Project 065(A) Fuel Testing Approaches for Rapid Jet Fuel Prescreening

University of Dayton*
*The project will continue at Washington State University under the direction of Joshua Heyne (Joshua.heyne@wsu.edu).

Project Lead Investigator

Joshua Heyne
Associate Professor
Mechanical Engineering
University of Dayton
300 College Park, Dayton, OH 45458
937 229-5319
Jheyne1@udayton.edu

University Participants

University of Dayton
- P.I.: Joshua Heyne
- FAA Award Number: 13-C-AJFE-UD, Amendments 26, 31, 34 and 42
- Period of Performance: June 5, 2020 to September 30, 2022
- Tasks:
    1. Prescreen sustainable aviation fuels (SAFs).
    2. Develop novel testing methods for the evaluation of SAF candidates.

Project Funding Level

Amendment No. 026: $159,998 (June 5, 2020 to June 4, 2021)
Amendment No. 031: $250,000 (August 11, 2020 to August 10, 2021)
Amendment No. 034: No cost extension (August 10, 2021 to February 10, 2022)
Amendment No. 042: $195,000 (June 14, 2022 to June 13, 2023)
Cost share is from the University of Dayton, VUV Analytics, Greenfield Global, and DLR Germany.

Investigation Team

- Joshua Heyne (University of Dayton) is the P.I., coordinating all team members (both ASCENT and non-ASCENT efforts) and communicating prescreening results with SAF producers.
- Linda Shafer (University of Dayton Research Institute) is a senior research engineer responsible for two-dimensional gas chromatography (GCxGC) measurements of SAFs.
- David Bell (University of Dayton) is a Ph.D. student developing GCxGC vacuum ultraviolet (VUV) detection methods and software.
- John Feldhausen (University of Dayton) is an M.S. student developing GCxGC-VUV methods and software.
- Zhibin (Harrison) Yang (University of Dayton) is a Ph.D. student conducting Tier Alpha prediction and Tier Beta measurements.

Project Overview

This project focuses on further developing the Tier Alpha and Beta test methods, which can help minimize the fuel volume needed for testing and improve a fuel's potential for meeting ASTM approval criteria. Tier Alpha refers to low-volume analytical testing approaches (i.e., GCxGC, nuclear magnetic resonance, and infrared analytical testing). Tier Beta tests focus on directly assessing the physical and chemical properties of a fuel rather than predicting them from GCxGC methods.

Task 1 - Prescreening of Sustainable Aviation Fuels
University of Dayton

Objective
The objective of this task is to develop a tiered prescreening process for new alternative jet fuels that uses low fuel volumes and will improve a fuel's potential to meet ASTM approval criteria. This work facilitates the flow of meaningful information to fuel producers when their production processes are at a low technology readiness level while simultaneously strengthening a producer's readiness for the approval process.

Research Approach
Previous annual reports summarized significant progress toward prescreening SAF candidates. The motivation, conceptual application, detailed description, and examples of this effort were described in publications in peer-reviewed journals in the first 18 months of this project. This report documents four additional peer-reviewed journal articles, published between September 30, 2021 and October 1, 2022. Citations for the articles are listed in the *Publications* section below and are provided as Appendices 1-4:

- Paper 5-. Threshold Sooting Index of Sustainable Aviation Fuel Candidates from Composition Input Alone: Progress Toward Uncertainty Quantification
- Paper 6- Synthetic aromatic kerosene property prediction improvements with isomer specific characterization via GCxGC and vacuum ultraviolet spectroscopy
- Paper 7- Blend Prediction Model for Freeze Point of Jet Fuel Range Hydrocarbons
- Paper 8- A Dataset Comparison Method Using Noise Statistics Applied to VUV Spectrum Match Determinations

In addition to these articles, the team is currently working on documenting efforts completed between June 1, 2020 and August 30, 2022, which has advanced toward our goal of estimating the impact of fuel composition on elastomer/fuel compatibility.

Milestones
- Tier Alpha was performed a total of 99 times.
- Tier Beta was performed a total of 81 times.
- The maximum blending ratio was determined for 32 SAF candidates.
- A refinement strategy was developed for six fully synthetic SAF candidates.

Major Accomplishments
- Determined specific isomer composition down to 0.2% by mass in samples containing less than 5 µL of fuel
- Developed a numerical analysis methodology to conclusively determine whether a chromatogram peak contains more than one analyte
- Developed means to reliably estimate the sooting propensity of candidate fuels
- Developed and validated means to predict the freeze point of mixtures from first principles

Publications
Peer-reviewed Publications

Bell, D. C., Boehm, R. C., Feldhausen, J., & Heyne, J. S. (2022). A data set comparison method using noise statistics applied to vuv spectrum match determinations. *Analytical Chemistry*, *94*(43), 14861–14868. https://doi.org/10.1021/acs.analchem.2c01931

Boehm, R. C., Coburn, A. A., Yang, Z., Wanstall, C. T., & Heyne, J. S. (2022). Blend prediction model for the freeze point of jet fuel range hydrocarbons. *Energy & Fuels*, *36*(19), 12046–12053. https://doi.org/10.1021/acs.energyfuels.2c02063

Feldhausen, J., Bell, D. C., Yang, Z., Faulhaber, C., Boehm, R., & Heyne, J. (2022). Synthetic aromatic kerosene property prediction improvements with isomer specific characterization via GCxGC and vacuum ultraviolet spectroscopy. *Fuel*, *326*, 125002. https://doi.org/10.1016/j.fuel.2022.125002

Boehm, R. C., Yang, Z., & Heyne, J. S. (2022). Threshold sooting index of sustainable aviation fuel candidates from composition input alone: Progress toward uncertainty quantification. *Energy & Fuels*, *36*(4), 1916–1928. https://doi.org/10.1021/acs.energyfuels.1c03794

Boehm, R. C., Yang, Z., Bell, D. C., Feldhausen, J., & Heyne, J. S. (2022). Lower heating value of jet fuel from hydrocarbon class concentration data and thermo-chemical reference data: An uncertainty quantification. *Fuel*, *311*, 122542. https://doi.org/10.1016/j.fuel.2021.122542

Heyne, J., Bell, D., Feldhausen, J., Yang, Z., & Boehm, R. (2022). Towards fuel composition and properties from Two-dimensional gas chromatography with flame ionization and vacuum ultraviolet spectroscopy. *Fuel*, *312*, 122709. https://doi.org/10.1016/j.fuel.2021.122709

Huq, N. A., Hafenstine, G. R., Huo, X., Nguyen, H., Tifft, S. M., Conklin, D. R., Stück, D., Stunkel, J., Yang, Z., Heyne, J. S., Wiatrowski, M. R., Zhang, Y., Tao, L., Zhu, J., McEnally, C. S., Christensen, E. D., Hays, C., Van Allsburg, K. M., Unocic, K. A., … Vardon, D. R. (2021). Toward net-zero sustainable aviation fuel with wet waste–derived volatile fatty acids. *Proceedings of the National Academy of Sciences*, *118*(13), e2023008118. https://doi.org/10.1073/pnas.2023008118

Outreach Efforts
Conference presentation
American Chemical Society Fall 2021 National Meeting & Exposition in San Diego, CA

Awards
Shane Kosir
- Washington State University/Pacific Northwest National Laboratory Distinguished Graduate Research Program Recipient

Joshua Heyne
- 2022 Achievement & Leadership Award, Commercial Aviation Alternative Fuels Initiative
- 2021 Net Good Summit on sustainable travel, honored guest
- 2021 U.S. Frontiers of Engineering Symposium, National Academies of Engineering, selected participant
- 2021 Vision Award for Excellence in Scholarship, School of Engineering, University of Dayton

Student Involvement
David Bell, Ph.D. student, leads this effort.
John Feldhausen, M.S. student, participates in this effort.
Zhibin (Harrison) Yang, Ph.D. student, participates in this effort.
Shane Kosir, M.S. graduate (2021), participated in this effort.
Steven Ivec, M.S. student, participated in this effort.
Christopher Borland, M.S. student, participated in this effort.
Aaron Spelies, undergraduate student, participated in this effort.
Allison Coburn, undergraduate student, participated in this effort.
Conor Faulhaber, undergraduate student, participated in this effort.

Plans for Next Period
Finalize the publications in progress, improve Tier Alpha prediction accuracy, test dielectric constant values for various fuels, and reduce the volume required for Tier Beta measurement.

This project will continue at Washington State University with P.I. Joshua S. Heyne (Joshua.heyne@wsu.edu).

Appendix 1

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

Paper 5: Threshold Sooting Index of Sustainable Aviation Fuel Candidates from Composition Input Alone: Progress Toward Uncertainty Quantification

1.   Introduction

The aviation industry is in the process of reducing its carbon footprint. To date, six processes that convert a specified biological feedstock into a jet fuel blend stock 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 to 50%) with petroleum-derived fuel to create a so-called, drop-in fuel. A drop-in fuel is one that behaves like 100% petroleum fuel within the detectability limits of the normal delivery and consumption processes that exist within the industry. For example, if an arbitrary hydrocarbon mixture were to absorb certain chemical species that have already been absorbed by certain polymeric materials (e.g. O-rings) that exist within the fuel system of an aircraft, those species could be depleted from the polymeric material, causing it to shrink and a corresponding joint that should be sealed by that polymeric material could spring a leak. That would not be acceptable, so there is a requirement for ASTM D7566 fuels to contain at least eight percent 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 as heat is converted to mechanical energy via compressed gas expansion through a turbine [3], the seal compatibility issue is the only one driving aromatics into fuel. 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], include smoking propensity which is added to tier a [5] by this work. Typically, aromatics have an adverse impact on aircraft energy efficiency arising from their low energy density [6] and from a combustion perspective, aromatics are believed to correlate with decreased thermal stability [7] and higher smoking propensity [8] and particulate emissions [9], while particulate emissions are believed to correlate with contrails, which may collectively have a larger impact on radiative forcing than CO2 emissions from aircraft [10]. Moreover, not all aromatics have the same impact on smoke [11,12], thermal stability, or seal swell. Similarly, not all alkanes have the same impact on smoke [13], and it follows that different fuels would have a different smoking propensity. The first revision of the standard test method for smoke point of kerosene and aviation turbine fuel, ASTM D1322 [14] 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 [14], and an overview of it is provided in the experimental methods section of this paper, but for now it is important to understand some logistics. To get a single datapoint requires ~20 minutes of labor to properly prepare the lamp and take the reading. Additionally, labor time is required to get repeat points, acquire samples, manage the inventory of samples, etc. For the sake of discussion, let's say each datapoint costs two hours of labor in total. If the database contains one million hydrocarbon mixtures for which data is desired, which is the case for pre-screening of candidate high-performance fuels, the cost to acquire that data would be two million hours. To get that done in one year would require one thousand full-time employees and ~500 lamps. These are staggering numbers. Even if we used a model to reduce the number of desired datapoints from one million to one thousand, it would still require a full labor-year to get that data. It is desired therefore 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 to predict the sooting tendency of mixtures based on the sooting tendencies of its constituents was introduced by Gill and Olson [15] in 1984 and has since been validated by Yan et al. [16] and Mensch et al. [17]. Their method leverages the threshold sooting index (TSI) which was introduced by Calcote and Manos [12] a year earlier as a tool to normalize smoke point data from different experiments. Equations 1 and 2 depict the blending rule and sooting index, respectively.

$$TSI_{mix} = \sum_i x_i * TSI_i \qquad\qquad (1)$$

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

In these equations, x_i is the mole fraction of the ith component and TSI_i 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 Equation 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'. We do not agree with Li and Sunderland [18], who argued that the dependence on molecular weight is non-physical and their approach to normalize smoke point data would be better than TSI. Additional discussion around establishing a best practice for deriving these coefficients will be presented later in this report.

Generally, the blending rule, Equation 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 influence the uncertainty in TSI_{mix} as a function of mole fraction. The example of preferential wicking or evaporation leading to a 5% off-set between the liquid and vapor phase mole fractions leads to a TSI verses 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 x=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 were used in the blending rule instead of the measured TSI of the pure component [19,20]. However, that approach could lead to several different virtual TSI 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 Equation 1 is not with its neglect of potential differences between vapor and liquid phase mole fraction or possibly synergistic combustion kinetics [21,22], but rather with the dearth of benchmark quality data for pure components, whether directly measured or derived from regression of mixtures (virtual). To date, twelve datasets [8,13,29,30,15,17,23–28] have been published representing a total of 112 hydrocarbons. 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 us. Thirty-nine molecules with trusted smoke point data are saturated hydrocarbons, and twenty-six are alkylated aromatics. Of these sixty-five data points, fifty-nine can be stitched together by using TSI, Equation 3 as a tool to normalize data from three different datasets [13,17,25] into a common basis. Just twenty-five of these molecules are included within our internal database of 1,128 molecules that are to be considered as blend stock for sustainable aviation fuel. Existing data therefore represents 2.2% (25/1128) of desired data and if we include all the available, relevant data then 5% (59/1162) of the combined database could potentially serve as training data to predict the remaining 95%.

Figure 1. TSI Blending Rule Error - Shape Factors for Binary Mixtures. The horizontal axis is the liquid-phase mole fraction of component A, scale 0 to 1. The vertical axis is the threshold sooting index, scale -10 to 70. The upper curves are the TSI of the mixture, where the black line is the blending rule prediction and the the other curves represent the what actual data would look like given certain sources of error. (A) The random error at either end point is ±1 or ±5 TSI. (B) The gas phase mole fraction of component A is 5% 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% benefit for a 50/50 blend. The lower curves are the difference between the two upper curves. The vertical bars in (A) represent random error at each point.

$$100 = a * \left(\frac{142.201}{Sp(1MN)} \right) + b \quad \text{and} \quad 5 = a * \left(\frac{98.189}{Sp(MCH)} \right) + b \qquad (3)$$

Yan et al. [16] used quantitative structure-property relationships (QSPR) employing structural fragment contributions to create models [31] to estimate 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 5thorder 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 declines 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-tert-butylbenzene is negative 31.3 while the predicted TSI of [1,4-dimethyl-1-(3-methylbutyl)pentyl]benzene is 95.8, using the so-called Jobak method. Even within the range of the correlated data, the three methods did not show qualitative agreement with each other. For example, the predicted TSI of 2,2,4,4,6,6,8-heptamethylnonane was 80.4 by the Jobak method and 14.6 by the modified Jobak 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 more to 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 re-scaling, we have generated a new model to predict the TSI of arbitrary molecules that is considerably more versatile than those published by Yan [16].

Other groups have used a QSPR approach to predict [32,33] the threshold sooting index [12], 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 twenty-nine un-oxidized carbon groups, eleven 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.

Within the experimental section of this report a basic description of the ASTM D1322 standard is provided and is followed by a discussion the additional controls (best practices) that are 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 three legacy datasets into a common basis with new data measured in our laboratory. The numerical methods section includes 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 includes a comparison of model results with the data used to train it, as well as a check of model results against measured smoke points for two pure compounds, n-butylbenzene and dimethylcyclooctane, and one simple surrogate fuel with 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 tier-a [5] 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. This model creates the database employed by the tier-a methodology.

2.   Experimental Methods

The experimental apparatus and procedures used in this work and that of Mensch et al. are described thoroughly in ASTM D1322-19. The apparatus consists of a wick-fed lamp that can be purchased from any of a variety of 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 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 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 four 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.

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 [37]. Three years later Terry et al. [38] published details of an improved Factor lamp, which was the experimental apparatus used by Hunt [25] who published in 1953, the most extensive database of smoke point data to date. While there is a variety of small differences between this experiment and the current ASTM standard, the main features of the experiments are common and, in theory [12], their respective smoke point data, once converted to a common TSI scale should be comparable.

Olson et al. [13] had observed a local feature (a shoulder or dip) in plots of fuel consumption rate verses flame height showing up at heights corresponding to the measured smoke points, and 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 showed good correlation (r^2 = 0.94) between their measured TSI's of 28 compounds and those reported by Calcote and Manos [12] who had compiled TSI data from 6 different sources, most extensively the data from Hunt [25].

For this work smoke points were measured 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 supplementary material to this report, Table 1 provides a list of each molecule used, its purity and source. All measurements were taken numerous times, including at least one change of wick throughout the progression of repeat data points.

Table 1. Chemicals Used[†]

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

[†] 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.   Numerical Methods

In previously published articles, the experiment constants labeled as 'a' and 'b' in Equation 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 units scale (analogous to Fahrenheit or Celsius temperature units scale) for every dataset and 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 [12] suggested pure hexane (TSI=2) and pure 1-methylnaphthalene (TSI=100) while Mensch et al. 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 that of Mensch, the earlier works [13,25] included just 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 smoke point or TSI values of interest. While we adopt without endorsement the convention of Mensch et al. [17] to define the TSI scale by setting its value for 1-methylnapthalene to 100 and its value for methylcyclohexane to 5, 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 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 experimental 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, which is given by Equation 4 [14], where 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. Even if we are correct in our suspicion that Equation 4 is off by a factor of two, that level of uncertainty in smoke point measurement of 1-methylnaphthalene (1-MN) is still much higher than desired. 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 ~10 repeat points for most of the data taken to support this project.

$$r = 0.0684 * (\overline{Sp} + 16)\qquad\qquad (4)$$
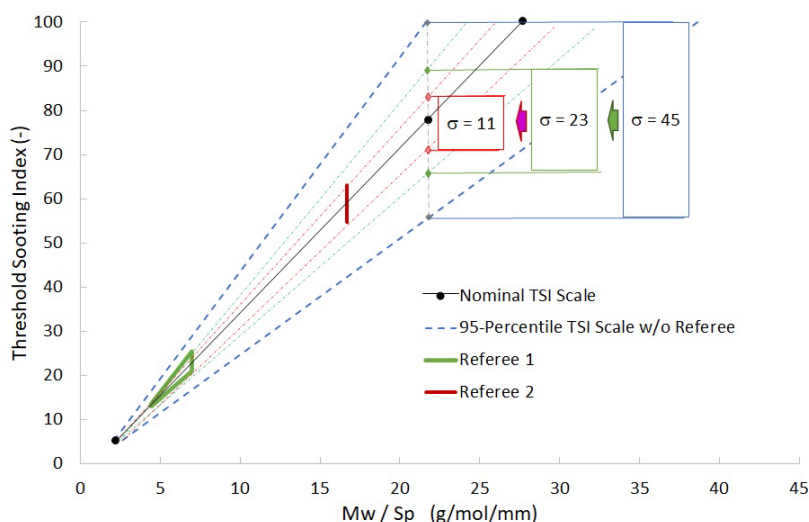
Figure 2. Impact of Reference Data Uncertainty on Experiment's TSI Scale Coefficients Nominal points are taken from this work. The horizontal line at the TSI = 100 was constructed from the nominal point and the 95-percentile repeatability quote from ASTM D1322-19.

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 (Equation 4) for a single point measurement of the smoke point for 1-MN. The dashed, gray, 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 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 ~ ±10 TSI. If we require the TSI of the first and fourth reference fuel blend of ASTM D1322 to be 23.2 ± 2.2 and 13.2 ± 0.3, 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 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.

If the referee control standards are not met initially then 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 Equation 3. Additionally, it should be confirmed that each of the referee criteria are met. The datasets from this work and that of Mensch meet these criteria and can be compared directly. To drive virtual compliance with these criteria for the datasets of Hunt [25] and Olson [13] 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 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), and 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 as Figure 3.

Table 2. Data Summary for Highlighted Fuels

Fuel	Experiment	Sp (mm)	TSI	Experiment	Sp (vSp)[1] (mm)	TSI
1-methylnaphthalene	This work	5.12	100	Olson [13]	5 (4.6)	100
Methylcyclohexane	[2] a = -3.40	43.56	5	a = -2.68	42	5
Toluene	[2] b = 3.727	7.59	41.8	b = 3.285	7	40.6
iso-octane		-	-		38	7.2
1,3,5-trimethylbenzene		7.19	58.9		6	63.1
40/60 %v toluene /iso-octane		14.88	22.4		-	-
1-methylnaphthalene	Mensch [20]	5.5	100	Hunt [25]	5 (4.43)	100
Methylcyclohexane	a = -4.75	40.8	5	a = 1.80	94	5
Toluene	b = 4.051	8.4	39.7	b = 3.059	6	48.8
iso-octane		40.0	6.8		86	5.9
1,3,5-trimethylbenzene		7.3	62.0		6	63.1
40/60 %v toluene /iso-octane		14.57	22.4		-	-

1-methylnaphthalene	Averages	4.9	100	[3] Standard Deviations	0.5	0
Methylcyclohexane		55	5		26	0
Toluene		7.2	42.7		1.0	4.1
iso-octane		55	6.6		27	0.7
1,3,5-trimethylbenzene		6.6	61.8		0.7	2.0
40/60 %v toluene /iso-octane		14.7	23.2		0.2	1.1

1. The number in paratheses corresponds to the virtual smoke point used to normalize legacy datasets.
2. 'a' and 'b' are the TSI scaling coefficients for each experiment and are defined by Equation 2.
3. Standard deviations (reproducibility) based on one value per experimental campaign.

Figure 3. Full TSI Database Supporting QSPR Model Development.

The data shown in Figure 4 has been extracted from the thesis of Mensch [20] and suggests the repeatability of the experiment is an order of magnitude tighter than suggested by Equation 4. While some of this discrepancy can be attributed to signal to noise improvements afforded by taking repeat data points, we suspect there is some error in Equation 4 because we cannot believe that the same operator of the same lamp, employing similarly prepared wicks of the same material would read the flame tip position at the smoke point as 4.0 to 7.0 mm, assuming its nominal height is 5.5 mm. We think the 95 percentiles should be plus or minus one half the result of Equation 4. In the same document, the quoted reproducibility is only 37% higher than the repeatability, but we think the reproducibility, which 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, benchtop levelness) should be substantially higher than the repeatability. If it were 2.7 times higher, that would seem about right and would be consistent with our suspicion that Equation 4 is too high by a factor of two.

The power of the TSI transformation resides in two important points. Systematic differences in smoke point data between experimental campaigns are attenuated appreciably by the transformation and the blending rule described by Equation 1

holds true [15,39]. Figure 4 provides one example comparing the result of Equation 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 threshold sooting index. The data provided by Mensch 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 different vent hole diameter). For that dataset, the imposed 1 mm offset in smoke point results in an average smoke point error of 11%, while the average error in TSI is 3.4%. The imposed 7% offset in smoke point results in a 1 mm smoke point difference on average and no difference at all in TSI.
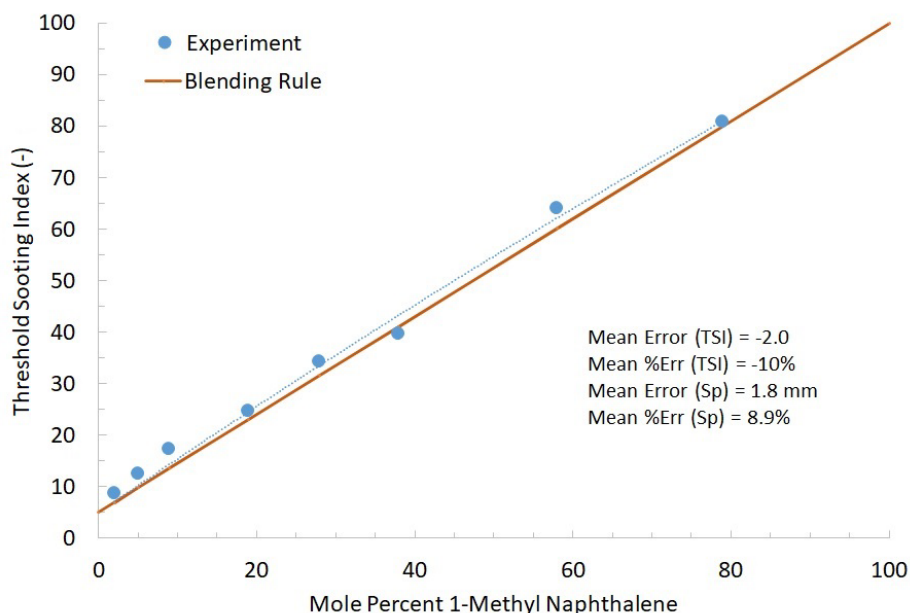
Figure 4. Evaluation of Linear Blending Rule, Equation 1 Data from Mensch [20] for binary mixtures of 1-methylnaphthalene and methylcyclohexane.
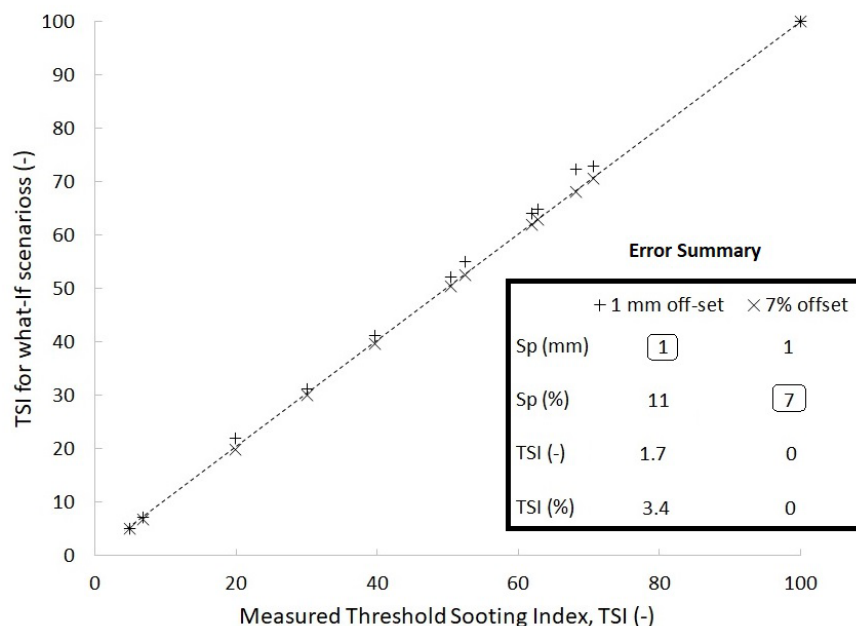
Figure 5. Partial Erasure of Systematic Differences Between Datasets Operations performed on data published by Mensch [20]. The dashed line is the original data. The + symbols are the result of adding 1 mm to each reported smoke point. The x symbols are the result of adding 7% to each reported smoke 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, Equation 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 at 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 wicking rate is described by the Washburn law [40], 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 it would be if the wicking rate were to be increased or 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 of pure A that could be used in Equation 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 the effect that a component has 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 Equation 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 have observed the following generic trend for sooting propensity of hydrocarbons. Normal-alkanes have the lowest sooting propensity followed by iso-alkanes and cyclo-alkanes then benzene followed by alkylated aromatics and finally naphthalenes, and within each class the sooting propensity increases with carbon number. Based on these observations, the constraints documented in Table 3 are imposed on the structural fragment contributions to the threshold sooting index. The coefficient, a_0 is an unconstrained scale shift applied to all predictions. The two anchor point constraints (MCH & 1-MN) remove two degrees of freedom (d_5 and d_8), and another pair of constraints corresponding to the referee fuels are applied globally. The regression employs eight variables (six degrees of freedom) to fit the combined datasets which contains 65 molecules and 124 data points. For the model of Lemaire et al. [33], no constraints were applied and the terms we call a_0, a_5 and a_6 do not exist, but four additional terms exist that distinguish an aliphatic carbon with an aromatic group attached to if from an aliphatic carbon with only other aliphatic groups attached to it. The training database compiled by Lemaire also differs substantially from this work, as it included many oxygenated species as well as data from different types of experiments. Their so-called unified sooting index is linearly correlated with TSI. For our model, the predicted TSI for molecules is given by Equation 5 where n_i is the number of each fragment type in the molecule. By combining Equation 5 with Equation 1 the final model is derived, and it is shown in Equation 6. In this equation, the index, j refers to molecules in the mixture and i refers to fragments in the molecule and the remaining terms are as defined by Equations 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 TSI.

$$TSI = a_0 + \sum_{i=1}^{9} n_i * a_i \qquad (5)$$

$$TSI_{mix} = a_0 + \sum_i a_i * \sum_j x_j * n_{ij} \qquad (6)$$

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

Index	Fragment	Coefficient	Constraint	Molecules that contain the fragment (%)
1	-CH3	$a_1 = d_1$	$0 < d_1 < 1$	86%
2	-CH2-	$a_2 = a_1 + d_2$	$0 < d_2 < 1$	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% (5 molecules)
5	-CH2- (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[††]	$a_7 = d_7$	$0 < d_7 < 20$	6% (4 molecules)
8	=cH-(aro)	$a_8 = (100 - d_1 - d_7 - 3*d_9 - d_0) / 10$		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)

[†]The two constraints of the referee controls are applied globally. [††]One naphthenic fragment is assigned to a molecule if the sum of the =cH-(aro) and =c<(aro) fragments is ten.

By summing over j for each datapoint, a system of 124 equations with 'ten' unknowns is set up, where 124 is the number of data points, including mixtures and pure molecules, and the set of QSPR model coefficients {a_i} could be determined by any multi-linear regression software package. However, for the convenience of implementing the constraints that have been discussed, the 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_i (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 [41] 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 63.1 TSI (Lemaire's model predicts 56.1). For the 40%/60% toluene/iso-octane blend the recommended control range is 21.0-25.4 TSI and our model predicts 25.4 TSI (Lemaire's model predicts 22.2), so the model is barely in control. In the discussion to follow some of the reasons behind why the model predicts the upper limit of the referee fuels control specification will be presented, along with its implications relative to 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 3.6 TSI and its rms error is 4.7 TSI (the rms error of Lemaire's model is 5.6). Relative to 1162 molecules relevant to this work, an overall comparison between the predictions made by each of the QSPR models is presented in Figure 6. While the trends predicted by this work and that of Lemaire are qualitatively similar, clearly there are significant quantitative differences between all three models. 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.

Table 4. QSPR Model Coefficients.

| Fragment	Coefficient
-CH3	$a_1 = 0.188$
-CH2-	$a_2 = 1.188$
>CH-	$a_3 = 5.188$
>C<	$a_4 = 7.174$
-CH2- (ring)	$a_5 = 2.191$
>CH- (ring)	$a_6 = 3.855$
naphthenic	$a_7 = 6.047$
=cH-(aro)	$a_8 = 7.807$
=c<(aro)	$a_9 = 16.372$
Scale shifter	$a_0 = -10.00$

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 weather it is reflected about the average data or the model result. Another point that is evident from these plots that that the data of Hunt [25] (especially) and Olson [13] trend higher than the data of Mensch [17] and this work. These trends are also evident in Figure 8 which shows a comparison between 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 within the original dataset by Hunt, p-cymene, t-butylbenzene, and triethylbenzenes were excluded based on inconsistency of their recorded smoke points relative to similar molecules within the same dataset and one point, tripentylbenzenes was excluded based on its molecular weight. The next worst match (also from the Hunt dataset) corresponds to n-butylbenzene. That datapoint was retained because it was within 10 TSI of its duplicate from the Olson 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 average of the reported data for n-butylbenzene is 65.3 TSI, compared to our model prediction of 49.2 and Lemaire's model prediction of 43.0.
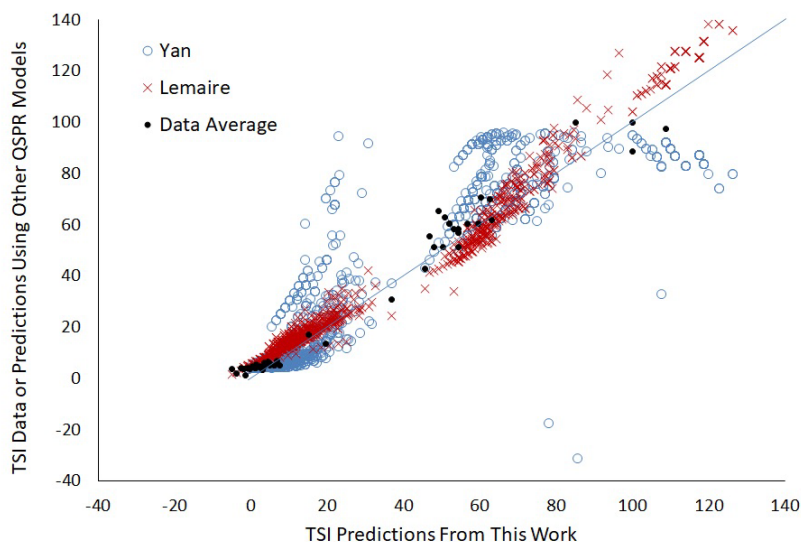
Figure 6. Comparison of QSPR Model Predictions Over The Target Database.
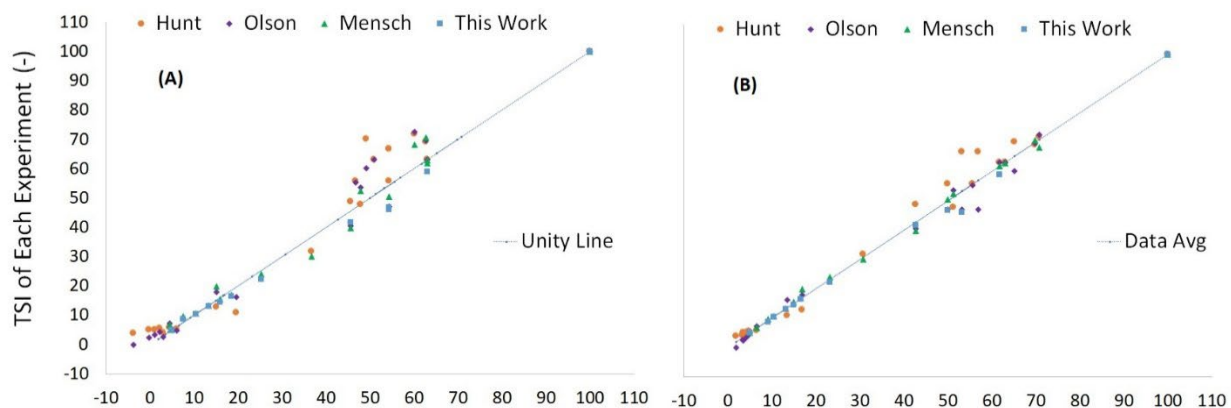
Figure 7. Model Prediction to Data in Context with Data Scatter Only points for which more than one measurement exists are presented, and the measured results are presented along the vertical axis. (A) QSPR model results are presented along the horizontal axis. (B) Average data are presented along the horizontal axis.
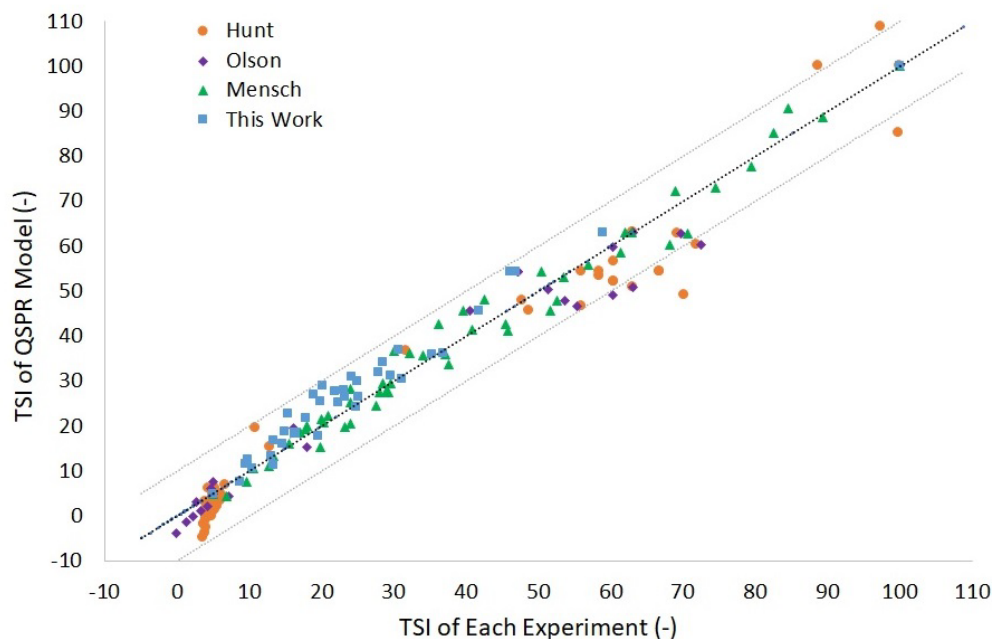
Figure 8. Model to Data Comparison All data and predictions are shown. The dashed lines correspond to the unity line and it ±10 TSI.

Although it is possible that a systematic difference in flame height readings contributes to the opposing trends (see Figure 4) of the Hunt dataset relative to this work, the more likely source is random error in the datapoint for 1-MN which is propagated via the experiment's TSI scaling coefficients, 'a' and 'b' in Equation 2, as 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, 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) 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, Equation 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 smoke point and total aromatics, both models are somewhat conservative relative to the data. The QSPR model, Equation 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 QSPR model, which is a regressed 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 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 overprediction for toluene. At 20%v toluene, a reference fuel which matches the smoke point of average petroleum-derived jet fuel [36], both models are conservative by 1.5 TSI. At 10%v toluene, a reference fuel which matches the smoke point of best-case petroleum-derived jet fuel [25], both models are still conservative. At 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 a_4 is supported by just 5 molecules within the database that contain the >C< (chain) fragment.
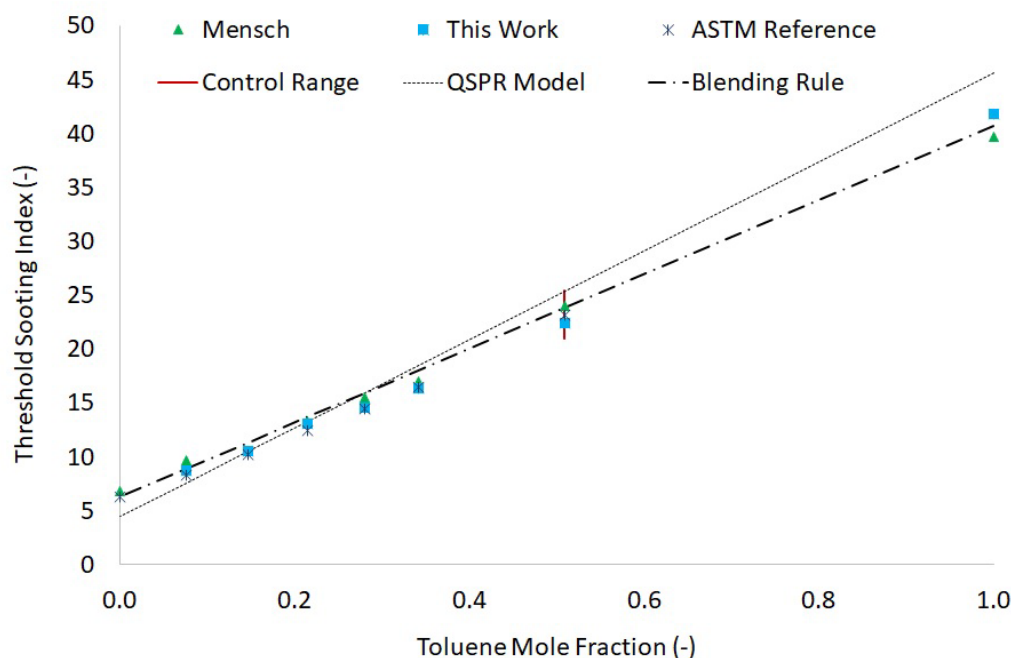
Figure 9. Measured and Predicted TSI of ASTM Reference Fuels.

In Figure 10, our QSPR model is compared with measured data for a total of five sets of binary mixtures that include toluene as one of the blended components, and in Figure 11 it is compared with three sets of measured data for binary mixtures that include iso-octane. The toluene/iso-octane blends are included in Figures 9, 10 and 11. While the data for each binary mixture lay on a straight line, confirming the validity of Equation 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 isooctane 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 shown also 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 with respect to the number of -CH2- fragments they contain, in isolation this comparison suggests the modeled contribution to TSI from the -CH2- 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 inclusion of additional dependent variables, it may be that alkyl fragments generally contribute more to 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 immeasurable impact on its smoke point. For example, the three isomers of xylene have the same measured smoke points within a given experimental campaign.
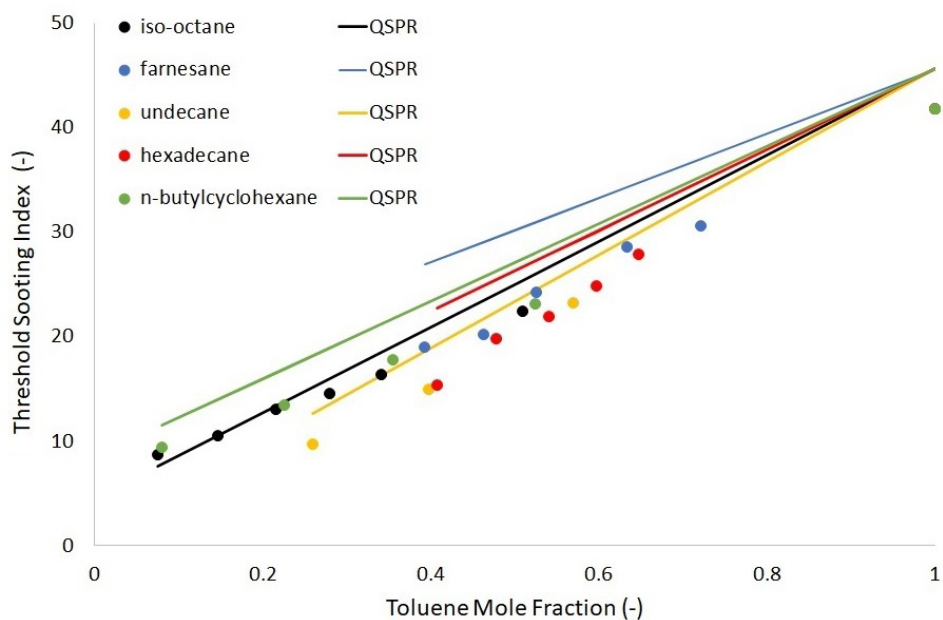
Figure 10. QSPR Model Assessment for Binary Mixtures with Toluene Filled circles represent data points. Solid lines represent model results.
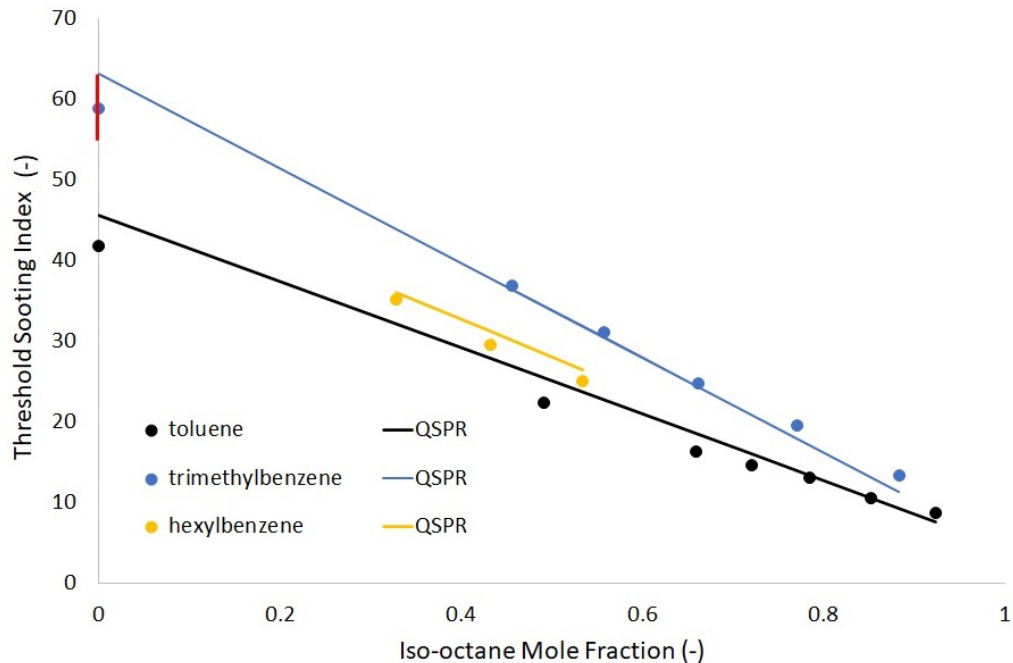
Figure 11. QSPR Model Assessment for Binary Mixtures with Iso-octane Filled circles represent data points. Solid lines represent model results.

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 63.1, 4.2 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 recorded in this work because of excess reading noise caused by flickering of the flame at its smoke point.

To validate Equation 6, measurements and predictions were made corresponding to (1) pure n-butylcyclohexane, (2) dimethylcyclooctane, and (3) a simple surrogate fuel. The results of these test cases are summarized in Table 5. The predicted TSI of n-butylcyclohexane, dimethylcyclooctane and a six-component surrogate jet fuel were 8.6, 11.2 and 17.7 TSI, respectively compared to measurements of 8.2, 10.8 and 18.5. In each case the predicted smoke point was well within the reproducibility 95-percentiles of the measurement method, which is 12.9 to 18.3 for fuel with a molecular weight of 150 g/mol and a nominal TSI of 15.2.

Another application of this model is to predict the smoke 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 smoke point directly. 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

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[†]	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	27.0	18.5	17.7
n-hexylbenzene	0.076				
m-xylene	0.117				
n-octane	0.118				
iso-cetane	0.166				
n-undecane	0.273				
cis decalin	0.249				

[†]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.

sooting),trimethylbenzene (most sooting), or any of the other structural isomers. By assuming a uniform distribution of isomers within this class, the tier-a methodology effectively assigns a value of 56.8 TSI to this class, while the minimum TSI in this class 48.0 and the maximum is 63.1. If this class had been represented exclusively by n-propylbenzene in the real sample, 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, our assumption would have introduced an error of -0.63 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 indeed significant relative to the model uncertainty term. Moreover, if the real sample consistently favors more/less branching across all classes present then the isomer uncertainty error terms will stack up as the composition is reconstructed from the ground up in the model, while the random QSPR model errors, weighted by 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 -CH3 fragments and fewer -CH2- 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.
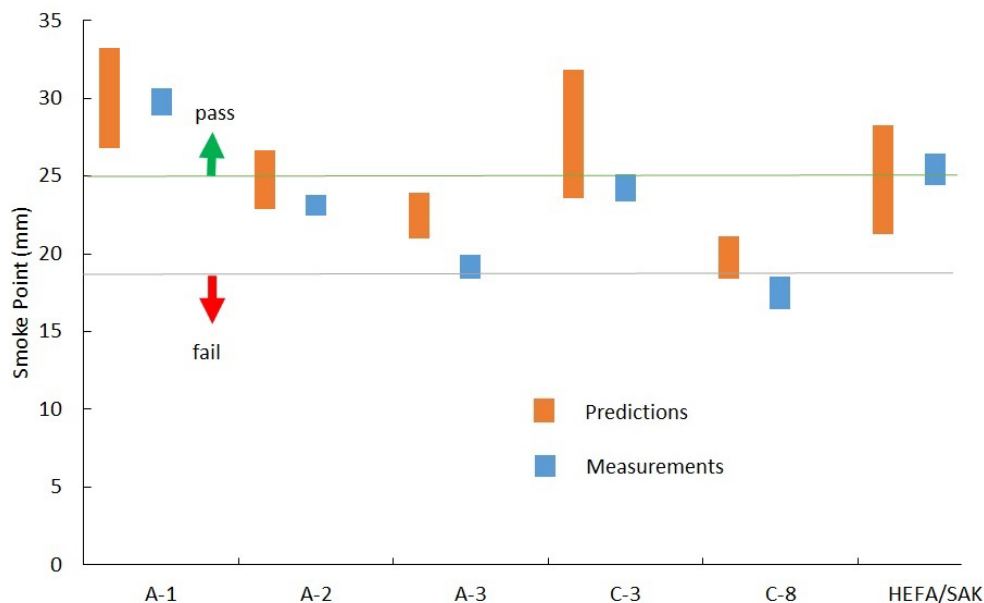
Figure 12. Predicted and Measured Smoke Points of Jet Fuels and Potential Sustainable Aviation Fuel.

5. Conclusion

A comprehensive model has been developed to the 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 which was constructed by stitching together data from four different experimental campaigns dating back 68 years. Each of 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 established commonality of the four datasets within acceptable boundaries by using the transformation from smoke point to TSI as described by Equation 4, and verification that the TSI of two referee fuels where 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 datapoints including simple mixtures, the model was found to match 95% of the data within 8.9 TSI. Validation of the model against n-butylcyclohexane, dimethylclyclooctane and a six-component surrogate jet fuel, resulted in predictions of 8.6, 11.2 and 17.7 TSI, respectively compared to measurements of 8.2, 10.8 and 18.5. 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 there is room for improvement in the quality / consistency of the data used, or 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 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 the data from Hunt trended 3.5 TSI higher while the data from this experiment trended 3.1 TSI lower, suggesting there are limits to the power of this data transformation and such inconsistency is likely the primary contributor to overall uncertainty in the model.

6.   Acknowledgements

7.   References

[1]   ASTM D7566 - 21 Standard Specification for Aviation Turbine Fuel Containing Synthesized Hydrocarbons, West Conshohocken, PA: ASTM International; 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. https://doi.org/10.1016/j.fuel.2020.117832.
[3]   Boehm RC, Scholla LC, Heyne JS. Sustainable alternative fuel effects on energy consumption of jet engines. Fuel 2021;304:121378. https://doi.org/10.1016/j.fuel.2021.121378.
[4]   Colket M, Heyne J. Fuel Effects on Operability of Aircraft Gas Turbine Combustors. August. AIAA, Progress in Astronautics and Aeronautics; 2021. https://doi.org/10.2514/4.106040.
[5]   Yang Z, Kosir S, Stachler R, Shafer L, Anderson C, Heyne JS. A GC × GC Tier α combustor operability prescreening method for sustainable aviation fuel candidates . Fuel 2021;292:120345. https://doi.org/10.1016/j.fuel.2021.120345.
[6]   Kosir S, Stachler R, Heyne J, Hauck F. High-performance jet fuel optimization and uncertainty analysis. Fuel 2020;281:118718. https://doi.org/10.1016/j.fuel.2020.118718.
[7]   DeWitt MJ, West Z, Zabarnick S, Shafer L, Striebich R, Higgins A, et al. Effect of Aromatics on the Thermal-Oxidative Stability of Synthetic Paraffinic Kerosene. Energy and Fuels 2014;28:3696–703. https://doi.org/10.1021/EF500456E.
[8]   Clarke AE, Hunter TG, Garner F. The tendency to smoke of organic substances on burning. J Inst Pet 1946;32:627–42.
[9]   Dodds WJ, Peters IE, Colket MB, Mellor AM. Preliminary Study of Smoke Formed in the Combustion of Various Jet Fuels. J ENERGY 1977;1:115–20. https://doi.org/10.2514/3.47933.
[10]   Voigt C, Kleine J, Sauer D, Moore RH, Bräuer T, Le Clercq P, et al. Cleaner burning aviation fuels can reduce contrail cloudiness. Commun Earth Environ 2021 21 2021;2:1–10. https://doi.org/10.1038/s43247-021-00174-y.
[11]   Chin JS, Lefebvre AH. Influence of Fuel Chemical Properties on Soot Emissions from Gas Turbine Combustors. Combust Sci Technol 1990;73:479–86. https://doi.org/10.1080/00102209008951664.
[12]   Calcote HF, Manos DM. Effect of molecular structure on incipient soot formation. Combust Flame 1983;49:289–304. https://doi.org/10.1016/0010-2180(83)90172-4.
[13]   Olson DB, Pickens JC, Gill RJ. The effects of molecular structure on soot formation II. Diffusion flames. Combust Flame 1985;62:43–60. https://doi.org/10.1016/0010-2180(85)90092-6.
[14]   ASTM D1322 - 19 Standard Test Method for Smoke Point of Kerosene and Aviation Turbine Fuel, West Conshohocken, PA: ASTM International; 2019.
[15]   Gill RJ, Olson DB. Estimation of Soot Thresholds for Fuel Mixtures. Combust Sci Technol 1984;40:307–15. https://doi.org/10.1080/00102208408923814.
[16]   Yan S, Eddings EG, Palotas AB, Pugmire RJ, Sarofim AF. Prediction of Sooting Tendency for Hydrocarbon Liquids in Diffusion Flames. Energy and Fuels 2005;19:2408–15. https://doi.org/10.1021/EF050107D.
[17]   Mensch A, Santoro RJ, Litzinger TA, Lee S-Y. Sooting characteristics of surrogates for jet fuels. Combust Flame 2010;157:1097–105. https://doi.org/10.1016/j.combustflame.2010.02.008.
[18]   Li L, Sunderland PB. An Improved Method of Smoke Point Normalization. Combust Sci Technol 2012;184:829–41. https://doi.org/10.1080/00102202.2012.670333.
[19]   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. https://doi.org/10.2514/6.2014-3468.
[20]   Mensch A. A Study on the Sooting Tendency of Jet Fuel Surrogates Using the Threshold Soot Index. Pennsylvania State University, 2009.
[21]   Wu J, Song KH, Litzinger T, Lee SY, Santoro R, Linevsky M, et al. Reduction of PAH and soot in premixed ethylene–air flames by addition of ethanol. Combust Flame 2006;144:675–87. https://doi.org/10.1016/J.COMBUSTFLAME.2005.08.036.
[22]   Wang H. Formation of nascent soot and other condensed-phase materials in flames. Proc Combust Inst 2011;33:41–67. https://doi.org/10.1016/J.PROCI.2010.09.009.

[23]   Minchin S. Luminous stationary flames: The quantitative relationship between flame dimensions at the sooting point and chemical composition, with special reference to petroleum hydrocarbons. J Inst Pet Technol 1931;17:102–20.

[24]   Schalla RL, McDonald GE. Variation in Smoking Tendency Among Hydrocarbons of Low Molecular Weight. Ind Eng Chem 2002;45:1497–500. https://doi.org/10.1021/IE50523A038.

[25]   Hunt RA. Relation of Smoke Point to Molecular Structure. Ind Eng Chem 2002;45:602–6. https://doi.org/10.1021/IE50519A039.

[26]   Van Treuren KW. Sooting Characteristics of Liquid Pool Diffusion Flames. Princeton, 1978.

[27]   Schug KP, Manheimer-Timnat Y, Yaccarino P, Glassman I. Sooting Behavior of Gaseous Hydrocarbon Diffusion Flames and the Influence of Additives. Combust Flame 1980;22:235–50. https://doi.org/10.1080/00102208008952387.

[28]   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.

[29]   Gülder ÖL. Influence of hydrocarbon fuel structural constitution and flame temperature on soot formation in laminar diffusion flames. Combust Flame 1989;78:179–94. https://doi.org/10.1016/0010-2180(89)90124-7.

[30]   Ladommatos N, Rubenstein P, Bennett P. Some effects of molecular structure of single hydrocarbons on sooting tendency. Fuel 1996;75:114–24. https://doi.org/10.1016/0016-2361(94)00251-7.

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

[32]   Barrientos EJ, Lapuerta M, Boehman AL. Group additivity in soot formation for the example of C-5 oxygenated hydrocarbon fuels. Combust Flame 2013;160:1484–98. https://doi.org/10.1016/J.COMBUSTFLAME.2013.02.024.

[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. https://doi.org/10.1016/J.FUEL.2021.121104.

[34]   McEnally CS, Pfefferle LD. Improved sooting tendency measurements for aromatic hydrocarbons and their implications for naphthalene formation pathways. Combust Flame 2007;148:210–22. https://doi.org/10.1016/j.combustflame.2006.11.003.

[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;162:3140–55. https://doi.org/10.1016/J.COMBUSTFLAME.2015.03.018.

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

[37]   Woodrow WA. Development of Methods of Measurement of Tendency to Smoke. 1st World Pet. Congr., London, UK: OnePetro; 1933, p. WPC-244.

[38]   Terry JB, Field E. . Ind Eng Chem Anal Ed 1936;8:293.

[39]   Yang Y, Boehman AL, Santoro RJ. A study of jet fuel sooting tendency using the threshold sooting index (TSI) model. Combust Flame 2007;149:191–205. https://doi.org/10.1016/j.combustflame.2006.11.007.

[40]   Washburn EW. The Dynamics of Capillary Flow. Phys Rev 1921;17:273. https://doi.org/10.1103/PhysRev.17.273.

[41]   Weininger D. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. J Chem Inf Comput Sci 1988;28:31–6. https://doi.org/10.1021/CI00057A005.

Appendix 2

Synthetic aromatic kerosene property prediction improvements with isomer specific characterization via GCxGC and vacuum ultraviolet spectroscopy

Paper 6: Synthetic aromatic kerosene property prediction improvements with isomer specific characterization via GCxGC and vacuum ultraviolet spectroscopy

Nomenclature

A_i :	area percentage of ith peak per ChromSpace
A_{class}:	area percentage of all peaks within a hydrocarbon group or class per ChromSpace
ASTM:	ASTM International
ATJ:	alcohol to jet
CI:	confidence interval
FID:	flame ionization detector
FTIR:	Fourier transform infrared spectroscopy
GC:	one-dimensional gas chromatography or gas chromatography
GC x GC:	two-dimensional gas chromatography
HEFA:	hydro processed esters fatty acids
LHV:	lower heating value or heat of combustion
m:	modulation number
MS:	mass spectroscopy
n:	representative (average) modulation
nvPM:	non-volatile particulate matter
NIST:	National Institute of Standards and Technology
NJFCP:	National Jet Fuel Combustion Program
NMR:	nuclear magnetic resonance
PIONA:	paraffins, olefins, naphthenes, and aromatics
QSPR:	quantitative structure property relationships
r^2:	correlation coefficient squared
R^2:	coefficient of determination
SAF:	sustainable aviation fuel
SAK:	synthetic aromatic kerosene
T:	temperature
TSI:	threshold sooting index
VUV:	vacuum ultraviolet light detector
Y_i:	mass percentage of ith analyte
%m:	mass percentage
%v:	volume percentage
δ:	fixed time interval over which absorbance was averaged
ρ:	density
σ:	surface tension
σ_{Y_i}:	analyte quantification uncertainty
σ_z:	root property data uncertainty
σ_{isomer}:	isomeric uncertainty

1.   Introduction

The reduction of anthropogenic emissions from the transportation sector has increased interest in recent years [1]. The aviation industry consumed ~400 billion liters of jet fuel globally in 2019, comprising ~10% of greenhouse gas emissions from transportation [1–3]. Current predictions show flight demand doubling from 2010 levels by 2050 [2,4], while simultaneously, airlines continue to pledge to 50% carbon reductions by 2050 [2]. Sustainable aviation fuels (SAFs) have been identified as the most viable option to achieve these desired carbon displacements [5] to address the rising airline industry and environmental goals.

On 1 December 2021, the United Airlines 737 MAX 8 aircraft flew from Chicago to Washington D.C. marking the first passenger flight, with one of two engines powered by 100% SAF [6]. The plane was propelled on a fuel blend of (1) World Energy's hydroprocessed esters fatty acids (HEFA) (ASTM D7566 A2) [7] and (2) Virent's synthetic aromatic kerosene (SAK) blended at 79 %v and 21 %v, respectively. As the name suggests, SAK is composed primarily of aromatics, which from a compositional standpoint, sharply contrasts with the composition of several other qualified SAFs. SAK is one of the only SAFs to include any aromatic component. Aromatics are associated with higher non-volatile particulate matter (nvPM) or soot

emissions, and nvPM is believed to be the primary nucleation source for aviation contrails [8]. Contrails, in turn, are suggested to be the dominant radiative forcing agent of aviation instead of CO_2 emissions alone [1]. However, not all aromatics are equivalent, with naphthalenes having higher sooting potentials [9]. Relatedly, within the SAF community, there is broad interest in compositions that can remain 'drop-in' while minimizing nvPM.

Fuels are required to have 8.4%v aromatics per ASTM D7566 to remain fungible with existing aircraft fueling infrastructure. This requirement, among others (e.g. density), often limits the amount of SAF that can be blended with petroleum-derived Jet A because SAF historically has no aromatic content. Alternative to Jet A blending, SAK provides an entirely sustainable option to achieve the aromatics requirement. The primary concern that aromatics address is material compatibility [10,11]. Aromatics offer higher density, enabling blending with lower density fuels like hydroprocessed esters fatty acids (HEFA) and Alcohol to Jet (ATJ) fuels [12], and potentially a blended dielectric constant in line with conventional fuels.

Numerous low volume (<1mL) hydrocarbon compositional analysis methods reduce SAF scale-up risks and streamline the various qualification processes [3,13–19]. Collectively, these technologies provide producers with critical property predictions that can guide feedstock-conversion engineering at earlier technology readiness levels thereby streamlining production up-scaling investment decisions. For example, a wet waste volatile fatty production process was recently guided in part by these analyses, leading to a fuel technology readiness level, in less than one year, sufficient to support ASTM D4064 testing requirements [3].

Multidimensional gas chromatography is a relatively mature technology capable of separating analytes in complex solutions and remains common in the fuel characterization community. Superposing columns with a modulator between columns affords greater separations than those afforded through single-column configurations [13,18]. Kilaz et al. provide a complete review of various analytical techniques [13], e.g., flame ionization detection (FID), mass spectroscopy (MS), Fourier transform infrared spectroscopy (FTIR), nuclear magnetic resonance (NMR), and concluded that multidimensional chromatography offers the best information to aid jet fuel composition to property relationships. These GC x GC configurations require 'stencil' calibration using MS results and reference samples. Once calibrated, stencils can determine the boundaries of various hydrocarbon classes and carbon numbers [20]. However, stencil techniques, even when coupled with MS detectors, cannot determine structural differences between most isomers [13]. For this reason, vacuum ultraviolet spectroscopy (VUV) has gained popularity in the food, forensics, environmental, and fuel research communities [18,21–24]. GC (x GC)-VUV research has demonstrated capability to identify structural and stereoisomers such as p-xylene, o-xylene, m-xylene [21], or the isomers of cis-decalin and trans-decalin [18], which were indistinguishable by MS systems.

Schug et al. explored the benchtop VUV detector in 2014 [21], suggesting its potential as a universal detector. Few species, separated from a gasoline sample, were resolved at the isomeric level in that work [21]. A later publication demonstrated the use of GC-VUV for hydrocarbon group type analysis (PIONA) of gasoline which was verified through various existing ASTM compositional methods [25]. However, in the case of a higher molecular weight jet fuel such as SAK, with higher isomerization, the second chromatographic dimension is necessary to sufficiently separate analytes as shown in this work. More recently, Wang explored a diesel fuel sample in a GCxGC-VUV configuration, but like Walsh et al. [25] , was only concerned with group type analysis between cycloalkanes and alkenes, rather than exact isomeric structures [26].

Property predictions from the compositional analysis have long been of interest within the fuel community [13,14,18,27]. Before Yang et al. [16], most fuel property predictions could be classified as 'top-down' approaches [19,28], where models are developed by regression of measured data, serving as both the independent and dependent variables. Such models risk extrapolating to non-physical results (regardless of the statistical method) and require a substantial quantity of data to quantify uncertainties [18]. Conversely, 'bottom-up' approaches [16,18,29] leverage composition data via GC x GC as one category of input, along with a library of property data corresponding to potential fuel constituents as the other category of input. These inputs are related to fuel properties by simple, and usually physically based, blending rules [16–18,30]. While the physically-based blending rules afford confidence to extrapolate to compositions beyond the historical record, all such bottom-up models afford traceability of errors, enabling a comprehensive uncertainty analysis [17,18]. The four sources of error that contribute to the uncertainty of predictions made by a bottom-up approach include the following: constituent (aka chromatogram peak, or deconvoluted peak) mass concentration measurement error, assignment of chromatogram peaks to specific isomers, isomeric properties data uncertainty, and blending rule accuracy. This work addresses the precision of the chromatogram peak assignments to specific isomers, as well as identification and deconvolution of chromatogram peaks comprised of 2 or 3 species.

Having no prior knowledge of composition or properties, a contemporary SAF candidate (Virent SAK) is investigated, starting with GCxGC/FID-VUV. Much of its composition (>71 %m) is found in coeluting peaks. Here, an isomeric identification

approach is detailed with four novel contributions: (1) definitive determination of species count (one or more than one) within any peak on the chromatogram, (2) two-dimensional VUV deconvolution with up to three analytes, (3) greater than 93%m in a real jet fuel is assigned to specific isomers, and (4) precision improvement of property predictions. Previous coelution detection methodologies leverage the completeness of reference libraries [25,26], whereas the method presented here relies solely on the measured signal. This definitive pre-processing of peaks distinguishes single elution from coelution peaks and categorizes them based on the presence of multiple unique spectra within the same peak. Finally, material compatibility and dielectric constant calculations are reported for 79/21 %v HEFA/SAK blend, further documenting the potential of SAK as a keystone blend component for potential 100% SAF.

2.   Methodology

A GC x GC-FID/VUV method was employed to identify the hydrocarbon species in SAK, similar to the method used in Heyne et al. [18]. This work builds on that methodology by adding deconvolution capabilities, showing a dramatic increase in predictive accuracy due to isomeric information, and compares them to predictions done with traditional hydrocarbon group type information [16].

2.1   Reference samples

The SAK fuel sample was provided courtesy of Virent, Inc, and the HEFA fuel sample was provided by World Energy. Additional materials that composed the majority of the SAK were procured to predict SAK properties more accurately and characterize the mixture further. A summary of measured properties for each component can be found in Tables 2 and 3 in the Supplementary Material. Three reference fuels from the National Jet Fuel Combustion Program (NJFCP) [27,31,32], A-1 (POSF 10264), A-2 (POSF 10325), and A-3 (POSF 10289) were also analyzed via GC x GC-FID/MS and used as an aromatic benchmark for several temperature-independent properties.

2.2   Gas chromatography, flame ionization detector, and vacuum ultraviolet light detector

The experimental setup was arranged as GC x GC-FID/VUV, where the VUV system supported the hydrocarbon isomeric identification and the FID supported quantification. The system included two columns separated by a modulator connected to a split plate after the second column where the analytes were directed to either the FID or the VUV. The system included a SepSolve INSIGHT flow modulator and VGA-101 Vacuum Ultraviolet light detector. A graphical overview of the GC x GC-FID/VUV is displayed in Figure 1, illustrating the major components of the GC x GC-FID/VUV system.
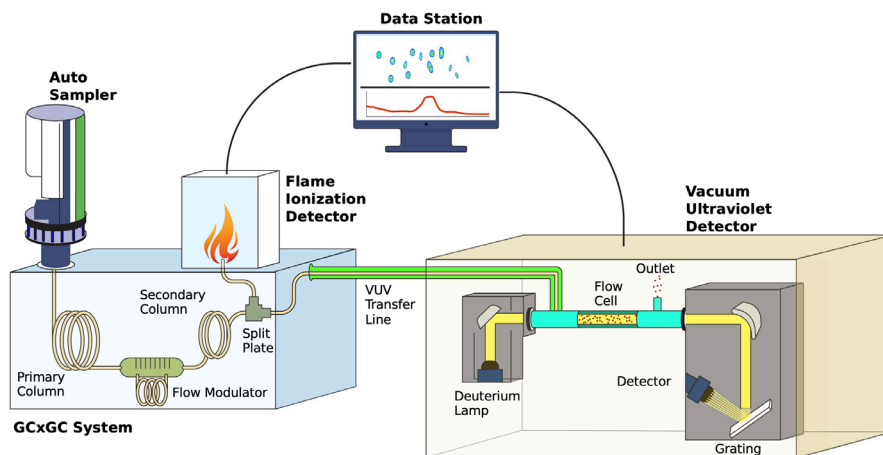
Figure 1. GC x GC-FID/VUV Test Setup Diagram

The Agilent 7693A Automatic Liquid Sampler (ALS) injected 5 μL of sample into the Agilent 8890 GC. The inlet temperature, pressure, and split ratio were maintained at 250 ˚C, 55.04 psi, and 100:1, respectively. A reverse column arrangement was chosen to achieve the desired separations for this study. Specifically, in their respective order, a Rxi-17Sil MS 60m x 0.32mm x 0.5µm and Rxi-1 15ms x 0.32mm x 0.5µm columns were utilized, both of which were manufactured by Restek. Constant flow rates for the first and second columns of 1.2 mL/min and 48 mL/min were held throughout the run with a Helium carrier

gas (grade 5.0), which passed through a Restek Triple Filter before entering the GC system. The GC oven was initialized at a temperature of 40 °C for 30 seconds, with a ramp rate of 1 °C / min until achieving a final temperature of 280 °C, where the temperature was held for 10 minutes. The GC ran for a total time of 250.5 minutes. Two modulation times and injection volumes were used. A 110-second modulation time with 5 μL injection and a 5-second modulation time with 1 μL injection was employed. The combination of long modulation times and high injection volumes enabled higher concentrations of trace analytes to be identified. While the shorter modulation time and lower injection volume facilitated deconvolution of peaks with high concentrations.

As mentioned previously and demonstrated in Figure 1, the analyte is divided at the split plate after traveling through the secondary column. One of the lines leaving the split plate junction feeds the FID where flow rates are applied to the air (ultra-zero grade) flow, H_2 (grade 6.0), and N_2 (grade 5.0) at flow rates of 400, 40, and 25 mL/min, respectively. The FID operated at a fixed temperature of 300 °C while recording data at a frequency of 50 Hz, which was processed by INSIGHT ChromSpace software (Version 1.5.1).

As illustrated in Figure 1, a second line parts from the split plate and directly connects to the VGA-101 transfer line. The sample was preferentially directed to the VUV detector relative to the FID. The length of the internal transfer line from the split plate to the VUV was shorter than the transfer line from the split plate to the FID to increase the amount of sample directed to the VUV. The transfer line from the GC to the flow cell was maintained at 250 °C. With N_2 (grade 5.0) being used as the system gas, data acquisition continuously occurred at 76.92 Hz over a wavelength range of 125 to 430 nm.

2.3   Identification and Quantification

The overall identification procedure includes the following steps: timestamp alignment between FID and VUV, removal of oversaturated VUV data, local background signal subtraction and noise reductions, coelution screening, identification of analyte(s) in the considered peak, and conversion to mass fractions. Each step of the procedure described above leveraged in-house Python (Version 3.8.5) code, which can be provided upon request.

2.4   FID and VUV Alignment

Synchronization of the FID signal to the VUV signal was completed by aligning the max signal for each peak across the experiment duration. A single offset value could be found and applied to the entire VUV dataset by minimizing the offset between the local maxima found with the separate detectors. With the offset applied, SepSolve ChromSpace FID area determinations were associated with the identified analytes and VUV data.

2.5   Analyte Identification

Pre-processing of VUV spectra included removing the full spectra at certain time stamps corresponding to cases where the absorbance at any wavelength exceeded one. Then local background subtraction and signal averaging, akin to the approach described in Lelevic et al. [33] and Heyne et al. [18], were done to attain the sample spectra used for matching against cataloged reference spectra. Single analyte identification followed the work discussed in Heyne et al. [18].

Determination of species count (one or more than one) within each peak on the chromatogram was the final pre-processing step. Prior state-of-the-art techniques [25] relied on the quality (e.g. R^2) of a multiple species match significantly exceeding that of the top single species matches in order to confirm or reject the presence of coelution. In this work, it is shown how this determination can be established prior to any matching exercise, which eliminates reliance on the completeness of spectral reference libraries. This insight is especially useful in downstream steps. For example, by demonstrating that given sample spectra originate from a single species subsequent comparison to reference spectra will either identify the most probable match or prove that the observed, sample spectra are not present in the reference library. Without first proving that sample spectra originated from a single species, any number of linear combinations of multiple spectra in the reference library could (incorrectly) meet the acceptable match criteria. Conceptually, the coelution check is done by comparing sets of spectra that are each averaged over the full width of one retention time axis (i.e. t_1 or t_2) and segments of the other retention time axis. If the spectral profile is static while sweeping through the segmented time axis, in both dimensions, the peak is comprised of a single analyte. If the normalized spectra change while traveling across the chromatogram peak, then it is known to be comprised of multiple analytes.

For chromatogram peaks shown to consist of more than one species, the sample VUV spectra at each point in time (t_1 , t_2) within that peak was matched by a linear combination of 2 or 3 reference spectra, leveraging the non-negative least-squares

optimization algorithm from Python SciPy, maximizing R^2. The decision logic around which 2 or 3 reference spectra to use as the basis functions was partially manual with the goal being to select the ones that yielded the best overall match throughout the peak. The overall concentration of each analyte was determined by summing the product of the reference spectra scale factors at each time point with the mass fraction attributed to each time point, where the integrated areas of sample VUV spectra were used to determine the mass fraction at each time point.

2.6 Quantification

Hydrocarbon type analysis was performed by generating a stencil with a method like the one described by Vozka et al. [34]. Chromatogram peak areas of each peak (A_i) were determined through the ChromSpace integration software and attributed to corresponding hydrocarbon groups (A_{class}) based on retention time, where the FID signal serves as the z-axis (or the color scale). Previous research [16–18] has leveraged the well tested hydrocarbon template from Striebich et al. for hydrocarbon group type mass fraction determinations [20], as were the mass fractions (Y_i) presented herein. To arrive at a single species mass faction, The group mass fraction per the Striebich et. al template (Y_{class}), was scaled by the area percentage of th selected peak (A_i) relative to the total area (A_{class}) of the respective hydrocarbon group. Repeatability for hydrocarbon class quantification is taken as RSD% <1.5% (n=3), as reported in a recent repeatability study [35].

$$Y_i = \frac{A_i}{A_{class}} * Y_{class} \qquad\qquad\qquad (1)$$
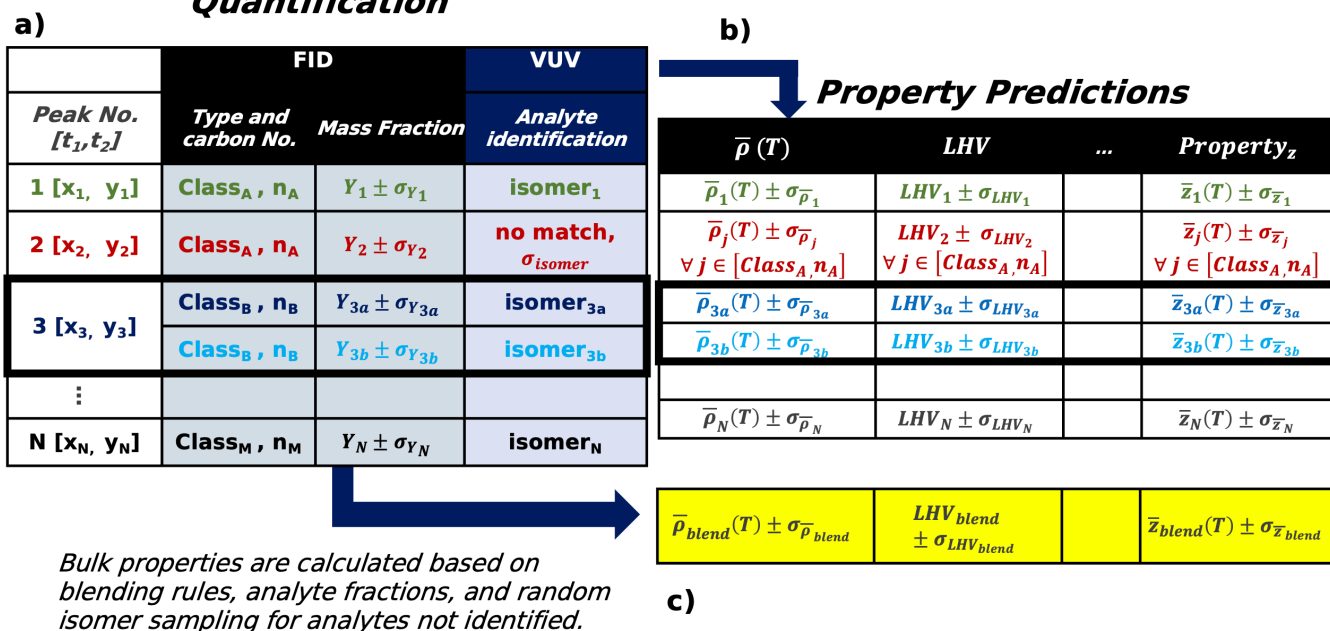
2.7 Property Measurements

Threshold Sooting Index (TSI) values of each species identified in the report were estimated via the QSPR model of Boehm et al. [9] and the blending rule detailed in that report was used to predict the TSI of SAK. An o-ring volumetric swell study was performed with optical dilatometry techniques detailed in Faulhaber et al. [36]. Here, two acrylonitrile-butadiene o-ring materials were submerged at room temperature into separate fuel-filled vials. The first contained neat HEFA, and the second contained a 79/21 %v HEFA/SAK blend. Refractive index measurements were taken at room temperature using a Reichert TS Meter. These measurements were taken with a light emission source at 589 nm and converted into dielectric constant values for neat HEFA, SAK, and the 79/21 %v HEFA/SAK using $K=n^2$ (K is dielectric constant, n is refractive index). The accuracy of the Reichert TS Meter was +- 0.0001 nD. Other properties such as flash point, freeze point, viscosity, density, and surface tension, were experimentally determined for several of the samples using aviation fuel specification tests [7]. Both the ASTM methods names and corresponding repeatabilities and reproducibilities are listed in Table 1 of the Supplementary Material, and served to evaluate the accuracy of the predictions. [9]. All other properties required but not measured for this study were sourced by the NIST Web Thermo Tables [37].

2.8 Tier Alpha Approach for Property Predictions

The Tier Alpha approach for property predictions employs three different pieces of information: empirically derived algebraic blending rules, an extensive database of pure molecules, and the best available composition data. Where composition data is limited to hydrocarbon group level mass concentrations, a random selection of a representative member of that group is made as part of a Monte Carlo simulation that also includes uncertainties in determined mass fractions and database properties. By far, the largest source of uncertainty (precision) in these determinations has been, until recently [18], the underdetermined speciation within each of the hydrocarbon groups[17]. The coupling of VUV spectroscopy with GCxGC/FID chromatography attacks this primary source of property prediction uncertainty and now, with this work, advances in sampled VUV spectra deconvolution further attack this primary source of property prediction uncertainty. With these improvements, blending rule accuracy, rather than prediction precision may be the largest source of overall prediction uncertainty for some properties of some hydrocarbon mixtures with volatility within the jet fuel range. A flow chart describing this process is provided in Figure 2 to help conceptualize the details of this approach.

Identification & Quantification

a)

Peak No. $[t_1, t_2]$	FID			VUV
	Type and carbon No.	Mass Fraction		Analyte identification
1 $[x_1, y_1]$	Class$_A$, n_A	$Y_1 \pm \sigma_{Y_1}$		isomer$_1$
2 $[x_2, y_2]$	Class$_A$, n_A	$Y_2 \pm \sigma_{Y_2}$		no match, σ_{isomer}
3 $[x_3, y_3]$	Class$_B$, n_B	$Y_{3a} \pm \sigma_{Y_{3a}}$		isomer$_{3a}$
	Class$_B$, n_B	$Y_{3b} \pm \sigma_{Y_{3b}}$		isomer$_{3b}$
⋮				
N $[x_N, y_N]$	Class$_M$, n_M	$Y_N \pm \sigma_{Y_N}$		isomer$_N$

b) *Property Predictions*

$\bar{\rho}(T)$	LHV	…	Property$_z$
$\bar{\rho}_1(T) \pm \sigma_{\bar{\rho}_1}$	$LHV_1 \pm \sigma_{LHV_1}$		$\bar{z}_1(T) \pm \sigma_{\bar{z}_1}$
$\bar{\rho}_j(T) \pm \sigma_{\bar{\rho}_j}$ $\forall j \in [Class_A, n_A]$	$LHV_2 \pm \sigma_{LHV_2}$ $\forall j \in [Class_A, n_A]$		$\bar{z}_j(T) \pm \sigma_{\bar{z}_j}$ $\forall j \in [Class_A, n_A]$
$\bar{\rho}_{3a}(T) \pm \sigma_{\bar{\rho}_{3a}}$	$LHV_{3a} \pm \sigma_{LHV_{3a}}$		$\bar{z}_{3a}(T) \pm \sigma_{\bar{z}_{3a}}$
$\bar{\rho}_{3b}(T) \pm \sigma_{\bar{\rho}_{3b}}$	$LHV_{3b} \pm \sigma_{LHV_{3b}}$		$\bar{z}_{3b}(T) \pm \sigma_{\bar{z}_{3b}}$
$\bar{\rho}_N(T) \pm \sigma_{\bar{\rho}_N}$	$LHV_N \pm \sigma_{LHV_N}$		$\bar{z}_N(T) \pm \sigma_{\bar{z}_N}$

$\bar{\rho}_{blend}(T) \pm \sigma_{\bar{\rho}_{blend}}$	$LHV_{blend} \pm \sigma_{LHV_{blend}}$		$\bar{z}_{blend}(T) \pm \sigma_{\bar{z}_{blend}}$

Bulk properties are calculated based on blending rules, analyte fractions, and random isomer sampling for analytes not identified.

c)

Figure 2. Illustration of the 'Tier Alpha' approach including: identification and quantification (a), component property data (b), and mixture property determinations (c). Deconvolution is depicted with a bold black outlined box.

3.   Results and discussion

3.1   Identification Results

The hydrocarbon group type analysis from the GCxGC chromatograph template classifies 93.2 %m into three alkylbenzene bins; C8 (3.7%m), C9 (62.0 %m), and C10 alkylbenzenes (27.5%m). Other minor fractions are distributed across C9 (0.8 %m) and C10 cycloaromatics (1.2 %m), along with various cycloalkane and isoalkane groups, each <1.5%m, across a carbon range of C8 to C10. While that approach classifies analytes into a specific hydrocarbon group, the goal of this study is to further separate the group mass fraction into isomer-specific fractions using VUV data. Illustrated in Figure 3, the SAK chromatograph has been labeled with English letters to inform which peaks were identified through the use of the VUV detector. These English letters, summarized in Table 1, delineate 26 isomers along with their mass fractions amounting to 93.6 %m of the SAK sample. All possible isomers of C8 and C9 alkylbenzenes are identified. In contrast, C9 cycloaromatics and C10 alkylbenzenes include incomplete isomer identification due to some peaks having low signal to noise, or due to detected isomers not matching against the available reference data. All fractions from the hydrocarbon group type analysis that were not identified on the isomeric level have been summarized in Table 4 of the Supplementary Material.
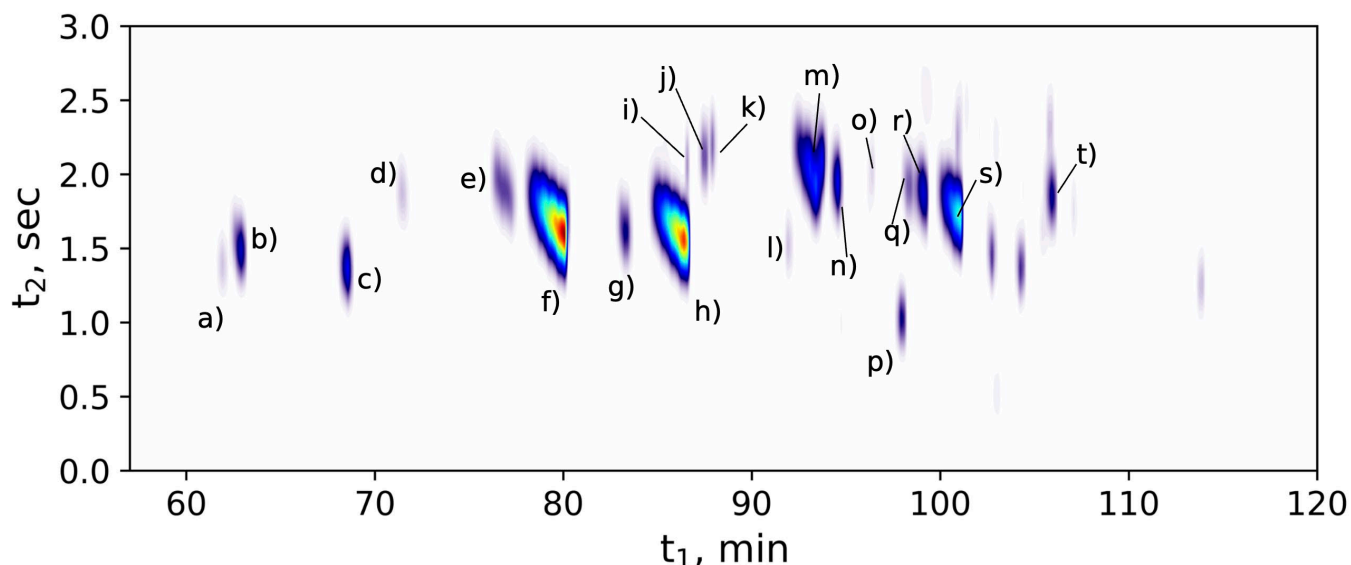
Figure 3. FID chromatograph (5 μL and 10 sec modulation time) with provided letters that correspond to identified species laid out in Table 1. The RGB color scheme is used where blue is low concentration and red is high concentration.

The two most prominent analytes present in SAK are 1-methyl-3-ethylbenzene and 1,2,4-trimethylbenzene, 20.6%m and 26.7 %m, respectively. Conversely, 1-methyl 2-n-propyl benzene (0.15 %m) and 1,3,5-trimethyl benzene (0.15 %m) are the smallest identifiable fractions. All peaks comprising at least 0.68%m of the SAK sample were identified by matching their corresponding VUV spectra with reference spectra in our library; $R^2 > 0.999$. At lower analyte concentration, the R^2 between sample and reference spectra of the same species decreases due to the low signal to noise ratio of the sample spectra. For 1-methyl 2-propylbenzene, the R^2 was 0.978.

The VUV performed excellently in the hydrocarbon group regions where the reference library was completely defined, yet limitations persist in making identifications in regions with less library coverage. This is especially true in heavy regions (>C10) where known possible isomers exponentially increases. For example, two prominent cycloaromatic peaks (lower right diagonal of the chromatogram peak 's', Fig. 3) are not identified beyond the hydrocarbon group type analysis. Conversely, all C8, C9, and C10 alkylbenzene isomer spectra are present within the reference library, which enabled comprehensive identifications. Additionally, as demonstrated by Lelevic et al., lower carbon number alkylbenzenes exhibit higher absorption (higher response factor) relative to other hydrocarbon class species. Alkanes, in contrast with low molecular weight alkyl benzenes, require higher molar concentrations for equivalent identification fidelity [38].

Table 1. Summary of the identification results organized by hydrocarbon groups. Identified species are listed along with their corresponding chromatograph peak labels in Figure 2.

Group Type	Species Name	Labels	%Mass
C9 Cycloaromatics	indane	p	0.799
C8 Alkylbenzenes	p-xylene	b*	0.384
	m-xylene	b*	1.274
	ethylbenzene	a	0.181
	o-xylene	c	1.844
C9 Alkylbenzenes	1-methyl-3-ethylbenzene	f*	20.604
	1-methyl-4-ethylbenzene	f*	10.879
	1,3,5-trimethylbenzene	f*	0.148
	1,2,4-trimethylbenzene	h*	26.686
	isopropylbenzene	d	0.275
	n-propylbenzene	e	1.92
	1-methyl-2-ethylbenzene	g	1.356
	1,2,3-trimethylbenzene	l	0.163
C10 Alkylbenzenes	sec-butylbenzene	i*	0.156
	1,3-diethylbenzene	m*	4.445
	1-methyl-3-n-propylbenzene	m*	3.213
	1-methyl-4-n-propylbenzene	m*	1.535
	1,3-dimethyl-5-ethylbenzene	n*	1.122
	1,4-diethylbenzene	n*	0.811
	1-methyl-3-isopropylbenzene	j	0.739
	1-methyl-4-isopropylbenzene	k	0.332
	1-methyl-2-n-propylbenzene	o	0.145
	1,4-dimethyl-2-ethylbenzene	q	0.678
	1,3-dimethyl-4-ethylbenzene	r	2.443
	1,2-dimethyl-4-ethylbenzene	s	10.089
	1,2,4,5-tetramethylbenzene	t	1.36

3.2   Two-Dimensional Deconvolution

Paramount to the success of identifying isomers in SAK is deconvolution capability. Coeluting species are marked with an asterisk (*) next to their peak letter label in Table 1. Collectively, 71.3 %m is contained in 6 peaks that required deconvolution, while 22.3 %m contained in 14 peaks did not require deconvolution. Nearly half of the identified isomers in this analysis required deconvolution for identification.

Figure 4 illustrates the deconvolution analysis of the largest peak in the SAK sample (labeled 'f*' in Table 2 and 'f' in Fig. 3). The bulk mass fraction of the peak is determined to be 31.6 %m of the total SAK composition. Namely, 1-methyl-3-ethylbenzene, 1-methyl-4-ethylbenzene, and 1,3,5-trimethyl benzene are found in this peak. The majority of this peak (99.54%m) is comprised of 1-methyl-3-ethylbenzene and 1-methyl-4-ethylbenzene. Also, 1,3,5-trimethyl benzene is found in

low concentrations here, 0.15%m total SAK mass or 0.46%m of the peak, respectively. Linear combinations of the scaled reference spectra for the three analytes of interest here achieve an R^2>0.999 at each timestamp reported.

The analytes at peak 'f' in Figure 2 exhibit a saturated VUV signal (absorbance exceeding 1.0) at some wavelengths, making its deconvolution less rigorous. To avoid that complication, the reduced injection volume method as described in the method section was applied. Figure 3 illustrates a blow-up of this peak after applying the reduced injection volume method. In contrast to Figure 2, the image shown in Figure 3c is from the VUV signal rather than the FID signal.

Figure 3a and 3d report the relative absorbance of 1-methyl-3-ethylbenzene and 1-methyl-4-ethylbenzene. Mathematically the absorbance signatures are dramatically different and distinguishable. The first dimension of separation, t_1, is the recorded time corresponding to modulation number, 'm', that a given analyte entered the secondary column. The second dimension, t_2, corresponds to the detection time in the VUV for a given modulation. Figure 3a illustrates the relative mass fractions of 1-methyl-3-ethylbenzene (dashed black lines) and 1-methyl-4-ethylbenzene (solid blue lines) over a given modulation index ('m'). Figure 3d compares the relative mass fractions of 1-methyl-3-ethylbenzene and 1-methyl-4-ethylbenzene for selected bins of t_2 values, with 'n' representing the first bin, and δ representing the width of each bin over t_2 for which absorbance was averaged. By parsing out the concentrations of each analyte across all of the peaks timesteps, the accuracy of the aggregate concentrations (across the whole peak) improves. Furthermore, it is possible to detect the presence minor peak concentrations, that are otherwise unidentifiable if looking at the average signal or summed signal of the entire peak, e.g., this deconvolution method reveals the presence of 1,3,5-trimethylbenzene at the low concentration of 0.15 %m.
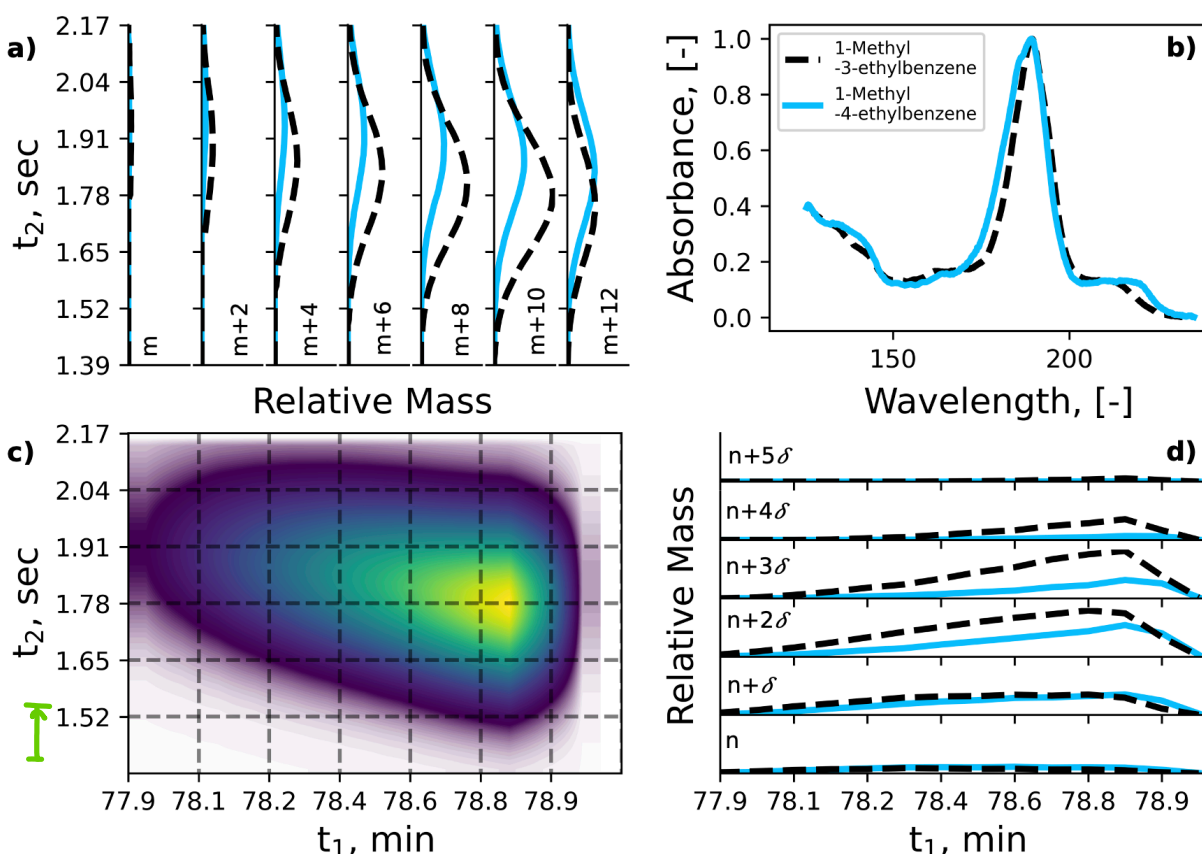
Figure 4. An illustration of the relative masses across peak 'f' (Fig. 3), determined through two-dimensional deconvolution (a and d) of 1-methyl-3-ethylbenzene (dashed black line) and 1-methyl-4-ethylbenzene (solid blue line). Subplots (c and d) depict the integrated VUV signal as a chromatographic peak and the normalized reference spectra of each species, respectively.

3.3 Fuel properties

Two sets of predictions, derived from our 'Tier Alpha' [16] methodology, are made and compared with laboratory measurements. In one set of predictions ('Tier Alpha + VUV'), the information conveyed in Table 1 is utilized in the calculations. For the other set of predictions ('Tier Alpha'), the total mass fraction of each hydrocarbon group type is randomly assigned to a specific isomer of the group in a Monte Carlo simulation. Either way, the properties of each specific component in the mixture-as-modeled are necessary input. The following mixture properties were measured and/or predicted in this work: surface tension at 22 °C, lower heating value (LHV), flash point, threshold sooting index (TSI), smoke point, freeze point, density as a function of temperature, and viscosity as a function of temperature, seal swell, and dielectric constant at 22°C.

Several temperature-independent property results are presented in Figure 5. Measurement data is represented by black-filled circles and lines. Predictions without (Table 1) isomer specificity are represented by blue open triangles and lines. Predictions that leverage the data presented in Table 1 are represented by red open circles and lines. Uncertainties for each determination are represented with 68 % confidence intervals (CI), solid lines, and 95 % CI, capped dashed lines. The accuracy of the applicable blending rule is not captured in the displayed confidence intervals. The green shaded region is derived from the aromatic fraction of the three reference fuels from the NJFCP: A-1 (POSF 10264), A-2 (POSF 10325), A-3 (POSF 10289). They are each the union of three 95% CIs as determined by the 'Tier Alpha' predictions and are intended to provide an additional context of where property values typically lie for the aromatic fraction of conventional fuels.

As evident in Figure 5, the confidence intervals (precision) of the predictions are markedly decreased when 93.6%m is attributable to specific isomers, which is not surprising since isomer uncertainty has been previously identified as the leading source of uncertainty in 'Tier Alpha' predictions when mass concentrations are lumped by hydrocarbon group [17,18]. With the improved prediction precision via specific isomer identifications, the prediction accuracy can be assessed more clearly. In these examples, however, the confidence intervals of the measurement overlap with those of both sets of predictions. Essentially no change in prediction accuracy is observed. The small shift in the prediction mean of LHV is consistent with the expectation that actual population distributions of isomers within any given hydrocarbon group are skewed toward lower heats of formation, not uniform [17]. The larger shift in the prediction mean of surface tension (σ) as well as the 4-5x improvement in the prediction precision underscores the value of the greater specificity of species identification afforded by the GCxGC/FID-VUV method relative to other separation methods where the analytes cannot be interrogated by a spectrographic method, such as VUV. The offset between the two mean surface tension predictions is due largely to the difference between the actual population distribution of C9 alkylbenzenes (accounting for 62.02 %m of the sample SAK) relative to a presumed uniform population distribution of these isomers. The mass fraction weighted average surface tension of the C9 alkylbenzenes is 0.9 mN/m less than the average surface tension of the C9 alkylbenzenes bin.

Viscosities and densities for SAK are illustrated in Figure 6 over a temperature range important to the safe operation of jet engines [27]. Specifically, Figure 6 includes SAK measurements, black circles, 'Tier Alpha' predictions without Table 1 inputs, blue lines, and 'Tier Alpha + VUV', predictions with Table 1 inputs, red lines. Uncertainty regions for each of the predicted methods are also included, where the light regions represent the 95% CIs, and the darker shaded regions represent the 68 % CIs. Device-reported uncertainties are also reported with error bars but are mostly masked due to scaling. Viscosity predictions, with Table 1 inputs, outperformed the standard 'Tier Alpha' by achieving reductions of 90% and 93% for mean error and 95 % CI, respectively. For density, mean error and 95 % CI reductions of 75 % and 89 % are observed, respectively. These accuracy and precision improvements are credited to the removal of the isomeric uncertainty gained by identifying the specific isomers that comprise more than 93%m of the SAK sample. The tier alpha method with either set of inputs, applied to viscosity, demonstrates better agreement with experimental data between 0 °C and 20 °C than they do below 0 °C. For viscosity, a portion of the error at the lower temperatures is due to error imparted through extrapolation (ASTM D341)[39] to lower temperatures than those available through NIST Thermo Tables [38]. This decrease in predictive accuracy between 0 °C and 20 °C was also observed in Heyne et al. [18]. The temperature sensitivity of density ($\Delta\rho/\Delta T$) is well captured by 'Tier Alpha + VUV' (with Table 1 inputs); deviating from the data by just 0.39%. A full tabulation of the neat material property measurements is available in Tables 2 and 3 of the Supplementary Material.

High carbon balance can be achieved with the techniques described in Section 2. The VUV absorption spectra for all known structural isomers of the major aromatic regions in this study (C8, C9 and C10 alkylbenzenes) were fully catalogued in our reference spectra library, thus eliminating concerns of encountering a false-positive match. Generally, as reference libraries become more complete at higher carbon numbers and other hydrocarbon types, the analysis demonstrated here will illuminate even more complicated and heavier fuels. To date, however, incomplete spectra libraries are relevant in instances of higher carbon numbers. For example, the two peaks to the lower right of 1,2-dimethyl-4-ethylbenzene (peak 's') in Figure 3 are unable to be resolved, likely due to incompletely catalogued VUV reference spectra for C10 cycloaromatics.
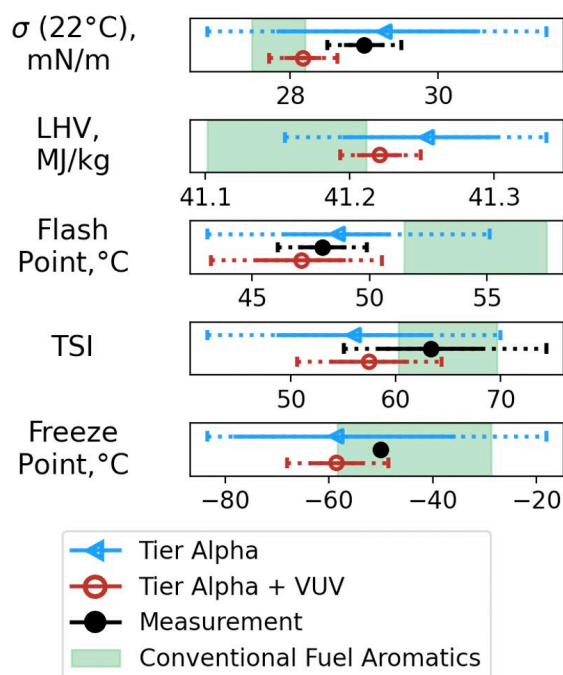
Figure 5. Comparison of predictions of 'Tier Alpha' (blue symbols and lines) to 'Tier Alpha + VUV' (red symbols and lines) in relation to nominal values determined through direct property measurement under ASTM standard methods. Conventional fuel (Jet A - only aromatic components) 95 % CI plotted as well in light green for reference.
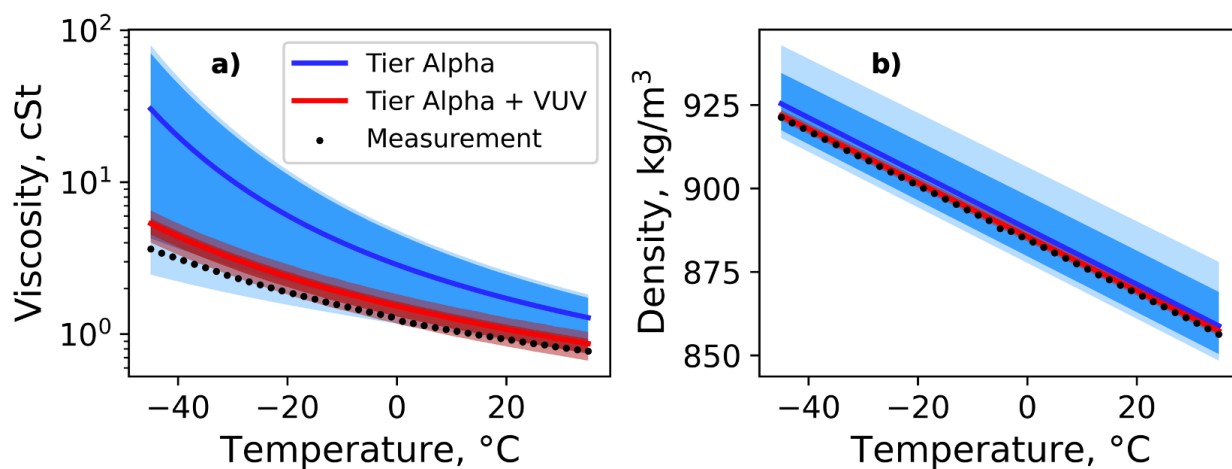
Figure 6. Predictions of viscosity and density with respect to temperature for Tier Alpha (blue line and shaded region) and Tier Alpha + VUV predictions (red line and shaded regions) in relation to measurement (black data points).

3.4 HEFA/SAK blend and HEFA properties

As mentioned in the introduction, aromatics facilitate compliance with several key properties, including material compatibility and dielectric constant. Compositions that are non-compliant with these two additional constraints, or any other fit-for-purpose or spec'd property (ASTM D7566), are not viable candidates for a 100% synthetic SAFs. Figure 7 reports the o-ring swelling, calculated dielectric constant values, and other important operability properties for a 79/21 %v HEFA/SAK blend and neat HEFA and neat SAK where available. Consistent with previous plots, the measurements, filled symbols, and uncertainties, error bars, are reported. These values are compared against a conventional fuel range (shaded green) and the specification limits, red lines, and shaded regions, described in ASTM D7566 [7] The blended HEFA/SAK composition is within the observed range of conventional fuels for each property considered in this work. Interestingly, the viscosity of the HEFA/SAK blend is significantly lower than the typical viscosities of conventional fuels. This is advantageous to engine operability as low viscosity often leads to finer and more uniformly distributed sprays at engine operating conditions that are consistent with altitude relight, ground start, and transitions in or out of flight idle [40,41]. Specifically, both the volume swell (12.3%) and predicted dielectric constant (2.096) are within the conventional fuel range.
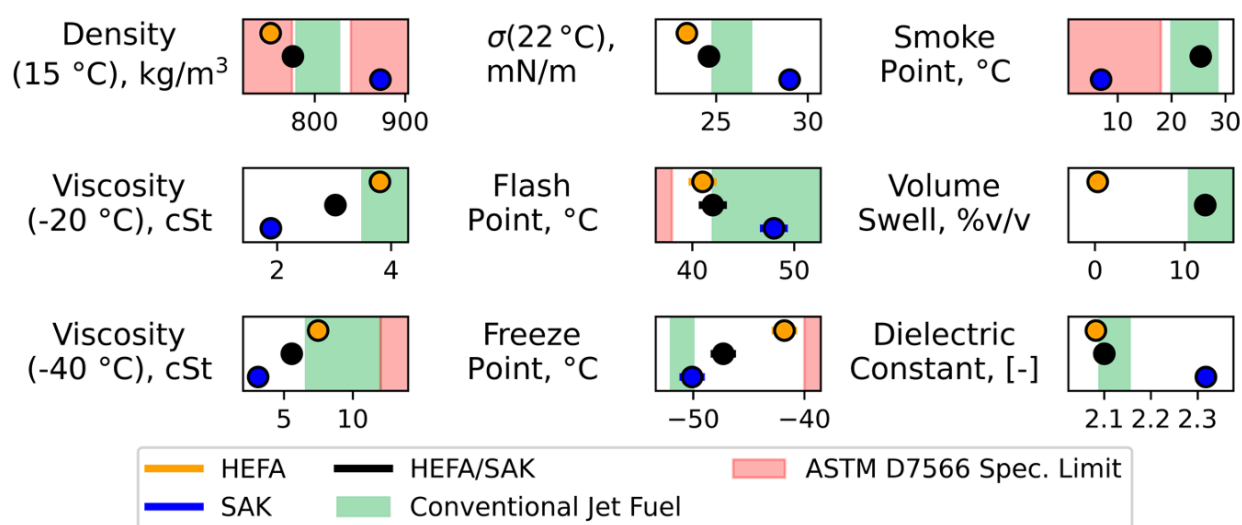
Figure 7. Selected 'fit for purpose' operability property measurements (black symbols) with measurement error (black lines), along with o-ring volume swell percentage and dielectric constant for 79/21 HEFA/SAK blend. Neat HEFA (yellow symbol and line) and neat SAK (blue symbol and line) are plotted. Conventional Jet Fuel (shaded green) and ASTM D7566 (shaded red) illustrate compliance.

4. Conclusions

A carbon balance greater than 93%m consisting of 26 separate analytes was identified in a relevant SAF (SAK) candidate. Of these, 73.1%m relied on the deconvolution method described herein. With added specific isomeric information, confidence in the property predictions improves relative to predictions predicated on conventional hydrocarbon group type analysis. Most notable, viscosity absolute error is reduced by 90% and 95-percentile confidence interval is reduced by 93%. For the properties measured, a HEFA/SAK blend illustrates a path for 100% SAF to remain drop-in for additional aviation properties (o-ring swelling and dielectric constant) while reducing nvPM with high relative aromatic contents and exhibiting other advantaged properties. Approximately 50%m of a typical conventional Jet-A is composed of in 44 peaks. Meaning the limit of detection for a VUV is not the limiting factor in using it for property evaluations. While the number of analyte spectra is the limiting bottleneck in GC x GC -VUV to property predictions, the potential for this method remains exceptionally high.

Funding

Bibliography

[1]   Lee DS, Fahey DW, Skowron A, Allen MR, Burkhardt U, Chen Q, et al. The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018. Atmospheric Environment 2021;244. https://doi.org/10.1016/j.atmosenv.2020.117834.
[2]   Vardon DR, Sherbacow BJ, Guan K, Heyne JS, Abdullah Z. Realizing "net-zero-carbon" sustainable aviation fuel. Joule 2022;6:16–21. https://doi.org/10.1016/j.joule.2021.12.013.
[3]   Huq NA, Hafenstine GR, Huo X, Nguyen H, Tifft SM, Conklin DR, et al. Toward net-zero sustainable aviation fuel with wet waste–derived volatile fatty acids. Proc Natl Acad Sci U S A 2021;118. https://doi.org/10.1073/PNAS.2023008118/-/DCSUPPLEMENTAL.
[4]   U.S. Energy Information Administration. Annual Energy Outlook 2021 2021. https:// www.eia.gov/outlooks/aeo/. (accessed January 26, 2022).
[5]   de Jong S, Antonissen K, Hoefnagels R, Lonza L, Wang M, Faaij A, et al. Life-cycle analysis of greenhouse gas emissions from renewable jet fuel production. Biotechnology for Biofuels 2017;10:1–18. https://doi.org/10.1186/S13068-017-0739-7/TABLES/4.
[6]   World's first passenger flight with 100% renewable fuel takes off—thanks to biotech | BIO n.d. https://www.bio.org/blogs/worlds-first-passenger-flight-100-renewable-fuel-takes-thanks-biotech (accessed May 9, 2022).
[7]   Designation: D7566 − 19b Standard Specification for Aviation Turbine Fuel Containing Synthesized Hydrocarbons 1 n.d. https://doi.org/10.1520/D7566-19B.
[8]   Kärcher B. Formation and radiative forcing of contrail cirrus. Nat Commun n.d.;9:1824.
[9]   Boehm RC, Yang Z, Heyne JS. Threshold Sooting Index of Sustainable Aviation Fuel Candidates from Composition Input Alone: Progress toward Uncertainty Quantification. Energy & Fuels 2022:acs.energyfuels.1c03794. https://doi.org/10.1021/acs.energyfuels.1c03794.
[10]   Kosir S, Heyne J, Graham J. A machine learning framework for drop-in volume swell characteristics of sustainable aviation fuel. Fuel 2020;274. https://doi.org/10.1016/j.fuel.2020.117832.
[11]   Romanczyk M, Ramirez Velasco JH, Xu L, Vozka P, Dissanayake P, Wehde KE, et al. The capability of organic compounds to swell acrylonitrile butadiene O-rings and their effects on O-ring mechanical properties. Fuel 2019;238:483–92. https://doi.org/10.1016/j.fuel.2018.10.103.
[12]   Edwards T. Reference jet fuels for combustion testing. AIAA SciTech Forum - 55th AIAA Aerospace Sciences Meeting 2017. https://doi.org/10.2514/6.2017-0146.
[13]   Vozka P, Kilaz G. A review of aviation turbine fuel chemical composition-property relations. Fuel 2020;268. https://doi.org/10.1016/j.fuel.2020.117391.
[14]   Vozka P, Modereger BA, Park AC, Zhang WTJ, Trice RW, Kenttämaa HI, et al. Jet fuel density via GC × GC-FID. Fuel 2019;235:1052–60. https://doi.org/10.1016/J.FUEL.2018.08.110.
[15]   Wang Y, Ding Y, Wei W, Cao Y, Davidson DF, Hanson RK. On estimating physical and chemical properties of hydrocarbon fuels using mid-infrared FTIR spectra and regularized linear models. Fuel 2019;255. https://doi.org/10.1016/j.fuel.2019.115715.
[16]   Yang Z, Kosir S, Stachler R, Shafer L, Anderson C, Heyne JS. A GC × GC Tier α combustor operability prescreening method for sustainable aviation fuel candidates. Fuel 2021;292. https://doi.org/10.1016/j.fuel.2021.120345.
[17]   Boehm RC, Yang Z, Bell DC, Feldhausen J, Heyne JS. Lower heating value of jet fuel from hydrocarbon class concentration data and thermo-chemical reference data: An uncertainty quantification. Fuel 2022;311. https://doi.org/10.1016/j.fuel.2021.122542.
[18]   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. https://doi.org/10.1016/J.FUEL.2021.122709.
[19]   Wang Y, Wei W, Zhang Y, Hanson RK. A new strategy of characterizing hydrocarbon fuels using FTIR spectra and generalized linear model with grouped-Lasso regularization. Fuel 2021;287. https://doi.org/10.1016/j.fuel.2020.119419.
[20]   Striebich RC, Shafer LM, Adams RK, West ZJ, DeWitt MJ, Zabarnick S. Hydrocarbon group-type analysis of petroleum-derived and synthetic fuels using two-dimensional gas chromatography. Energy and Fuels 2014;28:5696–706. https://doi.org/10.1021/ef500813x.
[21]   Schug KA, Sawicki I, Carlton DD, Fan H, McNair HM, Nimmo JP, et al. Vacuum ultraviolet detector for gas chromatography. Analytical Chemistry 2014;86:8329–35. https://doi.org/10.1021/ac5018343.

[22]   Roberson ZR, Gordon HC, Goodpaster J v. Instrumental and chemometric analysis of opiates via gas chromatography–vacuum ultraviolet spectrophotometry (GC-VUV). Analytical and Bioanalytical Chemistry 2020;412:1123–8. https://doi.org/10.1007/s00216-019-02337-5.

[23]   Kranenburg RF, Lukken CK, Schoenmakers PJ, van Asten AC. Spotting isomer mixtures in forensic illicit drug casework with GC–VUV using automated coelution detection and spectral deconvolution. Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences 2021;1173. https://doi.org/10.1016/j.jchromb.2021.122675.

[24]   Skultety L, Frycak P, Qiu C, Smuts J, Shear-Laude L, Lemr K, et al. Resolution of isomeric new designer stimulants using gas chromatography – Vacuum ultraviolet spectroscopy and theoretical computations. Analytica Chimica Acta 2017;971:55–67. https://doi.org/10.1016/j.aca.2017.03.023.

[25]   Walsh P, Garbalena M, Schug KA. Rapid Analysis and Time Interval Deconvolution for Comprehensive Fuel Compound Group Classification and Speciation Using Gas Chromatography-Vacuum Ultraviolet Spectroscopy. Analytical Chemistry 2016;88:11130–8. https://doi.org/10.1021/acs.analchem.6b03226.

[26]   Wang FCY. Comprehensive Two-Dimensional Gas Chromatography Hyphenated with a Vacuum Ultraviolet Spectrometer to Analyze Diesel-A Three-Dimensional Separation (GC × GC × VUV) Approach. Energy and Fuels 2020;34:8012–7. https://doi.org/10.1021/acs.energyfuels.0c00688.

[27]   Colket M, Heyne J, Rumizen M, Gupta M, Edwards T, Roquemore WM, et al. Overview of the national jet fuels combustion program. AIAA Journal 2017;55:1087–104. https://doi.org/10.2514/1.J055361.

[28]   Berrier KL, Freye CE, Billingsley MC, Synovec RE. Predictive Modeling of Aerospace Fuel Properties Using Comprehensive Two-Dimensional Gas Chromatography with Time-Of-Flight Mass Spectrometry and Partial Least Squares Analysis. Energy and Fuels 2020;34:4084–94. https://doi.org/10.1021/ACS.ENERGYFUELS.9B04108/SUPPL_FILE/EF9B04108_SI_001.PDF.

[29]   Boehm RC, Scholla LC, Heyne JS. Sustainable alternative fuel effects on energy consumption of jet engines. Fuel 2021;304. https://doi.org/10.1016/j.fuel.2021.121378.

[30]   Hall C, Rauch B, Bauder U, le Clercq P, Aigner M. Predictive Capability Assessment of Probabilistic Machine Learning Models for Density Prediction of Conventional and Synthetic Jet Fuels. Energy and Fuels 2021;35:2520–30. https://doi.org/10.1021/acs.energyfuels.0c03779.

[31]   Heyne JS, Colket M, Gupta M, Jardines A, Moder J, Edwards JT, et al. Year 2 of the national jet fuels combustion program: Moving towards a streamlined alternative jet fuels qualification and certification process. AIAA SciTech Forum - 55th AIAA Aerospace Sciences Meeting 2017. https://doi.org/10.2514/6.2017-0145.

[32]   Heyne JS, Peiffer E, Colket M, Moder J, Edwards JT, Roquemore WM, et al. Year 3 of the national jet fuels combustion program: Practical and scientific impacts. AIAA Aerospace Sciences Meeting, 2018 2018. https://doi.org/10.2514/6.2018-1619.

[33]   Lelevic A, Souchon V, Geantet C, Lorentz C, Moreaud M. Advanced data preprocessing for comprehensive two-dimensional gas chromatography with vacuum ultraviolet spectroscopy detection. Journal of Separation Science 2021;44:4141–50. https://doi.org/10.1002/jssc.202100528.

[34]   Vozka P, Kilaz G. How to obtain a detailed chemical composition for middle distillates via GC×GC-FID without the need of GC×GC-TOF/MS. Fuel 2019;247:368–77. https://doi.org/10.1016/j.fuel.2019.03.009.

[35]   Striebich R, Shafer L, West Z, Zabarnick S. UDR-TR-2021-159 Alternative Jet Fuel Evaluation and Specification. Dayton: 2021.

[36]   Faulhaber C, Kosir ST, Borland C, Gawelek K, Boehm R, Heyne JS. Optical Dilatometry Measurements for the Quantification of Sustainable Aviation Fuel Materials Compatibility. n.d.

[37]   Kroenlein K, Muzny C, Diky V, Chirico R, Magee J, Abdulagatov I, et al. NIST/TRC Web Thermo Tables (WTT) NIST Standard Reference Subscription Database 2 - Lite Edition Version 2 2011.

[38]   Lelevic A, Geantet C, Moreaud M, Lorentz C, Souchon V. Quantitative Analysis of Hydrocarbons in Gas Oils by Two-Dimensional Comprehensive Gas Chromatography with Vacuum Ultraviolet Detection. Energy and Fuels 2021;35:13766–75. https://doi.org/10.1021/ACS.ENERGYFUELS.1C01910.

[39]   Designation: D341 – 17 Standard Practice for Viscosity-Temperature Charts for Liquid Petroleum Products 1 n.d. https://doi.org/10.1520/D0341-17.

[40]   Colket M, Heyne J. Fuel Effects on Operability of Aircraft Gas Turbine Combustors. Fuel Effects on Operability of Aircraft Gas Turbine Combustors 2021. https://doi.org/10.2514/4.106040.

[41]   Hendershott T, Stouffer S, Monfort J, Diemer J, Busby K, Corporan E, et al. Ignition of conventional and alternative fuels at low temperatures in a single-cup swirl-stabilized combustor. AIAA Aerospace Sciences Meeting, 2018 2018. https://doi.org/10.2514/6.2018-1422.

Appendix 3

Blend Prediction Model for Freeze Point of Jet Fuel Range Hydrocarbons

Paper 7: Blend Prediction Model for Freeze Point of Jet Fuel Range Hydrocarbons

Nomenclature

<u>Acronyms</u>

ASTM	ASTM International, formerly known as American Society for Testing and Materials
DSC	Differential scanning colorimetry
SAF	Sustainable aviation fuel or fuels

<u>Equation symbols</u>

C_p	Constant pressure heat capacity
G	Gibbs free energy (G = H – TS)
H	Enthalpy (H = U + PV)
P	Pressure
S	Entropy
T	Temperature
N, n	Number of molecules (upper case), moles (lower case)
U	Internal energy
V	Volume fraction
x	Mole fraction
Δ	Property difference between two sets of defining environmental conditions

<u>Subscripts</u>

f	at the freeze point
i or j	arbitrary component of the mixture
liq	liquid
m or mix	mixture or mixing
T	total

1.   Introduction

Aviation operations alone produced roughly 2% of the total carbon emissions in 2019, equating to over 900 million tons of carbon dioxide [1]. Technologies such as electric motors powered by fuel cells appear to be impossible for medium and long-haul flights and it is still unclear whether these technologies will positively impact reduction in carbon intensity [2]. A viable, short-term solution to reduce net carbon emissions are sustainable aviation fuels (SAF). SAF are derived from a renewable source and have demonstrated the ability to reduce carbon intensity to low and negative values [3]. Once a SAF is produced, it is thoroughly evaluated and endorsed by all stakeholders in an expensive and strenuous process detailed in ASTM D4054. The property and process specifications for each fuel must be approved and recorded in ASTM D7566 for quality regulation of SAF productions. Because the approval process is time and cost intensive, developing a prescreening process to identify or eliminate potential candidates is value added. Typical prescreening processes utilize very small sample volumes to measure or predict fuel properties that influence combustion figures of merit and other fit-for-purpose properties that are referred to here as key fuel properties. Recently, a prescreening method has been developed [4,5] in which the key chemical and physical properties of a potential SAF are predicted. The prescreening process provides valuable information to suppliers of possible risks to passing the evaluations of ASTM D4054. Prescreening expedites the evaluation process of alternative jet fuels by providing important feedback to potential suppliers even before a production-scale processing facility is built.

Freeze point is a key physical property of jet fuel that is essential in the prescreening process. The freeze point of jet fuel, per ASTM D1655 specification, must be below -40 ˚C and is typically much lower than the freeze point of several of its components. For example, n-tridecane is commonly found in both petroleum-derived jet fuel and SAF even though its freeze point is -4 ˚C, much higher than spec limit for the fuel. Understanding the relationship between the concentration of components in fuel and the temperature at which that component freezes out of solution is necessary toward the objective of setting concentration limits on certain components that may be introduced via SAF, for the purpose of SAF composition optimization.

The freeze point as defined by the ASTM D5972 is the temperature, upon heating, at which the last hydrocarbon crystal is detected [6]. Although the segment of the solid/liquid phase boundary that is relevant to jet fuel is analogous to the dew point curve of the liquid/vapor phase boundary, melting is used instead of freezing to measure points on this curve in order to avoid kinetic factors that create an inconsistent time lag, during heat extraction, between the conditions of thermodynamic

equilibrium and crystallization, resulting in supercooled liquid [7]. Similar kinetic factors do not exist upon melting. As defined, fuel freeze point provides an operation limit of the aircraft fuel system because fuel must remain in a liquid state to flow properly through the fuel system. If there is crystallization, the system will not operate as designed and jeopardize the operability and safety of the aircraft. Thus, an accurate knowledge of freeze point for potential SAF candidates is important.

Perhaps motivated by prior knowledge of the relatively high freeze point of n-alkanes compared to other hydrocarbons with a similar normal boiling point, early works on freeze point of various jet fuels sought to describe the relationship between n-alkanes composition variation and freeze point, with varying results. For example, Petrovic and Vitorovic determined a simple linear relationship between freeze point and the sum of the concentrations of the three longest n-alkanes present [8]. Then, Solash et al. determined a correlation between the inverse of the freeze point temperature and the logarithm of the concentration of hexadecane [9]. Affens et al. later confirmed this same behavior for several alkanes (C12-C17) with Isopar-M as well as the hexadecane in several solvents [10]. This was followed by Cookson et al. who determined a simple linear relationship (R^2 =0.91 and rmse = 1.3 $°C$) between freeze point and total n-alkane content for fuels in a relatively uniform in boiling range [11]. Differential scanning colorimetry (DSC) and cold-stage microscopy have been used by Zabarnick and others [12,13] to study phase transition behavior in jet fuels and showed that the effect of supercooling on crystallization was significant. More recently, Kuryakov and others [14–16] used a light scattering technique to detect phase transition behavior (melting, crystallization, rotator phase) in n-alkanes and showed that at low n-alkane concentrations, the sensitivity of the light scattering technique provided greater sensitivity than DSC to detect the paraffin phase transitions.

In this work, the freeze point of neat hydrocarbons as well as various blends of neat hydrocarbons with each other or with a petroleum-derived fuel are investigated. The objective of the study is to develop and validate a freeze point model for mixtures derived from first principles. The developed model is compared with existing freeze point models in the literature. The rest of the work is divided up into four sections. First, the experimental details for the freeze point measurements are provided. Second, a derivation of the freeze point model is given along with the other common models used in the literature. Next, the results of the model comparison to the experimental data are detailed. Lastly, the conclusions of the study are summarized.

2.   Methodology

2.1 Experimental Setup

The Phase Technology PAC 70Xi: Cloud, Pour, and Freeze Point Analyzer was used to measure the freeze point of 55 samples as defined in ASTM D5972 [6]. The apparatus has a test resolution of 0.1 $°C$ from -88 to 70°C and consists of a metallic sample chamber (holding 0.15 ml sample volume), a Peltier temperature controller, and a proprietary detection system based on diffusive light scattering. Upon initiation of the experiment, the sample is cooled at a rate of 15°C/min $\pm$ 5 °C/min until the detected opacity is above some threshold suggesting the surface of the sample is mostly solid. The system then signals the Peltier controller to warm the sample back to 20°C at a warming rate of 10°C/min $\pm$ 0.5 °C/min. When the opacity returns to its original value, the apparatus will display an indicated freeze point temperature, which is -12.5 $°C$ for the example measurement shown in Figure 1. According to the Phase Technology product brochure the reproducibility of such a measurement is ±0.8°C. Our independent evaluation of the repeatability, however, was ±10.4, motivating an alternative freeze point determination methodology. Instead of focusing on the temperature at which the opacity returns to its value at time zero, we developed a method of projecting the steepest slope on the heating side to an opacity value of zero; this point was taken as the freeze point and had an estimated repeatability 95 percentile confidence interval of ±1.5 $°C$. Readers interested in comparing the results of our method with those generated by the software provided by Phase Technology should contact the correspondence author. A zipped folder containing all raw data files, such as the example provided in Figure 1, is available upon request. In-between (repeat) measurements without a material change, the sample chamber is cleaned and dried using a cotton swab. In-between measurements with a material change, the sample chamber is rinsed with the new material prior to drying with a cotton swab.
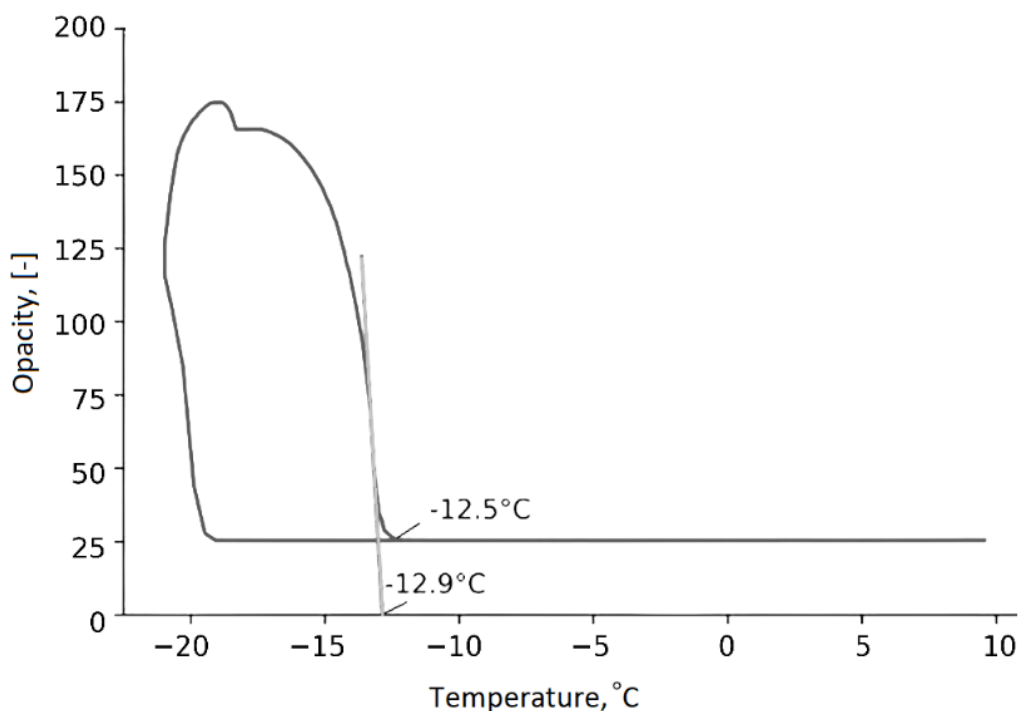
Figure 1. Representative freeze point determination by continuous opacity monitoring. The steep, straight line extending to zero opacity is an extension of the largest slope on the right side (heating) of the opacity verses temperature monitor.

2.2 Materials

Experimentally determined values were obtained from a total of seven neat liquids, 1 out-of-spec petroleum-derived fuel, and 47 binary or ternary blends of these materials for the purpose of evaluating blend freeze point prediction models. Bicyclohexyl and n-tridecane are particularly interesting because they freeze at temperatures well outside of the requirement specified in ASTM D7566 (-40 ˚C), but are known to exist at low concentration in fuels that do meet spec.[17,18] Fifteen mixtures, at various blend fractions, containing bicyclohexyl and 24 mixtures containing n-tridecane supported the primary focus of this investigation, which was to progress toward predictive capability relating to the fraction of high-freeze-point materials that a SAF may contain before it fails to meet the freeze point specification of ASTM D7566.

The supplier-reported purity and lab-measured freeze points of the neat liquids and petroleum fuel are listed in Table 1. While the reported impurities contribute to some freeze point depression of the neat liquids, three of the seven measured freeze points are higher than the values reported in the literature [19]. The primary cause of discrepancy between independent measurements is believed to be variation in sample purity rather than measurement methodology. The two neat liquids with the highest freeze points also have the highest purity, and these are the components that are most likely to freeze out of solution when they are blended with any of the other components. Relating specifically to the developed theory in later sections, accurate knowledge of the freeze point temperature and enthalpy (or entropy) of fusion of the component that is first-to-freeze (last-to-thaw) is especially important. The developed model uses this data as input and does not employ any information about the other components.

Table 1. Materials used and relevant properties.

Material	Molar Volume @ 15 °C , (cm³/mol)	Freeze Point, (°C) This Work / NIST	Purity as labeled
POSF 12968	187[†]	-34.2 / NA	NA
Bicyclohexyl	188.34	6.4 / 9.4	99%
n-Tridecane	243.87	-3.6 / -4.4	99%
Trans-decalin	154.30	-26.5 / -31.1	96.5%
1,2,3-trimethylbenzene	135.05	-29.5 / -25.3	>95%
1,2,4-trimethylbenzene	137.21	-47.5 / -44.4	98%
1,3,5-trimethylbenzene	139.11	-50.2 / -44.8	97%
Cis-1,2-dimethylcyclyhexane	140.97	-50.7 / -52	98%

[†] The molecular weight of sample POSF 12968 was determined from GCxGC/FID data.[18]

2.3 Blending Rules

Three blending rules are evaluated in this work. The wholly empirical model published by AlMulla et al. [20] was developed from a set of 33 data points corresponding to petroleum blends with freeze points ranging from -65 °C to +55 °C, which is substantially higher than jet fuel (-40 °C, maximum). This model is shown here as Equation 1 where V_i is the volume fraction of the ith component, $T_{f,i}$ is the freeze point of the ith component in Kelvin and $T_{f,m}$ is the freeze point of the mixture in Kelvin. Equation 2 is a rudimentary, volume-weighted model that seemed to capture much of the variation that had been observed in our lab prior to this study. The third model is derived from fundamental principles of thermodynamics as described next.

$$T_{f,m} = -0.639 + Ln(\sum_i V_i * (1.067)^{T_{f,i}}) \tag{1}$$

$$T_{f,m} = \sum_i V_i * T_{f,i} \tag{2}$$

For a pure solid in equilibrium with its liquid phase, the molar Gibbs free energy, G of the solid and liquid phases are equal, and the temperature is the freeze point (or melting point) regardless of the relative proportion of the liquid and solid phases in the system. The change in Gibbs free energy is nil when the proportion of the solid phase varies from 0 to 1 as enthalpy is extracted from the system while the temperature is constant. These fundamental observations, along with the definition of G, gives rise to equation 3, which is exact for pure materials, where H and S represent enthalpy and entropy, respectively and the subscript, fusion, represents the phase change.

$$T_f = \Delta H_{fusion}/\Delta S_{fusion} \tag{3}$$

For a mixture, Equation 3 can also be implemented, however, all terms refer to the freeze point, enthalpy, and entropy of fusion for the mixture and are generally unknown. To solve for these terms a Hess cycle, as illustrated in Figures 2 and 3, is constructed and analyzed. State 1 consists of two phases at a temperature equal to the freeze point of the mixture. The solid phase is assumed to consist of a single component and the liquid phase is assumed to consist of all components of the mixture except the one that freezes. State 2 also consists of two phases and differs from state 1 only with respect to system temperature which has been elevated to freeze point that the solid would have if it were isolated from the other components. To reach state 3 from state 2, the isolated solid has melted completely and the corresponding pure liquid remains isolated from the rest of the liquid components. State 4 consists of a single liquid mixture as the two isolated liquids of state 3 are mixed. Finally, to reach state 5, the liquid mixture is cooled until its temperature equals the freeze point of the mixture, the same temperature as state 1. Analogous to pure materials, states 1 and 5 represent different ends of the phase change process for which the change in G is zero. Therefore, the quantities of interest are the enthalpy and entropy differences between these two states as well as the mixture freeze point temperature. The change in enthalpy
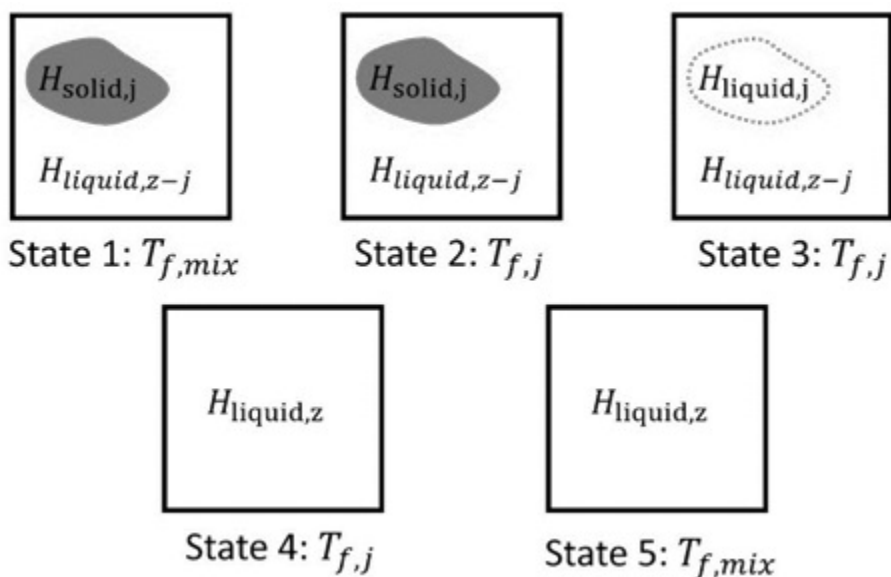
Figure 2. States of the system used in the Hess cycle. States 1 and 5 are the opposite ends of the phase transition isotherm throughout which the solid phase is in equilibrium with the mixed, liquid phase. The tracked processes are, from 1 to 5, heating, melting, mixing, cooling.
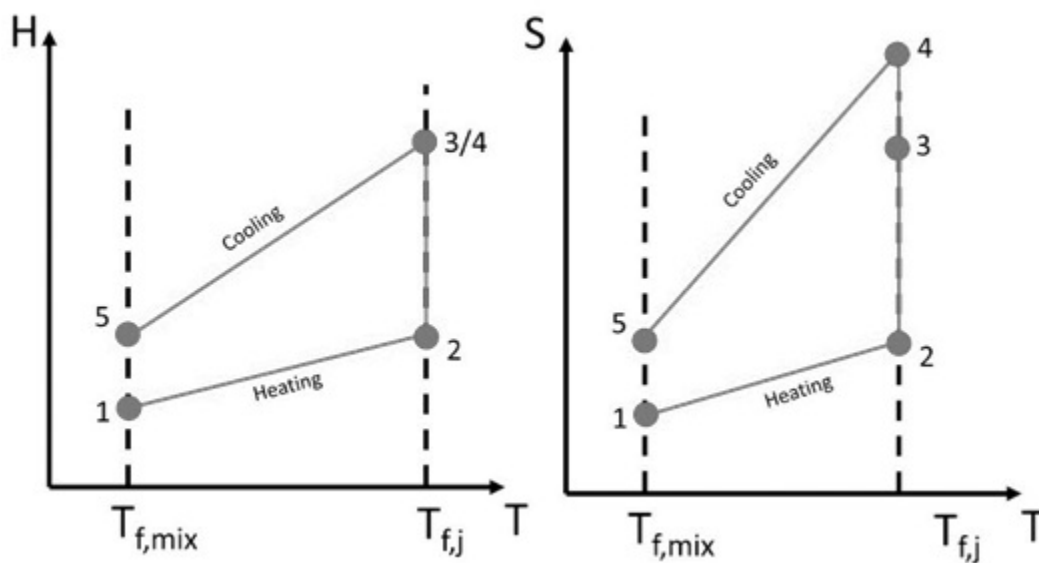
Figure 3. Enthalpy and entropy of states used in the Hess cycle.

and entropy is known or can be estimated for each leg of the Hess cycle. For one mole of component, j undergoing the phase change, Equations 5 through 14 apply, ultimately leading to Equation 4, which is the mixture analog of Equation 3. The heat capacity at constant pressure is represented by C_p in these equations.

$$T_{f,m} = \Delta H_{15}/\Delta S_{15} \tag{4}$$

$$\Delta H_{12} = C_{p,solid,j} * (T_{f,j} - T_{f,mix}) + (1/x_j - 1) * C_{p,liq,z-j} * (T_{f,j} - T_{f,mix}) \tag{5}$$

$$\Delta H_{23} = \Delta H_{fus,j} \text{ at } T = T_{f,j} \tag{6}$$

$$\Delta H_{34} = \Delta H_{mix} \tag{7}$$

$$\Delta H_{45} = (1/x_j) * C_{p,liq,z} * (T_{f,mix} - T_{f,j}) \tag{8}$$

$$\Delta H_{15} = \Delta H_{12} + \Delta H_{23} + \Delta H_{34} + \Delta H_{45} \tag{9}$$

$$\Delta S_{12} = C_{p,solid,j} * (Ln(T_{f,j}) - Ln(T_{f,mix})) + (1/x_j - 1) * C_{p,liq,z-j} * (Ln(T_{f,j}) - Ln(T_{f,mix})) \tag{10}$$

$$\Delta S_{23} = \Delta S_{fus,j} \text{ at } T = T_{f,j} \tag{11}$$

$$\Delta S_{34} = \Delta S_{mix} \tag{12}$$

$$\Delta S_{45} = (1/x_j) * C_{p,liq,z} * (Ln(T_{f,mix}) - Ln(T_{f,j})) \tag{13}$$

$$\Delta S_{15} = \Delta S_{12} + \Delta S_{23} + \Delta S_{34} + \Delta S_{45} \tag{14}$$

After expanding Equation 4, the following information is necessary to predict the freeze point of a given liquid mixture. The identity of the frozen component must be known as well as its enthalpy and entropy of fusion and the heat capacity of its solid phase. The heat capacity of the liquid mixture, with and without component j, must be estimated and the enthalpy and entropy of mixing must also be estimated. To estimate the heat capacity of the liquid mixtures, the blending rule for heat capacity [21] is used. This blending rule is reproduced here as Equation 15 where x_i is the mole fraction of the ith

$$C_{p,liq,z} = \sum_i x_i * C_{p,liq,i} \tag{15}$$

$$(n_T - 1) * C_{p,liq,(z-j)} - n_T * C_{p,liq,z} = -C_{p,liq,j} \tag{16}$$

component. Its application to this problem is represented by Equation 16 where n_T (or (1/xj)) is the total number of moles in the system, out of which one mole of component, j freezes. The work of Neaf [22] was leveraged to estimate the heat capacity of component, j in both the solid phase and the super-cooled liquid phase. The difference between these two estimates was scaled by x_j because in the limit of infinitesimal mole fraction, the freeze point temperature (of that component) approaches zero and the molecular-level structures of the respective phases are expected to converge. The enthalpy of mixing is estimated as zero because the changes to the internal energy (U) brought about by mixing are small relative to the changes brought about by the phase change and there is little change in volume as hydrocarbons are mixed. This approximation it written here as Equation 17, and is further justified empirically by the measurements of Lundberg [23].

$$\Delta H_{mix} = P \, \Delta V_{mix} + \Delta U_{mix} \approx P \, \Delta V_{mix} \approx 0 \tag{17}$$

The entropy of mixing is significant relative to the entropy of fusion. To estimate its magnitude, we start with the following set of approximations or assumptions, analogous to those used for gasses. We mix one mole of component, j into another liquid that contains $(n_T - 1)$ moles of molecules other than j. As a simplifying approximation, entropy is considered first on a basis of conformations and permutations of molecules rather than quantum energy levels available to the system. Moreover, the role of vacancies on the possible number of lattice conformations is, at first neglected as is the likely possibility that liquids mix on a cluster scale rather than a molecular scale. To correct for these three assumptions, an empirical scaling factor (α) is included to the computed entropy of mixing of an ideal solution. Within the ideal solution approximation, the entropy of the fluid is given by equation 18 where k is Boltzmann's constant and N is the number of molecules. This is applied to one mole of component j, $(n_T - 1)$ moles of the fluid before it is mixed with component j, and (n_T) moles of the mixed fluids, and the entropy of mixing is expressed as Equation 19, where the subscripts refer to component j (j), the final mixture (T), or the mixture without component j (T-j). By expressing the first occurrence of N_T as $(N_{T-j} + N_j)$, rearranging terms, and factoring out Avogadro's number, Equation 20 results.

$$S = k * \ln(N!) \tag{18}$$

$$\Delta S_{mixing} = k * [\, N_T \ln(N_T) - N_{T-j} \ln(N_{T-j}) - N_j \ln(N_j) \,] \tag{19}$$

$$\Delta S_{mixing} = -R * (1/X_j) * [\, (1 - X_j) * \ln(1 - X_j) + X_j * \ln(X_j) \,] \tag{20}$$

By substitution, Equation 4 can now be re-written as Equation 21, which is readily solvable by iteration: $y_n = f(y_{(n-1)})$. In this work the initial guess was determined by setting $x_i = 0$ and convergence was achieved within five iterations.

$$T_{f,mix} = \frac{\Delta H_{fusion,i} + x_i * (Cp_{sol,i} - Cp_{liq,i}) * (T_{f,i} - T_{f,mix})}{\Delta S_{fusion,i} + x_i * (Cp_{sol,i} - Cp_{liq,i}) * \ln\left(\frac{T_{f,i}}{T_{f,mix}}\right) + \alpha * \Delta S_{mixing}} \tag{21}$$
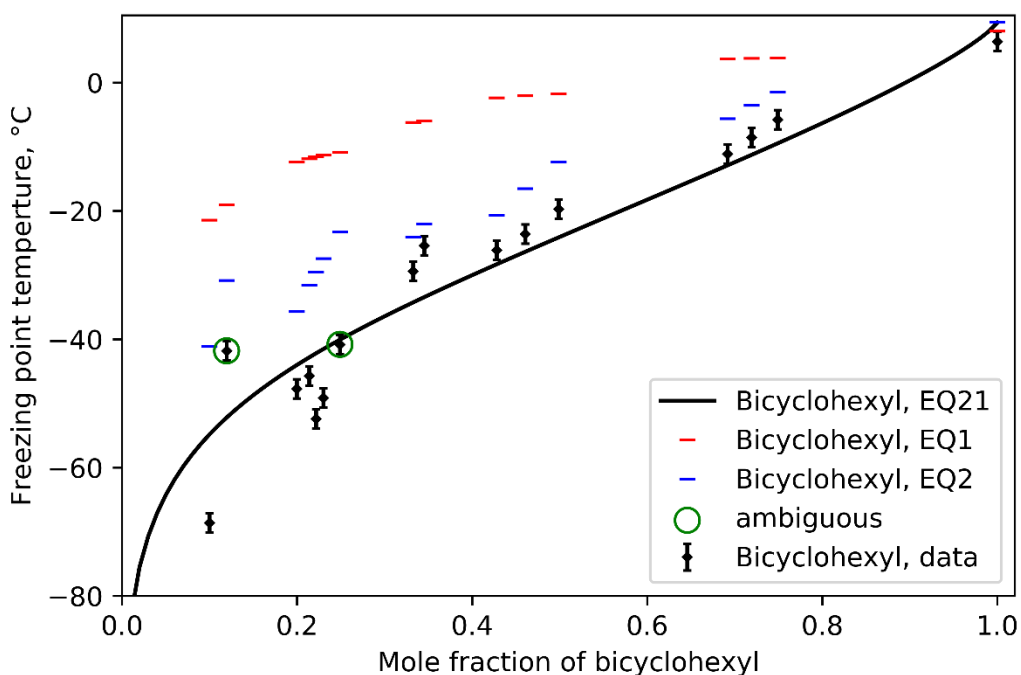
Figure 4. Measured and predicted freeze points of mixtures containing bicylcohexyl. The symbol height equals one standard deviation, measurement uncertainty. The error bar height equals two standard deviations.

3.   Results and Discussion

The experimental data and model predictions are presented within three figures and one table. In Figure 4, the freeze point temperature is displayed as a function of the mole fraction of bicyclohexyl (black diamonds) because this is the component that is most likely to freeze first (or thaw last). Apart from the two (circled) data points with a declared freeze point of bicyclohexyl of -40.8 and -41.8 ˚C, there is a very high degree of confidence that the last component to melt is as labeled in the figure. Those two data points refer to mixtures that contain 75%v petroleum fuel sample POSF 12968, which has a freeze point of -34.2 ˚C when it is not diluted by other solvents. The identity of the highest-freeze-point component in those two mixtures is not known with certainty. It could be bicyclohexyl or it could be an undetermined component found in the petroleum fuel sample. Assuming, for the sake of argument, that the highest-freeze-point component originated with the petroleum blend component, it is certainly possible that diluting its concentration by a factor of 0.75 is sufficient to cause a freeze point depression of 7 ˚C.

In Figure 5, the freeze point temperature of a different subset of data is displayed as a function of the mole fraction of n-tridecane (black diamonds) because this is the components that is most likely to freeze first (or thaw last) in those mixtures. Apart from the two (circled) data points with a declared freeze point of n-tridecane of -36.6 and -32.9 ˚C, there is a very high

degree of confidence that the last component to melt is as labeled in the figure. Those two data points refer to mixtures that contain 75%v trans decalin, which has a freeze point of -26.5 °C when it is not diluted by other solvents. The identity of the highest-freeze-point component in these mixtures is not known with certainty. It could be n-tridecane or it could be trans decalin. Assuming, for the sake of argument, that the highest-freeze-point component is trans decalin, it is certainly possible that diluting its concentration by a factor of 0.75 is sufficient to cause a freeze point depression of 6 to 10 °C. Among the displayed data, seven data points have a mole fraction of n-tridecane between 0.0156 and 0.0170, mixed with a variety of solvents. The measured freeze points of these mixtures are -32.9 ± 0.8 °C. The predicted freeze points of tridecane, via Equation 21, in these mixtures are -18.1 ± 0.2 °C. In contrast, the Equations 1 and 2 prediction ranges have standard deviations of 2.2°C and 6.0°C, respectively. While this is a particularly conspicuous example, closer examination of all the data collected in this investigation shows little to no solvent influence on the freeze point of the highest-freeze-point component, when plotted against the mole fraction of that component. In other words, the measured data shown in Figures 4 and 5 confirm that the freeze point of component j, where j refers to either bicyclohexyl or n-tridecane in these examples, depends on the mole fraction of the highest-freeze-point component, and does not show significant variation with the identity or number of other components present in the mixture. Consistent with this observation, Equation 21 predicts no solvent influence, while Equations 1 and 2 are both influenced significantly by solvents.
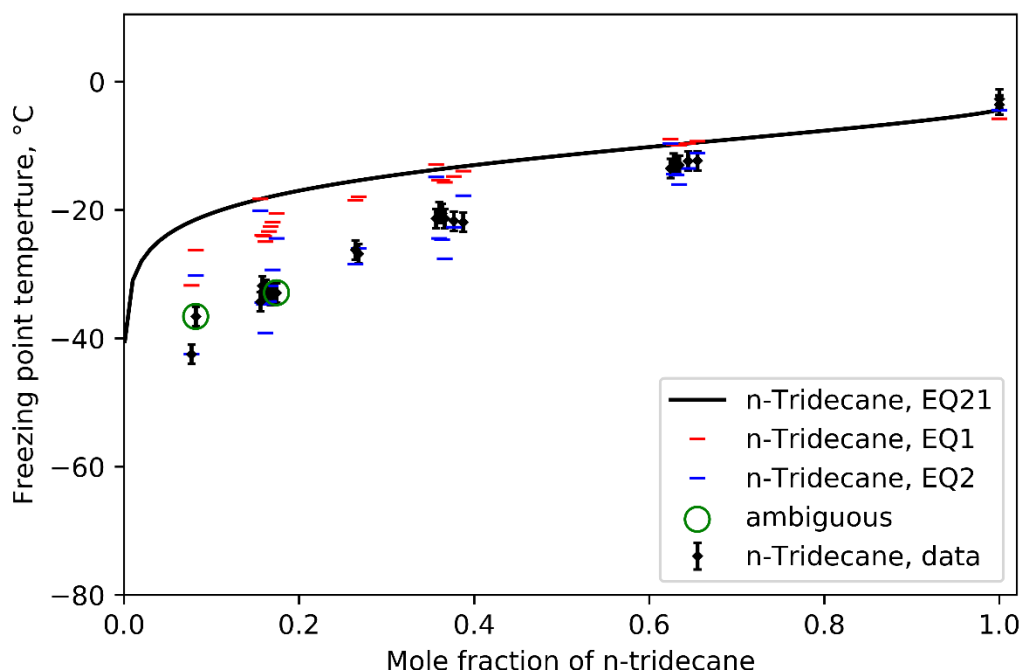

Figure 5. Measured and predicted freeze points of mixtures containing n-tridecane. The symbol height equals one standard deviation, measurement uncertainty. The error bar height equals two standard deviations.

Both sets of data show a steep decline in freeze point as the mole fraction of the highest-freeze-point component is taken below 0.25. While the model represented by Equation 21 also shows a steep roll off at low mole fraction, its knee is closer to a mole fraction of 0.10. Such an offset could be due to the assumption that the scale factor, α is independent of mole fraction and the identity of the highest-freeze-point component. We chose $\alpha = 0.25$ based on visual fit to bicyclohexyl datapoints. This selection results in a model that is conservative for both sets of data at mole fractions less than 0.25. Conservative estimates for freeze point at these low mole fractions provides a safety factor for system design. For detailed engineering work, to remove some of the conservatism, Equation 21 can be tuned to existing data by varying α; resulting in a refined design curve for the component of interest.

The modelling error is further discussed here and is summarized in Figures 6 and 7. Figure 6 is a unity plot, including the results for all three models, with filled/open symbols delineating mixtures containing bicyclohexyl or n-tridecane. The motivation for making this delineation is to underscore the observation made here that the identity of the highest-freeze-point component is particularly important to the freeze point of the mixture. The points previously identified as ambiguous in Figures 4 and 5 are included in this figure (and the next figure), which is relevant only to Equation 21 results as the predicted freeze point of bicyclohexyl (or n-tridecane) is irrelevant if that is not the highest-freeze-point component. Based on this figure alone, the superiority of Equation 21 over Equation 2 is not obvious, but clearly the extension of the wholly empirical model, Equation 1 to mixtures outside of the domain over which it was developed leads to undesirable error, especially at low freeze points.
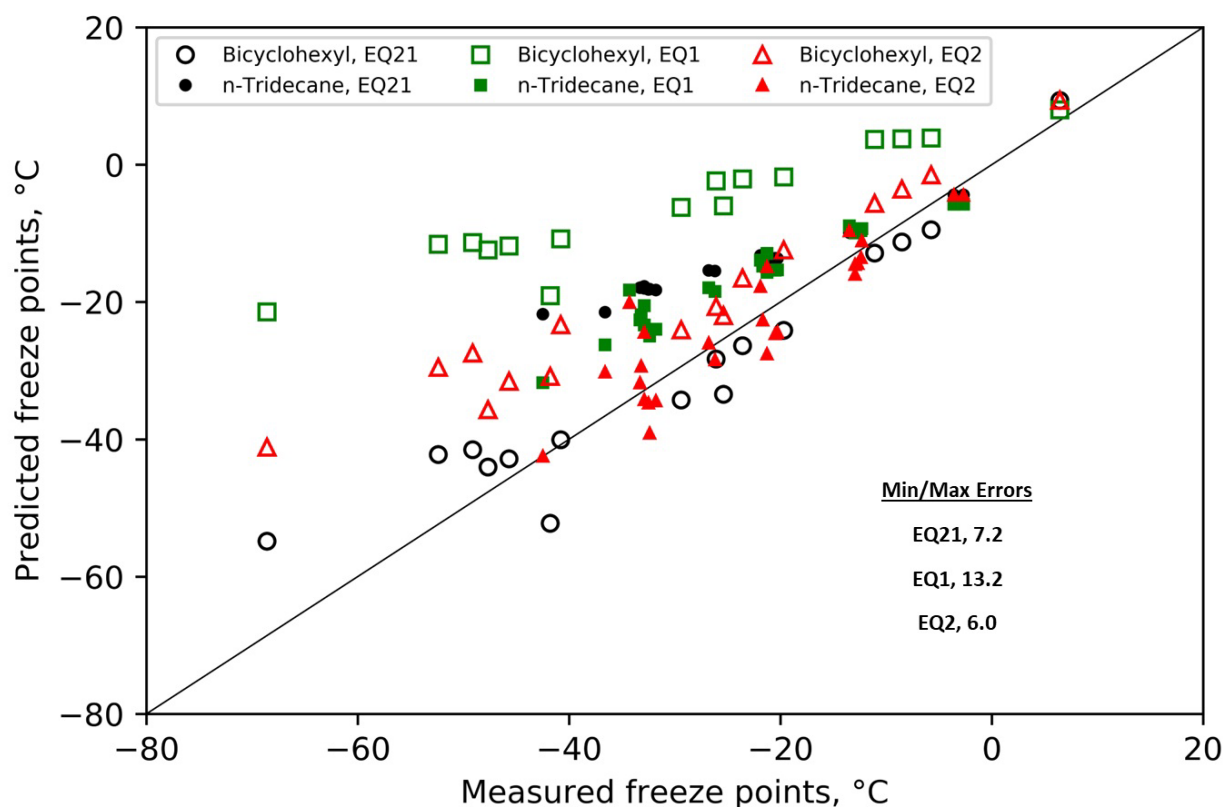
Figure 6. Unity plot assessment of modelling error.

Figure 7 shows the modeling error (prediction minus data) of each prediction method at each point, plotted as a function of the mole fraction of bicyclohexyl (open symbols) or the mole fraction of n-tridecane (filled symbols). All the models struggle to predict the freeze point of the highest-freeze-point component when it constitutes less than 25%$_{mol}$ of the mixture, which is unfortunate because real jet fuels rarely contain any component at a concentration above 25% because of volatility and freeze point restrictions[24]. (The only exception known: Annex A5 of ASTM D7566 does allow for a certain C12 isomer made from isobutanol to exist in jet fuel above this threshold.) Therefore, it is important to understand freeze point model prediction errors of relatively dilute solutes even though it is difficult to obtain a measured freeze point at lower mole fractions due to device limitations and because of competition between the various high-freeze-point components in the mixture. The results of this work indicate that the model developed here performs more consistently than the previous model. The un-tuned model (Equation 21) overpredicts by 0.7 to 20.1 °C at mole fractions between 0.065 and 0.25 (but can be tuned to match the measured data to within the 95 percent confidence interval). While the predicted temperature errors vary inversely and steeply with mole fraction in this region, the maximum observed offset in mole fraction is 0.07.
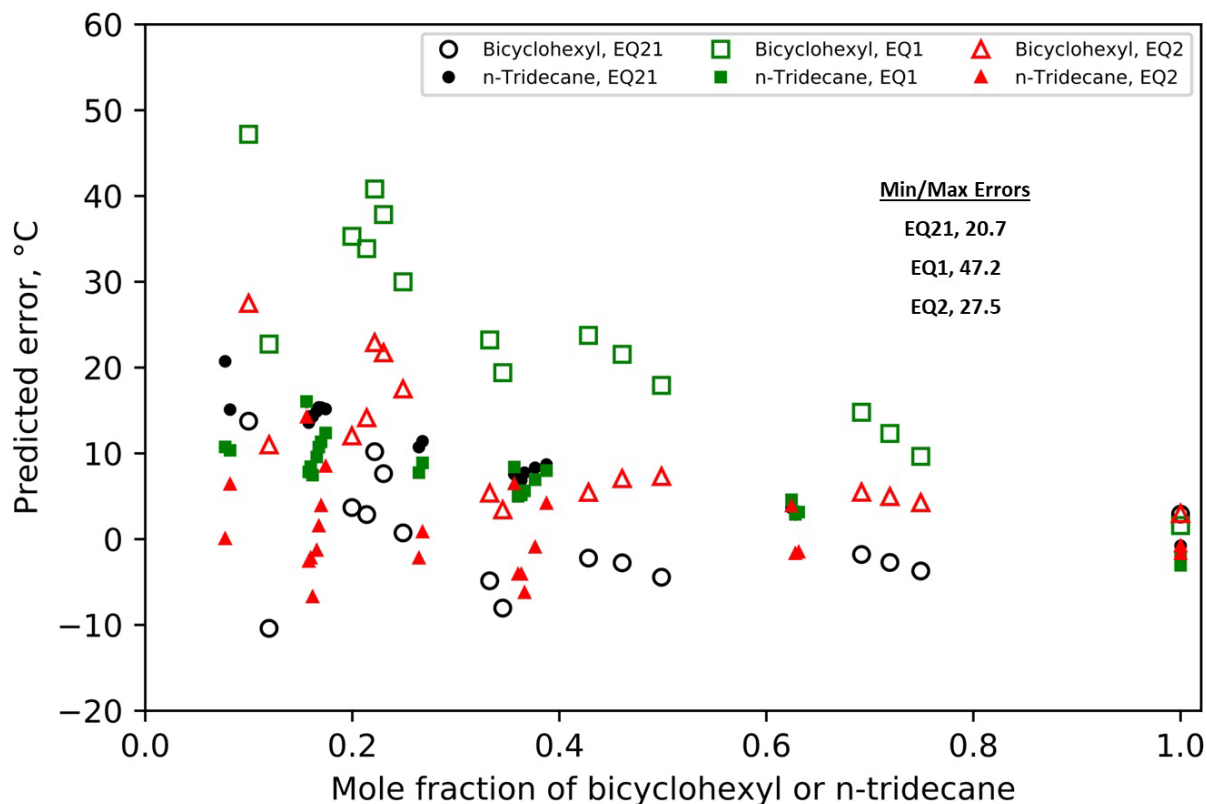
Figure 7. Modelling error as a function of high-freeze-point component mole fraction.

Quantifying allowable high-freeze-point component concentration in jet fuel is of more use to potential SAF developers. For example, the un-tuned model implies that n-tridecane can exist in jet fuel at a mole fraction up to 0.001 without risk to the freeze point specification (-40 ˚C), but the data shows that it can exist at a mole fraction up to 0.065. The un-tuned model therefore conservatively converts the specified maximum freeze point to a maximum allowed concentration of components based on risk to the to the freeze point specification. Most importantly, this model is known to be conservative. In contrast, and as seen in Figure 7, the errors stemming from the purely empirical models can be positive or negative and can have a magnitude up to 47 °C.

In problems where the composition and identity of each high-freeze-point component is known, Equation 21 is particularly convenient because it requires only two pieces of data (two of the three properties related by Equation 3) per component. In contrast, to use Equation 1 or 2, the freeze point and composition of every component is required. For example, a fuel developer may want to know how much 2-ethylnaphthalene (nBp=-7.4 ˚C) can be added to fuel (e.g., to meet the seal swell, fit-for-purpose requirement) before it presents a high risk to the freeze point specification. In this case, the evaluating engineer only needs to know, additionally, the enthalpy or entropy of fusion of 2-ethylnaphthalene and how much of this component may be already present in other blend stocks. In another potential application, a SAF developer may learn that its product freezes at, say -10 ˚C and desires to potentially alter the composition of their product to maximize the blending ratio into petroleum-derived fuel, resulting in a blended fuel with a freeze point less than -40 ˚C. The first step, in this example would be to identify each of the components in the SAF, via a technique such as GCxGC/FID-VUV[21], that present risk to the freeze point limit. The next step would be to employ the model developed here to understand how much reduction in concentration of each high-freeze-point component is necessary to drive its, as-mixed freeze point below -40 ˚C.

A potential concern with Equation 21, for species with an unusually low entropy of fusion such as trans-decalin (ΔS_{fusion} = 13.4 J/mol/K), is that it may not iterate to convergence at some values of mole fraction because the denominator approaches zero. The entropy of fusion of bicyclohexyl is 24 J/mol/K, which is towards the low end of its distribution for hydrocarbons,

and for n-tridecane it is 106 J/mol/K, which is towards the high end of its distribution for hydrocarbons. The rule-of-thumb estimate for hydrocarbons provided by Walden [25] is 56.5 J/mol/K. It is this term in Equation 21 (ΔS_{fusion}) that has the biggest impact on the slopes of the freeze point verses mole fraction curves shown in Figures 4 and 5. For samples of unknown composition, application of Equation 21 is not possible but Equations 1 and 2 can be used, provided the freeze point of each component is known, even if the so-called, "component" is really a complex mixture of unknown composition instead of a neat material.

The four data points that are labeled as ambiguous in Figures 4 and 5 are provided in Table 2 along with additional data as relevant to support the point of this paragraph. All other data, including average measured freeze point and model predictions are provided in the supporting material. Table 2 contains three sets of data points. One set is comprised of three blends of 1,2-dimethylcyclohexane with petroleum fuel sample, POSF 12968 in different blend ratios. Since the measured freezing point of two of these mixtures is greater than or equal to the freeze point of neat, 1,2-dimethylcyclohexane, we know that highest-freeze-point component in these mixtures originates with POSF 12968, which has a freeze point of -34.2 C. As the fraction of POSF 12968 decreases the freeze point of the mixture decreases. At the lowest blend ratio, the freeze point of the mixture is below the freeze point of both components. The observed trend between the mole fraction of POSF 12968 and the freeze point of the mixture is consistent with those of bicyclohexyl or n-tridecane mixtures, as discussed earlier. However, the predicted freeze point of 1,2-dimethylcyclohexane is also close to the observed freeze point of the mixture. The next set of datapoints include three different mixtures comprised of 75%v POSF 12968. The observed freeze point range of these mixtures is 41.6 ± 1.2. Since we know from theory that the freeze point of any component that is mixed at 12.5% should be significantly lower than it is at 25%, we conclude that highest-freeze-point component in the mixture that contains 12.5%v of bicyclohexyl is not bicyclohexyl. The highest-freeze-point component in the mixture that contains 25%v bicyclohexyl is not clear as its observed freeze point is only 1.3 °C higher than the average of the other two points in this set. The last subset included in this table are mixtures that contain 75%v trans-decalin, which has a freeze point of -26.5 °C when not mixed. All three of the mixtures have an observed freeze point that is significantly less than that of neat trans-decalin so it cannot be immediately ruled out as the highest-freeze-point component of any of these mixtures. However, the range of observed freeze points, -39.3 to -32.9 °C is sufficiently high to conclude that the highest-freeze-point component in the mixture that contains 25%v n-tridecane is n-tridecane, not trans-decalin. At the other end of the range, the highest-freeze-point component in the mixture containing 25%v dimethylcyclohexane must be trans-decalin, by elimination, because the freeze point of neat 1,2-dimethylcyclohexane is lower than the observed freeze point of the mixture. Regarding the mixture that contains 12.5%v n-tridecane and 75%v trans-decalin, it is not clear which of these two materials is the highest-freeze-point component.

Table 2. Freeze point of mixtures with uncertain highest-freeze-point component with supporting data .

Component A Blend Ratio, %v	Component B Blend Ratio, %v	Data, ºC	EQ21, ºC	EQ1, ºC	EQ2, ºC
12DMCH [a,b] 75	POSF 12968 [b] 25	-60.3	A, -55.8 [c] B, << -34	-66.9	-71.1
12DMCH [d] *50*	**POSF 12968** [e] **50**	-50.5	*A, -60.2* **B, < -34**	-60.2	-64.3
12DMCH *25*	**POSF 12968** **75**	-42.3	*A, -66.2* **B, < -34**	-55.5	-57.5
Bicyclohexyl 25	POSF 12968 75	-40.8	A, -40.1 B, < -34	-12.4	-35.7
Bicyclohexyl	**POSF 12968**	-41.8	*A, -52.2* **B, < -34**	-22.2	-46.6

12.5 [f]	75 [f]				
12DMCH	**Trans-decalin**	-39.3	*A, << -51* **B, -33.6** [c]	-36.6	-42.8
25	75				
n-tridecane	*Trans-decalin*	-32.9	**A, -19.1** *B, -35.8* [c]	-20.5	-24.4
25	75				
n-tridecane	Trans-decalin	-36.6	A, -22.4 B, -34.8 [c]	-26.5	-33.6
12.5 [f]	75 [f]				

[a] 12DMCH is short for cis- 1,2-dimethylcyclohexane.
[b] Where neither component is italicized or bolded, it is unclear which component is the last to thaw.
[c] Used Walden's value for entropy of fusion and measured freeze points of neat material as inputs.
[d] Components identified with italics are not the last to thaw.
[e] Components identified with bold font either are or contain the highest-freeze point species.
[f] The balance of the composition is 1,2-Dimethylcyclohexane.

4. Conclusion

A new model for freeze point prediction of sustainable aviation fuel candidates is presented along with experimental measurements of several hydrocarbon blends. The model is based on first principle thermodynamics and validated with experimental freeze data. The developed model is designed to require only the following inputs: the molecular identity of the components that present risk of freezing above -40 °C, their respective mole fractions, and their repective freeze points and enthalpies and entropies of fusion as pure components. It is not necessary to know the identity or mole fractions of the light fraction, or the identity of components that are present at very low concentration. The model captures the non-linear behavior verus mole fraction for mixture freeze points, and conservatively estimates the freeze point at low mole fractions. For fractions between 0.065 and 0.25 the model is shown to overpredict the measured freeze point by 0.7 to 20.1 °C, where the larger errors correspond to low mole fraction where the change in freeze point with mole fraction is steep. The practical implication of this error is that the model underpredicts the allowable mole fraction of the high-freeze-point components. For example, the acceptable mole fraction of n-tridecane is underestimated by 0.064.

5. Supporting Information

All the freeze point data, measured and predicted, taken in support of this manuscript is provided within a single worksheet of the attached document (XLSX).

Acknowledgments

References

[1]   Air Transportation Action Group: ATAG. *Aviation Beyond Borders*. Geneva, Switzerland, 2020.

[2]   Air Transportation Action Group: ATAG. *Waypoint 2050*. Geneva, Switzerland, 2020.

[3]   Huq, N. A., Hafenstine, G. R., Huo, X., Nguyen, H., M., Stephen, T., Conklin, D. R., Stück, D., Stunkel, J., Christensen, E. D., Cameron, Hays, Wiatrowski, M. R., Zhang, Y., Tao, L., Yang, Z., Heyne, J., Abdullah, Z., and Vardon, D. R. "Towards Net-Zero Sustainable Aviation Fuel with Wet Waste-Derived Volatile Fatty Acids." *Proceedings of the National Academies of Science*, 2021.

[4]   Yang, Z., Kosir, S., Stachler, R., Shafer, L., Anderson, C., and Heyne, J. S. " A GC × GC Tier α Combustor Operability Prescreening Method for Sustainable Aviation Fuel Candidates ." *Fuel*, Vol. 292, No. September 2020, 2021, p. 120345. https://doi.org/10.1016/j.fuel.2021.120345.

[5]   Heyne, J., Rauch, B., Le Clercq, P., and Colket, M. "Sustainable Aviation Fuel Prescreening Tools and Procedures." *Fuel*, Vol. 290, 2021, p. 120004. https://doi.org/https://doi.org/10.1016/j.fuel.2020.120004.

[6]   ASTM. "Standard Test Method for Freezing Point of Aviation Fuels." *Manual on Hydrocarbon Analysis, 6th Edition*, 2008, pp. 342-342–4. https://doi.org/10.1520/mnl10883m.

[7]   Moynihan, C., Shahriari, M., and Bardakci, T. "Thermal Analysis of Melting and Freezing of Jet and Diesel Fuels." *Thermochimica Acta*, Vol. 52, 1982, pp. 131–141.

[8]   Petrovic, K., and Vitorovic, D. "A New Method for the Estimation of the Freezing Point of Jet Fuels Based on the N-Paraffin Content." *J. Inst. Petroleum*, Vol. 59, No. 565, 1973, pp. 20–26.

[9]   Solash, J., Hazlett, R., Hall, J., and Nowack, C. "Relation between Fuel Properties and Chemical Composition. 1. Jet Fuels from Coal, Oil Shale and Tar Sands." *Fuel*, Vol. 57, No. 9, 1978, pp. 521–528. https://doi.org/https://doi.org/10.1016/0016-2361(78)90036-4.

[10]   Affens, W., Hall, J., Holt, S., and Hazlett, R. "Effect of Composition on Freezing Points of Model Hydrocarbon Fuels." *Fuel*, Vol. 63, No. 4, 1984, pp. 543–547. https://doi.org/https://doi.org/10.1016/0016-2361(84)90294-1.

[11]   Cookson, D. J., Lloyd, C. P., and Smith, B. E. "Investigation of the Chemical Basis of Kerosene (Jet Fuel) Specification Properties." *Energy and Fuels*, Vol. 1, No. 5, 1987, pp. 438–447. https://doi.org/10.1021/EF00005A011.

[12]   Zabarnick, S., and Widmor, N. "Studies of Jet Fuel Freezing by Differential Scanning Calorimetry." *Energy & Fuels*, Vol. 15, No. 6, 2001, pp. 1447–1453. https://doi.org/https://doi.org/10.1021/ef010074b.

[13]   Widmor, N., Ervin, J. S., Zabarnick, S., and Vangness, M. "Studies of Jet Fuel Freezing by Differential Scanning Calorimetry and Cold-Stage Microscopy." *Journal of Engineering for Gas Turbines and Power*, Vol. 125, No. 1, 2003, pp. 34–39. https://doi.org/10.1115/1.1492836.

[14]   Kuryakov, V., Zaripova, Y., Varfolomeev, M., De Sanctis Lucentini, P. G., Novikov, A., Semenov, A., Stoporev, A., Gushchin, P., and Ivanov, E. "Comparison of Micro-DSC and Light Scattering Methods for Studying the Phase Behavior of n-Alkane in the Oil-in-Water Dispersion." *Journal of Thermal Analysis and Calorimetry 2020 142:5*, Vol. 142, No. 5, 2020, pp. 2035–2041. https://doi.org/10.1007/S10973-020-10001-9.

[15]   Kuryakov, V. N., Ivanova, D. D., Tkachenko, A. N., and Sedenkov, P. N. "Determination of Phase Transition Temperatures (Melting, Crystallization, Rotator Phases) of n-Alkanes by the Optical Method." *IOP Conference Series: Materials Science and Engineering*, Vol. 848, No. 1, 2020, p. 012044. https://doi.org/10.1088/1757-899X/848/1/012044.

[16]   Kuryakov, V. N., Kuryakov, V. N., Ivanova, D. D., Semenov, A. P., Gushchin, P. A., Ivanov, E. V., Novikov, A. A., Yusupova, T. N., Shchukin, D., and Shchukin, D. "Study of Phase Transitions in N-Tricosane/Bitumen Aqueous Dispersions by the Optical Method." *Energy and Fuels*, Vol. 34, No. 5, 2020, pp. 5168–5175. https://doi.org/10.1021/ACS.ENERGYFUELS.9B03566.

[17]   Home | AJF:TD | U of I. https://altjetfuels.illinois.edu/. Accessed Jun. 5, 2021.

[18]   Edwards, T. Reference Jet Fuels for Combustion Testing. 2017.

[19]   Lemmon, E.W., Bell, I.H., Huber, M.L., McLinden, M. O. NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP.

[20]   AlMulla, H. A., and Albahri, T. A. "Predicting the Properties of Petroleum Blends." *Petroleum Science and Technology*, Vol. 35, No. 8, 2017, pp. 775–782. https://doi.org/https://doi.org/10.1080/10916466.2016.1277238.

[21]   Heyne, J., Bell, D., Feldhausen, J., Yang, Z., and Boehm, R. "Towards Fuel Composition and Properties from Two-Dimensional Gas Chromatography with Flame Ionization and Vacuum Ultraviolet Spectroscopy." *Fuel*, Vol. 312, 2022, p. 122709. https://doi.org/10.1016/J.FUEL.2021.122709.

[22]   Naef, R. "Calculation of the Isobaric Heat Capacities of the Liquid and Solid Phase of Organic Compounds at and around 298.15 K Based on Their 'True' Molecular Volume." *Molecules 2019, Vol. 24, Page 1626*, Vol. 24, No. 8, 2019, p. 1626. https://doi.org/10.3390/MOLECULES24081626.

[23]   Lundberg, G. W. "Thermodynamics of Solutions XI. Heats of Mixing of Hydrocarbons." *Journal of Chemical and Engineering Data*, Vol. 9, No. 2, 1964, pp. 193–198.

[24]   Colket, M., and Heyne, J. *Fuel Effects on Operability of Aircraft Gas Turbine Combustors*. AIAA, Progress in Astronautics and Aeronautics, 2021.

[25]   Walden, P. "Heat of Fusion, Specific Cohesion, and Molecular Complexity at the Melting Point." *Elektrochem*, Vol. 14, 1908, p. 713.

Appendix 4

A Dataset Comparison Method Using Noise Statistics Applied to VUV Spectrum Match Determinations

Paper 8: A Dataset Comparison Method Using Noise Statistics Applied to VUV Spectrum Match Determinations

1. Introduction

The ability to determine whether a given dataset is consistent with theory or some related dataset is critical across the sciences. Significant applications of this discipline range from criminal justice [1-3] to manufacturing process control [4,5] to name a few. In analytical chemistry, determining an analyte's identity is paramount. In this work, we show how the general notion of comparing the difference between pairs of datasets to white noise can be applied to spectral data to positively identify inconsistencies/mismatches between spectra and, in some cases, to positively identify an unknown, single molecular species.

Signature interactions (spectra) between molecules and electromagnet radiation are commonly used to identify an unknown analyte. The process includes experimentally collecting a spectrum of the unknown analyte (sample) and comparing it, one-by-one, to cataloged library of spectra of known molecules (reference) [6,7]. To score the goodness of fit between the sample spectra and an arbitrary reference spectrum, researchers typically use either the coefficient of determination (R^2), a related similarity term [8,9], or some combination of them. By summing some measure of the disparities, as done with an R^2 or mean absolute error determination, all these criteria forfeit some level of detail that may be readily detected through visual inspection. For example, common-sign disparities at consecutive wavelengths are more significant than randomly spaced disparities, and disparities in regions of low noise are more significant than disparities in regions of high noise. However, as will be discussed in this manuscript, numerical methods can illuminate the patterns which are naturally observed through visible methods, and numerical scoring facilitates automation of the match/no-match decision logic.

Electromagnetic radiation emission and absorption, mass spectrometry, nuclear magnetic resonance, characteristic chromatographic elution time, phase transition temperature, density, viscosity, refractive index, or any other observable property have been used to identify and quantify materials of interest. What began as a human-centered qualitative evaluation process has developed into the diverse field of chemometrics [10,11] where expected results, or hypotheses, are tested against measured information.

Advances in the diagnostic capabilities of spectrometers and the data processing capabilities of computers have enabled direct comparisons of reference spectra to sample spectra. For example, contemporary and widely available infrared (e.g. FTIR) and vacuum ultraviolet (VUV) light detectors can record spectra over more than a thousand wavelengths at high data acquisition frequencies.

Notably, the high acquisition rates of modern VUV detectors (77 Hz) have been capitalized in gas chromatography (GC) [12] and more recently two-dimensional gas chromatography (GC×GC).[13] In contrast to mass spectrometers, VUV detectors have the ability to discern isomeric structural differences between species [14] and show the potential to be quantitative due to the linearity of response factors.[15] The separation from GC combined with the spectral response from VUV enable isomer-specific identification potential.[12,16-22] These approaches currently 'identify' species by searching spectral libraries primarily for the highest R^2 match. This approach relies heavily on the completeness of the spectral library. The confidence in these results, although effective in simple exercises, becomes increasingly questionable as more uncertainties are included.

For the application discussed herein, high spectral and temporal resolution is needed to distinguish between structural and stereoisomers of hydrocarbons in sustainable aviation fuel (SAF) candidates. The motivation for knowing these isomers has been documented elsewhere.[19,23,24] To put it briefly here, knowing the isomeric structures minimizes the risk to candidate SAF pathways, as isomers can significantly impact aviation fuel properties.

In this work, experimental procedures and data processing are used to demonstrate that the analyte in the sample chamber and the detector noise are the only factors contributing to the observed VUV signal. The detector noise at each wavelength closely follows a Gaussian distribution and is unaffected by the presence of analyte in the detector which enables the following hypothesis: if the signal observed is not a match with reference spectra, then the residuals will have characteristics that are inconsistent with that of noise alone. For reasons that will be made clear in the methods section, the residuals are adjusted proportional to the variance of the background noise at each wavelength. The described methods are applied to three examples and discussed; a clear match and a clear mismatch with the same R^2 (0.976) and analysis of n-alkane spectra

demonstrating that unambiguous identification of analytes with spectra that are visually similar ($R^2 \geq 0.9997$) to that of several reference spectra can be accomplished. Additional examples are provided in section 2 of the supporting materials.

2. Methods

Experimental Setup. In this experiment, vacuum ultraviolet spectroscopy is coupled with comprehensive two-dimensional gas chromatography (GC×GC). A flame ionization detector is also equipped, but unused in this work. A graphical overview of the GC×GC-FID/VUV is available in the supporting materials, figure S1 illustrating the major components experimental setup. Details of this system and configuration have been described previously.[19] Key points are repeated here for convenience. A vacuum ultraviolet spectrometer (VUV Analytics, VGA-101) with an Agilent 8890 and a SepSolve GC×GC flow modulation system was used for this work. The system includes two columns with a flow modulator and a split plate connected in series after the second column. Parallel sample streams are sent to a flame ionization detector (FID) – not used for this work - and the VUV spectrometer. The columns employed were a Restek Rxi-17Sil MS (60m x 0.32mm x 0.5μm) and a Restek Rxi-1 (15m x 0.32mm x 0.5μm), respectively. The carrier gas for the separation was grade 5.0 helium, which passed through a Restek Triple Filter before entering the instrument. The carrier gas flow rate, modulation rate, oven temperature profile, sample injection volume, and duration of the experiment all contribute to the details of the separation of different species that may be in any sample. However, these items do not significantly impact the quality of spectra collected over periods during which 1 or 0 analytes flow into the VUV spectrometer. The temperature ramp rate used for the sample spectra starts at 40°C for 30 seconds, followed by a 1°C/minute ramp rate until 280°C is reached, where it remains for 10 minutes. This experimental setup uses longer than typical columns, and therefore requires a longer temperature program for optimal separations. Data is sent to the computer throughout the experiment over a range of wavelengths from 115 to 430 nm at 76.92 Hz.

Chemicals Used. Sample spectra originate from analytes that were separated using GC×GC from a petroleum-derived Jet-A fuel sample labeled as A-2 (POSF 10325) by the fuels committee of the National Jet Fuel Combustion Program [25] and Synthetic Aromatic Kerosene (SAK) by Virent.[26] Reference spectra originate from the Jet Range VUV database from VUV Analytics and supplemented with many internally measured spectra. Internally measured spectra were recorded by injecting samples of single analytes into the GC×GC system and recording the lone peak response. The spectrum for toluene was measured using a sample from Fischer Chemical (>99.5%). The spectrum for 1,3-diethylbenzene was from the VUV Analytics spectral library.

Data and Storage. Large data files are created over the course of each experiment, which last for ~250 minutes. Spectral response at ~2800 wavelengths are acquired at 76.92 Hz (F_{acq}) and typically stored at some lower frequency (F_{sto}), in this work 10 Hz, resulting in data files with sizes on the order of one gigabyte each. Absorption intensity (I) is determined by Equation 1. In this equation, "dark" is the number of counts detected when the light source is turned off and L_0 is the number of counts detected with the lamp on, immediately prior to sample injection. The subscript reflects the fact that each of these signals vary with wavelength, and L is the number of counts detected at 76.92 Hz. The stored values of I_λ are sums of Z snapshots, where Z is the ratio of the acquisition frequency to storage frequency.

$$I_\lambda = \Sigma[\log(L_{0,\lambda} - dark_\lambda) - \log(L_\lambda - dark_\lambda)]_i \tag{1}$$

The variation in light intensity as a function of wavelength is driven by the VUV light source and is significant. The intensity of light at, say 160 nm is much higher than the intensity of light at, say 300 nm, so the impact of a few random counts on I_{160} is significantly lower than the impact of the same few random counts on I_{300}. At this level of data processing the magnitude of the noise varies with wavelength. Subsequent steps are taken to remove this sensitivity. A plot showing the variability of the light source at different wavelengths is included in supporting materials, figure S2.

Spectra Acquisition. All spectra ($I_{sam,\lambda}$) are extracted from the stored data points. Raw sample intensity at each wavelength ($\overline{I_{raw,\lambda}}$) is averaged over a modulated period of time containing N_{sam} data points and bracketing the highest intensity of the analyte peak of the chromatogram. From this raw sample average, a background signal is subtracted as defined by Equation 2, where the background signal ($\overline{I_{BG,\lambda}}$) is averaged over a period of time that is at least double that used for the raw sample and throughout an elution period that is free of analyte yet closely coupled with the analyte time stamp. By doing this, any drift in mean background signal throughout an experiment is removed from the spectrum to be compared. The integration windows for the raw sample and the background are selected manually during visual inspection and the variation in rms noise introduced by the manual process is estimated to be 5% of the rms background noise.

$$I_{sam,\lambda} = N_{sam} * (\overline{I_{raw,\lambda}} - \overline{I_{BG,\lambda}}) \qquad (2)$$

Spectrum Scaling. Intuitively, we all make two datasets (i.e. spectra or images) size consistent to facilitate comparisons. Typically for spectra, the intensity of either the sample or the reference would be adjusted to make the mean average difference between them equal to zero. This ostensibly minimizes R^2. However, minimizing the sum of the square divided by the background variance at each wavelength (σ_λ^2) helps to deemphasize noisier portions of the spectrum in favor of regions with sharper signal. Equation 3 is provided to clarify this processing step. In this equation, a scaler, α, is varied to minimize the objective function, Q. In this context, σ_λ^2 can be determined from any set of partitions of the data where no analyte is detected, and these need not be from the same run. In practice, the intensity as defined by equation 1, meaning no additional averaging as in equation 2, is used to determine σ_λ^2. This measure of noise variance does change over time as a product of lamp decay or other factors so periodic updates to σ_λ^2 are made.

$$Q = \sum_\lambda \frac{\left(I_{sam,\lambda} - \alpha * I_{ref,\lambda}\right)^2}{\sigma_\lambda^2} \qquad (3)$$

Residuals. The difference between a pair of size-consistent spectra (their residuals, r) can be more indicative of similarity between them than an overlay. In this work, the residuals are adjusted proportional to the standard deviation of the background noise (σ_λ) to remove sensitivity to wavelength, as defined by Equation 4. By doing this, the distributed characteristics (or moments) of the residual should be the same as the moments of the noise, provided the two spectra match. The value in this normalization scheme will be further highlighted later in this paper.

$$\epsilon_\lambda = \frac{I_{sam,\lambda}}{\sigma_\lambda} - \alpha * \frac{I_{ref,\lambda}}{\sigma_\lambda} \qquad (4)$$

First Moment. The first moment of ϵ, the mean value ($\bar{\epsilon}$) is the sum of ϵ_λ over all wavelengths, which should be zero for matching spectra, divided by the total number of wavelengths, N.

$$\bar{\epsilon} = \frac{\sum \epsilon_\lambda}{N} \qquad (5)$$

We recommend trimming the wavelength range of the spectrum to exclude everything except the union of ranges that are active with respect to each possible match to the sample spectrum, where 'active' implies absorption evident upon visual inspection with intensity ~2x the background noise. When a sample spectrum is compared against a reference spectrum that does not match it, the first moment will not be zero because α is determined by Equation 3 rather than minimization of the first moment. By excluding portions of the spectrum where the signal is ostensibly zero, the mean of the residual is not dampened by extraneous points and therefore easier to detect.

Second Moment. The second moment of ϵ is its variance (σ_ϵ^2) as defined in Equation 6. It depends primarily on three details of the experiments and data processing procedure: N_{sam}, N_{ref}, and α which are not consistent from experiment to experiment. To get around these dependencies, we recognize and exploit the fact that all these influences have the same impact on the signal variance whether or not the signal reflects an absorbance. For convenience, the tail (or inactive region) of the residual, meaning the region of wavelengths where ostensibly no absorbance occurs for either spectrum, sample or reference, is used to define the expected variance for matching spectra throughout the entire spectral range. Using the variance in the tail has the added advantage of capturing any noise existing in the reference spectrum. Spectrum splitting rather than wavelength trimming is recommended for evaluation of the variance, where the active portion of the spectral range is used to identify mismatching spectra and the inactive portion of the spectral range is used to define the expected variance.

$$\sigma_\epsilon^2 \equiv \left(\frac{1}{N}\right) * \sum_\lambda (\epsilon_\lambda - \bar{\epsilon})^2 \qquad (6)$$

Third Moment. The third moment of ϵ, shown by Equation 7, is its skew (*Sk*), which should be zero provided the two spectra match one another, consistent with Gaussian noise distribution. Unlike the lower moments, inclusion of ostensibly zero signal over much of the spectrum helps to establish the shape of the distribution which in turn helps to clarify features that are introduced via mismatches in the active region of the spectrum. In the mathematical representation, when there is a mismatch, σ_ϵ is reduced by the inclusion of the ostensibly zeroes data. Moreover, the sum is larger because it includes more terms and $\bar{\epsilon}$ is driven closer to zero and further from ϵ_λ over a preponderance of the active region of the spectrum, where that difference contributes most to *Sk*.

$$Sk = \left(\frac{1}{N * \sigma_\epsilon^3}\right) * \sum_\lambda (\epsilon_\lambda - \bar{\epsilon})^3 \tag{7}$$

Fourth Moment. The fourth moment of the distribution of ϵ, shown by Equation 8, is its kurtosis (κ). If the two spectra match one another the kurtosis of the distribution of the residual should be equal to that of a spectrum obtained from a blank run (3.0). Once again, inclusion of ostensibly zero signal in the inactive region of the spectral range helps to establish a baseline shape, which in turn helps to highlight departures (mismatches) from a normal distribution.

$$\kappa = \left(\frac{1}{\sigma_\epsilon^4}\right) * \sum_\lambda (\epsilon_\lambda - \bar{\epsilon})^4 \tag{8}$$

Quantile-Quantile (Q-Q) Plots. A plot of the quantiles of ϵ against the quantiles of a Gaussian distribution provides an alternative (or additional) measure of its distribution. The coefficient of determination (R^2) of this curve can be used to quantify the plots. For matching spectra, R^2 should be 1.0, as the noise in the measured absorption and the normalized residuals of known matches has been determined separately to be Gaussian. For mismatching spectra, the R^2 of this curve is generally much lower than the R^2 obtained by direct comparison (overlays) of the two spectra, which renders the R^2 of a Q-Q plot easier to interpret than the R^2 between two spectra (overlays).

Consecutive Signs. Random noise about a mean signal is equally likely to fall above or below the mean. The probability of C consecutive signals falling above the mean is $0.5^C \equiv p$. For a residual of two spectra comprised of N wavelengths, there are $(N + 1 - C) \equiv N'$ opportunities to achieve C consecutive signs. Generally, Equation 9 represents the probability (P) of observing exactly k occurrences of exactly C consecutive signs in a residual between two matching spectra comprised of N wavelengths. For any comparison between two spectra, it is straightforward to determine the maximum length of consecutive signs (C'), or the number of occurrences (k') of C consecutive signs. These numbers (1 & C') or (k' & C) can be inserted into equation 8, serving as one end of a summation over k, with 0 or N' serving as the other end, to get the probability that the residual in question is truly just random noise about zero, as it should be for matching spectra.

$$P_k = \left(\frac{N'!}{k! * (N' - k)!}\right) * p^k * (1 - p)^{N'-k} \tag{9}$$

Evaluation of consecutive signs provides a particularly convenient way to assess whether the process described herein results in truly random distributions when comparisons are made between spectra originating from the same exact sample or the same sample compound. It has been determined from 144 such comparisons of trimmed spectra containing 744 points each that a small bias exists, as we found 9-28 consecutive signs in the processed experimental data, compared to 7-15 consecutive signs when fully synthesized random data was used to simulate 144 residuals. While it is possible this bias could be introduced by one of the numerical processes, such as scaling for example, when compared against the number of consecutive signs found in the residuals between mismatching spectra, 68-289 consecutive signs in a field of 2448 comparisons, it is quite evident that the small bias observed in the processed experimental data is much less than the observed difference between known matches and known mismatches.

Moment Probabilities. Just as probability can be assigned to a particular determination of consecutive signs, the probability of an observed mean being part of a Gaussian distribution can be readily determined from established statistical methods. For the higher moments, however the formulae necessary to make these determinations are not generally available. To create a guide to serve a purpose like that of equation 9, random sampling from an ideal Gaussian distribution was used to create 25,000 batches of N points. The moments of each batch of N points were tallied and analyzed as a set of 25,000 items. The 95th percentile was extracted from this analysis for 30 different values of N to create the contours of the shaded regions shown in Figure 1.
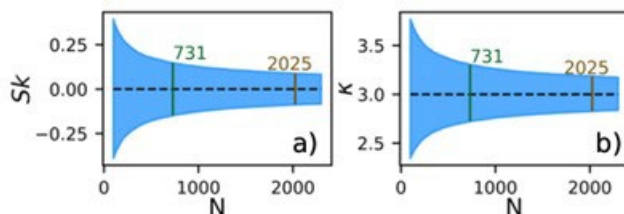
Figure 1. Random point selection simulations from a true Gaussian distribution demonstrating 95th percentile ranges of the (a) third and (b) forth moments as a function of population size. Any result within these ranges can be explained as noise alone.

A summary of the 95th percentile for each moment is provided in Table 1. When the numerical processing of experimental data comparison reveals a moment that is significantly outside of the 95th percentile of an ideal Gaussian distribution, there is a high degree of confidence that the corresponding sample and reference species are not the same. Definitive matches are harder to determine because that requires proof that only one out of all plausible matches cannot be ruled out. Because some analytes, especially those toward the lower volatility range of jet fuel may be isomers for which the reference spectra do not exist in the library there will be cases for which this process can only narrow the range of possible matches. That said, when the analyte happens to be one of the isomers for which a reference spectrum does exist, this process will identify that reference molecule as a probable match to the analyte.

3. Examples and Discussion

Background Signal. The images shown in Figure 2 were extracted from data collected at all timestamps during the first 20 minutes of an experiment, during which no analyte elutes from the chromatograph. The plotted result for each wavelength was derived by analyzing the distribution of the signals recorded at each timestamp. The skew is approximately 0.0 at all wavelengths and the kurtosis is nearly 3.0, illustrating that the background noise follows a Gaussian distribution at all
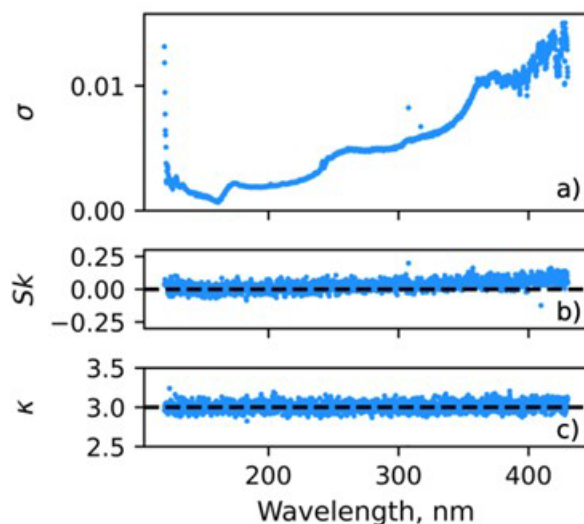
Figure 2. Statistics of VUV response with no analyte present in the detector at each independent wavelength. (a) Standard deviation, (b) skew, and (c) kurtosis at all wavelengths demonstrate an approximate Gaussian distribution with varying standard deviations. Dividing by standard deviation at each wavelength makes all measurements comparable for statistical analysis.

wavelengths. The mean background signal is also nearly zero at all wavelengths, confirming there is no change in absorption relative to time zero, and no significant change in light source intensity occurs throughout the 20-minute period selected for this background. However, the magnitude of the noise as measured by relative to its standard deviation is dependent on

the wavelength. This data is represented by the term σ_λ in Equations 3-5. By dividing each raw signal input at a given wavelength by σ_λ its standard deviation goes to one, but its skew and kurtosis remains the same, indicative of a Gaussian distribution. By normalizing the Gaussian distributions at each wavelength to the same size (via division by σ_λ) any collection of points sampled from each of the distributions, such as a single time stamp of the background noise across the spectrum, will also follow a Gaussian distribution with a standard deviation of one.

Table 1. Summary of Residual Moments Distribution Features Relevant to the Spectra Comparisons of this Work.

| Moment	Mean	N	95th Percentile
$\bar{\epsilon}$	0.00	N/A	$\pm 2\sigma_\epsilon/N^{0.5}$
†	1.00	731/1294	± 0.064
Sk	0.00	2025	± 0.088
κ	3.00	2025	$2.83 - 3.19$

† The ratio of the variances between the two partitions of the spectra is expected to equal one.

Matching Spectra. Figure 3 provides an example where the measured spectrum of toluene in A-2 is compared against its reference spectrum. The chromatogram peak corresponding to the sample analyte elutes at the same time as a toluene standard which provides supporting information to the statistics-driven conclusion that the analyte is toluene. Visually, the overlay of the spectra (Figure 3a) does look like a match, but the R^2 is low because of the low signal-to-noise ratio of the sample. A scale factor (α) of 1.026 was applied to render the two spectra size consistent. Figure 3b shows the residuals or difference between the measured and reference signal. The 2σ shaded region is the expected range the data will fall in based off the background noise profile (from figure 2a). For further detail on this process, see section 1 of the supporting material. Figure 3c shows the residuals after scaling for noise, demonstrating the equivalence of noise across the spectrum. Figure 3d shows the histogram of the residuals is very similar to a Gaussian distribution with a standard deviation as suggested by the tail data. Figure 3e shows another confirmation the residuals are consistent with a Gaussian distribution, evidenced by the linearity and high R^2 in the Q-Q plot. The moments of the noise-adjusted residual distribution being consistent with random sampling from a Gaussian distribution, which is summarized in Table 1, correctly suggests that the reference spectrum is a probable match to the sample spectrum. The high degree of linearity observed in the Q-Q plot of the noise-adjusted residuals further confirms that it is Gaussian and the relatively low count of consecutive wavelengths at which the sign of the residual is unchanged is also consistent with a match. The residuals therefore exhibit characteristics that are consistent with that of noise alone and can be logically attributed to noise alone. It is worth noting however that as signal-to-noise ratio decreases, several of the suggested measures of match approach that of noise alone (a Gaussian distribution) even when the sample and reference spectra do not correspond to the same species. The correct interpretation of this note is that fewer plausible matching species can be ruled out as signal strength decreases. For the one correct match, 6 out of the 7 suggested measures are not impacted by the signal-to-noise ratio.

Clearly Mismatched Spectra. Aviation fuel and sustainable aviation fuel usually contains a significant fraction of molecules with 12 or more carbon atoms for which hundreds of isomers are possible. A complete library of reference spectra of isomers of such hydrocarbons does not yet exist. As such, it is not yet possible to determine with absolute certainty that any given sample matches one of the reference spectra that happens to exist in the library. In some cases, the overlay of the sample spectra with one or more of the available reference spectra will be visually indistinguishable. The comparison between an unknown sample separated from Virent SAK by GC×GC is one such example and is highlighted in Figure 4. Figure 4 follows the same template as Figure 3, contrasting the difference in these criterion and subplots for a non-match. In this case the elution time of the analyte suggested that the analyte was an aromatic with 10 carbons. The highest R^2 to our library was to 1,3-diethylbenzene, 0.976, which is consistent with those of known matches, like the toluene example from Figure 3. While it is already clear from a plot of the residuals in Figures 4b and 4c that the R^2 of the overlay is misleading, each of the other measures of match suggested in this work also prove, even in isolation, that this sample is not 1,3-diethylbenzene. The appropriate conclusion to draw from these results is that the sample is not a species included in our library. Such eliminations help to reduce the range of fuel properties that would otherwise be ascribed to this peak on the chromatogram for property estimations by the tier α methodology.[24]
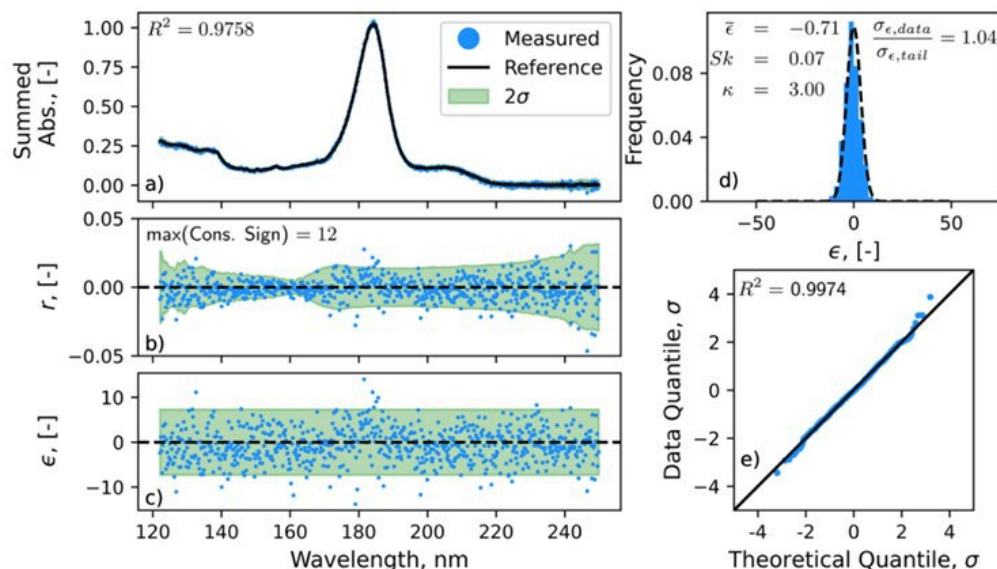
Figure 3. Example results of a definitive positive identification of toluene in petroleum derived A-2. (a) Shows the measured spectrum of a peak in A-2 and the reference spectrum for Toluene. (b) Residuals or differences between the measured and reference spectra and (c) the results after dividing through by wavelength specific σ. In both (b) and (c), the green region represents the expected 95th percentile range based on the background noise and tail data statistics. (d) The histogram of normalized residuals, ϵ, approximate a normal distribution, also evidenced in (e) a Q-Q plot of the normalized residuals showing strong linearity.
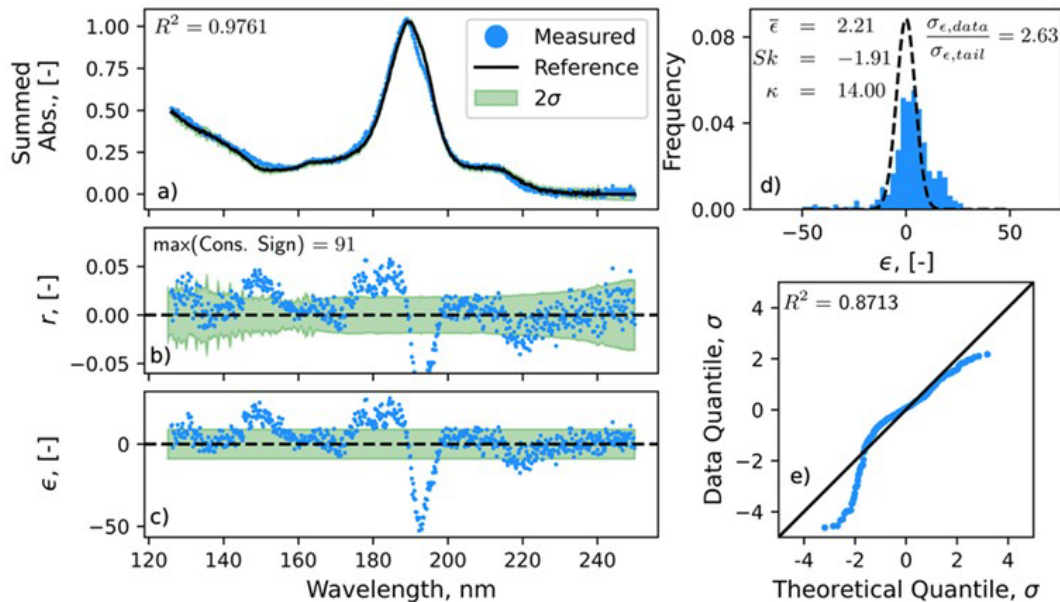
Figure 4. Example of a definitive negative identification of 1,3-diethylbenzene in Virent SAK. (a) Shows the measured spectrum of an analyte in SAK and the reference spectrum for 1,3-diethylbenzene. (b) Residuals or differences between the measured and reference spectra and (c) the results after dividing through by wavelength specific σ. In both (b) and (c), the green region represents the expected 95th percentile range based on the background noise and tail data statistics. (d) The histogram of normalized residuals does not approximate a normal distribution, also evidenced in (e) a Q-Q plot of the normalized residuals showing poor linearity.

Spectra Comparisons with High R^2. This paragraph provides a summary of comparisons between spectra of that are known to exhibit among the smallest differences in absorption known to these authors: n-undecane, n-dodecane, and n-tridecane.[20] Generally, for n-alkanes it is not unusual for the highest R^2 of a set of sample/reference spectra overlays to occur, misleadingly, between the sample, having *n* carbon atoms, and a reference, having *m≠n* carbon atoms. In this work, the three sample spectra were measured from analytes separated from A-2 by GC×GC and the three reference spectra were extracted from our library. An overlay plot of the three reference spectra is provided in Figure 5, while Figure 6 provides a summary of nine pairwise comparisons between three samples and three reference spectra. The lowest R^2 out of the nine comparisons is 0.9997 and visually, in Figure 5a, it is difficult to see any offsets between the lines. Figure 5b is included to highlight the area with the greatest divergence between spectra. The reference spectra were used in Figure 5 to stand-in for the sample spectra so the point could be made with one plot instead of nine. The words across the top of Figure 6 represent sample spectra and the symbols represent reference spectra, where the colors are consistent between sample and reference. For undecane and tridecane, the correct match shows a mean, standard deviation ratio, skew, kurtosis, Q-Q plot R^2, and maximum count of consecutive signs, each closest to the corresponding theoretical value for a Gaussian distribution. For dodecane the mean and the maximum count of consecutive signs are each closest to their respective theoretical values, while the other five measures do not clearly distinguish the dodecane sample from the tridecane reference. These statistics should be employed in aggregate. Each statistic should demonstrate results plausibly consistent with noise alone. The numerical result from R^2 of the spectra overlay comparison, alone would not be able to correctly determine the best match, in part because it is driven by the noise in the reference and sample spectra. In contrast, the moments of ϵ and the linearity of the Q-Q plot are not impacted the measurement noise.
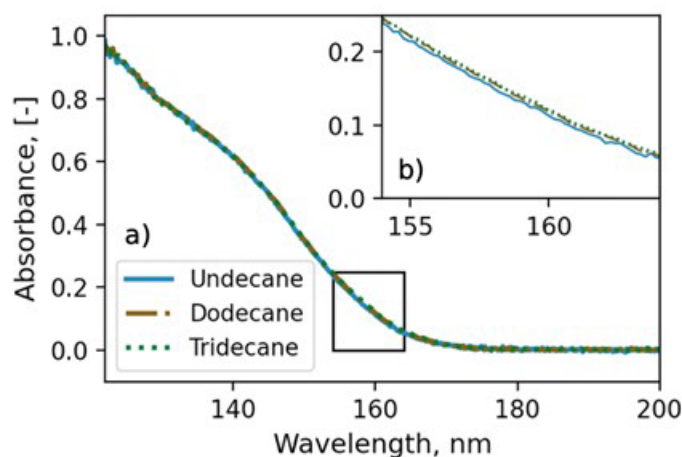
Figure 5. (a) Spectra overlay of three normal alkanes for typical wavelengths of interest and (b) magnification of the area of greatest difference in absorbance for n-undecane, n-dodecane, and n-tridecane
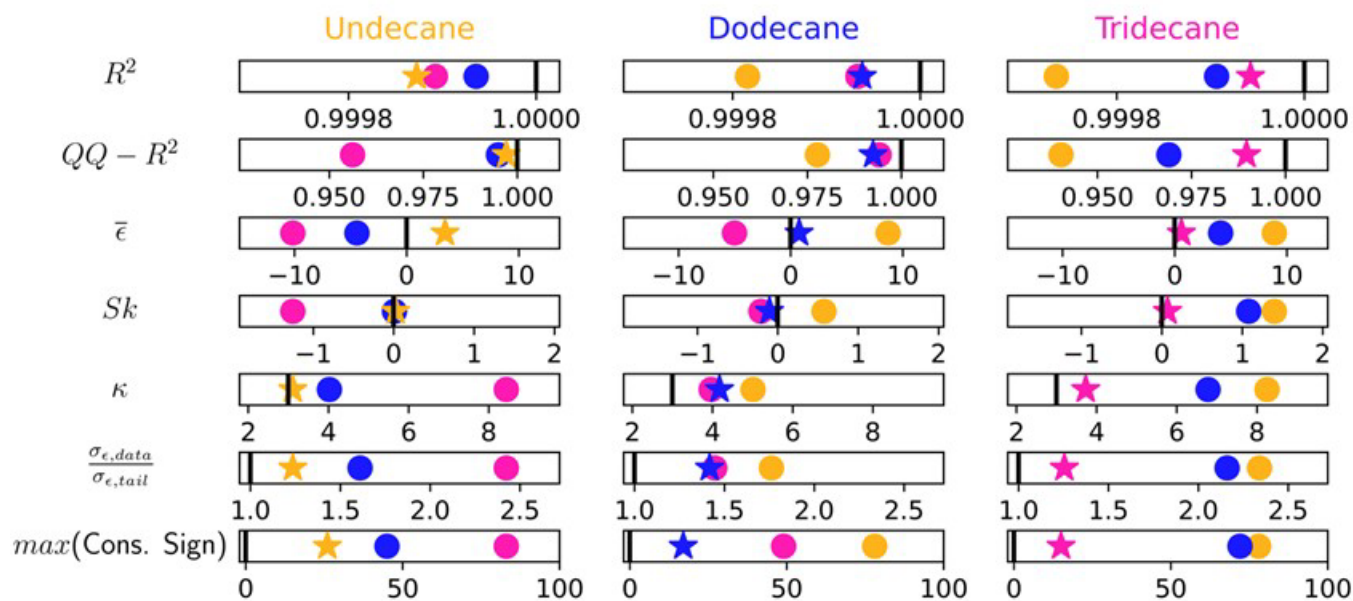
Figure 6. Scorecard of numerical analyses of residuals between measured and reference spectra of three normal alkanes. Stars rep-resent the correct match and circles represent incorrect matches. Undecane is incorrectly identified using R2 alone, but with other statistics considered the correct match can be observed. In all three examples shown, for all of the statistics, the correct match per-forms strongly, even if not always the best. The statistics should be interpreted collectively.

4. Conclusions

Six measures of the distribution of the residual between numerically adjusted spectra have been shown to be invaluable for comparing plausible matches of reference spectra to sample spectra of analytes eluted from a two-dimensional gas chromatograph. These measures include its first four moments (mean, variance, skew and kurtosis), the R^2 of a Q-Q plot with the distribution of the residual on one axis and an ideal Gaussian on the other, and the maximum count of consecutive (by wavelength) signs within the residual.

These measures, taken in combination, facilitate a logical ranking of all plausible matches, rendering positive identification of a single-component analyte possible provided a reference spectrum exists for all plausible matches. Unknown analytes, for which reference spectra are not available for all plausible matches (isomers), can be partially identified by definitive elimination of many of the otherwise plausible matches. Such eliminations help to reduce the uncertainty in fuel component property calculations. In contrast, the R^2 of the spectra overlay, by itself, is often ambiguous as many reference spectra will have an R^2 close to the maximum of the set.

Several numerical processing steps are described that remove differences in the background signal and noise characteristics at different wavelengths. Without these operations the residuals would not exhibit a distribution of values similar in shape to the noise at any wavelength. For our experiment the background noise has been shown to have a kurtosis value of three (Gaussian) at all wavelengths, and the numerically adjusted residuals of known matches have been shown to exhibit the same characteristics.

One exemplar application of the methodology is a definitive match of n-undecane, n-dodecane and n-tridecane sample spectra to their corresponding reference spectrum. While this example is impressive, the greatest power of the methodology generally is in its ability to unambiguously identify mismatches because the distributions of residuals between mismatching spectra are very clearly not Gaussian, and have a high consecutive sign count, even in cases where the R^2 between the compared spectra is ambiguous.

Supporting Information

In the supporting information there is a schematic of the GC×GC-VUV experimental setup, a figure of the VUV lamp counts vs wavelength, and there are the chromatograms of A-2 and Virent SAK. In addition, a set of comparisons between sample and reference spectra of three isomers of C_8H_{18} is provided along with a brief description.

Acknowledgements

References

[1]   Moses, K. R.; Higgins, P.; Mccabe, M.; Prabhakar, S.; Swann, S. Chapter: Automated Fingerprint Identification System.

[2]   Uenuma, F. The First Criminal Trial That Used Fingerprints as Evidence | History| Smithsonian Magazine. *Smithsonian Magazine.* 2018.

[3]   Pandey, F.; Dash, P.; Samanta, D.; Sarma, M. ASRA: Automatic Singular Value Decomposition-Based Robust Fingerprint Image Alignment. *Multimed. Tools Appl. 2021 8010* **2021**, *80* (10), 15647–15675. https://doi.org/10.1007/S11042-021-10560-5.

[4]   MacCarthy, B. L.; Wasusri, T. A Review of Non-Standard Applications of Statistical Process Control (SPC) Charts. *Int. J. Qual. Reliab. Manag.* **2002**, *19* (3), 295–320. https://doi.org/10.1108/02656710210415695/FULL/PDF.

[5]   Lim, S. A. H.; Antony, J.; Albliwi, S. Statistical Process Control (SPC) in the Food Industry – A Systematic Review and Future Research Agenda. *Trends Food Sci. Technol.* **2014**, *37* (2), 137–151. https://doi.org/10.1016/J.TIFS.2014.03.010.

[6]   Streitwieser, A. J.; Heathcock, C. H. *Introduction to Organic Chemistry*, Second.; Macmillan Publishing Co., Inc.: New York, 1981.

[7]   Atkins, P. W. *Physical Chemistry*, Third.; Freeman and Company: New York, 1986.

[8]   Zachariah Samuel, A.; Mukojima, R.; Horii, S.; Ando, M.; Egashira, S.; Nakashima, T.; Iwatsuki, M.; Takeyama, H. On Selecting a Suitable Spectral Matching Method for Automated Analytical Applications of Raman Spectroscopy. *ACS Omega* **2021**, *6*, 26. https://doi.org/10.1021/acsomega.0c05041.

[9]   Li, J.; Hibbert, D. B.; Fuller, S.; Vaughn, G. A Comparative Study of Point-to-Point Algorithms for Matching Spectra. *Chemom. Intell. Lab. Syst.* **2006**, *82* (1–2), 50–58. https://doi.org/10.1016/J.CHEMOLAB.2005.05.015.

[10]   Brereton, R. G.; Jansen, J.; Lopes, J.; Marini, F.; Pomerantsev, A.; Rodionova, O.; Roger, J. M.; Walczak, B.; Tauler, R. Chemometrics in Analytical Chemistry—Part I: History, Experimental Design and Data Analysis Tools. *Anal. Bioanal. Chem. 2017 40925* **2017**, *409* (25), 5891–5899. https://doi.org/10.1007/S00216-017-0517-1.

[11]   Keithley, R. B.; Mark Wightman, R.; Heien, M. L. Multivariate Concentration Determination Using Principal Component Regression with Residual Analysis. *TrAC Trends Anal. Chem.* **2009**, *28* (9), 1127–1136. https://doi.org/10.1016/J.TRAC.2009.07.002.

[12]   A. Schug, K.; Sawicki, I.; D. Carlton, D.; Fan, H.; M. McNair, H.; P. Nimmo, J.; Kroll, P.; Smuts, J.; Walsh, P.; Harrison, D. Vacuum Ultraviolet Detector for Gas Chromatography. *Anal. Chem.* **2014**, *86* (16), 8329–8335. https://doi.org/10.1021/ac5018343.

[13]   Gröger, T.; Gruber, B.; Harrison, D.; Saraji-Bozorgzad, M.; Mthembu, M.; Sutherland, A.; Zimmermann, R. A Vacuum Ultraviolet Absorption Array Spectrometer as a Selective Detector for Comprehensive Two-Dimensional Gas Chromatography: Concept and First Results. *Anal. Chem.* **2016**, *88* (6), 3031–3039. https://doi.org/10.1021/acs.analchem.5b02472.

[14]   Wang, F. C. Y. Comprehensive Two-Dimensional Gas Chromatography Hyphenated with a Vacuum Ultraviolet Spectrometer to Analyze Diesel-A Three-Dimensional Separation (GC × GC × VUV) Approach. *Energy and Fuels* **2020**, *34* (7), 8012–8017. https://doi.org/10.1021/acs.energyfuels.0c00688.

[15]   Lelevic, A.; Geantet, C.; Moreaud, M.; Lorentz, C.; Souchon, V. Quantitative Analysis of Hydrocarbons in Gas Oils by Two-Dimensional Comprehensive Gas Chromatography with Vacuum Ultraviolet Detection. *Energy and Fuels* **2021**, *35* (17), 13766–13775. https://doi.org/10.1021/acs.energyfuels.1c01910.

[16]   Roberson, Z. R.; Goodpaster, J. V. Differentiation of Structurally Similar Phenethylamines via Gas Chromatography–Vacuum Ultraviolet Spectroscopy (GC–VUV). *Forensic Chem.* **2019**, *15*, 100172. https://doi.org/10.1016/J.FORC.2019.100172.

[17]   Roberson, Z. R.; Gordon, H. C.; Goodpaster, J. V. Instrumental and Chemometric Analysis of Opiates via Gas Chromatography–Vacuum Ultraviolet Spectrophotometry (GC-VUV). *Anal. Bioanal. Chem. 2020 4125* **2020**, *412* (5), 1123–1128. https://doi.org/10.1007/S00216-019-02337-5.

[18]   Leghissa, A.; Smuts, J.; Qiu, C.; Hildenbrand, Z.; Schug, K. Detection of Cannabinoids and Cannabinoid Metabolites Using Gas Chromatography with Vacuum Ultraviolet Spectroscopy. *Sep Sci plus* **2018**, *1*, 37– 42. https://doi.org/doi.org/10.1002/sscp.201700005.

[19]   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. https://doi.org/10.1016/J.FUEL.2021.122709.

[20]   Anthony, I. G. M.; Brantley, M. R.; Gaw, C. A.; Floyd, A. R.; Solouki, T. Vacuum Ultraviolet Spectroscopy and Mass Spectrometry: A Tandem Detection Approach for Improved Identification of Gas Chromatography-Eluting Compounds. *Anal. Chem.* **2018**, *90* (7), 4878–4885. https://doi.org/10.1021/ACS.ANALCHEM.8B00531/SUPPL_FILE/AC8B00531_SI_001.PDF.

[21]   G.M. Anthony, I.; R. Brantley, M.; R. Floyd, A.; A. Gaw, C.; Solouki, T. Improving Accuracy and Confidence of Chemical Identification by Gas Chromatography/Vacuum Ultraviolet Spectroscopy-Mass Spectrometry: Parallel Gas Chromatography, Vacuum Ultraviolet, and Mass Spectrometry Library Searches. *Anal. Chem.* **2018**, *90* (20), 12307–12313. https://doi.org/10.1021/acs.analchem.8b04028.

[22]   Feldhausen, J.; Bell, D. C.; Yang, Z.; Faulhaber, C.; Boehm, R.; Heyne, J. Synthetic Aromatic Kerosene Property Prediction Improvements with Isomer Specific Characterization via GCxGC and Vacuum Ultraviolet Spectroscopy. *Fuel* **2022**, *326*, 125002. https://doi.org/10.1016/J.FUEL.2022.125002.

[23]   Boehm, R. C.; Yang, Z.; Bell, D. C.; Feldhausen, J.; Heyne, J. S. Lower Heating Value of Jet Fuel from Hydrocarbon Class Concentration Data and Thermo-Chemical Reference Data: An Uncertainty Quantification. *Fuel* **2022**, *311*. https://doi.org/10.1016/j.fuel.2021.122542.

[24]   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* (September 2020), 120345. https://doi.org/10.1016/j.fuel.2021.120345.

[25]   Edwards, T. Reference Jet Fuels for Combustion Testing. In *AIAA SciTech Forum - 55th AIAA Aerospace Sciences Meeting*; American Institute of Aeronautics and Astronautics Inc.: Grapevine, TX, 2017; pp 1–58. https://doi.org/10.2514/6.2017-0146.

[26]   Colket, M.; Heyne, J. *Fuel Effects on Operability of Aircraft Gas Turbine Combustors*, August.; AIAA, Progress in Astronautics and Aeronautics, 2021. https://doi.org/10.2514/4.106040.