energy 16 (2022) 277–291.

[125] K. Gleason, A. Gomez, A simple method for the quantitative assessment of soot production rate, Combust. Flame 258 (2023) 113043.

[126] R. Jalain, J. Bonnety, G. Legros, A. Matynia, Sooting tendencies of ethylene diffusion flame doped by C_3-C_5 alcohols, Proc. Combust. Inst. 39 (2023) 979–987.

[127] E. Monroe, J.S. Carlson, R. Dhaoui, M. Sarwar, P.T. Benavides, J. Zhu, C.S. McEnally, L. Pfefferle, A. George, N. Sizemore, R.W. Davis, Application of alkoxyalkanoates (AOAs) as renewable diesel blendstocks from chemical coupling of high-yield fermentation products, Energy Fuels 37 (2023) 2091–2099.

[128] A. Singh, N. Tsolas, Sooting tendency of isopropanol-butanol-ethanol (IBE)/diesel surrogate blends in laminar diffusion flames, Combust. Flame 250 (2023) 112630.

[129] J. Zhu, J.V. Alegre-Requena, P. Cherry, D. Curtis, B.G. Harvey, M.A. Jabed, S. Kim, C.S. McEnally, L.D. Pfefferle, J.-D. Woodroffe, Sooting tendencies of terpenes and hydrogenated terpenes as sustainable transportation biofuels, Proc. Combust. Inst. 39 (2023) 877–887.

[130] M.A. Arellano-Treviño, F.G. Baddour, A.T. To, T.L. Alleman, C. Hays, J. Luecke, J. Zhu, C.S. McEnally, L.D. Pfefferle, T.D. Foust, D.A. Ruddy, Diesel fuel properties of renewable polyoxymethylene ethers with structural diversity, Fuel 358 (2024) 130353.

[131] H. Jung, J. Cho, Y. Kim, Z. Xiang, S. Kumar, P. Barnard, C.S. McEnally, L.D. Pfefferle, S. Kim, Sooting tendency of substituted aromatic oxygenates: The role of functional groups and positional isomerism in vanillin isomers, Proc. Combust. Symp. 40 (2024) 105669.

[132] Z. Xiang, M.B. Acikel, C.J. Hansen, G.-U. Jeong, R. Pérez-Soto, D.Z. Wang, V.C. Whoriskey, S. Kim, C.S. McEnally, L.D. Pfefferle, Y. Xuan, Experimental and computational study of soot formation from a wide range of lactones, Sustainable Energy Fuels (submitted). Available at https://doi.org/10.26434/chemrxiv-2024-lnc5x.

[133] R.L. Axelbaum, W.L. Flower, C.K. Law, Dilution and temperature effects of inert addition on soot formation in counterflow diffusion flames, Combust. Sci. Technol. 61 (1988) 51–73.

[134] M.J. Montgomery, H. Kwon, A.L. Kastengren, L.D. Pfefferle, T. Sikes, R.S. Tranter, Y. Xuan, C.S. McEnally, In situ temperature measurements in sooting methane/air flames using synchrotron x-ray fluorescence of seeded krypton atoms, Sci. Adv. 8 (2022) eabm7947.

[135] Y. Xuan, G. Blanquart, Numerical modeling of sooting tendencies in a laminar co-flow diffusion flame, Combust. Flame 160 (2013) 1657–1666.
[136] H. Kwon, A. Jain, C.S. McEnally, L.D. Pfefferle, Y. Xuan, Numerical investigation of the pressure-dependence of yield sooting indices for n-alkane and aromatic species, Fuel 254 (2019) 115574.
[137] 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, Harvard Dataverse (2017). https://doi.org/10.7910/DVN/7HGFT8.
[138] J. Melder, J. Zinsmeister, T. Grein, S. Jürgens, M. Köhler, P. Oßwald, Comprehensive two-dimensional gas chromatography: A universal method for composition-based prediction of emission characteristics of complex fuels, Energy Fuels 37 (2023) 4580–4595.
[139] H. Kwon, S. Shabnam, A.C. van Duin, Y. Xuan, Numerical simulations of yield-based sooting tendencies of aromatic fuels using ReaxFF molecular dynamics, Fuel 262 (2020) 116545.
[140] C.S. McEnally, L.D. Pfefferle, Fuel decomposition and hydrocarbon growth processes for oxygenated hydrocarbons: Butyl alcohols, Proc. Combust. Inst. 30 (2005) 1363–1370.
[141] C. Hall, D.C. Bell, J. Feldhausen, B. Rauch, J. Heyne, Quantifying isomeric effects: A key factor in aviation fuel assessment and design, Fuel (2024) 129912.
[142] A.R. Katritzky, M. Kuanar, S. Slavov, C.D. Hall, M. Karelson, I. Kahn, D.A. Dobchev, Quantitative correlation of physical and chemical properties with chemical structure: Utility for prediction, Chem. Rev. 110 (2010) 5714–5789.
[143] M. Ihme, W.T. Chung, A.A. Mishra, Combustion machine learning: Principles, progress and prospects, Prog. Energy Combust. Sci. 91 (2022) 101010.
[144] K.G. Joback, R.C. Reid, Estimation of pure-component properties from group-contributions, Chem. Eng. Comm. 57 (1987) 233–243.
[145] L. Constantinou, R. Gani, New group contribution method for estimating properties of pure compounds, AIChE J. 40 (1994) 1697–1710.
[146] M.P. Hanson, D.H. Rouvray, Novel applications of topological indices. 2. Prediction of the threshold soot index for hydrocarbon fuels, J. Phys. Chem. 91 (1987) 2981–2985.
[147] P.C. St. John, P. Kairys, D.D. Das, C.S. McEnally, L.D. Pfefferle, D.J. Robichaud, M.R. Nimlos, B.T. Zigler, R.L. McCormick, T.D. Foust, Y.J. Bomble, S. Kim, A quantitative model for the prediction of sooting tendency from molecular structure, Energy Fuels 31 (2017) 9983–9990.
[148] Z. Gao, X. Zou, Z. Huang, L. Zhu, Predicting sooting tendencies of oxygenated hydrocarbon fuels with machine learning algorithms, Fuel (2019) 438–446.

[149] A.G. Abdul Jameel, Predicting sooting propensity of oxygenated fuels using artificial neural networks, Processes 9 (2021) 1070.

[150] T. Kessler, P.C. St. John, J. Zhu, C.S. McEnally, L.D. Pfefferle, J.H. Mack, A comparison of computational models for predicting yield sooting index, Proc. Combust. Inst. 38 (2021) 1385–1393.

[151] R. Lemaire, G. Le Corre, M. Nakouri, Predicting the propensity to soot of hydrocarbons and oxygenated molecules by means of structural group contribution factors derived from the processing of unified sooting indexes, Fuel 302 (2021) 121104.

[152] R.C. Boehm, Z. Yang, J.S. Heyne, Threshold sooting index of sustainable aviation fuel candidates from composition input alone: Progress toward uncertainty quantification, Energy Fuels 36 (2022) 1916–1928.

[153] A.E. Comesana, T.T. Huntington, C.D. Scown, K.E. Niemeyer, V.H. Rapp, A systematic method for selecting molecular descriptors as features when training models for predicting physiochemical properties, Fuel 321 (2022) 123836.

[154] T. Kessler, A. SubLaban, J.H. Mack, Predicting the cetane number, sooting tendency, and energy density of terpene fuel additives, *Proceedings of the ASME 2022 ICE Forward Conference*. *ASME 2022 ICE Forward Conference*. Indianapolis, Indiana, USA. October 16–19, 2022. V001T02A011. ASME. https://doi.org/10.1115/ICEF2022-91163.

[155] R. Li, J.M. Herreros, A. Tsolakis, W. Yang, Machine learning and deep learning enabled fuel sooting tendency prediction from molecular structure, J. Mol. Graphics Modell. 111 (2022) 108083.

[156] M.A. Ahmed Qasem, E.M. Al-Mutairi, A.G. Abdul-Jameel, Smoke point prediction of oxygenated fuels using neural networks, Fuel 332 (2023) 126026.

[157] F.D. Alboqami, A.A. Pasha, M.I. Alam, A. Abdulraheem, A.G. Abdul Jameel, Prediction of yield sooting index utilizing artificial neural networks and adaptive-network-based fuzzy inference systems, Arab. J. Sci. Eng. 48 (2023) 8901–8909.

[158] Z. Chen, F. Vom Lehn, H. Pitsch, L. Cai, Prediction of sooting index of fuel compounds for spark-ignition engine applications based on a machine learning approach, J. Therm. Sci. 32 (2023) 521–530.

[159] S.V. Shree Sowndarya, Y. Kim, S. Kim, P.C. St. John, R.S. Paton, Expansion of bond dissociation prediction with machine learning to medicinally and environmentally relevant chemical space, Digital Discovery 2 (2023) 1900–1910.

[160] Y. Kim, H. Jung, S. Kumar, R.S. Paton, S. Kim, Designing solvent systems using self-evolving solubility databases and graph neural networks, Chem. Sci. 15 (2024) 923–939.

[161] L. Breiman, Random Forests, Machine Learning 45 (2001) 5–32.

[162] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACS SIGKDD International Conference on Knowledge Discovery and Data Mining (2016) 785–794.

[163] L.S. Whitmire, 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 (2016) 8410–8418.

[164] C.W. Yap, PaDEL-Descriptor: An open-source software to calculate molecular descriptors and fingerprints, J. Comput. Chem. 32 (2011) 1466–1474.

[165] T.T. Le, W. Fu, J.H. Moore, Scaling tree-based automated machine learning to biomedical big data with a feature set selector, Bioinformatics 36 (2020) 250–256.

[166] H. Hua, K.C. Das, H. Wang, On atom-bond connectivity index of graphs, J. Mathematical Analysis and Applications, 479 (2019) 1099–1114.

[167] T. Kessler, E.R. Sacia, A.T. Bell, J.H. Mack, Artificial neural network based predictions of cetane number for furanic biofuel additives, Fuel 206 (2017) 171–179.

[168] Alvascience, alvaDesc (software for calculating molecular descriptors) version 1.0.22 (2020).

[169] Kode, Dragon 7.0

[170] P.B. Jørgensen, K.W. Jacobsen, M.N. Schmidt, Neural message passing with edge updates for predicting properties of molecules and materials, arXiv 1806.03146 [cs, stat] 2018.

[171] Y. Kim, J. Cho, N. Naser, S. Kumar, R.L. McCormick, P.C. St. John, S. Kim, Physics-informed graph neural networks for predicting cetane number with systematic data quality analysis, Proc. Combust. Inst. 39 (2023) 4969–4978.

[172] H. Moriwaki, Y.-S. Tian, N. Kawashita, T. Takagi, Mordred: A molecular descriptor calculator, J. Cheminformatics, 10 (2018) 4.

[173] H. Böhm, K. Kohse-Höinghaus, F. Lacas, C. Rolon, N. Darabiha, S. Candel, On PAH formation in strained counterflow diffusion flames, Combust. Flame 124 (2001) 127–136.

[174] P.H. Joo, Y. Wang, A. Raj, S.H. Chung, Sooting limit in counterflow diffusion flames of ethylene/propane fuels and implication to threshold soot index, Proc. Combust. Inst. 34 (2013) 1803–1809.

[175] Z. Li, H.M. Amin, P. Liu, Y. Wang, S.H. Chung, W.L. Roberts, Effect of dimethyl ether (DME) addition on sooting limits in counterflow diffusion flames of ethylene at elevated pressures, Combust. Flame 197 (2018) 463–470.

[176] K.T. Kang, J.Y. Hwang, S.H. Chung, W. Lee, Soot zone structure and sooting limit in diffusion flames: Comparison of counterflow and co-flow flames, Combust. Flame 109 (1997) 266–281.
[177] Y. Wang, S.H. Chung, Formation of soot in counterflow diffusion flames with carbon dioxide dilution, Combust. Sci. Technol. 188 (2016) 805–817.
[178] Y. Wang, S.H. Chung, Strain rate effect on sooting characteristics in laminar counterflow diffusion flames, Combust. Flame 165 (2016) 433–444.
[179] L. Xu, F. Yan, M. Zhou, Y. Wang, S.H. Chung, Experimental and soot modeling studies of ethylene counterflow diffusion flames: Non-monotonic influence of the oxidizer composition on soot formation. Combust. Flame 197 (2018) 304–318.
[180] M.A. Delichatsios, A phenomenological model for smoke-point and soot formation in laminar flames, Combust. Sci. Technol. 100 (1994) 283–298.
[181] T. Beji, J.P. Zhang, W. Yao, M. Delichatsios, A novel soot model for fires: validation in a laminar non-premixed flame. Combust. Flame 158 (2011) 281–290.
[182] Y. Niu, Y. Zhang, X. Chen, D. Zhou, Application of the global soot model based on smoke point in simulation of multiple laminar diffusion flames, Combust. Sci. Technol. 190 (2018) 1060–1074.
[183] A. Arad, E. Sher, G. Enden, Modeling soot formation in diesel-biodiesel flames, Fuel 206 (2017) 437–452.
[184] S. Motaghian, T. Beji, A laminar smoke point-based soot model considering surface growth and soot reactions, Combust. Theory Modell. (2023) 1–26. https://doi.org/10.1080/13647830.2023.2267526.
[185] C.W. Lautenberger, J.L. de Ris, N.A. Dembsey, J.R. Barnett, H.R. Baum, A simplified model for soot formation and oxidation in CFD simulation of non-premixed hydrocarbon flames, Fire Saf. J. 40 (2005) 141–176.
[186] M. Smooke, C. McEnally, L. Pfefferle, R. Hall, M. Colket, Computational and experimental study of soot formation in a coflow, laminar diffusion flame, Combust. Flame 117 (1999) 117–139.
[187] M.D. Smooke, M.B. Long, B.C. Connelly, M.B. Colket, R.J. Hall, Soot formation in laminar diffusion flames, Combust. Flame 143 (2005) 613–628.
[188] M. Saffaripour, P. Zabeti, S.B. Dworkin, Q. Zhang, H. Guo, F. Liu, G.J. Smallwood, M.J. Thomson, A numerical and experimental study of a laminar sooting coflow Jet-A! diffusion flame, Proc. Combust. Inst 33 (2011) 601–608.

[189] S.B. Dworkin, Q. Zhang, M.J. Thomson, N.A. Slavinskaya, U. Riedel, Application of an enhanced PAH growth model to soot formation in a laminar coflow ethylene/air diffusion flame, Combust. Flame 158 (2011) 1682–1695.

[190] J.-L. Consalvi, F. Liu, J. Contreras, M. Kashif, G. Legros, S. Shuai, J. Wang, Numerical study of soot formation in laminar coflow diffusion flames of methane doped with primary reference fuels, Combust. Flame 162 (2015) 1153–1163.

[191] J.-L. Consalvi, F. Liu, M. Kashif, G. Legros, Numerical study of soot formation in laminar coflow methane/air diffusion flames doped by *n*-heptane/toluene and iso-octane/toluene blends, Combust. Flame 180 (2017) 167–174.

[192] H. Pitsch, A C++ Computer Program for 0D Combustion and 1D Laminar Flame Calculations, Technical Report, University of Technology (RWTH), Aachen, 1998.

[193] S. Lapointe, Y. Xuan, H. Kwon, R.A. Whitesides, M.J. McNenly, A computationally-efficient method for flamelet calculations, Combust. Flame 221 (2020) 94–102.

[194] M.J. McNenly, R.A. Whitesides, D.L. Flowers, Faster solvers for large kinetic mechanisms using adaptive preconditioners, Proc. Combust. Inst. 35 (2015) 581–587.

[195] M. McNenly, R. Whitesides, Zero-RK: zero order reaction kinetics. Available at https://github.com/llnl/zero-rk, 2019.

[196] A.C. van Duin, S. Dasgupta, F. Lorant, W.A. Goddard, ReaxFF: a reactive force field for hydrocarbons, J. Phys. Chem. A 105 (2001) 9396–9409.

[197] K. Chenoweth, A.C. van Duin, W.A. Goddard, ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation, J. Phys. Chem. A 112 (2008) 1040–1053.

[198] Q. Mao, M. Feng, X. Zhuo Jiang, Y. Ren, K.H. Luo, A.C.T. van Duin, Classical and reactive molecular dynamics: principles and applications in combustion and energy systems, Prog. Energy Combust. Sci. 97 (2023) 101084.

[199] Q.D. Wang, J.B. Wang, J.Q. Li, N.X. Tan, X.Y. Li, Reactive molecular dynamics simulation and chemical kinetic modeling of pyrolysis and combustion of n-dodecane, Combust. Flame 158 (2011) 217–226.

[200] F. Castro-Marcano, A.C. van Duin, Comparison of thermal and catalytic cracking of 1-heptene from ReaxFF reactive molecular dynamics simulations. Combust. Flame, 160 (2013) 766–775.

[201] A. Bharti, T. Banerjee, Reactive force field simulation studies on the combustion behavior of n-octanol. Fuel Process. Technol. 152 (2016) 132–139.

[202] X.M. Cheng, Q.D. Wang, J.Q. Li, J.B. Wang, X.Y. Li, ReaxFF molecular dynamics simulations of oxidation of toluene at high temperatures, J. Phys. Chem. A 116 (2012) 9811–9818.

[203] C. Ashraf, S. Shabnam, A. Jain, Y. Xuan, A.C. van Duin, Pyrolysis of binary fuel mixtures at supercritical conditions: A ReaxFF molecular dynamics study. Fuel 235 (2019) 194–207.
[204] J. Liu, X. Guo, ReaxFF molecular dynamics simulation of pyrolysis and combustion of pyridine, Fuel Process. Technol. 161 (2017) 107–115.
[205] H. Yuan, W. Kong, F. Liu, D. Chen, Study on soot nucleation and growth from PAHs and some reactive species at flame temperatures by ReaxFF molecular dynamics, Chem. Eng. Sci. 195 (2019) 748–757.
[206] Q. Mao, J. Zhou, K.H. Luo, A.C. van Duin, Atomistic insights into the dynamics of binary collisions between gaseous molecules and polycyclic aromatic hydrocarbon dimers. Phys. Chem. Chem. Phys. 21 (2019) 3849–3856.
[207] P. Zhang, K. Zhang, X. Cheng, Y. Liu, H. Wu, Analysis of inhibitory mechanisms of ammonia addition on soot formation: A combined ReaxFF MD simulations and experimental study. Energy & Fuels, 36 (2022) 12350–12364.
[208] H. Kwon, A. Lele, J. Zhu, C.S. McEnally, L.D. Pfefferle, Y. Xuan, A.C. van Duin, ReaxFF-based molecular dynamics study of bio-derived polycyclic alkanes as potential alternative jet fuels, Fuel 279 (2020) 118548.
[209] H. Kwon, Y. Xuan, Pyrolysis of bio-derived dioxolane fuels: A ReaxFF molecular dynamics study, Fuel 306 (2021) 121616.
[210] P. Zhang, H. Wu, X. Song, Y. Liu, X. Cheng, X., Fuel molecular structure influences the polycyclic aromatic hydrocarbons formation of butanol/butane isomers: A ReaxFF molecular dynamics study, Fuel 310 (2022) 122460.
[211] Y. Wang, M. Gu, Y. Zhu, L. Cao, J. Wu, Y. Lin, X. Huang, Analysis of soot formation of CH_4 and C_2H_4 with H_2 addition via ReaxFF molecular dynamics and pyrolysis–gas chromatography/mass spectrometry, J. Energy Inst. 100 (2022) 177–188.
[212] H. Kwon, B.D. Etz, M.J. Montgomery, R. Messerly, S. Shabnam, S. Vyas, A.C. van Duin, C.S. McEnally, L.D. Pfefferle, S. Kim, Y. Xuan, Reactive molecular dynamics simulations and quantum chemistry calculations to investigate soot-relevant reaction pathways for hexylamine isomers, J. Phys. Chem. A 124 (2020) 4290–4304.
[213] C. Chen, X. Jiang, Molecular dynamics simulation of soot formation during diesel combustion with oxygenated fuel addition. Phys. Chem. Chem. Phys. 22 (2020) 20829–20836.