

## Supplemental Information

### **Prediction of Sustainable Aviation Fuel Properties for Liquid Hydrocarbons from Hydrotreating Biomass Catalytic Fast Pyrolysis derived Organic Intermediates**

Suphat Watanasiri <sup>a</sup>, Eugene Paulechka <sup>a,b</sup>, Kristiina Iisa <sup>a</sup>, Earl Christensen <sup>a</sup>, Chris Muzny <sup>b</sup>, Abhijit Dutta <sup>a</sup>

<sup>a</sup> Catalytic Carbon Transformation & Scale-Up Center, National Renewable Energy Laboratory, Golden, CO, USA

<sup>b</sup> Thermodynamics Research Center, Applied Chemicals and Materials Division, National Institute of Standards and Technology, Boulder, CO, USA

**Table S1. Correlation coefficients for calculating test method D86 correlated data from test method D2887 data.**

Coefficients for the conversion of D2887 data to D86 temperatures using equation (12) in the main article were not reproduced here to comply with ASTM copyright requirements. The parameters can be obtained from the ASTM publication (reference 32 in the main article):

*Standard test method for boiling range distribution of petroleum fractions by gas chromatography. ASTM D2887 – 19a. ASTM International, Pennsylvania. Table X4.1.*

**Table S2. GCxGC-TOFMS-FID measured mass percentages of compounds found in the jet range fuel from hydrotreated ZSM-5 CFP oil.**

Compound name <sup>a</sup>	CAS #	Formula	Mass %
Cyclohexane, methyl-	108-87-2	C7H14	0.087
Cyclohexane, 1,3-dimethyl-, cis-	638-04-0	C8H16	0.161
Cyclohexane, 1,2,3-trimethyl-, (1.alpha.,2.beta.,3.alpha.)-	1678-81-5	C9H18	0.941
Heptane, 2,4-dimethyl-	2213-23-2	C9H20	0.223
Cyclohexane, ethyl-	1678-91-7	C8H16	1.576
Octane, 2-methyl-	3221-61-2	C9H20	0.366
Cyclohexane, 1,2,4-trimethyl-, (1.alpha.,2.beta.,4.beta.)-	7667-60-9	C9H18	0.232
Octane, 2,3-dimethyl-	7146-60-3	C10H22	0.224
Cyclohexane, 1,2,4-trimethyl-	2234-75-5	C9H18	0.370
Nonane	111-84-2	C9H20	0.447
Cyclohexane, 1,2,3-trimethyl-	1678-97-3	C9H18	2.868
Cyclohexane, 1-ethyl-4-methyl-, cis-	4926-78-7	C9H18	3.797
Cyclohexane, (1-methylethyl)-	696-29-7	C9H18	0.785
Octane, 4-ethyl-	15869-86-0	C10H22	0.634
Cyclohexane, propyl-	1678-92-8	C9H18	9.829
Octane, 3,5-dimethyl-	15869-93-9	C10H22	1.184
Nonane, 3-methyl-	005911-04-6	C10H22	0.110
Cyclohexane, 1-ethyl-1,3-dimethyl-, trans-	62238-29-3	C10H20	0.507
Bicyclo[2.2.1]heptane, 2-ethyl-	2146-41-0	C9H16	2.537
Cyclopentane, 1-methyl-3-(2-methylpropyl)-	29053-04-1	C10H20	8.178
Decane	124-18-5	C10H22	0.813
Cyclohexane, 1-methyl-2-propyl-	4291-79-6	C10H20	8.322
1H-Indene, octahydro-, cis-	4551-51-3	C9H16	6.611
Cyclohexane, 1-ethyl-1-methyl-	4926-90-3	C9H18	3.162
Pentalene, octahydro-2-methyl-	3868-64-2	C9H16	3.017
Cyclohexane, 1,2-diethyl-, cis-	824-43-1	C10H20	0.451
Cyclopentane, 2-isopropyl-1,3-dimethyl-	32281-85-9	C10H20	0.952
Cyclohexane, butyl-	1678-93-9	C10H20	2.151
Cyclohexane, 2-ethyl-1,3-dimethyl-	7045-67-2	C10H20	0.161
1,1'-Bicyclopentyl	1636-39-1	C10H18	1.611

Bicyclo[5.3.0]decane (cis)	16189-46-1	C10H18	0.351
Cyclohexane, 1-ethyl-2-propyl-	62238-33-9	C11H22	4.822
Naphthalene, decahydro-, trans-	493-02-7	C10H18	0.807
(2-Methylbutyl)cyclohexane	54105-77-0	C11H22	0.150
cis-Decalin, 2-syn-methyl-	14398-72-2	C11H20	3.649
Naphthalene, decahydro-2-methyl-	2958-76-1	C11H20	5.801
Cyclohexane, pentyl-	4292-92-6	C11H22	1.464
Spiro[4.5]decane	176-63-6	C10H18	0.688
1-Methyldecahydronaphthalene	2958-75-0	C11H20	3.258
Naphthalene, decahydro-2,6-dimethyl-	1618-22-0	C12H22	3.167
Cyclohexane, 1-methyl-4-(1-methylbutyl)-	54411-00-6	C12H24	1.242
Dodecane	112-40-3	C12H26	0.583
Cyclopentylcyclohexane	1606-08-2	C11H20	0.722
Cyclohexane, hexyl-	4292-75-5	C12H24	0.847
Naphthalene, 2-ethyldecahydro-	1618-23-1	C12H22	0.506
Spiro[5.6]dodecane	181-15-7	C12H22	0.173
1,1'-Bicyclohexyl	92-51-3	C12H22	0.284
Bicyclo[3.1.1]heptane, 2,6,6-trimethyl-	473-55-2	C10H18	1.411
Perhydrophenalene, (3a.alpha., 6a.alpha., 9a.alpha., 9b.beta.)-	40250-64-4	C13H22	2.778
Cyclohexane, (cyclopentylmethyl)-	4431-89-4	C12H22	0.142
Benzene, 1,3-dimethyl-	108-38-3	C8H10	0.681
Benzene, propyl-	103-65-1	C9H12	0.174
Benzene, 1-ethyl-4-methyl-	622-96-8	C9H12	0.759
Benzene, 1,2,3-trimethyl-	526-73-8	C9H12	0.191
Benzene, 1-methyl-3-propyl-	1074-43-7	C10H14	0.219
Indane	496-11-7	C9H10	0.413
1H-Indene, 2,3-dihydro-2-methyl-	824-63-5	C10H12	0.056
Indan, 1-methyl-	767-58-8	C10H12	0.160
1H-Indene, 2,3-dihydro-4-methyl-	824-22-6	C10H12	0.927
Naphthalene, 1,2,3,4-tetrahydro-	119-64-2	C10H12	0.704
1H-Indene, 2,3-dihydro-1,3-dimethyl-	4175-53-5	C11H14	0.092
1H-Indene, 2,3-dihydro-4,7-dimethyl-	6682-71-9	C11H14	0.472

<sup>a</sup> Components were identified based on mass spectral matching using LECO ChromaTOF<sup>®</sup> software [1] and the NIST 2014 Mass Spectral Library [2], characteristic spectral features, and retention times.

**Table S3. GC×GC-TOFMS-FID measured mass percentages of compounds found in the jet range fuel from hydrotreated Pt/TiO<sub>2</sub> CFP oil.**

Compound name <sup>a</sup>	CAS #	Formula	Mass %
Cyclohexane, 1,2-dimethyl-, cis-	2207-01-4	C8H16	0.195
Cyclohexane, ethyl-	1678-91-7	C8H16	3.504

Ethylbenzene	100-41-4	C8H10	0.100
Benzene, 1,3-dimethyl-	108-38-3	C8H10	0.147
o-Xylene	95-47-6	C8H10	0.097
Indane	496-11-7	C9H10	0.501
Naphthalene, 1,2,3,4-tetrahydro-	119-64-2	C10H12	0.149
Octane, 2-methyl-	3221-61-2	C9H20	1.123
1H-Indene, 2,3-dihydro-4-methyl-	824-22-6	C10H12	0.464
Benzene, 1-ethyl-2-methyl-	611-14-3	C9H12	0.084
Benzene, 1-ethyl-4-methyl-	622-96-8	C9H12	0.478
Indan, 1-methyl-	767-58-8	C10H12	0.161
Benzene, propyl-	103-65-1	C9H12	0.573
Benzene, 1,2,3-trimethyl-	526-73-8	C9H12	0.206
Naphthalene, 1,2,3,4-tetrahydro-5-methyl-	2809-64-5	C11H14	0.125
Pentalene, octahydro-	694-72-4	C8H14	0.044
1H-Indene, 2,3-dihydro-1,3-dimethyl-	4175-53-5	C11H14	0.171
Benzene, 1-methyl-3-(1-methylethyl)-	535-77-3	C10H14	0.089
Benzene, (1-methylpropyl)-	135-98-8	C10H14	0.096
Benzene, 1-methyl-3-propyl-	1074-43-7	C10H14	0.275
Cyclohexane, 1,2,4-trimethyl-	2234-75-5	C9H18	0.403
Benzene, butyl-	104-51-8	C10H14	0.113
Cyclohexane, 1,2,3-trimethyl-	1678-97-3	C9H18	2.097
Cyclohexane, 1-ethyl-2-methyl-	3728-54-9	C9H18	3.203
Cyclohexane, 1-ethyl-4-methyl-, cis-	4926-78-7	C9H18	2.998
Nonane	111-84-2	C9H20	0.655
1H-Indene, octahydro-, cis-	4551-51-3	C9H16	8.591
Bicyclo[2.2.1]heptane, 2-ethyl-	2146-41-0	C9H16	1.244
Cyclohexane, (1-methylethyl)-	696-29-7	C9H18	0.576
Spiro[4.5]decane	176-63-6	C10H18	0.327
Cyclohexane, propyl-	1678-92-8	C9H18	21.290
Naphthalene, decahydro-, trans-	493-02-7	C10H18	2.728
9-Methylbicyclo[3.3.1]nonane	25107-01-1	C10H18	1.391
Bicyclo[5.3.0]decane (cis)	16189-46-1	C10H18	2.428
Perhydrophenalene, (3a.alpha., 6a.alpha., 9a.alpha., 9b.beta.)-	40250-64-4	C13H22	0.215
Cyclopentylcyclohexane	1606-08-2	C11H20	0.242
Pentalene, octahydro-2-methyl-	3868-64-2	C9H16	0.189
Naphthalene, decahydro-2-methyl-	2958-76-1	C11H20	2.759
cis-Decalin, 2-syn-methyl-	14398-72-2	C11H20	5.688
1,1'-Bicyclohexyl	92-51-3	C12H22	0.201
1-Methyldecahydronaphthalene	2958-75-0	C11H20	1.319
Cyclohexane, (cyclopentylmethyl)-	4431-89-4	C12H22	0.338
Naphthalene, 2-ethyldecahydro-	1618-23-1	C12H22	1.442
Cyclohexane, 1,2-diethyl-, cis-	824-43-1	C10H20	2.090

1-Methyl-4-(1-methylethyl)-cyclohexane	99-82-1	C10H20	1.230
Cyclohexane, 1-ethyl-1-methyl-	4926-90-3	C9H18	3.816
Cyclohexane, 1,1'-(1,2-ethanediyl)bis-	3321-50-4	C14H26	0.098
Cyclohexane, 1-ethyl-2-methyl-, cis-	4923-77-7	C9H18	1.032
Cyclohexane, butyl-	1678-93-9	C10H20	3.431
Cyclohexane, 1-methyl-2-propyl-	4291-79-6	C10H20	5.324
Bicyclo[3.1.1]heptane, 2,6,6-trimethyl-	473-55-2	C10H18	0.379
Cyclohexane, 1-ethyl-2-propyl-	62238-33-9	C11H22	2.473
Cyclohexane, hexyl-	4292-75-5	C12H24	2.228
(2-Methylbutyl)cyclohexane	54105-77-0	C11H22	1.405
Cyclopentane, 1,2-dipropyl-	91242-57-8	C11H22	0.502
Cyclohexane, (2-methylpropyl)-	1678-98-4	C10H20	0.301
Nonane, 3-methyl-	005911-04-6	C10H22	0.276
Nonane, 2-methyl-	871-83-0	C10H22	0.732
Decane	124-18-5	C10H22	1.699
Cyclohexane, 2-ethyl-1,3-dimethyl-	7045-67-2	C10H20	1.318
Cyclopentane, 1-methyl-3-(2-methylpropyl)-	29053-04-1	C10H20	1.532
Cyclopentane, 2-isopropyl-1,3-dimethyl-	32281-85-9	C10H20	0.895
Tridecane	629-50-5	C13H28	0.219

<sup>a</sup> Components were identified based on mass spectral matching using LECO ChromaTOF<sup>®</sup> software [1] and the NIST 2014 Mass Spectral Library [2], characteristic spectral features, and retention times.

Table S4 can be found in the accompanying Excel file.

**Table S5. Composition matrix of jet range fuel from hydrotreated ZSM-5 CFP oil. Values (in mass percent) were used for the Shi et al. method [4]**

Class <sup>a</sup>	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19
1	0	0	0.447	0.814	0	0.583	0	0	0	0	0	0	0
2	0	0	0.366	0.744	0	0	0	0	0	0	0	0	0
3	0	0	0.223	1.408	0	0	0	0	0	0	0	0	0
4	0.0874	1.576	10.614	2.151	1.464	0.847	0	0	0	0	0	0	0
5	0	0.161	11.371	18.571	4.972	1.242	0	0	0	0	0	0	0
6	0	0	12.165	4.869	13.43	4.272	2.778	0	0	0	0	0	0
7	0	0	0.174	0	0	0	0	0	0	0	0	0	0
8	0	0.681	0.949	0.219	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0.413	1.847	0.564	0	0	0	0	0	0	0	0

<sup>a</sup> Classes are defined in the description of the Shi et al. methods below.

**Table S6. Composition matrix of jet range fuel from hydrotreated Pt/TiO<sub>2</sub> CFP oil. Values (in mass percent) were used for the Shi et al. method [4]**

Class <sup>a</sup>	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19
1	0	0	0.655	1.699	0	0	0.219	0	0	0	0	0	0
2	0	0	1.123	1.008	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	3.504	21.866	3.733	0	2.228	0	0	0	0	0	0	0
5	0	0.195	13.549	12.39	4.379	0	0	0	0	0	0	0	0
6	0	0.0445	10.023	7.254	10.009	1.981	0.215	0.0979	0	0	0	0	0
7	0	0.1	0.573	0.113	0	0	0	0	0	0	0	0	0
8	0	0.244	0.767	0.461	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0.501	0.774	0.296	0	0	0	0	0	0	0	0

<sup>a</sup>Classes are defined in the description of the Shi et al. methods below.

**Table S7. Composition matrix of jet range fuel from hydrotreated ZSM-5 CFP oil. Values (in mass percent) were used for the Vozka et al. method [6]**

Class <sup>a</sup>	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
1	0	0	0.447	0.814	0	0.583	0	0	0	0	0	0	0	0
2	0	0	0.589	2.152	0	0	0	0	0	0	0	0	0	0
3	0.0874	1.737	21.985	20.722	6.435	2.088	0	0	0	0	0	0	0	0
4	0	0	12.165	4.869	13.430	4.272	2.778	0	0	0	0	0	0	0
5	0	0.681	1.124	0.219	0	0	0	0	0	0	0	0	0	0
6	0	0	0.413	1.847	0.564	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0

<sup>a</sup>Classes are defined in the description of the Vozka et al. methods below.

**Table S8. Composition matrix of jet range fuel from hydrotreated Pt/TiO<sub>2</sub> CFP oil. Values (in mass percent) were used for the Vozka et al. method [6]**

Class <sup>a</sup>	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
1	0	0	0.655	1.699	0	0	0.219	0	0	0	0	0	0	0
2	0	0	1.123	1.008	0	0	0	0	0	0	0	0	0	0
3	0	3.699	35.415	16.123	4.379	2.228	0	0	0	0	0	0	0	0
4	0	0.0445	10.023	7.254	10.009	1.981	0.215	0.0979	0	0	0	0	0	0
5	0	0.344	1.340	0.574	0	0	0	0	0	0	0	0	0	0
6	0	0	0.501	0.774	0.296	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0

<sup>a</sup>Classes are defined in the description of the Vozka et al. methods below.

## Description of GC x GC analysis method

Detailed characterization of the SAF fractions was conducted by comprehensive two-dimensional gas chromatography with simultaneous time-of-flight mass spectrometry and flame ionization detection (GCxGC TOFMS-FID). Analysis was conducted using a LECO Pegasus 4D system (LECO Corp., St. Joseph, MI) [1] equipped with liquid nitrogen cooled thermal modulator and a post column flow splitter. Method parameters are shown in Table S9. Samples were injected without dilution for GC analysis. Components were tentatively identified based on mass spectral matching using LECO ChromaTOF® software [1] and the NIST 2014 library of spectra [2]. Library matches of 80% or higher were considered strong matches while matches below 70% were considered suspect. For the ZSM-5 CFP, HT fuel and the Pt/TiO<sub>2</sub> CFP, HT fuel approximately 15 wt% and 5 wt% of the matches respectively had similarities of <80%, but all matches had similarities >75% for both samples. Compounds which could not be positively identified as an isomer were assigned a compound class based on characteristic spectral features and second dimension retention time comparison with a given class. In many cases, genuine standards are not available for verification of component IDs in the carbon number range of hydrocarbons analyzed; therefore, the analysis utilized GCxGC retention with mass spectral features for general component class assignments. A standard containing *n*-paraffins from C<sub>5</sub> to C<sub>30</sub> was used to determine retention time ranges for carbon number assignments of components binned by class. Standard components representing aromatics and cycloparaffins were used to establish second dimension retention times to aid in classifications.

For compounds identified by retention time and mass spectral features (not relying on library matching or similarity scores) the carbon number was determined from first dimension retention time and formula mass inferred from closest eluting known compound of that class. These compounds account for 6.5% in the case of the ZSM-5 derived sample and 2.4% of the Pt/TiO<sub>2</sub> sample. No fully unidentified compounds were encountered for these distillates; however, in this analysis if peaks are found for which a compound class cannot be determined these are assigned a response factor of *n*-paraffins (relative response of 1) and binned as unknowns, as is done with detailed hydrocarbon analysis (ASTM D6729/D6730). This approach is known to be in error as far as compound response factors and is used to provide a rough estimate of concentration for fully unidentified components. In the case of this analysis no peaks were detectable that could not be assigned to a hydrocarbon class.

Compounds were quantified from their theoretical response factors calculated via effective carbon numbers [3].

$$\text{Relative response factor} = \frac{M_i}{M_{\text{nonane}}} \times \frac{n_{\text{C,nonane}}}{n_{\text{C},i}}$$

Where  $M_i$  is the molar mass of compound  $i$ ,  $M_{\text{nonane}} = 128.26 \text{ g}\cdot\text{mol}^{-1}$  is the molar mass of nonane,  $n_{\text{C,nonane}} = 9$  is the effective carbon number for nonane, and  $n_{\text{C},i}$  is the effective carbon number for compound  $i$ .

The total mass detected was set to 100%.

**Table S9: GCxGC Method Parameters**

<b>Column</b>				
Primary	Rxi-17Sil, 20 m x 180 $\mu\text{m}$ x 0.18 $\mu\text{m}$			
Secondary	ZB-5HT, 1.0 m x 180 $\mu\text{m}$ x 0.18 $\mu\text{m}$			
<b>Injector</b>				
0.1 $\mu\text{L}$ injection, split 100:1	300°C			
<b>Oven</b>				
Primary	35°C, hold 5 min, ramp 3°C·min <sup>-1</sup> to 125°C, ramp 10°C·min <sup>-1</sup> to 340 °C hold 1 min			
Secondary	40°C offset from primary			
Modulator	15°C offset from secondary			
Modulator cycle timing		modulator period, s	hot time, s	cold time, s
0 to 800 s		8	1	3
800 s to end of run		8	2	2
<b>Mass Spectrometer</b>				
Transfer line	350°C			
TOF mass range (Da·e <sup>-1</sup> )	m/z 29-600			
TOF acquisition rate	200 spectra per second			
Delay	70 s			
<b>FID</b>				
Detector Temperature	350°C			
H <sub>2</sub> Flow	40 cm <sup>3</sup> ·min <sup>-1</sup>			
Air Flow	300 cm <sup>3</sup> ·min <sup>-1</sup>			
N <sub>2</sub> Makeup Flow	25 cm <sup>3</sup> ·min <sup>-1</sup>			

## Short description of Shi et al. methods

Shi et al. [4] developed four statistical algorithms to estimate properties of aviation fuels: weighted average (WA) method, partial least squares analysis (PLS), genetic algorithm (GA), and modified weighted average (MWA) method. The correlations were based on the total quantity (wt.%) of ten hydrocarbon classes for each carbon number. The detailed hydrocarbon composition was obtained from GC x GC – MS/FID measurements of aviation fuels. Their composition matrix was composed of 10 hydrocarbon classes and 13 carbon numbers, ranging from C7 to C19. The classes considered were: (1) *n*-paraffins, (2) mono-branched isoparaffins, (3) highly-branched isoparaffins, (4) mono-branched alkylmonocycloparaffins, (5) highly-branched alkylmonocycloparaffins, (6) di- and tri-cycloparaffins, (7) mono-branched alkylbenzenes, (8) highly-branched alkylbenzenes, (9) indene, alkylindene, naphthalenes, alkyl naphthalenes, and polycyclic aromatics, and (10) indane, alkylindane, tetralins, and alkyltetralins. The properties of interest were density at 20 °C, freezing point, flash point, and net heat of combustion.

The authors correlated the four properties with detailed composition of 17 aviation fuels. The fuel samples consisted of four commercial RP-3 aviation fuels, one Fischer-Tropsch synthetic fuel, four kerosene distillation cut samples at different temperatures, five aviation hydrocarbon fuels with higher



content in aromatics, and three laboratory synthetic aviation hydrocarbon fuels which were synthesized to ensure higher dicycloalkane content in samples [4].

The four algorithms are summarized below.

### Weighted average (WA) method

$$P = \sum_{i=1}^{10} \sum_{j=1}^{13} (x_{ij} u_{ij})$$

where  $P$  is the property of aviation fuel;  $x_{ij}$  is the element of composition matrix  $X$  (size = 10 x 13);  $u_{ij}$  is the element of property matrix  $U$  (size = 10 x 13). Property matrix for average density  $D$ , average freezing point  $M$ , average flash point  $F$ , and average net heat of combustion  $H$  were provided. Average property value was given for each class and carbon number.

### Partial least squares (PLS) method

Weight percentage of hydrocarbon classes was used as the independent variable of the PLS method and was obtained by transforming the composition matrix  $X$  as follows:

$$C = \begin{bmatrix} c_1 \\ \vdots \\ c_{10} \end{bmatrix} = X \cdot N = \begin{bmatrix} x_{1,1} & \cdots & x_{1,13} \\ \vdots & \ddots & \vdots \\ x_{10,1} & \cdots & x_{10,13} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

where  $c_i$  is the element of hydrocarbon classes column vector  $C$  (size = 10 x 1) and  $N$  is a column vector with all elements being unity (size = 13 x 1). The fuel property was calculated from:

$$P = C' A = \sum_{i=1}^{10} c_i a_i = A' C = [a_1 \cdots a_{10}] \begin{bmatrix} c_1 \\ \vdots \\ c_{10} \end{bmatrix}$$

where  $C'$  is the transposed matrix of  $C$ ;  $a_i$  is the element of PLS coefficients column vector  $A$  (size = 10 x 1). Coefficients vector  $A$  were provided for all four properties in the paper.

### Genetic algorithm (GA)

The property was calculated from:

$$P = L \cdot X \cdot K = [l_1 \cdots l_{10}] \begin{bmatrix} x_{1,1} & \cdots & x_{1,13} \\ \vdots & \ddots & \vdots \\ x_{10,1} & \cdots & x_{10,13} \end{bmatrix} \begin{bmatrix} k_1 \\ \vdots \\ k_{13} \end{bmatrix}$$

where  $L$  is the GA coefficient row vector (size = 1 x 10) and  $K$  is the coefficient column vector (size = 13 x 1). The  $L$  and  $K$  vectors were provided for all four properties in the paper.

### Modified weighted average method (MWA)

This method used the weighted property matrix  $W$ , as the independent variable. The weighted property matrix  $W$  was transformed from the composition matrix  $X$  and property matrix  $U$ . Each element of  $W$  is the weighted property defined as:

$$w_{ij} = x_{ij}u_{ij}$$

The property was correlated using the following equation:

$$P = M \cdot W \cdot N = [m_1 \cdots m_{10}] \begin{bmatrix} w_{1,1} & \cdots & w_{1,13} \\ \vdots & \ddots & \vdots \\ w_{10,1} & \cdots & w_{10,13} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

where  $M$  is the MWA coefficient row vector (size = 1 x 10);  $W$  is the weighted property matrix (size = 10 x 13);  $N$  is a column vector with all elements being unity (size = 13 x 1). The  $M$  coefficient vector was provided for all four properties in the paper.

In testing the MWA model, we found that our results for flash point and freezing point did not match those provided in the paper. On contacting the authors, we were informed that the  $M$  coefficients for flash point ( $T_F$ , K) and freezing point (FP, °C) in Table 9 of the paper were inadvertently switched [5]. The correct parameters are:

**Table S10. Corrected MWA model coefficients for flash point and freezing point [5]**

Model	Properties	Coefficients
MWA	$T_F$ (K)	$M = [1.1152; 0.7507; 0.9061; 1.3029; 0.8388; 1.0600; -4.851; 1.4785; 1.6361; 1.1350]$
	FP (°C)	$M = [-1.0978; 1.9496; 0.4752; 1.1457; 0.8571; 1.7976; -0.6740; 1.7006; 29.2301; 16.6297]$

### Short description of Vozka et al. density prediction methods

Vozka et al. [6] developed correlations between detailed chemical composition and density at 15 °C. The composition matrix obtained from GC × GC-TOF/MS and GC × GC-FID included seven hydrocarbon classes, each divided into carbon numbers or groups, ranging from C7 to C20. The hydrocarbon classes were: (1) *n*-paraffins, (2) isoparaffins, (3) monocycloparaffins, (4) di- and tri-cycloparaffins, (5) alkylbenzenes, (6) cycloaromatics compounds (indans, tetralins, indenenes, etc.), and (7) alkylnaphthalenes. The authors used 50 samples of petroleum-derived fuels and alternative fuel blending components. In addition, the density of a representative compound for each group was used in further calculations instead of the average density value for each class as utilized in the Shi et al. method [4]. In their work, in addition to the WA and PLS methods, they explored a high dimensional method using regularized support vector machines (SVM).

The methods are summarized below.

### Weighted average (WA) method

This method is the same as the WA method of Shi et al., except the density of a representative compound for each carbon number in a class was used instead of the average density value for each carbon number in a class as utilized in Shi et al.

### Partial least squares (PLS) and regularized support vector machines (SVM) methods

In this work, three matrices were defined. (1) a composition matrix X (size 7 x 14) containing weight fraction data of the seven classes and 14 groups representing carbon numbers C7 to C20. (2) a density matrix D (size 7 x 14) containing density values of the representative compounds for each group in the respective class and (3) a product matrix P (size 7 x 14) containing result of an elementwise multiplication of composition and density matrices, i.e.,

$$p_{ij} = x_{ij}d_{ij}$$

The product matrix was used to improve the predictive capabilities of the model.

PLS and SVM methods were applied to the composition matrix as well as the product matrix, resulting in two alternative correlations for each method. Therefore, a total of four correlations were presented: PLS composition, PLS product, SVM composition, and SVM product.

The algorithms utilized the composition matrix X in one of two ways: (1) weight fractions of each carbon numbers in a hydrocarbon class were summed and used a predictor resulting in seven predictors in total. (2) the weight fraction of each compound in the composition matrix was used resulting in 98 predictors in total.

The PLS composition and SVM composition correlations with 98 predictors used the following equation to predict density at 15 °C.

$$\rho = \beta_0 + \left( \sum_{i=1}^7 \sum_{j=1}^{14} \beta_{ij} x_{ij} \right)$$

where  $\beta_0$  is the intercept;  $\beta_{ij}$  is the correlation coefficient;  $x_{ij}$  is the element of the composition matrix.

When seven predictors were used, the equation can be simplified to:

$$\rho = \beta_0 + \left( \sum_{a=1}^7 \beta_a w_a \right)$$

where  $\beta_0$  is the intercept;  $\beta_a$  is the correlation coefficients of predictor  $a$ ;  $w_a$  is the sum of weight fractions of each carbon numbers in hydrocarbon class  $a$ .

The PLS product and SVM product correlations with 98 predictors used the following equation to predict density at 15 °C.

$$\rho = \beta_0 + \left( \sum_{i=1}^7 \sum_{j=1}^{14} \beta_{ij} p_{ij} \right)$$

where  $\beta_0$  is the intercept;  $\beta_{ij}$  is the correlation coefficient;  $p_{ij}$  is the element of the product matrix.

When seven predictors were used, the equation can be simplified to:

$$\rho = \beta_0 + \left( \sum_{a=1}^7 \beta_a p_a \right)$$

where  $\beta_0$  is the intercept;  $\beta_a$  is the correlation coefficients of predictor a;  $p_a$  is the sum product of weight fractions and densities of each carbon numbers in hydrocarbon class a.

The correlation coefficients  $\beta_0$  and  $\beta_{ij}$  for PLS and SVM using seven predictors were reported in the paper while those for the 98 predictors were given in the Supplemental Information [6].

---

<sup>1</sup> LECO Corporation. Pegasus® 4D GCxGC-MS (TOF), ChromaTOF software 4.51.6.0 optimized for Pegasus. <https://www.leco.com/product/pegasus-bt-4d/>. Website accessed March 7, 2023.

<sup>2</sup> NIST Standard Reference Database 1A, NIST/EPA/NIH Mass Spectral Library; Software Version: 2.2 (2014).

<sup>3</sup> Scanlon, J.T., D.E. Willis, "Calculation of Flame Ionization Detector Relative Response Factors Using the Effective Carbon Number Concept", *Journal of Chromatographic Science*, 23 (8), 1985, 333–340.

<sup>4</sup> Shi, X., H. Li, Z. Song, X. Zhang, and G. Liu, "Quantitative composition-property relationship of aviation hydrocarbon fuel based on comprehensive two-dimensional gas chromatography with mass spectrometry and flame ionization detector," *Fuel* 200 (2017) 395-406.

<sup>5</sup> Liu, G., Corresponding author of Shi et al, *Fuel* 200 (2017) 395-406 ([gliu@tju.edu.cn](mailto:gliu@tju.edu.cn)). Personal communication with lead author Suphat Watanasiri (Date: 05-07-2022).

<sup>6</sup> Vozka, P, B.A. Modereger, A.C. Park, W.T.J. Zhang, R.W. Trice, H.I. Kenttamaa, and G. Kilaz, "Jet fuel density via GC × GC-FID," *Fuel*, 2019, 235, 1052-1060.