Article

Threshold Sooting Index of Sustainable Aviation Fuel Candidates from Composition Input Alone: Progress toward Uncertainty Quantification

Randall C. Boehm,* Zhibin Yang, and Joshua S. Heyne

Cite This: https://doi.org/10.1021/acs.energyfuels.1c03794

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: A quantitative structure−property relationship model has been developed to predict the threshold sooting index (TSI) of arbitrary mixtures of aliphatic and aromatic hydrocarbons of known composition. The model employs contributions from eight molecular fragments plus a global shift and a penalty factor for naphthenic compounds. For each coefficient, the contributions were determined by a constrained regression to data from five different experimental campaigns, which were stitched together by setting the TSI of methylcyclohexane to 5 and the TSI of 1-methylnaphthalene to 100. Unique to this study; the TSI of 1,3,5-trimethylbenzene was restricted to the range of 54.7 and 63.1, which significantly constrains uncertainty. Within the composite training dataset, which contained 65 molecules and 124 data points, including simple mixtures, the model was found to match 95% of the data within 8.9 TSI. Validation of the model against n-butylcyclohexane, dimethylcyclooctane, and a six-component surrogate jet fuel shows prediction to be well within the 95-percentile confidence band of the experiment. This model is the first to integrate the linear blending rule for the TSI with a linear quantitative structure−property relationship model for the TSI, and the first time that referee controls have been applied to ensure that all datasets, experimental and modeled, are normalized to the same scale.
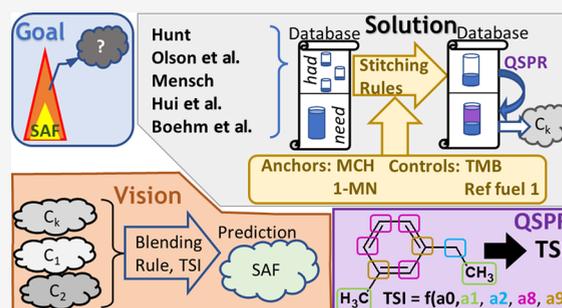
1. INTRODUCTION

The aviation industry is in the process of reducing its carbon footprint. To date, seven processes that convert a specified biological feedstock into a jet fuel blendstock have been incorporated into the standard specification for aviation turbine fuel containing synthesized hydrocarbons, ASTM D7566-21.[1] In each case, the synthetic component is to be blended up to some ratio (10−50%) with petroleum-derived fuel to create a so-called "drop-in" fuel. A drop-in fuel behaves like 100% petroleum fuel within the detectability limits of the typical delivery and consumption processes that exist within the industry. For example, suppose that an arbitrary hydro-carbon mixture was to absorb certain chemical species that have already been absorbed by certain polymeric materials (e.g., O-rings) that exist within an aircraft's fuel system. In that case, those species could be depleted from the polymeric material, causing the O-ring to shrink, and a corresponding joint that that polymeric material should seal could begin to leak. That would not be acceptable, so there is a requirement for ASTM D7566 fuels to contain at least 8% aromatics since aromatics have been identified as the class of hydrocarbons that is most involved with fuel-polymer material compatibility.[2]

Aside from their relatively low hydrogen-to-carbon ratio, which may offer a small energy efficiency benefit via increased gas expansion through a turbine,[3] the seal compatibility issue

and potentially the dielectric constant are the only two drivers maintaining the aromatic content in aviation fuels. As non-aromatic molecules are discovered to emulate aromatic/elastomer compatibility, such as dimethylcyclooctane, perhaps, it is important to assess their impact (good or bad) on other fit-for-purpose properties,[4] including smoking propensity. Typically, aromatics (i) harm aircraft energy efficiency arising from their low specific energies (energy per unit mass),[5] (ii) are believed to correlate with decreased thermal stability,[6] and (iii) increase the smoking propensity[7] and particulate emissions.[8] These particulate emissions, in particular, are believed to correlate with contrails, which may collectively have a larger impact on radiative forcing than CO_2 emissions from aircrafts.[9] Moreover, not all aromatics have the same impact on smoke,[10,11] thermal stability, or seal swell. Similarly, not all alkanes have the same impact on smoke,[12] and it follows that different fuel compositions would have a different smoking propensity. The first revision of the standard test method for

the smoke point of kerosene and aviation turbine fuel, ASTM D1322,[13] was published in 1954 as it was recognized that control of this combustion-related fuel property was necessary.

Full details of ASTM D1322 are published in the standard,[13] and an overview of it is provided in the Experimental Methods of Each Dataset section of this paper. Still, for now, it is important to understand some logistics. To get a single data point requires ~20 min of labor to prepare the lamp and take the reading correctly. Additionally, labor time is needed to get repeat points, acquire samples, manage the inventory of samples, etc. For the sake of discussion, let us consider that each data point costs 2 h of labor in total. Suppose that a database contains every point generated during an optimization of fuel composition against maximum lower heating value (LHV) and minimum engine-level specific fuel consumption.[14] In that case, data is desired for a million or more possible mixtures spawned from a database of 1124 molecules. The cost to acquire that data would be double that many hours. To get that done in 1 year would require over a thousand full-time employees and over 500 lamps. These are staggering numbers. Even if we used a trusted blending model to reduce the number of desired data points to 1124, it would still require a full labor year to get that data. Therefore, it is desired to use models to predict the smoke point (the result of the ASTM D1322 test) of any mixture of possible fuel constituents and to predict the smoke point of all the constituents for which data is not already available. Moreover, it is important to understand the accuracy of these models.

A method of predicting the sooting tendency of mixtures based on the sooting tendencies of its constituents was introduced by Gill and Olson[14] in 1984 and has since been validated by Yan et al.[15] and Mensch et al.[16] Their method leverages the threshold sooting index (TSI) introduced by Calcote and Manos[11] a year earlier to normalize smoke point data from different experiments. Equations 1 and 2 depict the blending rule and sooting index, respectively. In these equations, x_i is the mole fraction of the ith component and TSI_i

$$TSI_{mix} = \sum_i x_i \times TSI_i \tag{1}$$

$$TSI = a + b \times \left(\frac{Mw}{Sp} \right) \tag{2}$$

is its threshold sooting index. The determined coefficients, a and b, are device- and operator-dependent, Sp is the measured smoke point, and Mw is the molecular weight of the sample molecule or mixture. While it is unclear whether eq 2 truly puts data from different experiments into a common basis, at minimum, the definition needs two anchor points to derive the device- and operator-dependent coefficients, a and b. Additional discussion around establishing a best practice for deriving these coefficients will be presented later in this report.

Generally, the blending rule, eq 1, can produce inaccurate predictions in two ways. The TSI corresponding to any or all the components could be inaccurate, or the neglect of potential synergistic effects could be significant relative to component uncertainty. An example of a potential synergistic effect would be compounds that produce an unusual concentration of radicals that accelerate or hinder the kinetics of combustion. Somewhat different, preferential wicking rates could alter the vapor-phase mole fractions relative to the liquid-phase mole fractions in any experiment that employs a wick, and

preferential evaporation could alter this relationship in diffusion flames whether or not a wick is involved. In Figure 1, imagined data are plotted to illustrate how each of these error types influences the uncertainty in TSI_{mix} as a function of the mole fraction. The example of preferential wicking or evaporation leading to a 5% (or 20%) off-set between the liquid and vapor phase mole fractions leads to a TSI versus a
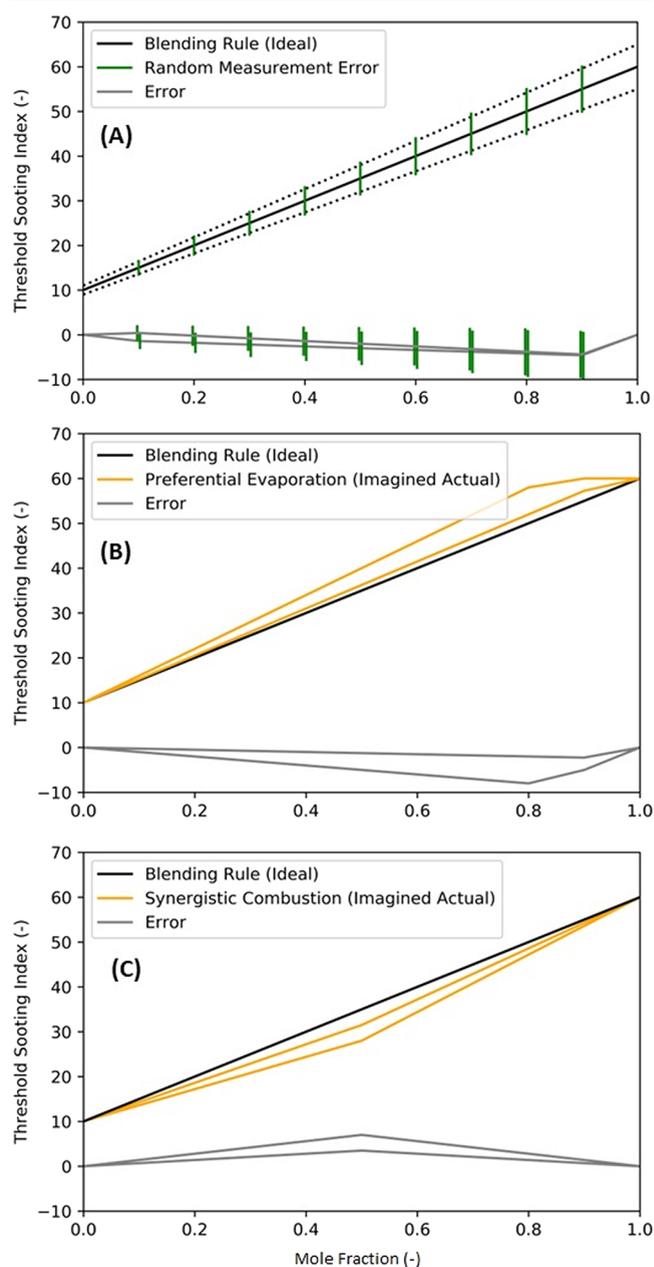

Figure 1. TSI blending rule error - shape factors for binary mixtures. The TSI of the mixture (upper curves): blending rule predictions (black lines) and the gold lines represent what the actual data would look like given certain sources of error. (A) The random error is ±1 at the low end point and ±5 TSI at the high end point. (B) Component A's gas-phase mole fraction is arbitrarily 5% (or 20%) higher than its liquid-phase mole fraction. (C) Species created through combustion of component B facilitate the combustion of soot precursors created through combustion of component A, an imagined 10% (or 20%) benefit for a 50/50 blend. The lower curves (gray lines) are the difference between the upper curves (gold and black).

mole fraction curve that is indistinguishable, without additional information, from the example of an error in the TSI of the component that is more prone to smoking. In each case, less error would result from the blending rule if the $x = 0$ and 1 intercepts of the best fit line through all the data were used to establish a virtual TSI of each pure component and that virtual TSI was used in the blending rule instead of the measured TSI of the pure component.[17,18] However, that approach could lead to several different virtual TSIs for each molecule depending on what other molecule(s) it is mixed with. The concept of virtual TSI is not without merit, but if it is used, the virtual TSI should be derived from a global regression of a variety of mixtures and this point will be addressed more thoroughly in the methods section of this report.

In our opinion, the primary issue with eq 1 is not with its neglect of potential differences between the vapor and liquid phase mole fraction or possibly synergistic combustion kinetics[19,20] but rather with the dearth of benchmark quality data for pure components, whether directly measured or derived from regression of mixtures (virtual). Among the datasets that have been published to date, those from refs 7, 12, 15, 16, and 21−29, representing a total of 112 hydrocarbons, have been considered therein. Some of these hydrocarbons are not liquids at standard temperature and pressure, some have wicking rates that are lower than their fuel consumption rate, and some are alkenes or alkynes, which are of little interest to the focus of this group (aviation fuels). Thirty-nine molecules with trusted smoke point data are saturated hydrocarbons, and 26 are alkylated aromatics. Of these 65 data points, 59 can be stitched together by using the TSI, eq 3 as a tool to normalize data from this work, and four published datasets[12,16,23,29] into a common basis. Just 25 of these molecules are included within our internal database of 1124 molecules, a subset of the NIST database,[31] which are to be considered the blendstock for sustainable aviation fuel.[30] Existing data, therefore, represents 2.2% (25/1124) of desired data. If we include all the available relevant data, then 5% (59/1158) of the combined database could potentially serve as training data to predict the remaining 95%.

$$100 = a \times \left(\frac{142.201}{\mathrm{Sp}(1MN)} \right) + b \text{ and } 5$$

$$= a \times \left(\frac{98.189}{\mathrm{Sp}(MCH)} \right) + b \tag{3}$$

Yan et al.[15] used quantitative structure−property relationships (QSPR) employing structural fragment contributions to create models to estimate the TSI. They executed a curve fit three times, employing a somewhat different set of dependent variables (the structural fragments) each time. They named these three sets the Joback method, modified Joback method, and SOL method. Each method employed a fifth-order polynomial (−1 to 4) to relate the TSI to a single independent variable, which was the sum of the contributions from the structural fragments. A finite segment of each polynomial was monotonic over the range of molecules used to train the model. However, for each method, the predicted TSI decreases sharply as the dependent variable increases to values higher than the maximum afforded by the training set. For example, the predicted TSI of 1,3,5-tri-tertbutylbenzene is −31.3 while the predicted TSI of [1,4-dimethyl-1-(3-methylbutyl)pentyl]-benzene is 95.8 using the so-called Joback method. Even within

the range of the correlated data, the three methods did not show qualitative agreement with each other. For example, the predicted TSIs of 2,2,4,4,6,6,8-heptamethylnonane were 80.4 by the Joback method and 14.6 by the modified Joback method.

By delving into these issues further, it was observed that the contributions from each molecular fragment, their relationships to each other, did not follow physical intuition. For example, the central carbon of neopentane should contribute to a higher TSI than the central carbon of isobutane, and the contribution from all fragments should be positive. In other words, it was obvious from the coefficients that the correlations were executed without logical constraints. By adding constraints onto the relationships between fragment contributions and eliminating the polynomial rescaling, we have generated a new model to predict the TSI of arbitrary molecules that is considerably more versatile than those published by Yan et al.[15]

Other groups have used a QSPR approach to predict[32,33] the threshold sooting index,[11] yield sooting index,[34] oxygen extended sooting index,[32] or fuel equivalent sooting index.[35] Barrientos et al.[32] reported OESI activities for seven non-oxygenated carbon groups, three of which are related to this work. However, significant differences between their training/target datasets and our target database render those activities unsuitable for our application. Very recently, Lemaire et al.[33] reported unified index activities for 29 unoxidized carbon groups, 11 of which are relevant to this work. As that work was published after the technical aspects of this work were already complete, a comparison of the two approaches, supporting data, and predictions will be made in the discussion section of this paper.

A basic description of the ASTM D1322 standard is provided within the Experimental Methods of Each Dataset section of this report. That description is followed by a discussion of the additional controls (best practices) necessary to parlay this experiment, which was designed to inspect aviation fuel with a smoke point of 25 ± 3 mm,[36] into a research tool. This is followed by a description of what has been done specifically in this work to bring four legacy datasets into a common basis with new data measured in our laboratory. The Data Scaling and Uncertainty Analysis section includes a discussion around the pros and cons of using data from mixtures to help train a model to predict the TSI of molecules that will ultimately be used to help predict the TSI of different (and arbitrary) mixtures. This section also includes a detailed description of our QSPR model, the constraints that were employed for its development, and justification for those constraints. The Results section compares model results with the data used to train it and a check of model results against measured smoke points for two pure compounds, n-butylbenzene and dimethylcyclooctane, and one simple surrogate fuel with a fully identified composition. Finally, a comparison is drawn between measured smoke points of seven complex hydrocarbon mixtures, including three conventional jet fuels, and our predicted smoke points for these mixtures. For these predictions, the so-called Tier α[37] methodology for estimating a representative property for an unknown mix of hydrocarbons with the same empirical formula was used to complement the model presented in this work. The Tier α methodology employs blending rules to estimate mixture properties based on molecular constituents' mole (mass or volume) fractions and a molecular property database. Where mole fraction inputs are indeterminant, the method employs

Monte Carlo simulations to sample the range of possible values. This model creates the sooting propensity piece of the Tier α methodology's database.

2. EXPERIMENTAL METHODS OF EACH DATASET

The experimental apparatus and procedures used in this work and Mensch et al.[16] are described thoroughly in ASTM D1322-19. The apparatus consists of a wick-fed lamp that can be purchased from various suppliers (keyword search: "smoke point lamp"). A Koehler lamp was used for this work. The base of the lamp should be placed on a horizontal surface, so its mounted candle section is vertical, and the flame it produces radiates vertically upward before a mounted ruler that is 5 cm in length. The flame is fully shielded on four sides. Twenty small (2.9 mm) intake holes regulate the flow of fresh air into the reaction zone, and a cylindrical chimney (40 mm diameter) is positioned 5 cm above the top of the lamp to allow vitiated air to escape from the reaction zone. The fuel consumption rate is controlled by adjusting the length of the wick that is exposed to air by raising or lowering the wick assembly through the candle body. The ASTM D1322 documentation also describes standardized procedures for wick preparation and defines the flame shape and tip position corresponding to the smoke point. Since 2012, revisions of the standard describe optional automation equipment, including a computer controller to adjust the length of the exposed wick and image processing software to determine the smoke point based on digitized images captured from a video camera that is mounted normal to the flame. The automation is reported to improve the reproducibility of the method by about a factor of 4 over the range of smoke points produced by the calibration fuels identified in the standard. However, most published smoke point data was collected prior to 2012, and the new data reported here was also collected as per the manual procedures. Future work concentrating on the expansion (~doubling) of the collective TSI database should consider automated data acquisition and alternative test methodologies[24,39−41] to improve data precision, accuracy, collection efficiency, and the range of molecules for which a direct measurement can be made.

While the first revision of ASTM D1322 was published in 1954, the Institution of Petroleum Technologists formed a "Standardization Sub-Committee on Tendency to Smoke" in 1931.[38] Three years later, Terry and Field[39] published details of an improved factor lamp, the experimental apparatus used by Hunt,[23] who published in 1953 the most extensive database of smoke point data to date. While there are various small differences between this experiment and the current ASTM standard, the main features of the experiments are common, and in theory,[11] their respective smoke point data, once converted to a common TSI scale, should be comparable.

Olson et al.[12] had observed a local feature (a shoulder or dip) in plots of fuel consumption rate versus flame height showing up at heights corresponding to the measured smoke points. They used this feature to define the smoke point in terms of fuel consumption rate to improve the repeatability of the measurements. The apparatus and procedures used for this experiment differ significantly from those described in ASTM D1322. Nonetheless, Olson et al. showed a good correlation (R^2 = 0.94) between their measured TSIs of 28 compounds and those reported by Calcote and Manos,[11] who had compiled TSI data from six different sources, most extensively the data from Hunt.[23] Hui et al.[29] used a wick-fed lamp that conformed to ASTM D1322 and followed the methodology described by Olson et al.[12] to define the smoke point.

This work measured smoke points for eight neat molecules, seven binary mixtures of varying composition, a six-component surrogate fuel, three conventional jet fuels, and three complex mixtures under consideration as sustainable aviation fuel. While all our smoke point data is provided as the Supporting Information to this report, Table 1 lists each molecule used, purity, and source. All measurements were taken 8−20 times, including at least one change of wick throughout the progression of repeat data points. This information is included in the data summary for each point and is provided in the Supporting Information.

Table 1. Chemicals Used[a]

name	purity	supplier
methylcyclohexane	99%	Sigma-Aldrich
1,3,5-trimethylbenene	>97%	TCI
iso-octane	>99%	Sigma-Aldrich
m-xylene	>99%	TCI
o-xylene	>98%	TCI
n-butylcyclohexane	99%	Alfa Aesar
n-octane	>99%	ACROS
toluene	99.5%	Fisher
cis-decalin	98%	TCI
1-methylnaphthalene	96%	Alfa Aesar
n-undecane	>99%	Sigma-Aldrich
hexylbenzene	98%	Alfa Aesar
iso-cetane	98%	Sigma-Aldrich
farnesane	>99%	Amyris
n-hexadecane	99%	Alfa Aesar
1,4-dimethylcyclooctane	98.5%	B.G. Harvey, NAWCWD and CleanJoule

[a]Conventional and two of the potential sustainable jet fuel samples were provided by T. Edwards, AFRL. HEFA was provided by World Energy, and SAK was provided by an anonymous supplier.

3. DATA SCALING AND UNCERTAINTY ANALYSIS

In previously published articles, the experiment constants labeled as a and b in eq 1 have been recalculated to minimize the collective difference between ostensibly common data points between the authors' work, which is taken as the gold standard, and any previously published results to which there was value in comparing. While that is certainly one way to normalize data, it creates a different TSI unit scale (analogous to the Fahrenheit or Celsius temperature unit scale) for every dataset. It lacks any formal rigor to ensure that any of the experiments were in control, i.e., properly calibrated. One step toward establishing a check on the process control is to define certain reference materials with defined TSI or smoke point values. Calcote and Manos[11] suggested pure hexane (TSI = 2) and pure 1-methylnaphthalene (TSI = 100), while Mensch et al.[16] suggested pure methylcyclohexane (MCH, TSI = 5) and pure 1-methylnaphthalene (1-MN, TSI = 100) as reference materials, but neither imposed these constraints onto previously published datasets to which a comparison was made. Indeed, several of the earlier datasets did not include these proposed reference materials. The ASTM D1322 standard calls out iso-octane (Sp = 42.8 mm) and six different blends of iso-octane with toluene as reference materials to be used for device calibration. While several of these reference fuels were included in this work and Mensch et al.'s,[16] the other works[12,24,27] included pure iso-octane.

For inspection of aviation fuel, the iso-octane/toluene reference fuels are sufficient to establish control because they bracket the smoking propensity of any aviation fuel sample that is likely to be inspected by this method. If not, the fuel is so good or so bad that accuracy ceases to be necessary to determine whether the fuel passes inspection. However, for our purposes, many of the molecules of interest have smoke points below that of 60%v iso-octane blended with 40%v toluene, which is 14.7 mm, or above that of pure iso-octane (42.8 mm). For our purposes, additional reference materials are necessary to establish calibration throughout the entire range of the smoke point or TSI values of interest. Here, we adopt, without endorsement, the convention of Mensch et al.[16] to define the

TSI scale by setting its value for 1-methylnapthalene to 100 and its value for methylcyclohexane to 5, and we suggest that the reference fuels of ASTM D1322 should also be used to establish control and, more importantly, that another molecule or mixture with a TSI of ~60 should also be used. We suggest 1,3,5-trimethylbenzene. The advantage of MCH relative to hexane is that its flame is less susceptible to flickering noise at its smoke point because it is shorter, but assigning it such a low TSI value, 5 leads to a negative TSI for some normal alkanes and lightly branched iso-alkanes, which could be confusing.

Figure 2 has been constructed to emphasize a motivation for introducing additional referee materials to establish exper-
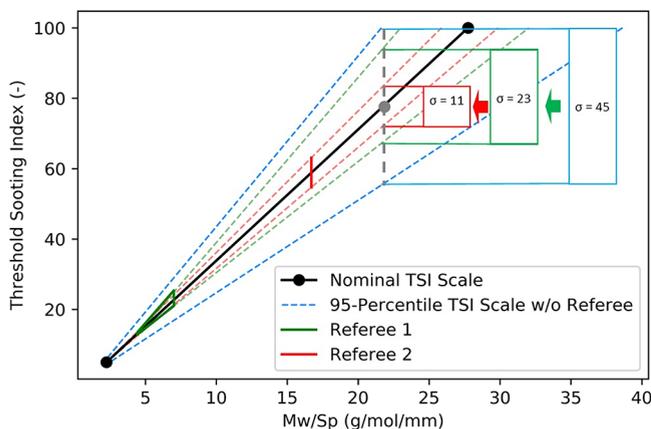
Figure 2. Impact of reference data uncertainty on the experiment's TSI scale coefficient. Nominal points are taken from this work. The horizontal line at TSI = 100 was constructed from the nominal point and the 95-percentile repeatability quote from ASTM D1322-19.

imental control. The filled circles shown in Figure 2 correspond to the measured data (this work) for 1-

methylnaphthalene and methylcyclohexane, which are connected by a solid black line. The dashed lines on either side of the solid black line correspond to plus or minus the quoted 95 percentile for smoke point repeatability,[13] which is based on an inter-laboratory study conducted by subcommittee D02.J0 ASTM and is provided here, as shown in eq 4. Lower case r is the repeatability and $\overline{Sp}$ is the average smoke point. At face value, this plot suggests that the TSI slope coefficient for our experiment could be anywhere from 2.60 to 4.95 and the intercept coefficient anywhere from −7.33 to −0.35, which is much higher than desired.

$$ r = 0.0684 \times (\overline{Sp} +16) \tag{4} $$

Of course, one way to help reduce the repeatability uncertainty is to average over multiple readings (N) of the same experiment because the effective repeatability scales with the inverse square root of N, and we have taken between 8 and 20 repeat points in this work.

Another way to drive sufficiently tight repeatability into the smoke point measurements of the scale-setting TSI anchor points (methylcyclohexane and especially 1-MN) is to establish TSI limits for referee fuels. The details within Figure 2 help to illustrate how this helps. The blue, horizontal line at the top of the plot follows directly from the 95-percentile confidence interval (eq 4) for a single-point measurement of the smoke point for 1-MN. The dashed, gray, and vertical line at 22 on the horizontal axis illustrates how that random error in smoke point determination at the 1-MN anchor point is transferred into a large uncertainty (77.5 ± 22.5) in the TSI for a fuel with a measured (Mw/Sp) of 22 g/mol/mm, which is within the range of data we might use to build a QSPR model. On top of that, the random error associated with smoke point measurement of the potential data point is approximately ± 10 TSI. If we require the TSIs of the first and fourth reference fuel blend of ASTM D1322 to be 23.2 ± 2.2 and 13.2 ± 0.3,

Table 2. Data Summary for Highlighted Fuels

Fuel Sample	Index
1-methylnaphthalene	A
Methylcyclohexane	B
Toluene	C
iso-octane	D
1,3,5-trimethylbenzene	E
40/60 %v toluene /iso-octane	F

Property	Experiment	Fuel A	Fuel B	Fuel C	Fuel D	Fuel E	Fuel F
Smoke Point (mm)	This work	5.12	43.56	7.59	–	7.19	14.88
	Mensch [18]	5.5	40.8	8.4	40.0	7.3	14.57
	Olson et al. [12]	5 (4.6)[a]	42	7	38	6	–
	Hunt [23]	5 (4.43)[a]	94	6	86	6	–
	Avg ± σ[b]	4.9 ± 0.5	55 ± 26	7.2 ± 1	55 ± 27	6.6 ± 0.7	14.7±0.2
TSI (-)	This work	100	5	41.8	–	58.9	22.4
	Mensch [18]	100	5	39.7	6.8	62.0	24.0
	Olson et al. [12]	100	5	40.6	7.2	63.1	–
	Hunt [23]	100	5	48.8	5.9	63.1	–
	Avg ± σ[b]	100 ± 0	5 ± 0	42.7 ±4.1	6.6 ± 0.7	61.8±2.0	23.2±1.1

[a]The numbers in parentheses correspond to the virtual smoke point used to normalize legacy datasets. [b]Standard deviations (or reproducibility) based on one value per experimental campaign.

respectively, then the actual value of (Mw/Sp) of 1-MN must lie in-between the two green lines at TSI = 100 and that reduces the transferred random error from the anchor point to the potential data point from ±22.5 to ±11.5. While the iso-octane/toluene reference fuels serve as convenient TSI referees because these data should exist for every modern test campaign, they do not force sufficient precision into the smoke point measurement of 1-MN. To tighten the precision further, we suggest using 1,3,5-trimethylbenzene (TSI = 58.9 ± 4.2) as a second referee. By doing that, the transferred uncertainty from the upper anchor point to the potential data point at TSI = 77.5 goes to ±5.5. This is about half as large as the uncertainty that comes directly from the random error of smoke point measurement of the data point and contributes ±1.2 out of the ±11.2 overall (quadrature addition) uncertainty in the TSI of the data point, which is tolerable.

Suppose that the referee control standards are not met initially. In that case, either additional repeat data points should be taken for the anchor fuels or the referee fuels, or the apparatus correction factor should be adjusted. Before stitching any dataset into a master database of TSI values, it should be put onto the same scale by adjusting its corresponding experiment TSI constants, a and b, to satisfy eq 3. Additionally, it should be confirmed that each of the referee criteria is met.

The datasets from this work and those from Mensch et al.[16] meet these criteria and can be compared directly, while the dataset from Hui et al.[29] technically does not meet these criteria. However, in the case of Hui et al., the recorded TSIs of pure iso-octane and pure toluene were within the range created by the other datasets. Additionally, a 60/40%v mathematical (eq 1) blend of these two readings does fall within our recommended acceptance range for referee 1. To drive virtual compliance with these criteria for the datasets of Hunt[23] and Olson et al.,[12] a virtual data point for 1-MN was created by adjusting the reported smoke point within its repeatability window until the experiment TSI coefficients resulted in a TSI of 58.9 ± 4.2 for 1,3,5- trimethylbenzene. For the Hunt dataset, the virtual smoke point of 1-MN was 4.43 mm, and for the Olson et al. dataset, it was 4.60 mm, each compared to a reported smoke point of 5 mm. A summary of the recorded smoke points, virtual smoke points (where applicable), the TSI for pure toluene, and each of the referee and anchor fuels is provided in Table 2, and a graphical representation of all TSI data used to support the model reported in this work is presented in Figure 3. There in Figure 3, the TSI for each
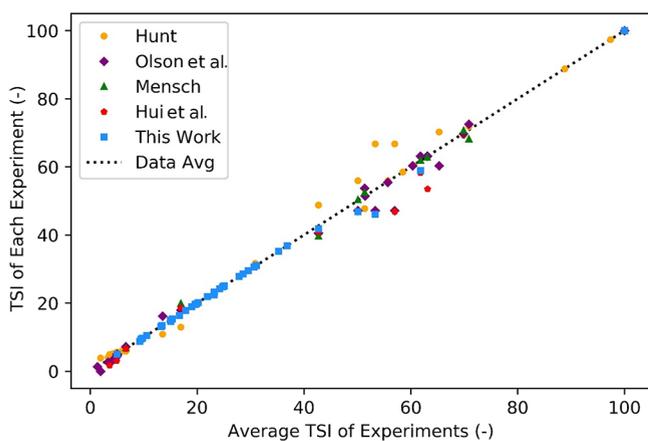
study is plotted on the Y axis, and where applicable, the average TSIs for materials with more than one reported observation are plotted on the X axis.

The data shown in Figure 4 has been extracted from the thesis of Mensch[18] and suggests that the repeatability of her
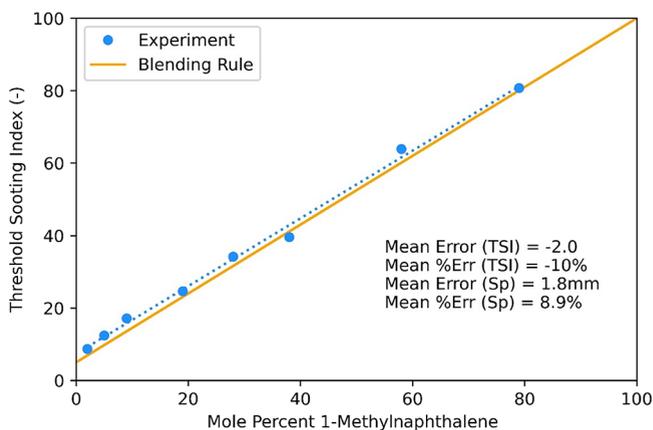

Figure 4. Evaluation of the linear blending rule (eq 1). Data from Mensch[18] for binary mixtures of 1-methylnaphthalene and methyl-cyclohexane.

experiment is an order of magnitude tighter than suggested in eq 4. In the data reported by these authors, we also have observed similar repeatability as Mensch.[18] Some of the discrepancies between our determined repeatability and eq 4 can be attributed to signal-to-noise improvements afforded by taking repeat data points, and we have elected to use same-lab data replication to represent repeatability instead of eq 4 wherever that is possible and applicable. The quoted[13] reproducibility of the method is consistent with our observations of differences between common molecules of published data but 37% higher than the quoted repeatability in the specification.[13] The reproducibility covers differing opinions about the exact shape of the flame at its smoke point, differing intuition regarding line-of-sight being perpendicular to the ruler, lamp hardware differences (e.g., intake hole diameter), fuel purity differences, and laboratory differences (e.g., humidity, ventilation, temperature, pressure, and bench-top levelness).

The power of the TSI transformation resides in two important points. The transformation attenuates systematic differences in smoke point data between experimental campaigns appreciably, and the blending rule described in eq 1 holds true.[14,40] Figure 4 provides one example comparing the result of eq 1 with measured data, and more examples (from this work) will be shown in the Results section. Figure 5 provides an example of systematic error attenuation caused by the transformation of smoke point data to the threshold sooting index. The data provided by Mensch[18] is plotted along the X axis, and the Y axis has contrived data. In one scenario, the contrived smoke point data is 1 mm higher than the actual data (simulating operator reading error), and in the other scenario, it is 7% higher (simulating a different vent hole diameter). For that dataset, the imposed 1 mm off-set in the smoke point results in an average smoke point error of 11%, while the average error in the TSI is 3.4%. The imposed 7% off-set in the smoke point results in a 1 mm smoke point difference on average and no difference at all in the TSI.
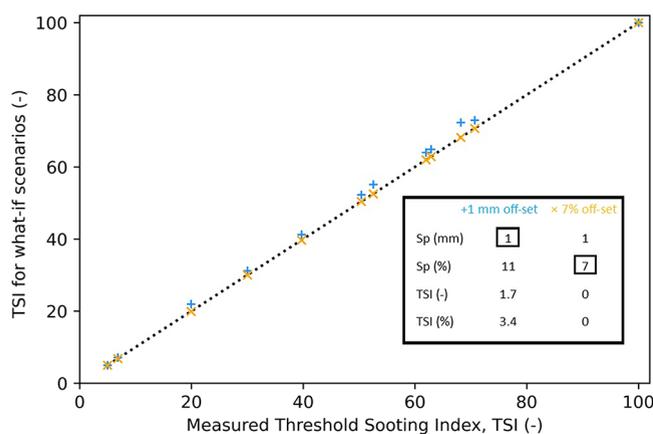

Figure 3. Full TSI database supporting QSPR model development.

Figure 5. Partial erasure of systematic differences between datasets. Operations performed on data published by Mensch.[18] The dashed line is the original data. The plus symbols are the result of adding 1 mm to each reported smoke point. The cross symbols are the result of adding 7% to each reported smoke point. The arbitrary errors assigned here are larger than is possible for experiments adhering to the calibration controls described in ASTM D1322.

The linearity of the TSI blending rule (eq 1) is particularly powerful because it opens the possibility of deriving useful smoking propensity information for materials that may not lend themselves to direct measurement. Some compounds result in a diffusion flame height that is too high (flickering) or too low (vision/optics limited) at their respective smoke points to read accurately. Some compounds are not liquids under standard ambient conditions but may exist in solution with other hydrocarbons up to some threshold concentration. Still, other compounds such as 2,6,10-trimethyldodecane (farnesane) and hexadecane cannot be evaluated as pure materials via ASTM D1322-19 because their wicking rate is less than their fuel consumption rate under the conditions of the experiment. The Washburn law describes the wicking rate[41] and scales inversely with dynamic viscosity and linearly with surface tension and the contact angle between the liquid and wick material. When the wicking rate is lower than the fuel consumption rate, the wick fabric starts to burn, and the apparent smoke point is lower than if the wicking rate was to be increased or the fuel consumption rate decreased (off-spec) by changing the conditions of the experiment. In all these cases, it is possible to measure the smoke point of a variety of mixtures that contain the problematic component A at some mixture fraction and to leverage that data to determine a virtual TSI or smoke point of pure A that could be used in eq 1 for estimating the TSI (or smoke point) of some other mixture of known composition.

Another point in favor of using mixtures to derive virtual smoke points of pure molecules is that we ultimately care more about a component's effect on the smoke point of a mixture than what its smoke point is as a pure material. As noted in the Introduction and the brief discussion around Figure 1, the TSI blending rule neglects the effects of certain physical processes such as preferential evaporation and synergistic flame chemistry that would lead to some non-linearity, but these effects are likely to show up (qualitatively) in a large variety of mixtures including the mixtures that we ultimately want a prediction for. The virtual smoke points will include some of these affects, while the smoke points of pure compounds obviously will not. The greater the variety of mixtures in the

training data, the more likely it is that a derived virtual smoke point will be representative of that compound's impact in sample mixtures. The linearity of eq 1 also facilitates and motivates the development of linear QSPR-type models of molecules for which there is no data, pure or blended. While the uncertainty of the blending rule could be fully integrated into (i.e., transferred to) the QSPR model uncertainty if the QSPR training dataset contained enough mixtures, for now, we simply recognize that the blending rule inaccuracy is small relative to the reproducibility of the data points and is therefore neglected.

Many authors, for example,[11,12,27,42] have observed general trends for the sooting propensity of hydrocarbons. This open literature informs us that, when TSI data is sorted by a hydrocarbon class or carbon type and plotted against a carbon number, there is a clear separation between the groups and a positive slope within each one. Normal alkanes have the lowest sooting propensity followed by iso-alkanes and cyclo-alkanes, benzenes, alkylated aromatics, and finally naphthalenes. Based on these observations and some manual initiation of the numerical optimization, the constraints documented in Table 3

Table 3. QSPR Fragments, Coefficients, and Constraints[a]

index	fragment	coefficient	constraint	molecules that contain the fragment (%)
1	$-CH_3$	$a_1 = d_1$	$0 < d_1 < 1$	86%
2	$-CH_2-$	$a_2 = a_1 + d_2$	$0 < d_2 < 1$	63%
3	$>CH-$	$a_3 = a_2 + d_3$	$2 < d_3 < 4$	32%
4	$>C<$	$a_4 = a_3 + d_4$	$0 < d_4 < 3$	8% (five molecules)
5	$-CH_2-$ (ring)	$a_5 = (5 - d_1 - d_6 - d_0)/6$		20%
6	$>CH-$ (ring)	$a_6 = a_5 + d_6$	$0 < d_6 < 3$	14%
7	naphthenic[b]	$a_7 = d_7$	$0 < d_7 < 20$	6% (four molecules)
8	$=CH-$ (aro)	$a_8 = (100 - d_1 - d_7 - 3 \times d_9 - d_0)/10$		38%
9	$=C<$ (aro)	$a_9 = a_8 + d_9$	$0 < d_9 < 12$	37%
0	scale shifter	$a_0 = d_0$	$-10 < d_0 < 0$	100% (65 molecules)

[a]The two constraints of the referee controls are applied globally. [b]One naphthenic fragment is assigned to a molecule if the sum of the $=CH-$ (aro) and $=C<$ (aro) fragments is 10.

are imposed on the structural fragment contributions to the threshold sooting index. A smaller range of allowable values was assigned to fragments that were clearly isolated in the prior works, main effects plots, while a larger range of allowable values was assigned to fragments that were less clearly characterized by prior works. For example, naphthenic compounds very clearly stand apart from alkylated benzenes and di-benzenes in prior works, but quantitatively, it was unclear how much. The coefficient a_0 is an unconstrained scale shift applied to all predictions. The two anchor point constraints (MCH and 1-MN) remove two degrees of freedom (d_5 and d_8), and another pair of constraints corresponding to the referee fuels is applied globally. The regression employs eight variables (six degrees of freedom) to fit the combined datasets, containing 65 molecules and 124 data points.

The terms we call a_0, a_5, and a_6 do not exist in the Lemaire et al.[33] model, which was developed through Bayesian linear regression as done in ref 43, but Lemaire et al. do include four

terms we do not. These terms distinguish an aliphatic carbon with an aromatic group attached to it, from an aliphatic carbon with exclusively other aliphatic groups attached. The training database compiled by Lemaire et al. also differs substantially from this work as it included many oxygenated species and data from different types of experiments. Their so-called unified sooting index is linearly correlated with the fuel equivalent sooting index, equal to the TSI for non-oxygenated molecules. For our model, the predicted TSI for molecules is given in eq 5 where n_i is the number of each fragment type in the molecule. By combining eq 5 with eq 1, the final model is derived, as shown in eq 6. In this equation, the index j refers to molecules in the mixture, i refers to fragments in the molecule, and the remaining terms are as defined in eqs 1 and 5. The model of Lemaire et al.[33] has a similar form, except a_0 does not exist, and their dependent variable is their unified sooting index, instead of the TSI. By summing

$$\text{TSI} = a_0 + \sum_{i=1}^{9} n_i \times a_i \tag{5}$$

$$\text{TSI}_{\text{mix}} = a_0 + \sum_i a_i \times \sum_j x_j \times n_{ij} \tag{6}$$

over j for each data point, a system of 124 equations with "10" unknowns is set up, where 124 is the number of data points, including mixtures and pure molecules, and any multilinear regression software package could determine the set of QSPR model coefficients $\{a_i\}$. However, for the convenience of implementing the constraints that have been discussed, the root mean square (rms) difference between the model and the data was minimized using the GRG nonlinear solver within Microsoft Excel with upper and lower bounds applied to each independent variable, d_j, (see Table 3) and each referee fuel acceptability criteria. For each molecule (j) in our database, an algorithm was used to derive $\{n_{ij}\}$ based on its SMILE[44] formula, and for each molecule within the training dataset, the result of this algorithm was verified manually. The mole fractions were derived from measured volume fractions, molecular weights, and known densities at room temperature.

4. RESULTS

The first step toward evaluating the integrated QSPR/blending model is to compare its predictions relative to the suggested ranges for each of the referee fuels. For 1,3,5-trimethylbenzene, the recommended control range is 54.7−63.1 TSI and our model predicts 63.0 TSI (Lemaire et al.'s model predicts 56.1). For the 40/60%v toluene/iso-octane blend, the recommended control range is 21.0−25.4 TSI, and our model predicts 25.4 TSI (Lemaire et al.'s model predicts 22.2), so the model is barely in control. In the discussion to follow, some of the reasons why the model predicts the upper limit of the referee fuels control specification will be presented, along with its implications relative to the application of the model. Overall, for the coefficients listed in Table 4, the QSPR model underpredicts the average data by 0.5 TSI (mean error). Its mean absolute error is 3.5 TSI, and its rms error is 4.6 TSI. By comparison, the more widely applicable model published by Lemaire et al.[33] throws an rms error of 5.4 when compared against the data used in this work and underpredicts the data by 0.7 TSI.

Figure 6 compares TSI predictions for 1162 molecules relevant to the ongoing work on aviation fuel work in the lab

Table 4. QSPR Model Coefficients

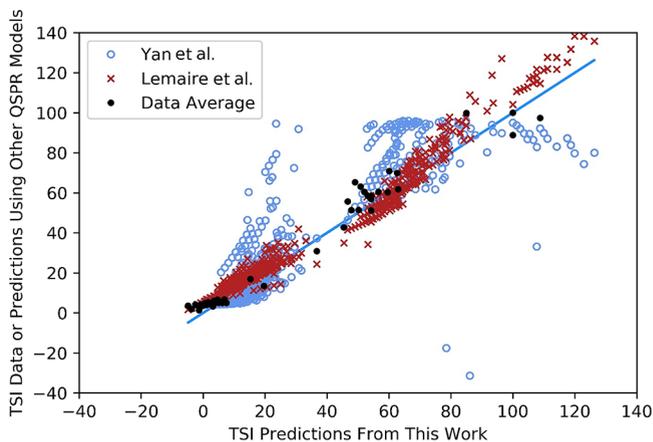| fragment	coefficient
−CH$_3$	$a_1 = 0.187$
−CH$_2$−	$a_2 = 1.187$
>CH−	$a_3 = 5.187$
>C<	$a_4 = 7.781$
−CH$_2$− (ring)	$a_5 = 2.190$
>CH− (ring)	$a_6 = 3.862$
naphthenic	$a_7 = 6.263$
=CH− (aro)	$a_8 = 7.781$
=C< (aro)	$a_9 = 16.362$
scale shifter	$a_0 = -10.00$

Figure 6. Comparison of QSPR model predictions over the target database.

for each of the QSPR models. The models of this work and those of Lemaire et al. are in reasonable agreement with each other, which validates both, but the model published by Yan et al. predicts significantly different TSIs, which validates our motivation for developing this model. Such differences could be the result of differing training data, differing model formulation, or differing regression constraints or objectives. All three models predict the available data reasonably well.

A subset of the data and model predictions corresponding to those points with measurements from more than one research group is plotted in Figure 7. Apart from *n*-butylbenzene, which has a modeled TSI of 49.2 and a measurement average TSI of 65.3, the scatter in the data looks about the same whether it is reflected in the average data or the model result. Another point that is evident from these plots is that the data of Hunt[23] (especially) and Olson et al.[12] trend higher than the data of Mensch et al.,[16] Hui et al.,[29] and this work. These trends are also evident in Figure 8, which compares our model result and all data. The model trends 3.5 TSI low compared to the data from Hunt and 3.1 TSI high compared to our data. Three data points from the original dataset by Hunt (*p*-cymene, *t*-butylbenzene, and triethylbenzenes) were excluded. These data were inconsistent with their recorded smoke points relative to similar molecules within the same dataset. Specifically, tripentylbenzene was excluded based on its molecular weight. The next worst match (also from the Hunt dataset) corresponds to *n*-butylbenzene. That data point was retained because it was within 10 TSI of its duplicate from the Olson et al. dataset and because there is a possibility that a special cause, not captured by this QSPR formulation, was partially responsible for its unusually high TSI measurement. The
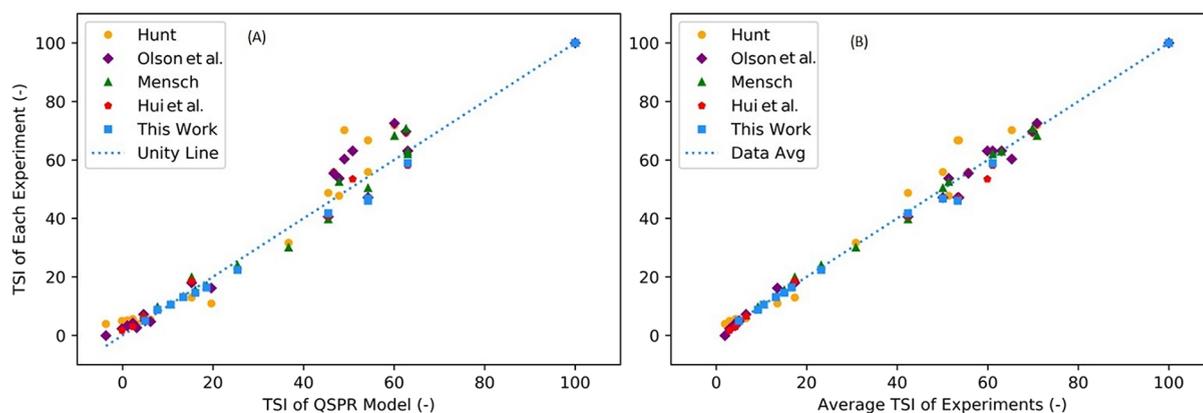
Figure 7. (A, B) Model prediction to data in context with data scatter. Only points for which more than one measurement exists are presented, and the measured results are presented along the vertical axis.
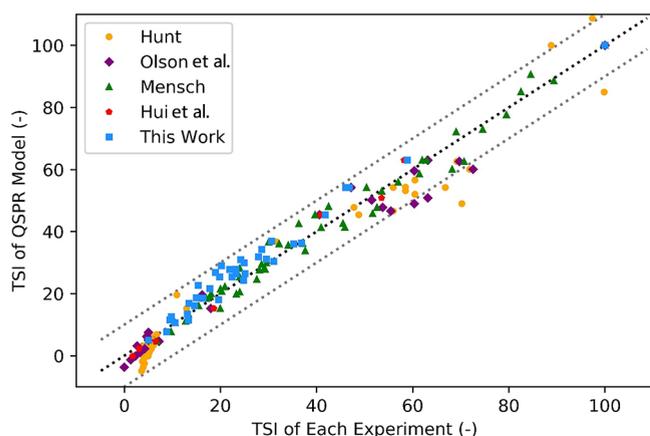

Figure 8. Model to data comparison. All data and predictions are shown. The dark dashed line corresponds to the unity line, while the light dashed lines correspond to ±10 TSI from unity.

Figure 9. Measured and predicted TSI of ASTM reference fuels.

average of the reported data for *n*-butylbenzene is 65.3 TSI, compared to our model prediction of 49.2 and Lemaire et al.'s model prediction of 43.0.

Systematic differences in flame height readings may contribute to the opposing trends (see Figure 4) of the Hunt dataset relative to this work. The more likely source of this difference is the random error in the data point for 1-MN, which is propagated via the experiment's TSI scaling coefficients, *a* and *b* in eq 2, as a seemingly systematic error in the remainder of the dataset (see Figure 2). For example, if the TSI scaling coefficients were derived from a smoke point for 1-MN that is 0.2 mm lower than was used for this work, then the mean TSI for the Hunt dataset would drop from 35.2 to 33.9. For a systematic error in flame height reading to cause that much shift in the mean for the dataset, each of the other readings would have to be too low by 0.27 mm. This observation underscores the need for tight control around the repeatability of the upper anchor point of the TSI scale, and it also underscores a need for more benchmark quality data. Without the Hunt dataset, the coefficient for a naphthalene-like (a_7) molecule could not be determined empirically, and others would lack sufficient experimental variety to justify the regression approach.

The measured and predicted TSI of toluene and each of the toluene/iso-octane reference fuels defined in ASTM D1322 are compared in Figure 9. The blending rule, eq 1 based on the
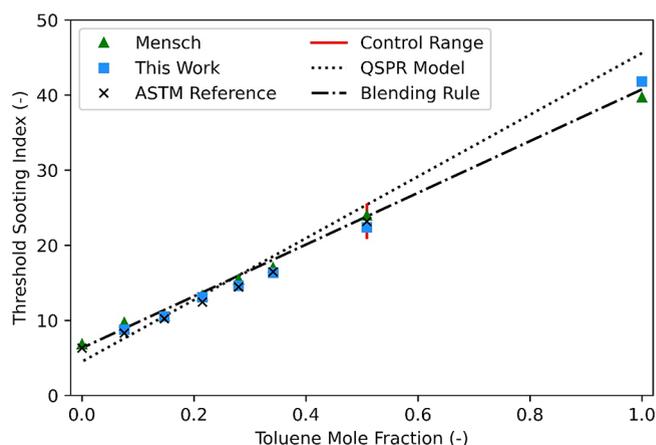
experimentally determined TSI for toluene and iso-octane, is within 1.6 TSI of the data at all points. At 25%v toluene, which strikes the fuel specification limit[1] for both the smoke point and total aromatics, both models are somewhat conservative relative to the data. The QSPR model (eq 6) is 2.0 TSI higher than the data at this important point. Above this point, its error grows more positive as driven by its error for toluene, where the regression QSPR model is driven high by trying to minimize the largest mismatches to data, such as *n*-butylbenzene. The modeled difference between toluene and *n*-butylbenzene is 3 times a_2, the contribution from $-CH_2-$ fragments, and there is a lot of data from alkanes and other alkylated benzenes that suggest that a_2 is small. Therefore, the most impactful way for the regression to reduce the underprediction for *n*-butylbenzene is to increase a_8 or a_9, the aromatic carbon coefficients, which drives the over-prediction for toluene. At 20%v toluene, a reference fuel that matches the smoke point of average petroleum-derived jet fuel,[36] both models are conservative by 1.5 TSI. Both models are still conservative at 10%v toluene, a reference fuel that matches the smoke point of best-case petroleum-derived jet fuel.[25] At a still lower concentration of toluene, the globally regressed QSPR model underpredicts the contribution from iso-octane, perhaps driven by a_3 or a_4, where just five molecules support a_4 within the database that contains the >C< (chain) fragment.

In Figure 10, our QSPR model is compared with measured data for a total of five sets of binary mixtures. Those mixtures
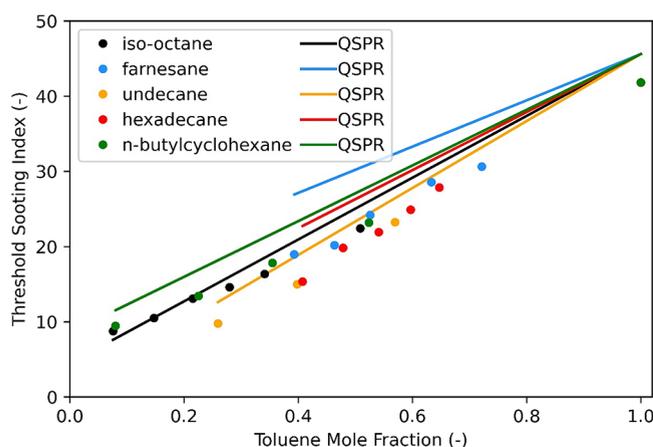
Figure 10. QSPR model assessment for binary mixtures with toluene. Filled circles represent data points. Solid lines represent model results.

include toluene as one of the blended components, and Figure 11 compares it with three sets of measured data for binary
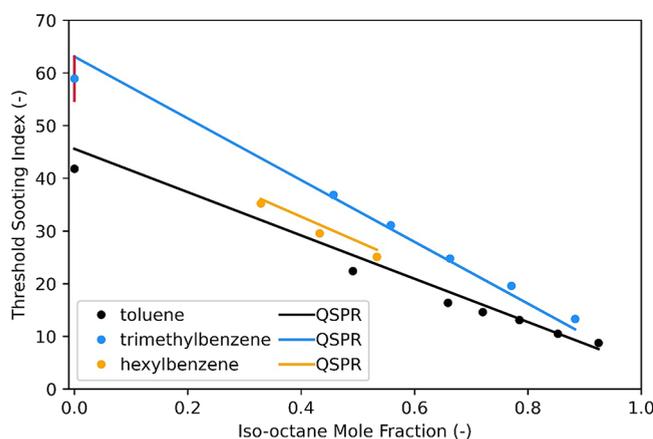

Figure 11. QSPR model assessment for binary mixtures with iso-octane. Filled circles represent data points. Solid lines represent model results.

mixtures that include iso-octane. The toluene/iso-octane blends are included in Figures 9−11. While the data for each binary mixture lay on a straight line, confirming the validity of eq 1, the difference between the QSPR model and the data is evident.

For the iso-octane/toluene blends, the model underpredicts the data at high iso-octane concentrations, but for farnesane/toluene blends, the model overpredicts the data more at high farnesane concentrations than it does at high toluene concentrations (see Figure 10). Taken in isolation, this would hint at a_4 being too low and a_3 being too high since iso-octane has one >C< fragment compared to none in farnesane and farnesane has three >CH− fragments compared to just one in iso-octane, but globally this is not the case. In fact, the incremental difference between a_2 and a_3 was driven to its intuitive maximum by the data regression. Turning now to the undecane/toluene and hexadecane/toluene blends, also shown in Figure 10, it is evident that the model overpredicts the TSI contribution from hexadecane while its contribution from undecane matches the data quite well. Since these two molecules differ only concerning the number of −CH$_2$− fragments they contain, in isolation, this comparison suggests that the modeled contribution to the TSI from the −CH$_2$− fragment is too high. Taken globally, however, the incremental difference between a_1 and a_2 also is regressed up to its intuitive upper boundary. The scale shifter coefficient, a_0, regressed to its lower intuitive limit, which was imposed to restrict the number of molecules for which the predicted TSI would be less than zero. The other five regressed coefficients were near the middle of their respective intuitive ranges. While there is insufficient data to support the inclusion of additional dependent variables, it may be that alkyl fragments generally contribute more to the TSI when they are part of a molecule that also has an aromatic group. That said, the dataset does contain a sufficient variety of methyl-substituted molecules of the same carbon number and class to ascertain that the position of the branch along a chain or ring has an immeasurable impact on its smoke point. For example, the three xylene isomers have the same measured smoke points within a given experimental campaign.

As illustrated in Figure 11, our QSPR model matches the measurements for all mixture fractions of three sets of binary aromatic/iso-octane mixtures that were investigated. The largest difference between the model and the measurement is with respect to 100% trimethylbenzene where the model predicted a TSI of 63.0, 4.1 TSI higher than the measurement. This point also corresponds to a constraint in the regression, where the range of acceptable values for the model prediction of the TSI for trimethylbenzene is shown as a red vertical line in Figure 11. A measurement for pure iso-octane was not

Table 5. Model Validation Summary

fuel/component	mole fraction (-)	measured smoke point (mm)	predicted smoke point (mm)	measured TSI (-)	predicted TSI (-)
n-butylcyclohexane	1.00	45.1 ± 1.3[a]	43.7	8.2	8.6
1,4-dimethylcyclooctane	1.00	36.8 ± 1.2	35.7	10.8	11.2
surrogate 1		26.0 ± 1.0	26.9	18.5	17.8
n-hexylbenzene	0.076				
m-xylene	0.117				
n-octane	0.118				
iso-cetane	0.166				
n-undecane	0.273				
cis decalin	0.249				

[a]The reported uncertainty intervals are the repeatability 95 percentiles based on measurements taken for this work. 1,4-Dimethylcyclooctane was not represented at any concentration within the training dataset. All other molecules were represented in the training dataset but in different mixtures and at different concentrations.

recorded in this work because of excess reading noise caused by the flickering of the flame at its smoke point.

To validate eq 6, measurements and predictions were made corresponding to (i) pure *n*-butylcyclohexane, (ii) dimethylcyclooctane, and (iii) a simple surrogate fuel. The results of these test cases are summarized in Table 5. The predicted TSIs of *n*-butylcyclohexane, dimethylcyclooctane, and a six-component surrogate jet fuel were 8.6, 11.2, and 17.8 TSI, respectively, compared to measurements of 8.2, 10.8, and 18.5, respectively. In each case, the predicted smoke point was well within the reproducibility 95 percentiles of the measurement method, which is 12.9−18.3 for fuel with a molecular weight of 150 g/mol and a nominal TSI of 15.2.

5. APPLICATION TO FUELS

Another application of this model is to predict the smoke point of potential sustainable aviation fuel based on a mix of specific isomer and hydrocarbon class concentration data, as determined by GCxGC/FID-VUV measurements of samples with insufficient volume to measure the smoke point directly.[37,45] For this application, another potentially large uncertainty term arises from undetermined isomer population distributions within any given class. For example, suppose we know that the mole fraction of C3-benzenes is 0.10, but we do not know how much of that is *n*-propylbenzene (least sooting), trimethylbenzene (most sooting), or any of the other structural isomers. By assuming a uniform distribution of isomers within this class, the Tier α methodology effectively assigns a value of 56.8 TSI to this class, while the minimum TSI in this class is 48.0 and the maximum is 63.0. If this class had been represented exclusively by *n*-propylbenzene in the real sample, then our (incorrect) assumption of a uniform distribution would have introduced an error of +0.88 TSI into the prediction, and if this class had been represented exclusively by trimethylbenzene in the real sample, then our assumption would have introduced an error of −0.62 TSI. The 95 percentile of the QSPR model predictions times the mole fraction for this class is ±0.89 TSI, so the isomer uncertainty term is significant relative to the model uncertainty term. Moreover, suppose that the real sample consistently favors more/less branching across all classes present. In that case, the isomer uncertainty error terms will stack up as the composition is reconstructed from the ground up in the model.

In contrast, the random QSPR model errors, weighted by the mole fraction, will sum in quadrature. It is, therefore, possible for the isomer error term to dominate, depending on the sample and how much is known about its composition. Figure 12 provides a comparison between predicted and measured smoke points of three conventional aviation fuels (labeled A-1, A-2, and A-3) and three complex mixtures that have received some attention as potential sustainable aviation fuel (labeled as C-3, C-8, and HEFA/SAK). As evident from the plot, the 95-percentile confidence intervals overlap for five of the six samples. We hypothesize that the miss for A-3 fuel is the result the real fuel having more =C< (aro) and −CH$_3$ fragments and fewer −CH$_2$− fragments than is predicted by our assumed uniform distribution of isomers. Work is already in progress to further utilize vacuum ultraviolet spectroscopy and calibrated time/time stencils to positively identify important isomers in samples.
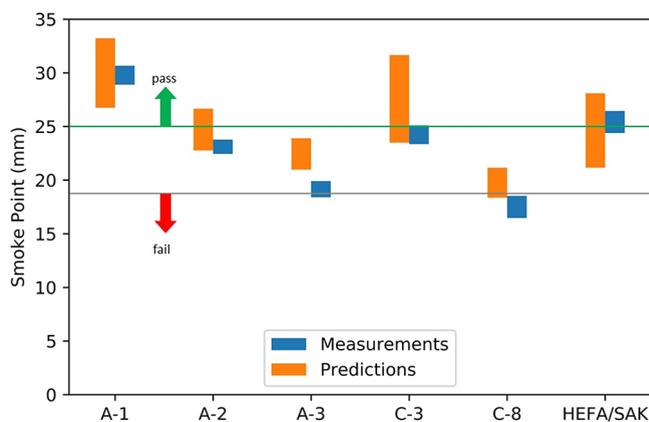

Figure 12. Predicted and measured smoke points of jet fuels and potential sustainable aviation fuel. The bands represent 95-percentile confidence intervals.

6. CONCLUSIONS

A quantitative structure−property relationship model has been developed to predict the threshold sooting index (TSI) of arbitrary mixtures of aliphatic and aromatic hydrocarbons of known composition. The model employs linear contributions from each of eight molecular fragments plus a global shift and a penalty factor for naphthenic compounds. It was constructed from a constrained regression over a composite database by stitching together data from five different experimental campaigns dating back 68 years. The datasets included smoke point data for 1-methylnaphthalene, methylcyclohexane, 1,3,5-trimethylbenzene, and points including toluene and iso-octane at some ratio between 0 and 1. This sub-set of data was used to establish the commonality of the five datasets within acceptable boundaries by transforming the smoke point to the TSI as described in eq 2 and verifying that the TSI of two referee fuels was within acceptable limits. The TSI of trimethylbenzene was controlled to fall between 54.7 and 63.1, and the TSI of 40/60%v toluene/iso-octane, where available, was verified to fall between 21.0 and 25.4. To establish conformance with these referee controls, it was necessary to fine-tune the reported smoke points of 1-methylnaphtalene in the two older datasets, within the boundaries of their respective experimental uncertainties.

Within the composite training dataset, which contained 65 molecules and 124 data points, including simple mixtures, the model was found to match 95% of the data within 8.9 TSI. Validation of the model against *n*-butylcyclohexane, dimethylcyclooctane, and a six-component surrogate jet fuel resulted in predictions of 8.6, 11.2, and 17.8 TSI, respectively, compared to measurements of 8.2, 10.8, and 18.5, respectively. While the agreement between model predictions and measured data for these points is as good or better than can be expected, the size of the 95-percentile band (±8.9 TSI) around the model predictions suggests that there is room for improvement in the quality/consistency of the data used, in the QSPR model, or both. While data from simple mixtures were also used to support regression to the QSPR model coefficients, the blending rule inaccuracy is less than 2 TSI, or 5%, or the total uncertainty.

While the transformation of smoke point data to the threshold sooting index is a powerful tool to erase constant percentage differences between smoke points measured in different experiments and to reduce the impact of constant off-

set type differences between measured smoke points, the tool is very sensitive to the precision of the smoke point measurement for 1-methylnaphthalene—or whatever reference fuel may be chosen to set the upper range of the scale. Using the model trendline as a guide, we found that the data from Hunt trended 3.5 TSI higher. In contrast, the data from this experiment trended 3.1 TSI lower, suggesting that there are limits to the power of this data transformation. Such inconsistency is likely the primary contributor to the overall uncertainty in the model.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.energyfuels.1c03794.

> A concise table of TSI scaling coefficients for each experiment, as used in this work (Table S1); the raw data, be it the smoke point and/or TSI, of each dataset (Tables S2−S6); the repeatability and number of readings taken in our work (Table S2) along with the averages; the reproducibility between the five experimental campaigns (Table S7); and finally, a cross-index of molecules and molecular fragments represented in the composite dataset (Table S8) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Randall C. Boehm − *Department of Mechanical and Aerospace Engineering, University of Dayton, Dayton, Ohio 45469, United States;* ⓞ orcid.org/0000-0003-2983-1337; Email: rboehm1@udayton.edu

Authors

Zhibin Yang − *Department of Mechanical and Aerospace Engineering, University of Dayton, Dayton, Ohio 45469, United States*

Joshua S. Heyne − *Department of Mechanical and Aerospace Engineering, University of Dayton, Dayton, Ohio 45469, United States;* ⓞ orcid.org/0000-0002-1782-9056

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.energyfuels.1c03794

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) ASTM *D7566−21 Standard Specification for Aviation Turbine Fuel Containing Synthesized Hydrocarbons*; ASTM International: West Conshohocken, PA, 2021.

(2) Kosir, S.; Heyne, J.; Graham, J. A machine learning framework for drop-in volume swell characteristics of sustainable aviation fuel. *Fuel* 2020, *274*, 117832.

(3) Boehm, R. C.; Scholla, L. C.; Heyne, J. S. Sustainable alternative fuel effects on energy consumption of jet engines. *Fuel* 2021, *304*, 121378.

(4) Colket, M; Heyne, J. Fuel Effects on Operability of Aircraft Gas Turbine Combustors. August. *Prog. Astronaut. Aeronaut.;* 2021, DOI: 10.2514/4.106040.

(5) Kosir, S.; Stachler, R.; Heyne, J.; Hauck, F. High-performance jet fuel optimization and uncertainty analysis. *Fuel* 2020, *281*, 118718.

(6) DeWitt, M. J.; West, Z.; Zabarnick, S.; Shafer, L.; Striebich, R.; Higgins, A.; Edwards, T. Effect of Aromatics on the Thermal-Oxidative Stability of Synthetic Paraffinic Kerosene. *Energy Fuels* 2014, *28*, 3696−3703.

(7) Clarke, A. E.; Hunter, T. G.; Garner, F. The tendency to smoke of organic substances on burning. *J Inst Pet* 1946, *32*, 627−642.

(8) Dodds, W. J.; Peters, J. E.; Colket, M. B., III; Mellor, A. M. Preliminary Study of Smoke Formed in the Combustion of Various Jet Fuels. *J. Energy* 1977, *1*, 115−120.

(9) Voigt, C.; Kleine, J.; Sauer, D.; Moore, R. H.; Bräuer, T.; Le Clercq, P.; et al. Cleaner burning aviation fuels can reduce contrail cloudiness. *Commun. Earth Environ.* 2021, *2*, 1−10.

(10) Chin, J. S.; Lefebvre, A. H. Influence of Fuel Chemical Properties on Soot Emissions from Gas Turbine Combustors. *Combust. Sci. Technol.* 1990, *73*, 479−486.

(11) Calcote, H. F.; Manos, D. M. Effect of molecular structure on incipient soot formation. *Combust. Flame* 1983, *49*, 289−304.

(12) Olson, D. B.; Pickens, J. C.; Gill, R. J. The effects of molecular structure on soot formation II. Diffusion flames. *Combust. Flame* 1985, *62*, 43−60.

(13) ASTM D1322−19 Standard Test Method for Smoke Point of Kerosene and Aviation Turbine Fuel, West Conshohocken, PA: ASTM International; 2019.

(14) Gill, R. J.; Olson, D. B. Estimation of Soot Thresholds for Fuel Mixtures. *Combust. Sci. Technol.* 1984, *40*, 307−315.

(15) Yan, S.; Eddings, E. G.; Palotas, A. B.; Pugmire, R. J.; Sarofim, A. F. Prediction of Sooting Tendency for Hydrocarbon Liquids in Diffusion Flames. *Energy Fuels* 2005, *19*, 2408−2415.

(16) Mensch, A.; Santoro, R. J.; Litzinger, T. A.; Lee, S.-Y. Sooting characteristics of surrogates for jet fuels. *Combust. Flame* 2010, *157*, 1097−1105.

(17) Haas, FM; Qin, A; Dryer, FL. *"Virtual" smoke point determination of alternative aviation kerosenes by threshold sooting index (TSI) methods.* 50th AIAA/ASME/SAE/ASEE Jt Propuls Conf 2014 2014, DOI: 10.2514/6.2014−3468.

(18) Mensch, A. *A Study on the Sooting Tendency of Jet Fuel Surrogates Using the Threshold Soot Index.* Pennsylvania State University, 2009.

(19) Wu, J.; Song, K. H.; Litzinger, T.; Lee, S. Y.; Santoro, R.; Linevsky, M.; et al. Reduction of PAH and soot in premixed ethylene−air flames by addition of ethanol. *Combust. Flame* 2006, *144*, 675−687.

(20) Wang, H. Formation of nascent soot and other condensed-phase materials in flames. *Proc Combust Inst* 2011, *33*, 41−67.

(21) Minchin, S. Luminous stationary flames: The quantitative relationship between flame dimensions at the sooting point and chemical composition, with special reference to petroleum hydro-carbons. *J Inst Pet Technol* 1931, *17*, 102−120.

(22) Schalla, R. L.; McDonald, G. E. Variation in Smoking Tendency Among Hydrocarbons of Low Molecular Weight. *Ind Eng Chem* 1953, *45*, 1497−1500.

(23) Hunt, R. A. Relation of Smoke Point to Molecular Structure. *Ind Eng Chem* 1953, *45*, 602−606.

(24) Van Treuren, KW. *Sooting Characteristics of Liquid Pool Diffusion Flames.* Princeton, 1978.

(25) Schug, K. P.; Manheimer-Timnat, Y.; Yaccarino, P.; Glassman, I. Sooting Behavior of Gaseous Hydrocarbon Diffusion Flames and the Influence of Additives. *Combust. Flame* 1980, *22*, 235−250.

(26) Tewarson, A. *Prediction of fire properties of materials : Part 1. Aliphatic and aromatic hydrocarbons and related polymers. NBS-GCR-86*. Norwood, CA: Factory Mutual Research; 1986.

(27) Gülder, Ö. L. Influence of hydrocarbon fuel structural constitution and flame temperature on soot formation in laminar diffusion flames. *Combust. Flame* **1989**, *78*, 179−194.

(28) Ladommatos, N.; Rubenstein, P.; Bennett, P. Some effects of molecular structure of single hydrocarbons on sooting tendency. *Fuel* **1996**, *75*, 114−124.

(29) Hui, X.; Liu, W.; Xue, X.; Sung, C. J. Sooting characteristics of hydrocarbon compounds and their blends relevant to aviation fuel applications. *Fuel* **2020**, *287*, 119522.

(30) Behnke, LC; Boehm, RC; Heyne, JS. *Optimization of Sustainable Alternative Fuel Composition for Improved Energy Consumption of Jet Engines*. AIAA Scitech 2022 Forum, 2022, p. submitted.

(31) Poling, BE; Prausnitz, JM, *O'Connell JP*. Properties of Gases and Liquids, Fifth Edition. Fifth Edit. McGraw-Hill Education; 2001.

(32) Barrientos, E. J.; Lapuerta, M.; Boehman, A. L. Group additivity in soot formation for the example of C-5 oxygenated hydrocarbon fuels. *Combust. Flame* **2013**, *160*, 1484−1498.

(33) Lemaire, R.; Le Corre, G.; Nakouri, M. 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* **2021**, *302*, 121104.

(34) McEnally, C. S.; Pfefferle, L. D. Improved sooting tendency measurements for aromatic hydrocarbons and their implications for naphthalene formation pathways. *Combust. Flame* **2007**, *148*, 210−222.

(35) Lemaire, R.; Lapalme, D.; Seers, P. 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* **2015**, *162*, 3140−3155.

(36) Martin, D; Wilkins, P. *Petroleum Quality Information System (PQIS) 2011 Annual Report*. Fort Belvoir, VA: Defense Technical Information Center; 2011.

(37) Yang, Z.; Kosir, S.; Stachler, R.; Shafer, L.; Anderson, C.; Heyne, J. S. A GC × GC Tier α combustor operability prescreening method for sustainable aviation fuel candidates. *Fuel* **2021**, *292*, 120345.

(38) Woodrow, WA. *Development of Methods of Measurement of Tendency to Smoke*. 1st World Pet. Congr., London, UK: OnePetro; 1933, p. WPC-244.

(39) Terry, J. B.; Field, E. *Ind. Eng. Chem., Anal. Ed.* **1936**, *8*, 293.

(40) Yang, Y.; Boehman, A. L.; Santoro, R. J. A study of jet fuel sooting tendency using the threshold sooting index (TSI) model. *Combust. Flame* **2007**, *149*, 191−205.

(41) Washburn, E. W. The Dynamics of Capillary Flow. *Phys Rev* **1921**, *17*, 273.

(42) Li, L.; Sunderland, P. B. An Improved Method of Smoke Point Normalization. *Combust. Sci. Technol.* **2012**, *184*, 829−841.

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

(44) Weininger, D. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.

(45) Heyne, J.; Bell, D.; Feldhausen, J.; Yang, Z.; Boehm, R. Towards fuel composition and properties from Two-dimensional gas chromatography with flame ionization and vacuum ultraviolet spectroscopy. *Fuel* **2022**, *312*, 122709.
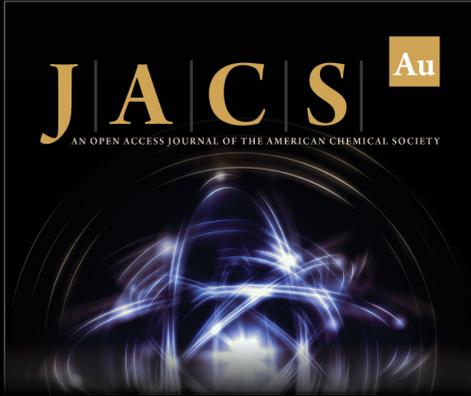
Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle		5. Report Date
		6. Performing Organization Code
7. Author(s)		8. Performing Organization Report No.
9. Performing Organization Name and Address		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplementary Notes		
16. Abstract		
17. Key Words	18. Distribution Statement	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified		

Form DOT F 1700.7 (8-72)   Reproduction of completed page authorized