energy&fuels

pubs.acs.org/EF

Article

Influence of Branching Degree-Derived *iso*-Alkane GC×GC Subgroups on Fuel Property Prediction

Published as part of Energy & Fuels special issue "18th International Conference on Stability, Handling and Use of Liquid Fuels (IASH2024)".

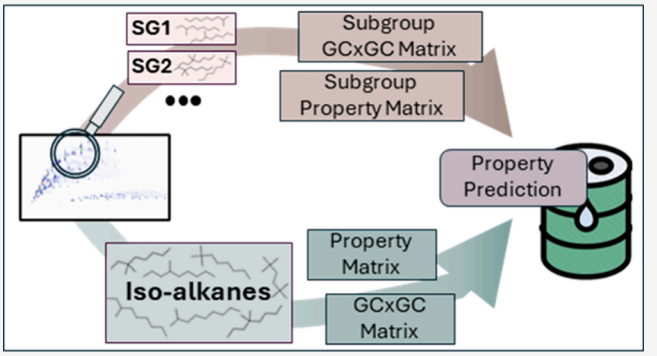
Florian Pütz,* Hannes Lüdtke, Astrid Ramirez Hernandez, Patrick Oßwald, Patrick Le Clercq, Uwe Bauder, Georg Eckel, and Andreas Huber

Cite This: *Energy Fuels* 2025, 39, 18001−18012

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: This study addresses a fundamental source of uncertainty in predicting the properties of sustainable aviation fuels (SAF) based on their composition: the unresolved distribution of structural isomers within iso-alkanes. Two complementary weighted average models are applied to analyze the influence of subgroups in composition analysis on the prediction of fuel properties. Both models are based exclusively on isomer properties and mixing rules without any training or fitting process. The first, the Mean Matrix model, represents fuels using conventional two-dimensional gas chromatography (GC×GC) resolution (carbon number by hydrocarbon family), whereas the second, the SubGroup Mean Matrix (SGMM) model, incorporates retention index (RI) based subgroups within the C7−C17 *iso*-alkane range to capture structural differences among isomers. Both models are built from a comprehensive isomer database augmented by Quantitative Structure−Property Relationship (QSPR) predictions where experimental data was unavailable. The model performance is evaluated on eight samples (Jet A-1; ATJ-SPK, two FT-SPK and four paraffinic solvents), comparing absolute errors and a normalized Relative Performance Change (RPC) metric. By incorporating higher compositional resolution via iso-alkane subgroups in the SGMM, the predictive accuracy for all evaluated properties of the ATJ-SPK fuel, which is characterized by a narrow and highly branched composition, was significantly improved. These properties included distillation temperatures, density, viscosity, net heat of combustion, and cetane number. In general, the improvements in volatile properties are most noticeable. Bulk properties such as density and net heat of combustion show only minor changes in prediction accuracy. Although the approach improves or maintains the predictive accuracy for the fuel samples, a deterioration can also be observed to some extent in the property prediction of the paraffinic solvents. Overall, the results indicate that incorporating subgroup-level resolution not only improves prediction accuracy for fuels with narrow, highly branched isomer profiles, but also can lead to an average performance improvement for most of the samples.

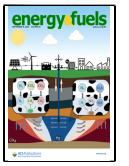


1. INTRODUCTION

The pathway toward climate neutral aviation through sustainable aviation fuel (SAF), which is being driven by political targets such as those of the European Union,[1] requires a rapid scale-up in fuel production capacities. Achieving these targets requires the utilization of a wide range of feedstocks and conversion pathways.[2−4] However, this diversity brings substantial challenges for certification of new production pathways. New synthetic blending component (SBC) candidates must undergo the ASTM D4054 approval process in order to being annexed to the ASTM D7566 standard specification to ensure a safe use. The certification process involves extensive experimental testing and requires significant cost, time, and resource investments.[5,6] These barriers often slow the transition from early laboratory-scale innovation to commercial deployment.

To accelerate development and derisk early stage investment, a variety of fuel property prediction models[6−14] have been developed to support research and development in the field of SAF. These models provide composition-based property predictions for candidate fuels in early stages. To

Received: May 28, 2025
Revised: August 29, 2025
Accepted: August 29, 2025
Published: September 8, 2025



address uncertainties in model predictions, researchers have adopted various strategies - ranging from competing models[10] and probabilistic frameworks[7] to predictive capability metrics that assess model reliability.[15]

While these strategies are effective in assessing and quantifying uncertainties, they do not reduce them at the root. One source of such uncertainty stems from structural isomers - molecules with identical molecular formula but different structures. These isomers can have different physical properties,[16] which limits the accuracy of predictive models. Conventional models often oversimplify fuel composition by assuming that either all isomers for which a measured property value is available or only one or two within a hydrocarbon family and carbon number might be present in the fuel.[7,17,18] These simplifications can lead to inaccuracies in predicting fuel properties and performance by neglecting the actual isomeric distribution present in real fuels. The problem is that even advanced techniques, such as comprehensive two-dimensional gas chromatography (GC×GC), are unable to resolve the composition of fuels on a structural isomer level.[12]

Shi et al.[18] assumed a limitation in predicting the cold weather properties of diesel fuel due to the reliance on compositional data based solely on carbon number and hydrocarbon family. To address this, they proposed distinguishing between mono- and highly branched structures, emphasizing that a more detailed understanding of isomeric composition could enhance predictive accuracy. Bell[4] identified isomeric variance as a major source of uncertainty in fuel property prediction and developed a method using two-dimensional gas chromatography with a vacuum ultraviolet detector (GC×GC−VUV) to resolve specific isomers in jet fuels. The study aimed to improve the reliability of property predictions by addressing the influence of isomeric diversity. The effectiveness of VUV-based detection, though, is limited by the availability of reference spectra in spectral libraries.[4] Feldhausen et al.[11] used GC×GC−VUV with a novel deconvolution technique to identify 26 structural isomers accounting for 93.6% of a Synthetic Aromatic Kerosene (SAK) mass. This high analytical resolution led to significant improvements in prediction accuracy, including a 90% reduction in viscosity prediction error. Yang et al.[9] developed Tier α property prediction tools and found that properties with low isomeric variance, such as density, are predicted more accurately than those with high variance, such as viscosity. All the studies highlight the importance of knowing the isomer distribution within a fuel for accurately predicting its properties.

To better resolve the distribution of isomers in GCxGC analysis, Lüdtke et al.[19] developed a method in which the *iso*-alkanes are divided into subgroups based on their retention indices (RI).[19] Although it does not fully eliminate isomer-related uncertainty, this advancement aims to reduce it and therefore addresses key limitations of earlier models. It has the potential to enable more accurate predictions of fuel properties by characterizing the composition in more detail. In this way, the dependence on reference spectra, which limits the applicability of the GC×GC-VUV method,[4] is avoided.

Building on these advancements in compositional analytics, this study investigates the influence of structural resolution on the prediction of fuel properties using two basic, training-free modeling approaches. The first, the Mean Matrix model is similar to the weighted average method proposed by of Shi et al.[18] and Vozka et al.[20] The Mean Matrix model extends the weighted average method[18,20] by using Quantitative Structure−Property Relationship (QSPR)[21] predictions in addition to the measured properties to calculate the average property values. It represents fuel composition at the level of carbon number and hydrocarbon family, reflecting the current state-of-the-art resolution typically available in compositional analyses. The second model, the SubGroup Mean Matrix (SGMM), extends this framework by including retention index (RI)-based subgroups and therefore uses additional information about the distribution of isomers within the fuel. This enhanced resolution captures systematic differences among isomers - particularly within *iso*-alkane families - and thus has the potential to enable more accurate modeling of bulk properties. The rule for forming subgroups was adopted for this work exactly as published by Lüdtke et al.[19] Since the underlying study of Lüdtke et al.[19] has so far been limited to *iso*-alkanes with a chain length C_7−C_{17}, this study is subject to the same limitations with regard to the iso-alkane range.

Both models are part of a broader spectrum of methodologies used for predicting fuel properties. This spectrum includes molecule-based models, which use the structures and properties of individual components along with mixing rules to estimate bulk properties.[9,10,18,20,21] Hybrid approaches, such as the M-QSPR model,[7] which combine isomer property data with the general composition and properties of fuels. This allows them to be trained on both isomer and fuel data sets, eliminating the need for mixing rules as these can be derived from the data itself. At the other end of the spectrum are empirical models that treat the fuel as a single entity.[19,20] These models can learn composition-property correlations directly from fuel data but they require large and diverse data sets for training. Within this context, the Mean Matrix and SGMM models explicitly represent fuels as mixtures of individual molecular components. This design enables their use as proof-of-concept frameworks, as they rely exclusively on measurements and QSPR-based[10,21−24] predictions of molecular properties without any training on fuel samples. As a result, there is no need for a large amount of subgroup GCxGC data, which is not available at the time of writing. Both models are based on the calculation of the mean property values for all combinations of carbon numbers and hydrocarbon families of a GCxGC analysis.[18] The resulting mean property matrix can then be used to predict the bulk property by simple multiplication of the matrix values with the composition, taking mixing rules into account. Saldana et al.[25] have previously shown that the combination of QSPR-predicted property values with appropriate mixing rules can provide accurate flash point predictions for binary and ternary mixtures. While their hybrid approach differs in implementation, the underlying concept of combining QSPR predictions and mixing rules is related to the methodology used in this study.

To evaluate the impact of subgroups on this weighted average method, a range of test fuels was selected, including conventional Jet A-1, synthetic Fischer−Tropsch (FT-SPK) and Alcohol-to-Jet (ATJ-SPK) fuels, as well as paraffinic solvents. These samples span a spectrum of isomeric complexity and compositional diversity, making them ideal for assessing the potential benefits and limitations of subgroup-informed modeling. Using this data set, the study investigates whether the enhanced compositional resolution provided by the SGMM, relative to the baseline Mean Matrix model, improves the predictive accuracy of fuel properties and thereby

https://doi.org/10.1021/acs.energyfuels.5c02762
Energy Fuels 2025, 39, 18001−18012

supports more reliable early stage prescreening in sustainable aviation fuel (SAF) development.

2. METHODOLOGY

2.1. Molecule Database. For the calculation of the mean matrix, an extensive isomer database is required. The isomers used in this work are all from commercially or freely available databases of single components. The majority of these isomers are from the NIST Standard Reference Database 103a[26] and DIPPIR 801 database[27] as well as the public databases Pubchem[28] and Chemspider.[29] Since it is not possible to determine all isomers in a fuel[19] and, at the time of writing, no information is available on the general possible isomers in such mixtures, this work follows the approach of Hall et al. and assumes that all molecules listed in physicochemical property databases can occur in a fuel.[10] In contrast to the approach of Yang et al.[30] or Shi et al.,[18] this method is not limited to the availability of measurements, as property values are predicted using structural information from the referenced databases whenever measurement data are unavailable.[10]

The need for isomer property prediction arises from the significant disparity between the number of theoretically possible constitutional isomers and those with available property data. For *iso*-alkanes in the C7−C17 range, 42900[31] isomers are theoretically possible, yet only 353 are included in the molecular database in this study (~0.82%), and only 42 of the entries have a measured cetane number values (~0.1%). This highlights the limitations of existing databases and underscores the importance of predictive methods to address these data gaps. (Supporting Information A)

2.2. Isomer Property Prediction with Quantitative Structure−Property Relationship Models. To increase the number of available property values for potential isomers for each carbon number, we use the Quantitative Structure−Property Relationship (QSPR) models developed by Hall et al.[7] to predict the properties of all isomers in our database for which no measurement is available. These are based on the research of Saldana et al.[22−24] and correlate the structure of a molecule with its physicochemical properties. The structure of a molecule is thereby described by quantifying molecular descriptors such as the carbon number or the number of methyl groups.[21] The research of Hall et al.[10,21] and Saldana et al.[22−24] has shown that QSPR models provide reliable property predictions for a wide range of hydrocarbons, and therefore provide the foundation for the calculation of the mean property matrix.

2.3. Mean Property Matrix. The mean property matrix P is the core element of the proposed modeling framework. As shown in Figure 1, each cell corresponds to the average property for a given combination of carbon number and hydrocarbon family. These average properties are calculated in three steps. First, all isomers corresponding to a given matrix cell - defined by hydrocarbon family and carbon number - are sampled from the isomer database (chapter 2.1). In the second step, the property value for each isomer is predicted individually using QSPR models (chapter 2.2) if no experimental measurement is available. Finally, the arithmetic mean of these predicted (and measured, if available) values is computed for each matrix entry. This means that the property for an average pseudo isomer is calculated for each cell of the matrix.

For the experimental validation carried out in chapter 3, the matrices for the following properties are calculated: density at 15 °C, kinematic viscosity at 20 °C and net heat of combustion. However, the principle can be extended to other properties and temperatures as long as QSPR models (or other reliable models) are available for these properties.

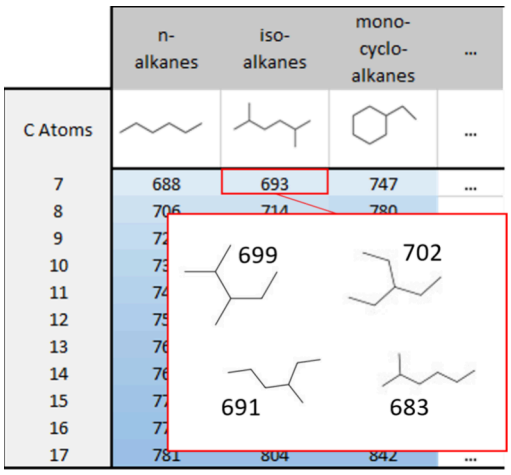


Figure 1. Calculation of the mean property matrix P. The figure shows exemplary values for density at 15 °C in kg/m^3.

2.4. Mean Matrix Model. The Mean Matrix model combines the two different matrices shown in Figure 2. The composition matrix $W = (w_{ij})_{ij}$ represents the detailed composition of the fuel based on a GCxGC analysis. It is organized according to hydrocarbon families and carbon chain length, where each element w_{ij} is the relative occurrence of a particular carbon number i for hydrocarbon family j. The same applies to the average property matrix $P = (p_{ij})_{ij}$ described in Chapter 2.3. Each element p_{ij} stores the mean property value per hydrocarbon family and carbon number.

The bulk property $p_{mixture}$ of the fuel is calculated by the summation of the individual mean properties $p_{i,j}$, weighted by their respective fractions $w_{i,j}$. A linear mixing rule is used for the properties density, cetane number and net heat of combustion (eq 1).

$$p_{mixture,lin} = \sum_{i,j} w_{i,j} \cdot p_{i,j} \tag{1}$$

An exponential mixing rule is used for the kinematic viscosity (eq 2).

$$p_{mixture,\ln} = \exp\left(\sum_{i,j} w_{i,j} \cdot \ln p_{i,j}\right) \tag{2}$$

During the validation against the CRC World Fuel Survey data set[32] (Supporting Information B), a consistent offset was observed in the predicted density values. This offset is likely due to the overrepresentation of subgroups 1(a and b) and 2 in the database. In the isomer database of this work (Supporting Information C), subgroup 1 comprise 49.0% of the data set (220 entries), whereas subgroups 3−7 account only for 21.8% (96 entries). According to the subgroup classification of Lüdtke et al.,[19] lower-branched isomers occur predominantly in smaller subgroups. Isomers with fewer branches exhibit slightly higher densities than those with a higher degree of branching (see also Figure 11, exemplary for C9). Consequently, when mean values are calculated, these isomers would receive disproportionate weighting resulting in elevated mean matrix values and consequently higher density. One

https://doi.org/10.1021/acs.energyfuels.5c02762
Energy Fuels 2025, 39, 18001−18012

C Atoms	n-alkanes	iso-alkanes	mono-cyclo-alkanes	...
7	0.0	0.0	0.0	...
8	1.1	1.8	0.1	...
9	2.4	5.8	0.5	...
10	2.0	5.7	0.3	...
11	2.3	6.4	0.4	...
12	1.0	6.3	0.2	...
13	4.2	6.3	0.1	...
14	2.8	7.0	0.1	...
15	2.7	8.9	0.0	...
16	0.1	6.4	0.0	...
17	0.0	1.1	0.0	...

C Atoms	n-alkanes	iso-alkanes	mono-cyclo-alkanes	...
7	688	693	747	...
8	706	714	780	...
9	722	725	795	...
10	735	746	808	...
11	743	766	822	...
12	752	768	823	...
13	760	777	823	...
14	767	786	826	...
15	772	789	833	...
16	777	795	836	...
17	781	804	842	...

Figure 2. Mean Matrix model combining composition matrix W (left, mass %) and mean property matrix P (right, density at 15 °C in kg/m^3).

C Atoms	n-alkanes	iso-alkanes	mono-cyclo-alkanes	bi-cyclo-alkanes	mono-aromatics
7	0.0	0.0	0.0	0.0	0.0
8	1.1	1.8	0.1	0.0	0.0
9	2.4	5.8	0.5	5.3	0.0
10	2.0	5.7	0.3	5.9	0.1
11	2.3	6.4	0.4	4.2	0.0
12	1.0	6.3	0.2	3.7	0.1
13	4.2	6.3	0.1	2.2	0.0
14	2.8	7.0	0.1	1.0	0.0
15	2.7	8.9	0.0	0.7	0.0
16	0.1	6.4	0.0	0.6	0.0
17	0.0	1.1	0.0	0.0	0.0
18	0.0	0.1	0.0	0.0	0.0

C Atoms	n-alkanes	iso-alkanes group1	iso-alkanes group2	...	mono-cyclo-alkanes	bi-cyclo-alkanes	mono-aromatics
7	0.0	0.0	0.0	...	0.0	0.0	0.0
8	1.1	0.4	0.5	...	0.1	0.0	0.0
9	2.4	1.2	1.7	...	0.5	5.3	0.0
10	2.0	1.1	1.7	...	0.3	5.9	0.1
11	2.3	1.3	1.9	...	0.4	4.2	0.0
12	1.0	1.3	1.9	...	0.2	3.7	0.1
13	4.2	1.3	1.9	...	0.1	2.2	0.0
14	2.8	1.4	2.1	...	0.1	1.0	0.0
15	2.7	1.8	2.7	...	0.0	0.7	0.0
16	0.1	1.3	1.9	...	0.0	0.6	0.0
17	0.0	0.2	0.3	...	0.0	0.0	0.0
18	0.0	0.0	0.0	...	0.0	0.0	0.0

Figure 3. Difference in GCxGC resolution - Mean Matrix model (left) vs SubGroup Mean Matrix model (right). Values are shown in mass %.

often used solution would be to fit the property matrix to the density data. Since this requires more data than is available, we account for this systematic deviation by subtracting a correction factor of 8.1 kg/m^3 from all predicted densities. This correction was applied uniformly for both the current model and the SGMM model.

The prediction of distillation curve properties - T_{10}, T_{50}, T_{90}, and the final boiling point (FBP) - deviates from the methodology applied to other fuel properties. Instead of relying on mean values from the property matrix, improved prediction accuracy was achieved by distributing the mass fraction to the individual boiling points of the isomers within each carbon number and hydrocarbon family. This method captures the natural variability of boiling behavior rather than a single average value, resulting in a more representative prediction of the distillation profile. The distillation values are predicted for ASTM D2887, and, when needed, a correlation according to Ford et al.[33] is applied to estimate atmospheric distillation values (ASTM D86), as suggested in the D2887 standard.

2.5. SubGroup Mean Matrix Model. The Subgroup Mean Matrix (SGMM) model extends the Mean Matrix model by integrating additional compositional details obtained from retention index (RI)-based subgrouping in two-dimensional gas chromatography (GC×GC). This enhancement leads to a more detailed characterization of the fuel composition, as the isomer distribution and its influence on the fuel properties can be reflected to some extent. As a result, the accuracy of the property prediction is expected to improve and its uncertainty to decrease.

In contrast to the conventional composition matrix, which aggregates hydrocarbons exclusively according to family and carbon number, the subgroup composition matrix contains an additional degree of separation (Figure 3). According to the rule for the formation of subgroups proposed by Lüdtke et al.,[19] *iso*-alkanes were sorted in up to eight subgroups, depending on carbon number and RI. The subgroup classification is structured so that subgroup number correlates with molecular branching, with lower subgroup numbers corresponding to a lower degree of branching.[19] The structure of the SGMM model remains essentially unchanged compared to the Mean Matrix model, except for the addition of these subgroups. From a modeling perspective, the change is equivalent to the introduction of new hydrocarbon families.

This change also requires an adjustment of the property matrix as well as the standard deviation matrix. Instead of sampling from all *iso*-alkanes within a given carbon number, the SGMM matrices are now calculated by sampling only the *iso*-alkanes that belong to the corresponding subgroup (an overview of the isomers per subgroup available in our database is provided in Supporting Information C). While this should lead to a more accurate prediction of property values, it also increases the likelihood of empty matrix cells if there is no molecule for a subgroup in the database (section 2.1). For these cases, where no isomer structure and therefore no property measurement or prediction is available, an adaptive

https://doi.org/10.1021/acs.energyfuels.5c02762
Energy Fuels 2025, 39, 18001−18012

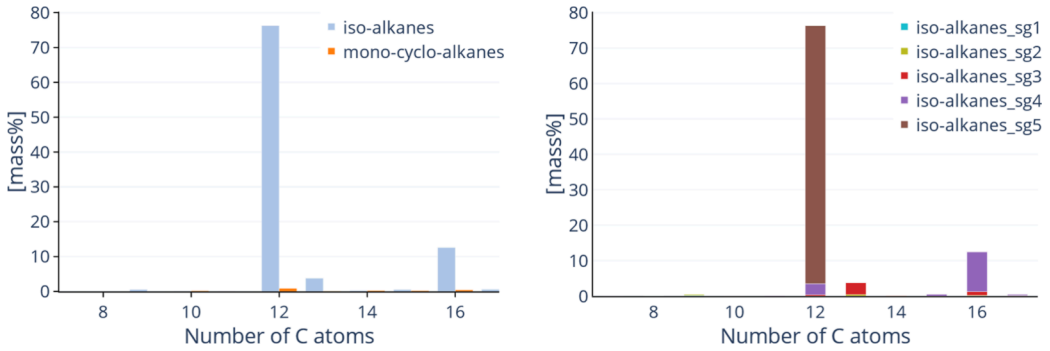


Figure 4. Fuel composition (left) and RI-based *iso*-alkane subgroup distribution (right) of ATJ-SPK. The compositional profile is resolved by hydrocarbon family and carbon number (color coded families), while the subgroup plot illustrates structural differentiation within the *iso*-alkane fraction by carbon number (color coded subgroups). Reproduced from ref 19. Available under a CC BY 4.0 license. Copyright 2025 Lüdtke et al.

property value assignment is carried out. In the first step, the subgroup distinction is removed and the entire *iso*-alkane family is considered. If there is still no matrix value available for the given carbon number, in the second step a search is made for property values within a range of ± 2 carbon numbers. For all other families (not *iso*-alkanes), step two is carried out directly if average property values are missing. All equations (eqs 1−4) from section 2.4 are also valid for the SGMM model. Since not enough measurement data with subgroup composition analyses is available at the time of writing, the scaling factor S_f determined for the Mean Matrix model is also applied to the SGMM model.

2.6. Relative Performance Change Metric. To quantitatively assess model performance, this study introduces the Relative Performance Change (RPC) metric. RPC quantifies the percentage change in absolute prediction error between an evaluated model and a baseline model, normalized by the error of the baseline model. The prediction error is defined as the difference between the measured and predicted values for a given fuel property pair. The RPC is calculated as follows:

$$RPC_{p,f} = \frac{|\varepsilon_{base\,p,f}| - |\varepsilon_{eval\,p,f}|}{\max_{\hat{f}\,\epsilon F}|\varepsilon_{base\,p,\hat{f}}|} \times 100\% \quad (3)$$

where:

- $\varepsilon_{basep,\,f}$ and $\varepsilon_{evalp,\,f}$ are the absolute prediction errors of the baseline and evaluated models, respectively for fuel f and property p,
- F denotes the set of all fuels,
- and $\max_{\hat{f}\epsilon F}|\varepsilon_{\mathrm{base}_{p,\hat{f}}}|$ is the maximum absolute error of the baseline model across all fuels for a specific property.

This normalization scheme ensures that performance differences are scaled by the maximum observed prediction error within each property, making the results comparable across properties with different physical units and magnitudes. Standardizing by the maximum baseline error across all fuels, instead of the baseline error of the evaluated fuel, reduces the risk of inflated relative differences when prediction errors are small and the baseline model already exhibits high accuracy. Thus, the RPC metric provides a more interpretable and balanced measure of comparative model performance for heterogeneous property types.

For the comparison of the mean matrix and the SGMM model, the RPC is calculated as

$$RPC_{p,f} = \frac{|\varepsilon_{MM\,p,f}| - |\varepsilon_{SGMM\,p,f}|}{\max_{\hat{f}\,\epsilon F}|\varepsilon_{MMp,\hat{f}}|} \times 100\% \quad (4)$$

This metric is a measure of the extent to which the introduction of compositional detail at the subgroup level impacts predictive accuracy. Positive RPC values represent an enhancement in prediction due to subgrouping, while negative values indicate performance decline.

3. EXPERIMENTAL EVALUATION

The effectiveness of the SGMM model is evaluated by comparing the model predictions with measured properties for the distillation curve, density at 15 °C, kinematic viscosity at −20 °C, net heat of combustion and cetane number. This analysis focuses on assessing how the inclusion of subgroup information affects property predictions compared to the traditional Mean Matrix model. A property-wise overview of how the subgroups affect the modeling assumptions can be found in the Supporting Information D. Due to the limited availability of detailed measurements of subgroup composition, this study should be considered a proof of concept demonstrating the potential of incorporating subgroup-level information into predictive modeling frameworks. The limited size of the data set is primarily due to two factors. First, subgroup composition analysis has been introduced recently, and currently, only one analytical laboratory performs these measurements. Second, as reported in Lüdtke et al.,[19] subgroups 1 and 2 dominate the composition of most samples, especially conventional jet fuels. However, for these subgroups, there is hardly any deviation between the property matrix of the SGMM and the Mean Matrix Model (Supporting Information D). The reason for similarity of the property matrices is the disproportionately high representation of subgroups 1 and 2 in the underlying isomer database (Supporting Information C). This study therefore focuses primarily on fuels whose composition differs significantly from conventional Jet A-1. Those fuels cause major challenges in terms of property prediction. These samples are rare for several reasons. One reason is that many production processes are still in development, sample data often cannot be published. Another reason is that the fuels themselves are usually only available in small quantities, which limits the possibility of comprehensive property measurements. A Predictive Capability Assessment as proposed by Hall et al.,[15] to determine the impact of subgroups on prediction accuracy is beyond the scope of the present study due to these limitations.

https://doi.org/10.1021/acs.energyfuels.5c02762
Energy Fuels 2025, 39, 18001−18012

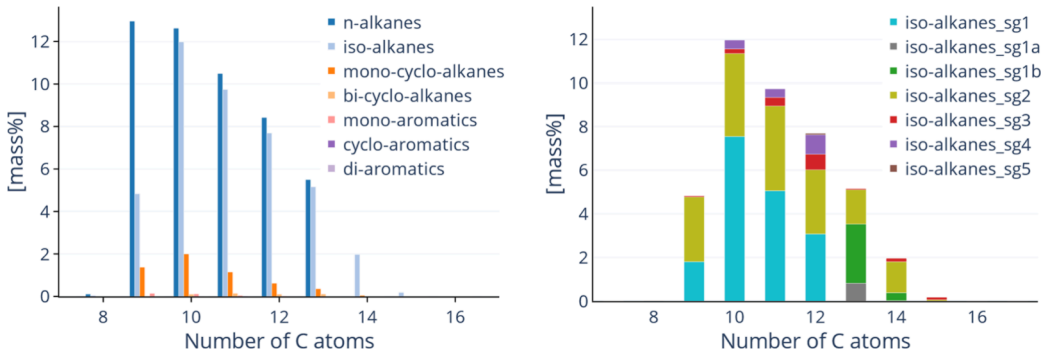


Figure 5. Fuel composition (left) and RI-based *iso*-alkane subgroup distribution (right) of FT-SPK_1. The compositional profile is resolved by hydrocarbon family and carbon number (color coded families), while the subgroup plot illustrates structural differentiation within the *iso*-alkane fraction by carbon number (color coded subgroups). Reproduced from ref 19. Available under a CC BY 4.0 license. Copyright 2025 Lüdtke et al.

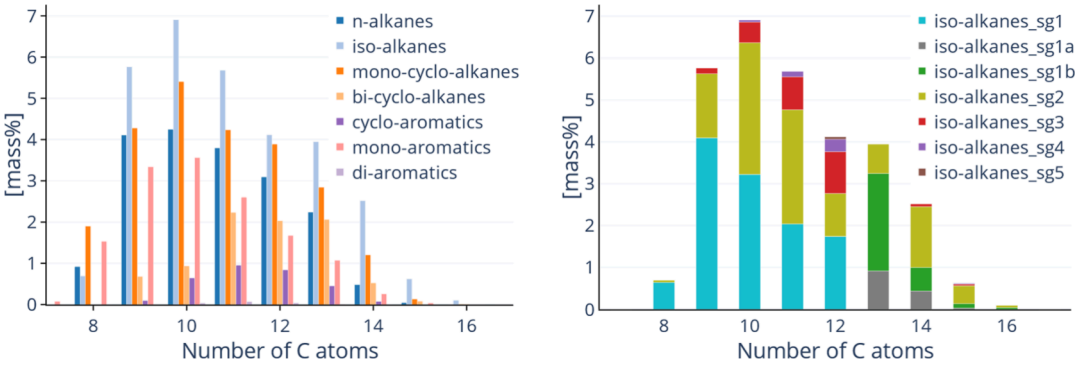


Figure 6. Fuel composition (left) and RI-based *iso*-alkane subgroup distribution (right) of Jet A-1. The compositional profile is resolved by hydrocarbon family and carbon number (color coded families), while the subgroup plot illustrates structural differentiation within the *iso*-alkane fraction by carbon number (color coded subgroups). Reproduced from ref 19. Available under a CC BY 4.0 license. Copyright 2025 Lüdtke et al.

3.1. Fuel Selection - Experimental Setup. The SGMM model is evaluated using a total of eight samples, including a conventional Jet A-1, three synthetic paraffinic blending components, including an Alcohol to Jet (ATJ-SPK) and two Fischer–Tropsch (FT-SPK_1, FT-SPK_2) as well as four commercially available paraffinic solvents (IsoPar G, IsoPar H, IsoPar J, IsoPar L). This distinction allows the performance of the SGMM model to be analyzed for different aspects. While the Jet A-1, the FT-SPK and the ATJ-SPK differ significantly in their composition, the paraffinic solvents are an ideal test case for the modeling approach, as the high content of *iso*-alkanes enables good applicability of the subgroup approach.

The ATJ-SPK shown in Figure 4 is a relevant test case to demonstrate the limitations of the mean matrix model and the benefits of adding additional subgroups to the modeling approach. The composition of the fuel has two distinct peeks in subgroup 5 of C_{12} and subgroup 4 of C_{16}. As explained by Lüdtke et al., a high degree of branching is therefore expected.[19] It is already reported in the literature, that these two peaks are caused by two isomers, which together account for 88.2% of the composition.[34,35] This is in clear contradiction to the core assumption of the mean matrix approach, that the fuel can be modeled by considering all isomers within a certain hydrocarbon family and carbon number.

In addition to its low molecular diversity, the composition of the fuel consists of approximately 95% *iso*-alkanes between carbon numbers C_7 and C_{17}. The subgroup method according to Lüdtke et al.[19] is only defined for *iso*-alkanes in this range, which means that a subgroup analysis is possible for almost the entire fuel.

Unlike the ATJ-SPK, the FT-SPK_1 (Figure 5) has a lower degree of branching. This is reflected in the dominance of subgroups 1 and 2. In addition, the *iso*-alkane content is significantly lower at around 42%. Another 50% of the composition are *n*-alkanes (*n*-alkanes + *iso*-alkanes ∼ 92%). The effect of the isomer distribution from nonparaffinic hydrocarbon families is therefore limited to 8% of the fuel.

The Jet A-1 (Figure 6) has the most complex composition of the selected jet fuels. Only 30% of the fuel composition is *iso*-alkanes. Due to the current restriction of the subgroup method to *iso*-alkanes from C_7 to C_{17}, a subgroup analysis is only possible for this share. In addition, the *iso*-alkane profile reveals that it is a fuel with low branching, with the dominant subgroups 1 and 2. In contrast to the FT-SPK_1, the isomer information is missing for 50% of the fuel composition.

The remaining samples (solvents and FT-SPK_2) are shown in Figure 7. The figure shows the respective composition of the samples including the subgroup profile, summed up by the carbon number. The exact compositions for the individual samples are provided in Supporting Information E.

4. RESULTS AND DISCUSSION

4.1. Impact of Subgroup Analysis on Property Prediction. To assess the influence of subgroup resolution on property prediction, several key fuel properties - namely distillation curve characteristics (T_{10}, T_{50}, T_{90}, and final boiling point), density, net heat of combustion, kinematic viscosity, and cetane number - were modeled using both the Mean Matrix and SGMM approaches. The model comparison in this study is based on deviations between predicted and

https://doi.org/10.1021/acs.energyfuels.5c02762
Energy Fuels 2025, 39, 18001−18012

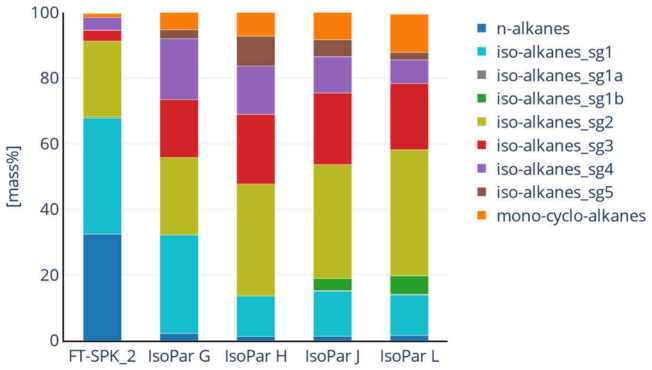


Figure 7. Composition FT-SPK_2, IsoPar G, H, J and L, with subgroup information, aggregated over the carbon number. Hydrocarbon families are only shown if their mass fraction is at least 0.2% of the total composition.

experimentally measured mean values across all available samples for both modeling methods. This approach enables an evaluation of relative prediction accuracy across fuels with differing subgroup profiles.

Figure 8 shows the Relative Performance Change (RPC) for each fuel-property pair, comparing the SGMM model to the

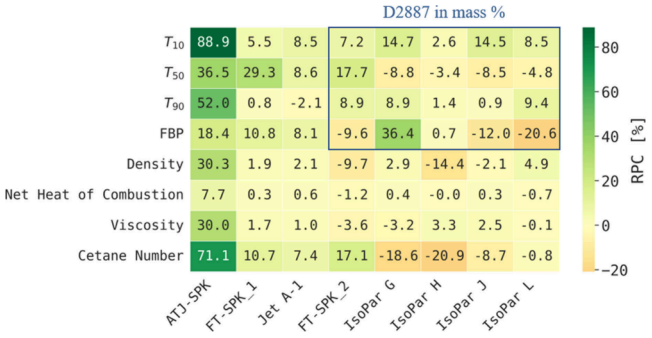


Figure 8. RPC of SGMM model compared to Mean Matrix baseline for T_{10}, T_{50}, T_{90}, FBP, density, net heat of combustion, kinematic viscosity and cetane number. Positive values indicate improved prediction accuracy with the SGMM model; negative values indicate reduced accuracy. The blue box indicates that the distillation values are not from an atmospheric distillation according to ASTM D86, but from a simulated distillation according to ASTM D2887.

Mean Matrix baseline. The heatmap visualizes the direction and magnitude of performance differences using a color gradient, with positive values indicating improved prediction accuracy and negative values indicating lower accuracy for the SGMM model. This representation facilitates direct comparison of model performance across different fuels and properties. For the distillation properties T_{10}, T_{50}, T_{90}, FBP only simulated distillation values were available for IsoPar G, H, J and L as well as the FT-SPK_2, which is indicated by the blue box.

The data presented in Figure 8 forms the foundation for an analysis of each individual property, which will be discussed in detail in the following sections. To make the results more accessible, the calculated RPC have been divided into five categories. RPC values greater than 15% are classified as a "significant improvement", values between 15% and 5% as a "moderate improvement", and values between 5% and −5% as "no significant change". Similarly, values less than −15% are classified as a "significant decline" and values between −15% and −5% as a "moderate decline". An overview of the classification of the evaluation metric and a graphical summary of the RPC matrix from Figure 8 are provided in Supporting Information F.

While the RPC formulation enables meaningful comparison of a model performance across heterogeneous properties, it remains sensitive to the characteristics of the fuel data set. In particular, when the maximum baseline error is very small, RPC values may appear disproportionately large, potentially overstating improvements. Conversely, if the maximum error is high, the metric may understate performance change.

Distillation. For T_{10}, the incorporation of subgroup information resulted in noticeable improvements in prediction accuracy for seven of the eight samples, as indicated by the absolute prediction errors in Figure 9. With the exception of IsoPar H, all samples exhibited a moderate improvement in prediction accuracy, while ATJ-SPK showed a significant improvement in agreement with the measured values (Figure 8).

For T_{50}, the incorporation of subgroup information led to improvements across all non IsoPar samples (Figure 9). The largest improvements were observed for ATJ-SPK, FT-SPK_1 and FT-SPK_2, which exhibited significant improvements in prediction accuracy. IsoPar G and IsoPar J showed moderate decline (Figure 8).

As shown in Figure 9 for T_{90}, predictive performance improved for half of the samples. IsoPar H and IsoPar J, which were previously overestimated by the baseline model, are now underestimated, resulting in comparable absolute errors. FT-SPK_1 and Jet A-1 exhibited no significant change in predictive accuracy. The remaining samples showed moderate improvements, while ATJ-SPK showed significant improvements as the prediction error was reduced by 52% (Figure 8).

For the final boiling point, the incorporation of subgroup information resulted in improved predictive accuracy for four out of the eight samples, while three were showing a decline in prediction accuracy (Figure 9). ATJ-SPK, which was previously overpredicted, is now slightly underpredicted, but the overall accuracy significantly improved. Similar levels of improvement were observed for IsoPar G. Jet A-1 and FT-SPK_1 exhibited moderate improvement in prediction accuracy, whereas IsoPar H remained almost unaffected. In contrast, a moderate decline in performance was noted for IsoPar J, FT-SPK_2 and a significant decline for IsoPar L (Figure 8). It is worth mentioning that within the distillation parameters, the final boiling point according to ASTM D86 and ASTM D2887 is the measurement value with the greatest uncertainty. The standard specifies a reproducibility of up to 10.4 °C. For this reason, the evaluation of the model prediction for this property should be treated with caution.

Overall, the inclusion of subgroup information improved the prediction performance for the distillation properties. This general improvement is consistent with the theoretical foundation of gas chromatography, where the retention strongly correlates to volatility, a key parameter in distillation behavior.[36,37] A plot showing boiling point measurements against RI with color coded subgroups is provided for *iso*-alkane C10 structures in Supporting Information G.

Density and Net Heat of Combustion. As shown in Figure 10, the incorporation of subgroup information did not result in significant changes in prediction accuracy for density and net heat of combustion across the majority of samples. This outcome reflects the inherently lower isomeric variance

https://doi.org/10.1021/acs.energyfuels.5c02762
Energy Fuels 2025, 39, 18001−18012

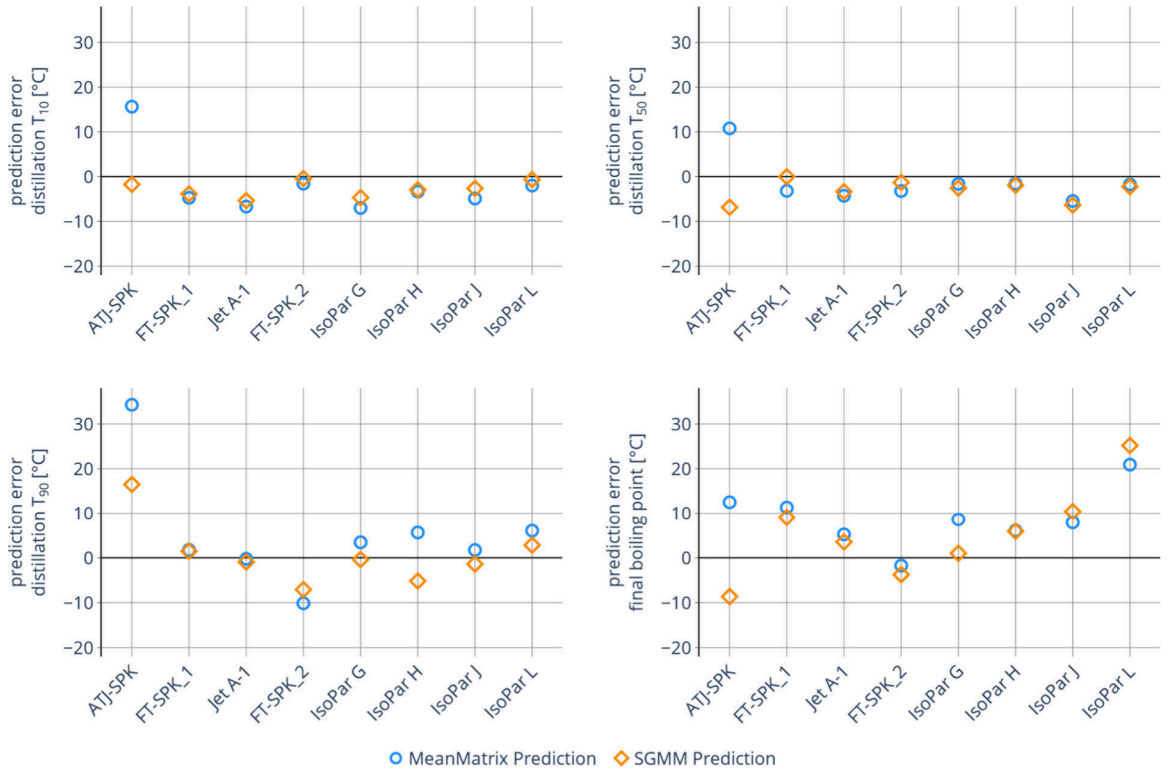


Figure 9. Absolute prediction errors for distillation temperatures T10 (top left), T50 (top right), T90 (bottom left), and final boiling point (bottom right) across all samples. Blue circles represent errors from the Mean Matrix model; orange diamonds represent errors from the SGMM model. The horizontal zero line indicates perfect agreement with experimental measurements. For FT-SPK_2 and IsoPar G, H, J and L the distillation values are not from an atmospheric distillation according to ASTM D86, but from a simulated distillation according to ASTM D2887.

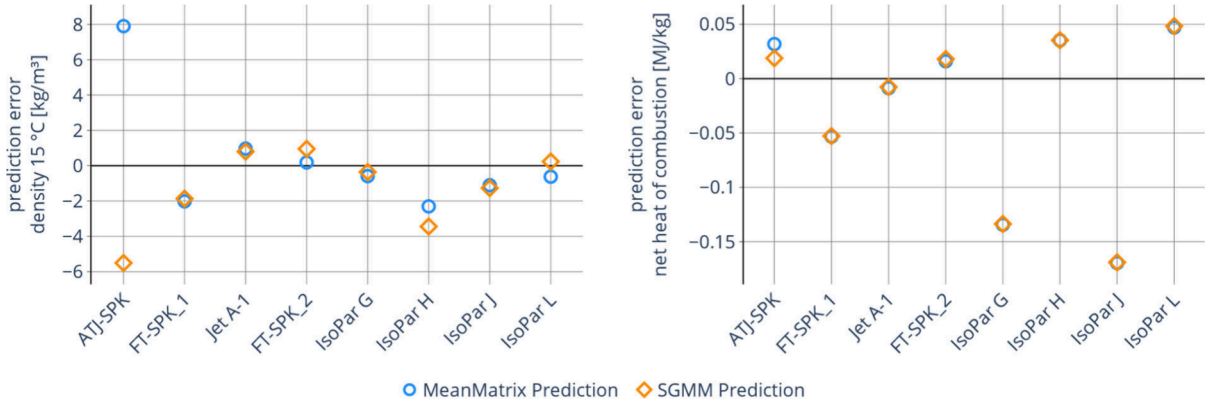


Figure 10. Absolute prediction errors for density (left) and net heat of combustion (right) across all samples. Blue circles represent errors from the Mean Matrix model; orange diamonds represent errors from the SGMM model. The horizontal zero line indicates perfect agreement with experimental measurements.

associated with these bulk properties,[16] which lowers the potential impact of increased compositional detail through subgrouping. A notable exception is ATJ-SPK, where the SGMM model produced a significant improvement in density prediction and a moderate improvement in net heat of combustion prediction (Figure 8). These improvements are due to the highly distinctive and compositionally narrow profile of the fuel, which violates the core assumption of the mean matrix model that the isomers are evenly distributed within each compositional bin. The SGMM model is better suited to capture such compositional specificity due to the inclusion of RI-based subgrouping. This is particularly relevant for the ATJ-SPK in this work, which contains mostly high branched isomers (subgroups 4 and 5) that are underrepresented in the isomer database.

For the remaining fuels, a moderate decline in prediction accuracy was observed for FT-SPK_2 and IsoPar H for the density prediction (Figure 8), whereas all other fuels exhibited negligible changes. These results are consistent with the trends shown in Figure 11, which indicates a relatively weak correlation between density and RI in GC×GC-based characterization. This is particularly evident in subgroups 1, 2 and 3, which dominate in the samples other than the ATJ-SPK.

Similarly, for net heat of combustion, the SGMM model yielded no accountable change in predictive performance for

https://doi.org/10.1021/acs.energyfuels.5c02762
Energy Fuels 2025, 39, 18001−18012

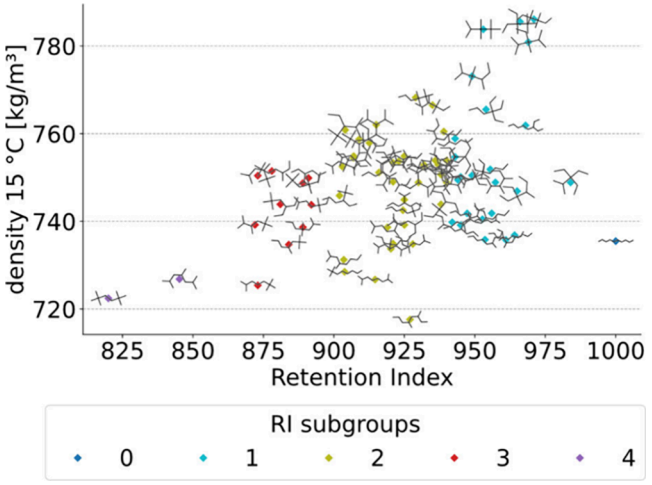


Figure 11. Correlation between retention index and density measurements at 15 °C for C_10 *iso*-alkanes. The individual isomers are color coded in their respective subgroups.

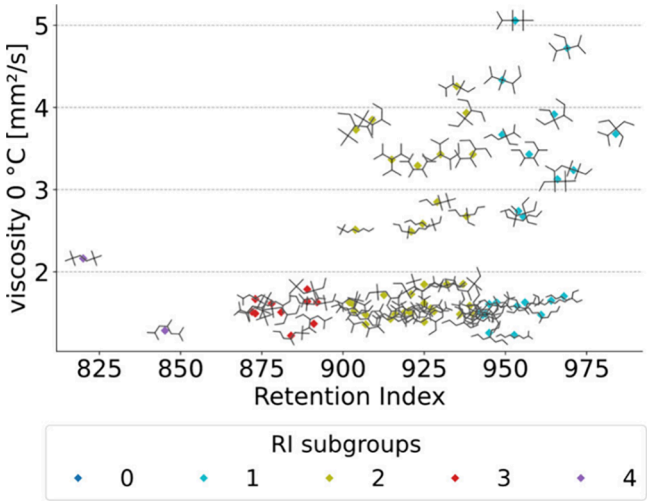


Figure 13. Correlation between retention index and kinematic viscosity measurements at 0 °C for C_10 *iso*-alkanes. The individual isomers are color coded in their respective subgroups. Reproduced from ref 19. Available under a CC BY 4.0 license. Copyright 2025 Lüdtke et al.

fuels other than ATJ-SPK (Figure 10). This observation is consistent with previous studies indicating that the net heat of combustion varies only slightly within the isomers of the *iso*-alkane family.[16] However, since the subgroups are only defined for this hydrocarbon family, the current more precise analysis within this family has little effect on the predicted results.

Overall, these results suggest that the SGMM approach offers limited benefit for the prediction of bulk properties like density and net heat of combustion - except in cases where the assumptions of the baseline model are clearly violated by the compositional structure of the fuel.

Kinematic Viscosity and Cetane Number. Although Hall et al.[16] have demonstrated a significant isomer effect on the kinematic viscosity this effect could not reflected by the implementation of subgroup information (Figure 12). This discrepancy is largely attributed to the weak correlation between viscosity and the RI derived from GC×GC measurements, as shown in Figure 13.

Nevertheless, a significant improvement in viscosity prediction was observed for ATJ-SPK (Figure 8). Due to its distinct molecular composition, which is dominated by a few highly branched *iso*-alkanes, the assumptions underlying the mean-matrix model (uniform isomer distribution within the compositional bins) are again invalid. The SGMM captures the high degree of branching more accurately, leading to a lower and more accurate viscosity prediction.

In contrast, a clear dependence of cetane number on molecular branching and thus RI has been reported in the literature,[38−41] supporting the value of subgroup analysis for this property. As visualized in Figure 14, the incorporation of subgroup information introduces systematic adjustments of the used mean values. While higher subgroups and thus highly branched isomers lower the cetane number, smaller subgroups lead to an increase in the cetane number when comparing the SGMM model assumption with the mean matrix baseline.

However, this does not result in consistent improvement across all samples (Figure 12). For the jet fuel and SBC samples, SGMM yields a better agreement with experimental data, while for the IsoPars, the conventional Mean Matrix model performs comparably or even better. This outcome confirms that the use of a linear mixing rule is not suitable for cetane number prediction, as nonlinear molecular interactions play a significant role. This is supported by the findings of

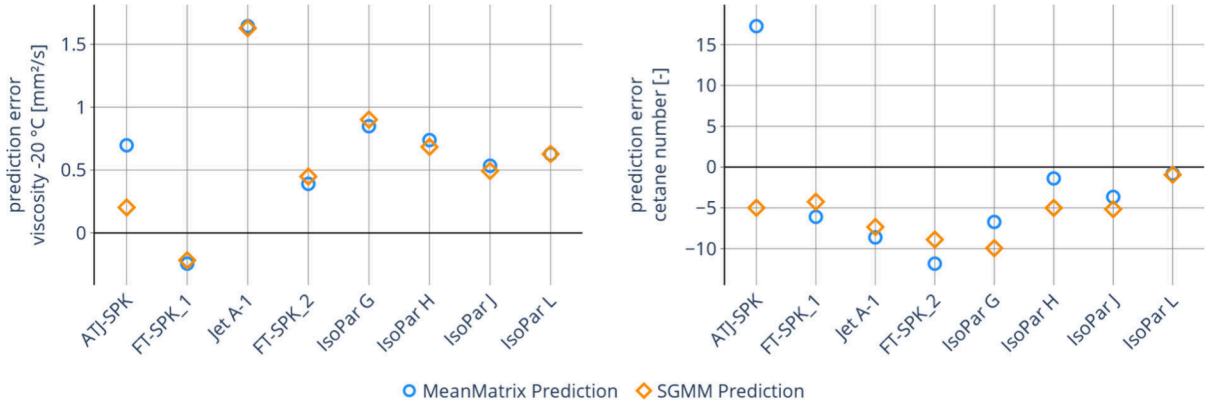


Figure 12. Absolute prediction errors for kinematic viscosity (left) and cetane number (right) across all samples. Blue circles represent errors from the Mean Matrix model; orange diamonds represent errors from the SGMM model. The horizontal zero line indicates perfect agreement with experimental measurements.

https://doi.org/10.1021/acs.energyfuels.5c02762
Energy Fuels 2025, 39, 18001−18012

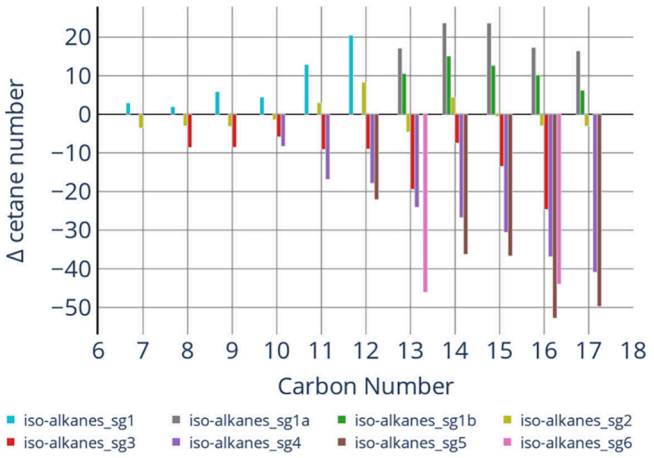


Figure 14. Subgroup-level deviations in cetane number predictions relative to the Mean Matrix model. The horizontal zero line reflects the baseline Mean Matrix assumptions; bars represent the absolute differences introduced by the SGMM approach for each subgroup.

Boehm et al.[42] and Kim et al.,[43] who observed nonlinear blending behavior in certain mixtures.

4.2. Comparative Assessment of Key Fuels. This chapter synthesizes the modeling results and evaluates the comparative performance of the SGMM approach across key the in chapter 3.1 highlighted fuel samples (ATJ-SPK, FT-SPK_1 and Jet A-1). Among the evaluated fuels, ATJ-SPK clearly proves to be a case that is particularly suitable for the SGMM approach. Due to its strong subgroup profile dominated by highly branched isomers, SGMM significantly improved the prediction accuracy for all properties. These results illustrate the importance of considering specific isomer contributions in fuels with low compositional variability, where the core assumption of the mean matrix model, a uniform isomer distribution, is clearly violated.

In contrast, the impact of subgroup incorporation on FT-SPK_1 was substantially lower. This outcome can be partially attributed to its lower *iso*-alkane content, which accounts for only 42% of the total composition, compared to 95% in ATJ-SPK. In addition, the subgroup distribution in FT-SPK_1 is dominated by less branched isomers whose properties are more consistent with the assumptions of the mean matrix and therefore have less influence on the divergence between the modeling approaches.

For Jet A-1, only moderate improvements were observed. This is likely due to the complex composition of the fuel, which includes a substantial proportion of *cyclo*-alkanes and aromatics, for which subgrouping has not yet been developed. The unresolved isomeric variance within these hydrocarbon families limits the effectiveness of the current SGMM approach, indicating the need for further subgrouping methods that include additional hydrocarbon families.

4.3. Limitations of the Modeling Approach. A key limitation of the modeling approach is the potential bias in the reference database toward lower-branched isomers, coupled with its reliance on QSPR-predicted property values. Since the database content determines which isomers contribute to the bulk property estimates, any underrepresentation of specific species can introduce systematic bias into the results. This issue also affects the subgrouping approach: although subgrouping mitigates this source of error to some extent, it cannot be fully eliminated until all isomers present in a given fuel are definitively identified. An optimization of the weighted average matrices, as proposed by Shi et al.,[18] is not possible due to the limited data with subgroup information available at the time of writing. Moreover, the approach underlies the assumption that the QSPR models used to supplement experimental property data are sufficiently accurate. As the availability of subgroup-resolved fuel samples grows, systematic evaluation of the QSPR contribution to predictive accuracy will be essential.

5. CONCLUSION AND OUTLOOK

This study shows the effect of retention index (RI)-based *iso*-alkane subgroups, derived from GCxGC compositional analysis, on fuel property prediction. The subgroup framework was implemented within a weighted average method,[18,20] represented by the mean matrix method. A comparison of property predictions from the Mean Matrix and SubGroup Mean Matrix (SGMM) approaches was conducted using eight samples, comprising a jet fuel, three synthetic blending components and four *iso*-paraffinic solvents. Both methods rely on isomer property measurements supplemented by quantitative structure−property relationship (QSPR)-predictions and mixing rules to compute the bulk property.

The SGMM approach was selected because it does not require training on fuel samples and relies solely on isomer-specific property measurements or QSPR models to estimate isomer properties. Its capacity to capture isomeric interaction effects is limited because it represents all isomers of a given carbon number within a hydrocarbon family as a linear combination. The approach collapses them into a single pseudoisomer defined by the mean property value for each hydrocarbon family−carbon number pair. This approach inherently introduces uncertainties. However, refining the resolution of GCxGC data, by subdividing iso-alkanes into multiple subgroups, is expected to reduce these uncertainties and improve accuracy. The SGMM operates independently of large training fuel data sets, which makes it well suited for an early stage integration of subgroup-level information. Despite its simplicity, the modeling results were overall positive: the introduction of subgroup-level analysis resolution led to improved prediction accuracy, especially for fuel samples dominated by highly branched subgroups. Moreover, the prediction of distillation properties, which are strongly correlated with the RI, was significantly improved.

Unexpectedly, the inclusion of subgroups led to both improvements and deteriorations in the predictive accuracy for paraffinic solvents. The reasons for these results are still unclear and should be investigated in future work. Deteriorations were most notably for the cetane number prediction accuracy, where a data-driven model capable of learning nonlinear mixing behavior directly from experimental data would likely benefit from the inclusion of subgroup descriptors. Alternatively, implementing an empirical mixing rule, such as that proposed by Ghosh and Jaffe[44] and later refined by Creton et al.,[45] may improve the accuracy of the cetane number predictions.

The utility of the subgroup framework for fuel prescreening models could be further enhanced by extending its applicability to additional hydrocarbon families, such as *cyclo*-alkanes and aromatics. Additionally, expanding the data set of subgroup-resolved samples is essential to strengthen the findings and enable the use of standardized predictive capability metrics, such as those proposed by Hall et al.[15] A larger data set would

https://doi.org/10.1021/acs.energyfuels.5c02762
Energy Fuels 2025, 39, 18001−18012

also enable the development and validation of more advanced modeling techniques. As more samples with subgroup information become available, the impact of subgroup boundary selection and isomer distribution on prediction accuracy should be evaluated.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.energyfuels.5c02762.

> Overview of isomer database relative to theoretically possible constitutional isomers and those with measured cetane numbers; analysis of density prediction offset using the Mean Matrix Model; comparison of database coverage with possible existing isomers by subgroup; assessment of subgroup effects on Mean Matrix Model assumptions; detailed compositions and iso-alkane subgroup distributions for FT-SPK_2 and Isopars (G, H, J, L); classification scheme for evaluation metrics; boiling point measurements of iso-alkanes C10 versus retention index by subgroup (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Florian Pütz − *DLR, German Aerospace Center, Institute of Combustion Technology, 70569 Stuttgart, Germany;* orcid.org/0009-0006-5270-2030; Email: Florian.Puetz@dlr.de

Authors

Hannes Lüdtke − *DLR, German Aerospace Center, Institute of Combustion Technology, 70569 Stuttgart, Germany;* orcid.org/0009-0002-4772-8245

Astrid Ramirez Hernandez − *DLR, German Aerospace Center, Institute of Combustion Technology, 70569 Stuttgart, Germany*

Patrick Oßwald − *DLR, German Aerospace Center, Institute of Combustion Technology, 70569 Stuttgart, Germany;* orcid.org/0000-0002-2257-2988

Patrick Le Clercq − *DLR, German Aerospace Center, Institute of Combustion Technology, 70569 Stuttgart, Germany;* orcid.org/0000-0001-6011-5625

Uwe Bauder − *DLR, German Aerospace Center, Institute of Combustion Technology, 70569 Stuttgart, Germany;* orcid.org/0000-0002-5019-6043

Georg Eckel − *DLR, German Aerospace Center, Institute of Combustion Technology, 70569 Stuttgart, Germany*

Andreas Huber − *DLR, German Aerospace Center, Institute of Combustion Technology, 70569 Stuttgart, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.energyfuels.5c02762

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

All content and arguments presented in this work were developed by the authors. Language editing support using AI tools was employed to improve clarity and readability, particularly in Section 4.

■ REFERENCES

(1) Fetting Constanze, *The European Green Deal, ESDN Report 2/2020*, ESDN Office, 2020.

(2) European Commission. *ReFuelEU Aviation*. https://transport.ec.europa.eu/transport-modes/air/environment/refueleu-aviation_en (accessed 2025−05−12).

(3) Griffin, M. B.; Iisa, K.; Dutta, A.; Chen, X.; Wrasman, C. J.; Mukarakate, C.; Yung, M. M.; Nimlos, M. R.; Tuxworth, L.; Baucherel, X.; Rowland, S. M.; Habas, S. E. Opening pathways for the conversion of woody biomass into sustainable aviation fuel via catalytic fast pyrolysis and hydrotreating. *Green Chem.* **2024**, *26* (18), 9768−9781.

(4) Bell, D. C. Isomer Identification in Sustainable Aviation Fuel Using Gas Chromatography and Vacuum Ultraviolet Spectroscopy. PhD Thesis, Washington State University, 2024. DOI: 10.7273/000006915.

(5) Hileman, J.. Fuel Approval Process & Status, ICAO SAF Stocktaking 2019, ICAO, 2019.

(6) Heyne, J.; Rauch, B.; Le Clercq, P.; Colket, M. Sustainable aviation fuel prescreening tools and procedures. *Fuel* **2021**, *290*, No. 120004.

(7) Hall, C.; Creton, B.; Rauch, B.; Bauder, U.; Aigner, M. Probabilistic Mean Quantitative Structure−Property Relationship Modeling of Jet Fuel Properties. *Energy Fuels* **2022**, *36* (1), 463−479.

(8) Shao, Y.; Yu, M.; Zhao, M.; Xue, K.; Zhang, X.; Zou, J.-J.; Pan, L. Comprehensive accurate prediction of critical jet fuel properties with multiple machine learning models. *Chem. Eng. Sci.* **2025**, *304*, No. 121018.

(9) 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*, No. 120345.

(10) Hall, C.; Rauch, B.; Bauder, U.; Aigner, M. Comparison of probabilistic jet fuel property models for the fuel screening and design. *Fuel* **2023**, *351*, No. 128965.

(11) 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*, No. 125002.

(12) Vozka, P.; Kilaz, G. A review of aviation turbine fuel chemical composition-property relations. *Fuel* **2020**, *268*, No. 117391.

(13) Wang, Y.; Ding, Y.; Wei, W.; Cao, Y.; Davidson, D. F.; Hanson, R. K. On estimating physical and chemical properties of hydrocarbon fuels using mid-infrared FTIR spectra and regularized linear models. *Fuel* **2019**, *255*, No. 115715.

(14) Wang, Y.; Cao, Y.; Wei, W.; Davidson, D. F.; Hanson, R. K. A new method of estimating derived cetane number for hydrocarbon fuels. *Fuel* **2019**, *241*, 319−326.

(15) 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 Fuels* **2021**, *35* (3), 2520−2530.

(16) Hall, C.; Bell, D. C.; Feldhausen, J.; Rauch, B.; Heyne, J. Quantifying isomeric effects: A key factor in aviation fuel assessment and design. *Fuel* **2024**, *357*, No. 129912.

(17) Villanueva, N.; Flaconnèche, B.; Creton, B. Prediction of Alternative Gasoline Sorption in a Semicrystalline Poly(ethylene). *ACS combinatorial science* **2015**, *17* (10), 631−640.

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

(19) Lüdtke, H.; Pütz, F.; Grams, S.; Bauder, U.; Köhler, M.; Huber, A.; Oßwald, P. GCxGC-Based iso -Alkane Subgrouping for Enhanced Compositional Analysis of Sustainable Aviation Fuels. *Energy Fuels* **2025**, *39* (9), 4434−4449.

(20) Vozka, P.; Modereger, B. A.; Park, A. C.; Zhang, W. T. J.; Trice, R. W.; Kenttämaa, H. I.; Kilaz, G. Jet fuel density via GC×GC-FID. *Fuel* **2019**, *235*, 1052−1060.

(21) Hall, C. A. Data-based methods for the screening and design of jet fuels. Dissertation, Universität Stuttgart, 2024.
(22) Saldana, D. A.; Starck, L.; Mougin, P.; Rousseau, B.; Pidol, L.; Jeuland, N.; Creton, B. Flash Point and Cetane Number Predictions for Fuel Compounds Using Quantitative Structure Property Relationship (QSPR) Methods. *Energy Fuels* **2011**, *25* (9), 3900−3908.
(23) Saldana, D. A.; Starck, L.; Mougin, P.; Rousseau, B.; Ferrando, N.; Creton, B. Prediction of Density and Viscosity of Biofuel Compounds Using Machine Learning Methods. *Energy Fuels* **2012**, *26* (4), 2416−2426.
(24) Saldana, D. A.; Starck, L.; Mougin, P.; Rousseau, B.; Creton, B. On the rational formulation of alternative fuels: melting point and net heat of combustion predictions for fuel compounds using machine learning methods. *SAR and QSAR in Environmental Research* **2013**, *24* (4), 259−277.
(25) Saldana, D. A.; Starck, L.; Mougin, P.; Rousseau, B.; Creton, B. Prediction of Flash Points for Fuel Mixtures Using Machine Learning and a Novel Equation. *Energy Fuels* **2013**, *27* (7), 3811−3820.
(26) National Institute of Standards and Technology. *NIST Standard Reference Database 103a:NIST ThermoData Engine Version 10.4.5 - Pure Compounds*. https://www.nist.gov/mml/acmd/trc/thermodata-engine/srd-nist-tde-103a (accessed 2022−01−17).
(27) Design Institute for Physical Properties. *DIPPR 801 Database*. https://www.aiche.org/dippr/events-products/801-database (accessed 2022−01−18).
(28) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2025 update. *Nucleic acids research* **2025**, *53* (D1), D1516−D1525.
(29) Royal Society of Chemistry. *ChemSpider*. https://www.chemspider.com/ (accessed 2022−03−04).
(30) Yang, Z.; Boehm, R. C.; Bell, D. C.; Heyne, J. S. Maximizing Sustainable aviation fuel usage through optimization of distillation cut points and blending. *Fuel* **2023**, *353*, No. 129136.
(31) Gugisch, R.; Kerber, A.; Kohnert, A.; Laue, R.; Meringer, M.; Rücker, C.; Wassermann, A. MOLGEN 5.0, A Molecular Structure Generator. In *Advances in Mathematical Chemistry and Applications: Revised Edition*; Bentham Books, 2016; Vol. *1*, pp 113−138.
(32) Hadaller, O. J.; Johnson, J. M. *World Fuel Sampling Program: Final Report*, CRC Report No. 647, Coordinating Research Council, Inc., 2006.
(33) Ford, D. C.; Miller, W. H.; Thren, R. C.; Wertzler, R. Correlation of ASTM Method D 2887−73 Boiling Range Distribution Data with ASTM Method D 86−67 Distillation Data. In *Calculation of physical properties of petroleum products from gas chromatographic analyses*, Online-Ausg; Green, L. E., Albert, D. K., Green, L. E., Eds.; ASTM STP 577; American Society for Testing and Materials; ASTM International, 1975; 20−20−11. DOI: 10.1520/STP35520S.
(34) Oßwald, P.; Zinsmeister, J.; Kathrotia, T.; Alves-Fortunato, M.; Burger, V.; van der Westhuizen, R.; Viljoen, C.; Lehto, K.; Sallinen, R.; Sandberg, K.; Aigner, M.; Le Clercq, P.; Köhler, M. Combustion kinetics of alternative jet fuels, Part-I: Experimental flow reactor study. *Fuel* **2021**, *302*, No. 120735.
(35) Zschocke, A.; Scheuermann, S.; Ortner, J. High Biofuel Blends in Aviation (HBBA), ENER/C2/2012/420−1 Final Report; Lufthansa & Wehrwissenschaftliches Institut für Werk- und Betriebsstoffe 2012.
(36) Dallüge, J.; Beens, J.; Brinkman, U. A. T. Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool. *Journal of chromatography. A* **2003**, *1000* (1−2), 69−108.
(37) Prodhan, M. A. I.; Yin, X.; Kim, S.; McClain, C.; Zhang, X. Surface fitting for calculating the second dimension retention index in comprehensive two-dimensional gas chromatography mass spectrometry. *Journal of chromatography. A* **2018**, *1539*, 62−70.
(38) Abdul Jameel, A. G. A functional group approach for predicting fuel properties, KAUST Research Repository, 2019.
(39) Abdul Jameel, A. G.; Naser, N.; Emwas, A.-H.; Dooley, S.; Sarathy, S. M. Predicting Fuel Ignition Quality Using 1 H NMR Spectroscopy and Multiple Linear Regression. *Energy Fuels* **2016**, *30* (11), 9819−9835.
(40) Sarathy, S. M.; Westbrook, C. K.; Mehl, M.; Pitz, W. J.; Togbe, C.; Dagaut, P.; Wang, H.; Oehlschlaeger, M. A.; Niemann, U.; Seshadri, K.; Veloo, P. S.; Ji, C.; Egolfopoulos, F. N.; Lu, T. Comprehensive chemical kinetic modeling of the oxidation of 2-methylalkanes from C7 to C20. *Combust. Flame* **2011**, *158* (12), 2338−2357.
(41) Won, S. H.; Dooley, S.; Veloo, P. S.; Wang, H.; Oehlschlaeger, M. A.; Dryer, F. L.; Ju, Y. The combustion properties of 2,6,10-trimethyl dodecane and a chemical functional group analysis. *Combust. Flame* **2014**, *161* (3), 826−834.
(42) Boehm, R. C.; Yang, Z.; Bell, D. C.; Faulhaber, C.; Mayhew, E.; Bauder, U.; Eckel, G.; Heyne, J. S. Perspective on Fuel Property Blending Rules for Design and Qualification of Aviation Fuels: A Review. *Energy Fuels* **2024**, *38* (18), 17128−17145.
(43) Kim, Y.; Kumar, S.; Cho, J.; Naser, N.; Ko, W.; St. John, P. C.; McCormick, R. L.; Kim, S. Designing high-performance fuels through graph neural networks for predicting cetane number of multi-component surrogate mixtures. In *SAE Technical Paper Series*; Kim, Y., Kumar, S., Cho, J., Naser, N., Ko, W., St. John, P. C., McCormick, R. L., Kim, S., Eds.; Society of Automotive Engineers of Japan, 10−2 Gobancho, Chiyoda-ku, Tokyo, Japan, 2023. DOI: 10.4271/2023-32-0052.
(44) Ghosh, P.; Jaffe, S. B. Detailed Composition-Based Model for Predicting the Cetane Number of Diesel Fuels. *Ind. Eng. Chem. Res.* **2006**, *45* (1), 346−351.
(45) Creton, B.; Brassart, N.; Herbaut, A.; Matrat, M. Numerical Approaches to Determine Cetane Number of Hydrocarbons and Oxygenated Compounds, Mixtures, and their Blends. *Energy Fuels* **2024**, *38* (16), 15652−15661.